Institute of Medical Microbiology

Comparative transcriptomics of intracellular survival of *Listeria*

A bioinformatics approach to determine the minimal genome required for intracellular survival of *Listeria monocytogenes*

Inaugural-Dissertation (Cumulative thesis)
Submitted to the Faculty of Medicine
in fulfilment of the requirements for the degree of Dr. biol. hom.
of the Faculty of Medicine
of the Justus-Liebig-University Giessen

by

André Billion

Gießen 2012

From the Institute of Medical Microbiology
Director: Prof. Dr. Trinad Chakraborty
Universitätsklinikum Gießen und Marburg GmbH
Standort Gießen

First supervisor and Committee Member: Prof. Dr. T. Chakraborty
Committee Member: Prof. Dr. K.T. Preissner
Date of Doctoral Defence: 23.04.2013

Index of Contents

1 Introduction

Over the past decade, there has been a fundamental technological shift from identification of individual nucleotides to the simultaneous detection of billions of nucleotides during massive-parallel short read sequencing. Due to the fact, that this technology allows scientists to digitally query genomes on a revolutionary precise scale, a paradigm shift in DNA and RNA analysis has occurred. Enhanced by technological advantages and the dramatic fall in sequencing costs have a great impact on research of many areas of science including host-pathogen interaction, genomic epidemiology, pathogen detection and metagenomics. However, the flood of data becoming available to the scientific community is huge and the challenge is to aggregate information into interpretable and well-curated datasets. Understanding this data rests fundamentally on well-curated up-to-date annotation of reference genomes which can then be leveraged to perform comparative and epidemiological genome analysis.

From the detour of cloning & amplicons to direct sequencing

Molecular cloning and sanger sequencing of amplified products was commonly used to reveal the sequences of whole genomes of virus or bacteria. This approach has serious limitations with regards to costs and speed and new and improved technologies for sequencing large numbers of genomes has now become available.

The ability to sequence genetic material easily and with a relatively small financial effort on a whole genome level within only a few hours as compared months and years previously, will change the types of questions that can be asked tremendously. In the past, over 1000 scientists working for more than 13 years spent at least 3 billion US dollars to sequence a single human genome. Today the same progress can be achieved by 3 scientists within one month and with a cost of less than 10.000 US dollars.

With this new technology, researchers are now in the position to e.g. unveil the genetic background information of our environment by sequencing an enormous amount of bacteria available. In this respect, sequencing of pathogenic and non-pathogenic isolates has been developed by us as an approach to identify genes involved in virulence, since genetic differences can be best obtained by comparative genomics and therefore results are more precise the more data becomes available.

In the year 2001, within a consortium, scientists from the Institute of Medical Microbiology sequenced and published the first pair of pathogenic and non-pathogenic member of the genus *Listeria* to enable better identification of genes involved in e.g. metabolism and intracellular survival. Concurrently we established and optimized genome analysis workflows to guarantee high quality annotation of protein function for further analysis. As an example the software platform AUGUR was developed to automatically annotate cell surface proteins of bacteria on the fly and to catalogue and categorize function. Because surface proteins are crucial for the interaction with the environment they represent major targets for the induction of immune response and can be used for diagnostic purposes.

The next level: From DNA to RNA

It is beyond debate that genome sequencing is essential for future research, however it can only provide the basic information as e.g. gene content. The next step is to validate that identified genes are expressed under various conditions which can be only achieved by transcriptomics, proteomics and functional studies.

For a long time, only three major types of RNA were generally known: messenger RNA (mRNA), ribosomal RNA (rRNA) and transfer RNA (tRNA). To investigate transcribed genes in the past, a complex and time-consuming blotting technology, originating from southern blotting, was used. In this technique, fragmented DNA is bound to a substrate, denatured, dried and then exposed to a labelled hybridization probe in an appropriate buffer. Over time, blotting has been largely replaced by new molecular techniques like marco- and microarrays.

First stop: microarrays

Microarrays evolved as scientists had to face the challenge to establish high-throughput methods to detect and identify the relative expression of genes in RNA samples. Briefly, a microarray has a numerous number (up 30k) of microscopic DNA spots attached to a solid (glass) surface. Each spot contains a specific DNA sequence which is used to hybridize a

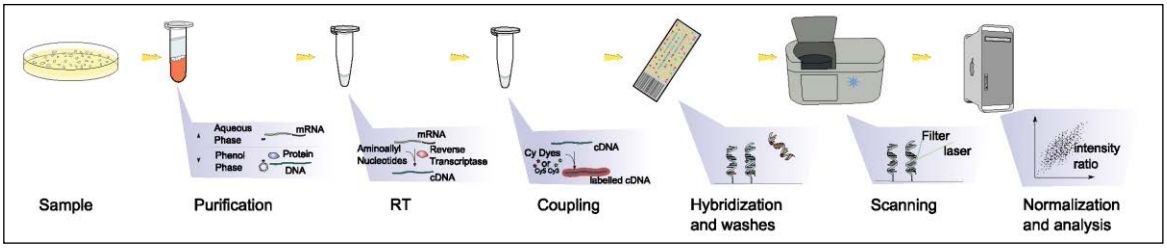


Fig. 1 Regular microarray workflow

cDNA or cRNA sample. Hybridization is usually detected by fluorophore-labelled targets to determine the amount of sample binding to each spot. After scanning and bioinformatics analysis, a relative quantification, based on comparison between two experiments, enables the researcher to measure the expression levels of large numbers of genes simultaneously. Previously, microarray technology was one of few methods which could provide expression information on a whole genome scale for bacteria. However, it is crucial to remember that microarrays require *a priori* knowledge of the queried genome which makes this technology unusable to discover the unknown and unexplored.

After revealing genome sequences of several *Listeria* genomes, the next logical step was the investigation of the transcriptome of these strains. For this purpose we developed a complete microarray workflow. This pipeline is composed of a section for manufacturing slides, including array design, spotting of slides, hybridization, and scanning as well as a complete analysis platform ranging from quality control, statistics, further detailed analysis, and storage of data, to an export into public databases like GEO or ArrayExpress to obtain accession numbers for publication. In the past, this pipeline was used for a number of experiments [3-8]. The pipeline itself will be described in more detail later.

Second stop: transcriptome sequencing (RNA-Seq)

Today, scientists have an even better method at one's disposal, namely transcriptome sequencing (RNA-Seq). This technology can be used for a wide spectrum of analysis like transcriptome profiling, miRNA profiling, DNA-protein interaction investigation (ChIP-Seq), or DNA methylation studies. For the first time it is possible to obtain precise information about type and quantity of RNA, present at a certain condition at whole genome scale. As already mentioned above, microarray technology only allows observations of expression of already known RNAs, while RNA-Seq is able to represent each RNA molecule even at relatively low concentrations or low copy-numbers respectively. With this technology, researchers get new insights into the transcriptomic and thus genetic diversity of an organism and the ability to explore functionalities at a complete new level of detail.

New technology, new insights

Before going further into the analysis of e.g. new types of regulatory RNAs, RNA-Seq should first be used to improve the results of genome sequencing and annotation to guarantee the optimal use of information for continuing research. For nearly all prokaryotic genomes annotated gene content is based on predictions. That includes the number of genes

as well as their start and termination positions. These predictions are error prone, have problems in detecting small genes and are incapable of forecasting UTRs. RNA-Seq can reveal new genes (Fig. 2a) and additionally supports re-annotation of transcription start sites (TSS) to refine actual genome information (Fig.2b). TSS analysis supports the identification of new or known promoter sequences respectively. Also multiple transcription start sites of a single gene can now be detected in a genome-wide manner.

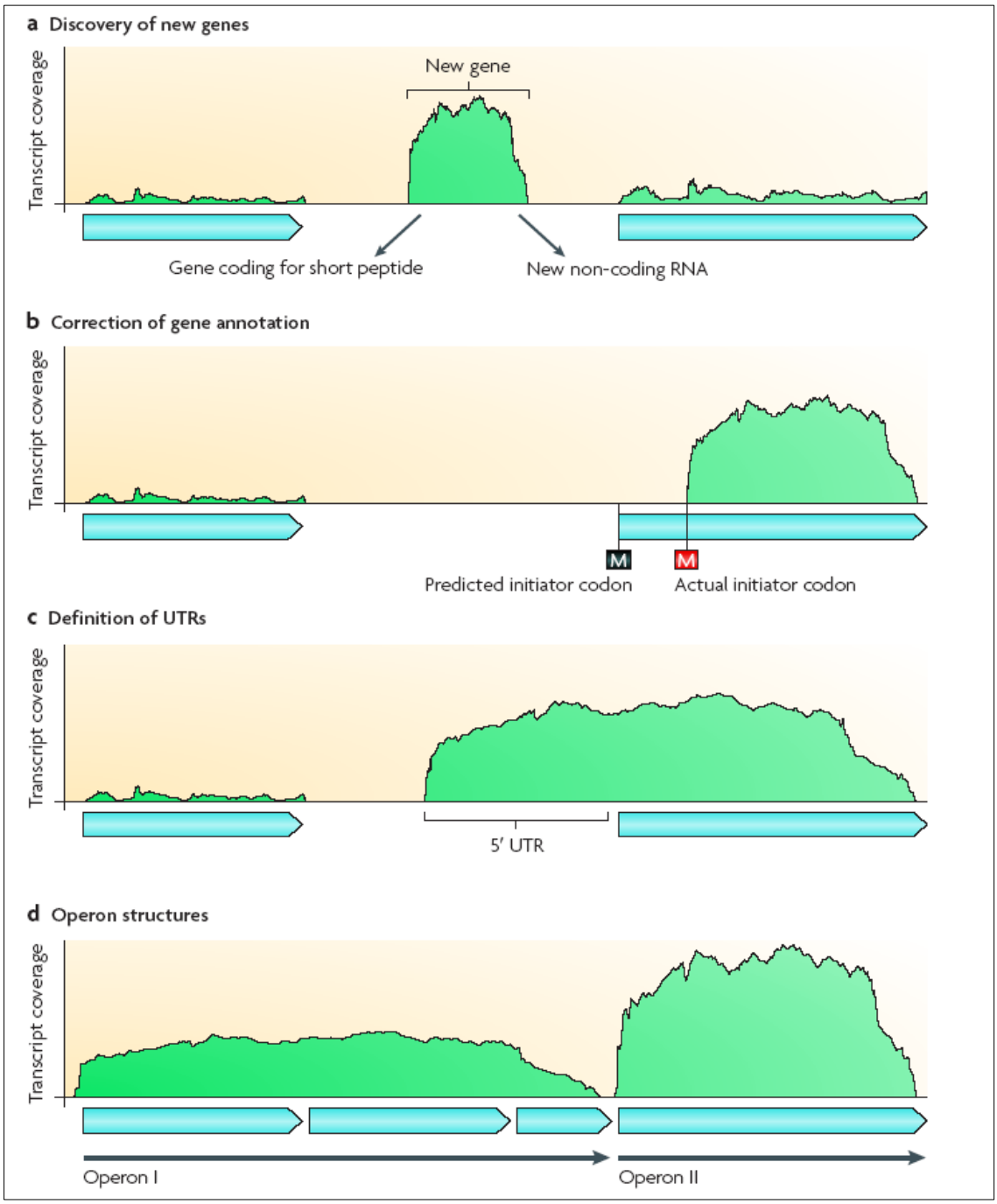


Fig.2 Novel information through RNA-Seq experiments [2]

Furthermore, to date, sequencing is the only available technology to reveal *untranslated regulatory regions* (UTRs) at whole-genome scale depending on examined conditions.

These UTRs contain important regulatory elements like riboswitches or binding sites for regulatory proteins and small RNAs while it was shown that for example for *Salmonella typhi* long UTRs in front of pathogenicity islands suggest a role in virulence regulation (Fig. 2c). Finally, RNA-Seq can be used to reveal the complete operon structure (Fig. 2d). Since 60 – 70% of all genes in bacteria are organized in operons, this task becomes more and more important. Recent studies created an operon map of *M. pneumoniae* grown under 173 different conditions [9]. These kinds of studies promote our understanding of a versatile regulation and evolution of operon structures in prokaryotes and enable the creation of comparative transcriptional maps for highly diverse conditions.

Primer on new types of RNA

A part from improvement of already existing knowledge, RNA-Seq experiments uncovered a variety of new types of RNAs such as small non-coding RNA (sRNA), small interfering RNA (siRNA), antisense RNA (asRNA), *cis*-regulatory RNA and riboswitches. These classes of regulatory RNAs modulate a wide range of physiological processes such as transcription, translation, mRNA stability, and DNA stability or DNA silencing. They achieve this by changes in conformation, protein binding, and base pairing with other RNAs and interaction with DNA. The majority of these regulators respond to changes in environmental conditions such as the availability of nutrients, temperature and stress response. In general, three different classes of small RNAs are most prevalent.

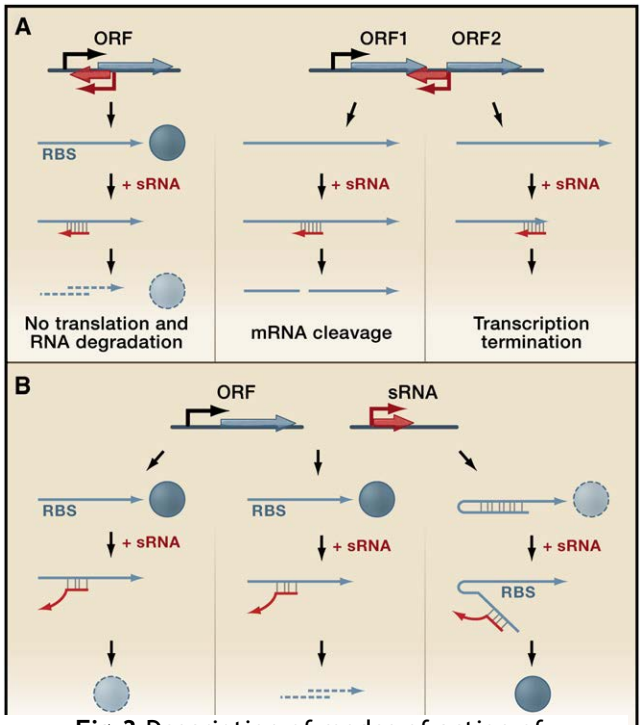


Fig.3 Description of modes of action of regulatory RNAs [1]

Class I: trans-acting RNAs (sRNAs)

Trans acting RNAs represent the regular type of small regulatory RNA (Fig. 3b). They generally originate within intergenic regions while their target is not necessarily located in the immediate vicinity. Their major role is to affect translation and/or stability of target mRNA by standard base pairing. Most of them regulate negatively by repression of protein levels through translational inhibition or mRNA degradation. They form a complex with a

target mRNA which is cleaved by RNAseE. Currently only a few sRNA <-> mRNA interactions are characterized, albeit these sRNA are analogous in structure and behaviour to miRNAs from eukaryotes.

Class II: antisense RNAs (asRNAs)

Most characterized asRNAs regulate gene expression by base pairing with their target mRNA (Fig. 3a). The most abundant fraction of this class is encoded antisense of the DNA strand opposite to their target RNA. They share around 50 to 100 nucleotides with complete complementary sequence to their target. Most of the known asRNAs were found in bacteriophages, plasmids, and transposons and regulate copy numbers of their mobile elements by inhibition of replication primers or transposase translation. It is known that asRNAs act also as antitoxins by repressing the translation of toxic proteins that may kill cells from which the origin has been lost. Nevertheless the catalogue of asRNAs is far from complete and needs further studies.

Class III: Riboswitches and cis-acting RNAs

Riboswitches are perhaps the simplest and best known regulatory RNAs. They change their conformations as a reaction to environmental signals such as nutrients, varied temperatures, and uncharged tRNAs. In general they act as repressors that affect expression of downstream genes usually by forming hairpin structures to interrupt transcription or translation in presence of e.g. metabolites. Since metabolites are essential for growth and adaptation, immediate and almost costless regulation at genome scale is an invaluable advantage for adaption to a hostile environment. On that account, they are often termed as "metabolite sensors".

A unique class of regulatory RNAs is represented by the so called CRISPR RNAs. **C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats enable bacteria to defend against bacteriophages and plasmid conjugation. CRISPRs could be found in half of all bacteria and nearly all archaea sequenced today. The system consists of an approximately 500 bp leader followed by an array of alternating repeat and highly specific spacer sequences. Functional CRISPR arrays are flanked by several **C**RISPR **as**sociated genes (Cas genes) which are responsible for maintenance, processing, and targeting of foreign RNA and

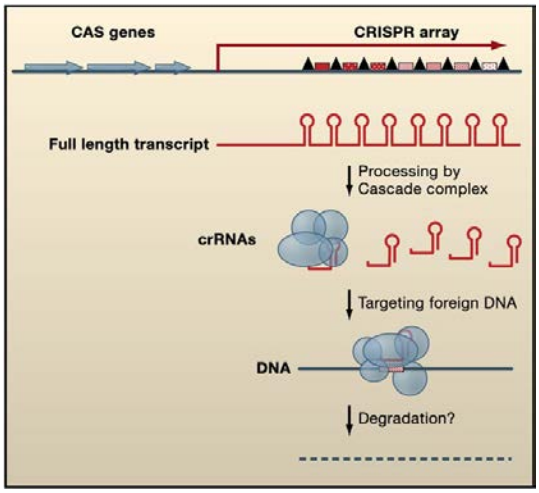


Fig.4 Schematic overview of CRISPR systems [1]

DNA (Fig. 4). After contact with putative hostile bacteriophages a specific sequence of the aggressor will be incorporated in the CRISPR array, providing immunity to bacteria that may never have encountered the phage itself before. So it is justifiable to claim that the CRISPR system is the only framework known so far, that represents an adaptive and inheritable immune system for bacteria. In consequence of its uniqueness the system can be used in broad applications. Due to the very high variability of the spacer array itself, CRISPR provide a useful tool to genotype strains, study horizontal gene transfer and investigate microevolution. While each CRISPR locus records the history of phage attacks over time, this kind of specific "bar code" can be used to differentiate between various strains and allows to identify and track the progression of pathogenic bacteria worldwide [10]. Furthermore, the CRISPR system can be used to protect industrial relevant strains in e.g. dairy production by this so called "self-vaccination". This can potentially prevent the entire bacterial culture and production batches from being destroyed. Additionally these bacteria are not regarded as genetically modified (Ref.no. 6790-05-01-60).

Unreached effectivity

In summary, it can be stated that RNA regulators have several advantages over regular protein regulation. Especially concerning their size, sRNAs are less costly and faster to produce than most mRNAs and need no extra step of translation. In case of riboswitches, the regulation speed itself is unmatched, because it allows the cell to respond to a signal in an extremely rapid and specific way regardless of number and location of regulated genes. Especially in hostile situations a cheap, fast and specific mechanism of fine-tuning can increase the ability of survival of bacteria in the host cell.

These mostly new facts are however only accessible through high throughput transcriptome sequencing. To uncover all these fascinating new insights in e.g. regulation during intracellular survival, the applied technology has to produce plenty of data to provide detailed information. This emerging flood of information is the major challenge in future sequencing based research, thus the demand on suitable software and bioinformatics is bigger than ever before.

RNA-Seq – Establishment of sequencing workflow

A primary goal of this thesis was the isolation and investigation of total RNA from intracellularly growing bacteria. This is a challenge because bacteria have to be isolated from infected cells and therefore the RNA is a mixture of large amounts of RNA deriving from

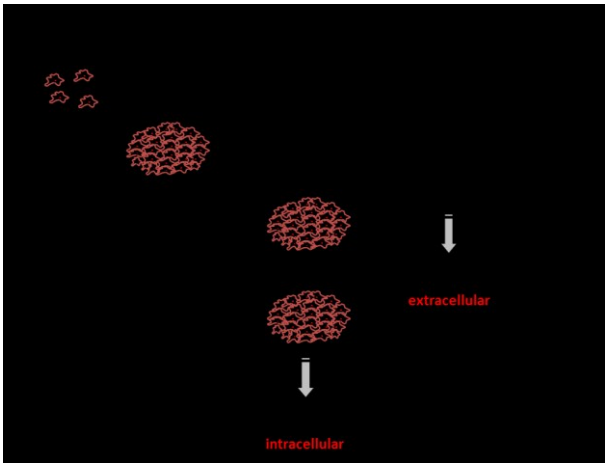


Fig.5 Probe preparation

the host cell and very small amounts of RNA that can be assigned to bacteria. In addition a large portion of bacterial RNA is ribosomal RNA, which complicates matters to determine the contribution of messenger- and small RNAs (mRNA & sRNA). To separate eukaryotic RNA from bacterial RNA we used a protocol developed in the laboratory [7], using acid-phenol extraction. In order to extract sRNAs, total bacterial RNA was fractionated by gel-electrophoresis and the region between 20 to 500 nt's was isolated. RNA was subsequently processed by reverse transcription and converted to cDNA and sequenced using 454 Roche pyro-seq. on GS-FLX (Roche).

Retrieved nucleotide sequences were processed after clipping of the 5'-linker and the polyA tail and all sequences less than 16 nt were removed prior to further analysis. For this purpose a new software program called *sncRAS* was written, which can process such data automatically i.e. without clipping. This pre-processed data is used directly to map the reads against a reference genome using the NCBI Blast algorithm. The identified reads were stored in a central database for further analysis.

The program *sncRAS* can process data deriving from any NGS technologies (454 Roche, Illumina, Ion Torrent, PacBio). A brief description is given below in section 1.1 Bioinformatics workflows.

To illustrate the steps required for processing of data from NGS platforms an example using data derived from Ion Torrent platform is presented. The NGS Ion Torrent platform was commercially introduced in October 2011 and is presently used to generate in-house sequencing data. Briefly, the complete process is divided into four steps starting with a standardized, library construction with offered library kits. Construction time of particular libraries could be optimized to approximately 3.5 hours. New instruments like the *Ion OneTouch* contribute to the efforts in optimization of the whole RNA-Seq workflow. Time required for Step 2 in which the template is prepared, initially took around 10 hours with a substantial amount of hands on time. Using the *OneTouch* instrument the entire template preparation process is fully automated and can be completed in about three hours. The actual sequencing itself is carried out on a *Personal Genome Machine* (PGM) and requires,

according to the used chip, approximately two hours and can produce up to 1 Gigabases of data per run. Time needed for base-calling of the data (Step 4) is also dependent on the used chip, but usually requires around one hour.

Downstream analysis is carried out by an *in-house* established software called *sncRAS*. *sncRNAs* was developed, because free as well as commercial software was not available like today and needed flexibility continues to be a problem. The decision about creation of an *in-house* workflow was based on various demands such as free choice of parameters for trimming/quality and mapping as well as storage of results in a database using to answer individual questions. Since flexibility was given the highest priority, *sncRAS* accepts data from different platforms like Roche 454, Illumina, Ion Torrent, and PacBio. Several quality control, trimming and sorting steps are implemented before data is mapped to the reference sequence via different algorithms such as Bowtie [11] or segemehl [12]. After the mapping step, all results are transferred into a central database from where a wide range of different statistics, predictions, analysis, and exports can be executed. For comprehensive analysis of transcriptomic data, a powerful visualization is of particular importance, supporting clearer understanding of this complex and extensive information.

Microarray vs. RNA-Seq

Technological developments during the lifetime of this thesis made it necessary to constantly adapt bioinformatics programs to be able to process very disparate methodologies to analyse data obtained using microarrays and NGS. Thus, in microarray analysis the sequence of a known genome is used to create oligonucleotide probes representing every

single ORF. The transcriptome of bacteria growing under different conditions is obtained by hybridizing total bacterial RNA to the oligonucleotides spotted onto a glass slide.
In contrast NGS sequencing, as in RNA-Seq, allows the determination of nucleotide sequences of individual transcripts and their relative quantities regardless of whether a genome sequence is available or not. With an existing reference genome, RNA-Seq provides a comprehensive view of all transcribed genes, operons, and regulatory RNAs.
The following table describes pros and cons of either method:

Microarray Analysis		**Next Generation Sequencing**	
Pros	Cons	Pros	Cons
relatively inexpensive	low sensitivity	low background, highly sensitive	relatively expensive
simple sample preparation	limited dynamic range	lage dynamic range	complex sample preparation
minimal mRNA preprocessing	not quantitative	quantitative	limited bioinformatics
mature informatics & statistics	only known sequence detectable	detection of completely present DNA/RNA	massive tech. Infrastructure required
	effectivity based on quality of annotation	less bias	
	large amounts of source material needed	only nanograms of source material required	
	competitive/cross hybridisation		
	limitations in reproducibility (to many possibilities)		

Table 1 Comparison between microarray and next generation sequencing technologies

Major drawback of NGS technology is the need for sophisticated information technology and bioinformatics which are able to handle the vast amount of data. Nevertheless, it is unavoidable that next-generation sequencing will supersede microarray technology for most applications. However, there are specific tasks such as target verification of e.g. small RNAs or experiments necessitating the use of untreated mRNA is necessary, where microarrays may find their niche in the future.

Describing the transcriptional landscape of Listeria monocytogenes

Pathogenic intracellular bacteria are the major cause for a wide range of severe diseases worldwide. Despite this fact, adaptation and strategies for replication and survival have

been studied in detail but are still not completely understood. Typical representatives of this kind are *Yersinia*, *Chlamydia* and *Rickettsia* which place very particular demands on e.g. cultivation.

On that account, *Listeria* represents a distinguished model for the study of intracellular infections. Especially *L. monocytogenes* is the causative agent of severe human infection (listeriosis). Infection with *Listeria*, mainly in certain well-defined high-risk groups, including pregnant woman, neonates, the elderly and immunosuppressed individuals [13] result in high mortality rates of up to 30%. The bacterium inhabits numerous ecological niches and is able to multiply at high salt concentrations (10% NaCl), at a broad range of pH (4.5-9.6) and temperature (0-45 C) [14]. The infection cycle itself has been elucidated in detail and virulence factors which are essential for infection have been extensively studied. The majority of these virulence factors are organized in a virulence gene cluster (VGC) which is regulated by *prfA,* localized at the very beginning of the cluster.

Its highly robust nature and versatility to grow under different conditions has led to it being classified as one of the most deadly food-borne pathogens responsible for numerous outbreaks in recent years. In Autumn of 2011 a total of 146 infected persons and 30 reported deaths, was reported in the United States i.e., a mortality rate 20,5%. Interestingly, the outbreak was spread over 28 states. This contrasts to the recent outbreak of HUSEC 2011 which infected around 3500 people with 53 deaths(mortality rate 1,5%), but of 214 registered patients, 119 were localized in only 3 cities (http://www.eurosurveillance.org).

Previous studies performed in this laboratory have led to an overview of intracellular gene expression of *Listeria monocytogenes*. The studies have been important in detection of new virulence factors and have provided us with an unsurpassed view of genes required for adaptation to intracellular growth. The purpose of this thesis was to refine this data and provide information on the role of sncRNAs in intracellular growth and survival. Despite extensive studies on other intracellular pathogens such as *Salmonella typhimurium* and *Mycobacteria tuberculosis* no information is available on the role of sncRNAs on intracellular growth and survival.

This study was therefore undertaken to understand the role sncRNAs during infection. Furthermore it was an aim to comprehensively map every single messenger RNA in terms of promoter and termination structures, to determine TSS with high precision and to examine if genes are transcribed individually or as operons. For this study it was necessary to interact closely with wet lab scientists as the success of the project was dependent on analysis of the data and constant feedback in terms of quality of the data produced.

1.1 Bioinformatics workflows created in this thesis

As indicated in the introduction, technological developments in the understanding of the transcriptome have been rapid and necessitated the constant development of new bioinformatics tools to cope and improve data processing. Thus, at the beginning of this thesis the focus was on the analysis of microarray data. This proceeded to the analysis of tiling array data and finally required the development of a new bioinformatics platform to analyse sequences from NGS technologies.

Microarray Workflow

We established an *in-house* workflow, in which first PCR amplicons are performed with HotStar*Taq* DNA Polymerase (QIAGEN) with cDNA as a template. After purification and inspection by agarose gels, DNA concentration will be determined by a SpectraMax GEMINI XS spectrofluorometer (Molecular Devices). Whole genome microarrays are then spotted using a Generation III microarray spotter from GE Health Care. After labeling probes with Cy-dyes (both Cy3 and Cy5 for each probe) hybridization with an ASP base Hybridizer (GE Health Care) is performed. Finally slides are then imaged with a Generation III array scanner (GE Health Care) and signal intensities are quantified using Spotfinder software from GE Health Care as well.

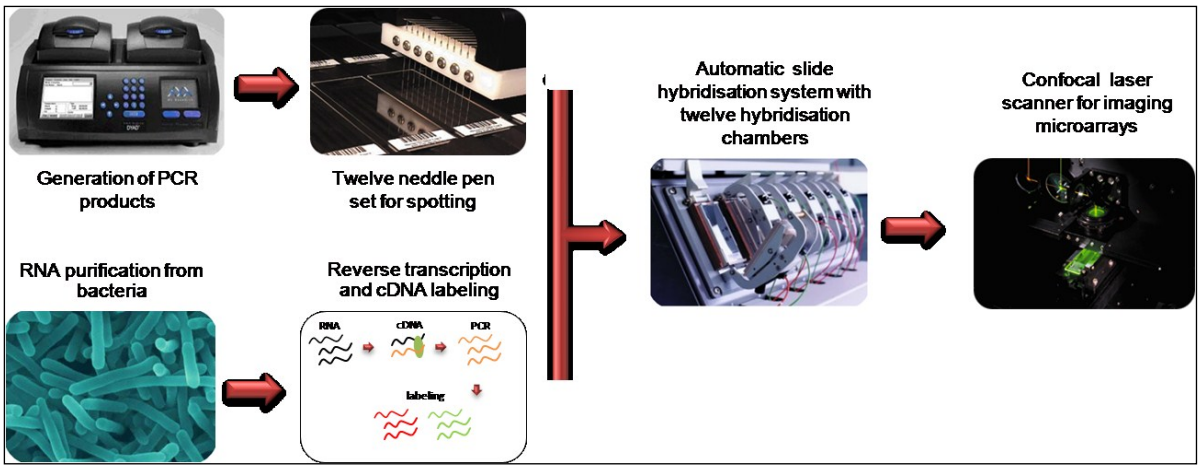


Fig.6 In-house microarray workflow; wet-lab

Since each technology has its own peculiarities and to obtain optimal results, we created our own analysis workflow called **MARS II** (**M**icroarray **A**nalysis, **R**etrieval, and **S**torage system) as well. Its backbone was adapted from another free software **MARS** from TU Graz[15].

Briefly, after measuring spot intensities, experimental data will be normalized using a quantile algorithm. Aligned data is then channelled thought a quality control (QC) module which first removes all genes with a high number of missing values. Secondly remaining missing values will be imputed using sequential K-nearest neighbour (SKNN) imputation following an additional normalization. Remaining genes are analysed using wrapped significance analysis of microarrays (SAM) software. Raw data, normalized data, and results are finally stored in MARSII database for further analysis. The MARSII pipeline enables the user to analyse, store, reanalyse with different parameters, view and export expression data to e.g. ArrayExpress Database to obtain an accession number for publication.

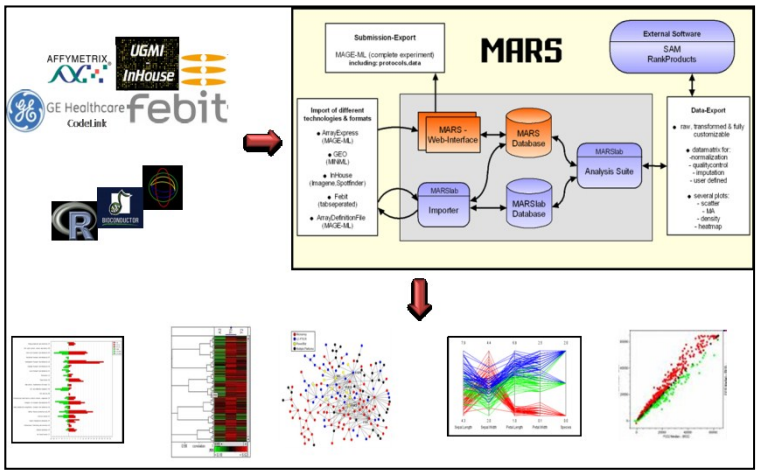


Fig.7 In-house microarray workflow; bioinformatics

Tiling array workflow

Presently microarray analysis only allows the examination of gene expression of predicted ORFs from known genome sequences. Each individual ORF is covered by two to three oligonucleotides placed at the 5'- and 3'-end of the predicted gene. For an average genome of 3 million nt we can estimate a total of three thousand genes and hence a total of 6 to 9 thousand oligonucleotides for a standardized microarray. Such a setup disregards the half-lives of mRNA and their degradation from the 3'-end and also does not consider RNA that initiate at intergenic regions.

To overcome this limitation tiling arrays have been introduced that cover the sequence of the entire genome. This leads to a quantitative increase in the number of oligonucleotides required and leads to a higher complexity of oligonucleotide sequences thereby increasing problems of e.g. cross-hybridization. For a standard genome using tiling arrays it can be estimated that more than one million oligonucleotides are needed to completely cover the genome. This requires specialized formats for in-situ synthesis of oligonucleotides and chip formats that allow the representation of this number of features. Therefore the costs for producing such arrays can be highly prohibitive.

We collaborated with a commercial provider of tiling arrays (Febit) in order to produce tiling array for dedicated analysis of sRNAs for the major gram-positive pathogens. Because the costs of whole genome tiling arrays were prohibitive, all arrays were designed to only cover the intergenic regions based on the assumption that sRNAs are generally produced within intergenic regions. For each genome this meant a total of ~30.000 oligonucleotides per chip.

Five array designs were calculated to cover the whole intergenic regions of the gram-positive pathogens investigated within the consortium *sncRNAomics* depicted.

	L. monocytogenes 1/2a EGD-e	S. pyogenes NZ131	S. aureus COL	E. faecalis V583	C. difficile 630
probe size, overlap	50, 30	50, 15	50, 15	50, 15	50, 23
chromosome size	2944528 bp	1815785 bp	2809422 bp	3218031 bp	4290252 bp
plasmidssizecombined	0	0	4400 bp (1)	141943 bp (3)	7881 bp (1)
non-controlprobes total	25504	17823	29198	29413	58561
control probes total	2105	174	111	254	211
arrays necessary	2 (1.978)	2 (1.2)	2 (1.961)	2 (1.995)	4 (3.96)

Table 2 Investigated bacteria within *sncRNAomics*

The increased complexity of tiling arrays and therefore the shift to another vendor with all its required modifications concerning format and analysis, necessitated on extension of the microarray pipeline described above.

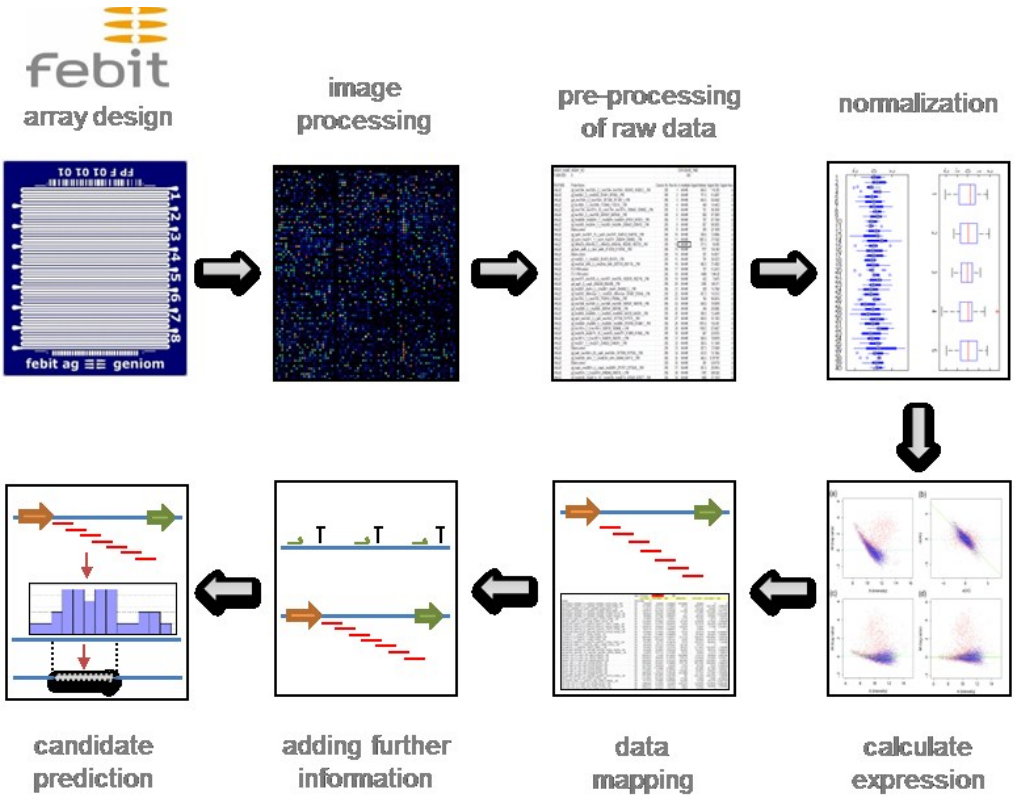


Fig.8 Flow chart of modified microarray workflow customized for tiling array experiments

Following modifications and extensions were accomplished: a) processing of hybridization images to increase quality b) pre-processing of raw data derived from scanning e.g. background correction and reassignment of identifiers etc. c) normalization of data by standard quantile normalization d) statistical analysis to determine expression values e) mapping of data onto the intergenic genome of the respective bacteria f) adding additional information such as promoters and terminators to enhance putative candidate identification g) prediction of putative candidates based on suitable inducted fragments.

Results mainly consist of formatted lists with putative sncRNA loci which have to be verified manually. Additional information such as promoters, terminators, external software predictions, integration of published sRNAs etc. supports identification.

RNA sequencing workflow

The advent of whole genome sequencing technologies opened up new opportunities for detailed examination of the transcriptome by direct sequencing of total RNA. Within a single run using e.g. current Illumina HiSeq sequencing technology, it is possible to obtain 600 Gb per run. This allows transcript levels to be determined from one to several million copies depending on the gene studied, by simply counting the extent of coverage per base per gene depending on the condition studied. In addition to quantitative analysis, qualitative data regarding the exact length of the transcript, sequence, and start and stop site can be described in detail.

The massive amount of data obtained per experiment made it imperative to develop new strategies to handle and analyse data. For example sizes of primary data files reach orders of magnitude of several hundreds of GBs which represents a limitation in working with normal desktop computers. Therefore it was necessary to implement new hard- and software structures to cope with this new dimension of data deluge. The implemented hardware environment was focused on necessities of NGS technologies considering main memory and storage capacity.

The current setup consists of approximately 100 CPUs, 300 GB of main memory and around 35 TB of disk space organised in different nodes which in turn where optimized for different types of tasks. The developed hardware structure in this thesis will be continually extended in the near future.

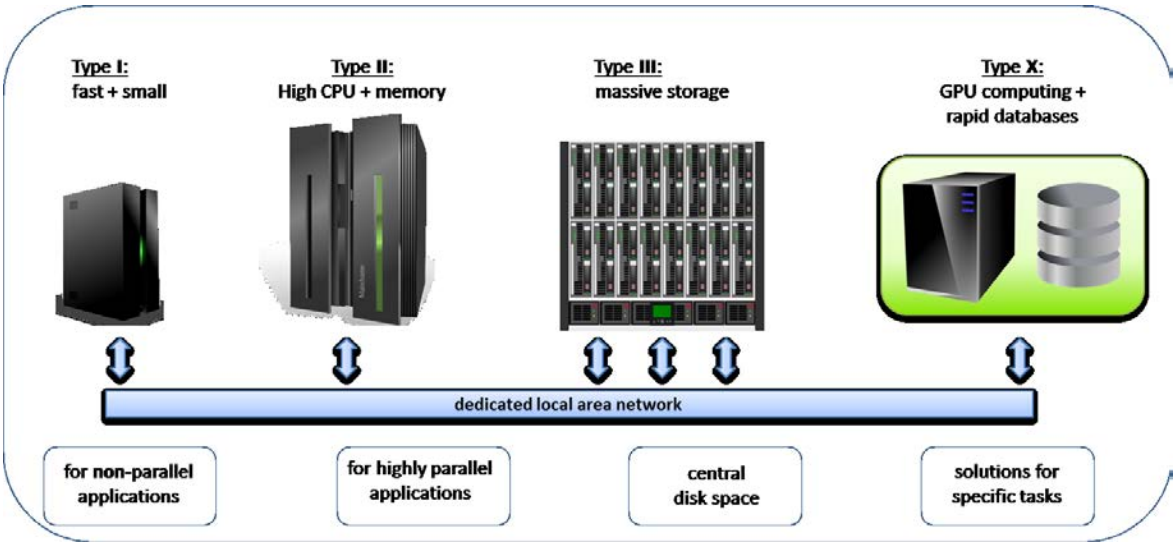


Fig.9 Hardware environment

In general three different classes of nodes are available (i) nodes with only a few but very fast CPUs, used for non-parallelized jobs, (ii) nodes with a large quantity of slower CPUs for parallelized tasks, and (iii) nodes constructed to store plenty of data, and finally class (x) which comprises single nodes which were furthermore optimized for e.g. very fast database queries or computing with graphic cards (GPU computing).

Alongside the development of new hardware structures, a completely new software pipeline had to be created. Since possible requirements and bottlenecks were not fully known in the beginning of this project, a constant re-evaluation and re-adjustment was necessary to increase the quality of the workflow.

Finally a multistep procedure to process primary data to figures for publication was implemented. Initially, primary data has to be retrieved from the current sequencing machine. In order to establish a software with a focus on flexibility, the first quality controls are rather basic. During this step low quality parts of the sequence, independent of type of technology, will be clipped based on standard FASTQ quality values (Fig.10 A+B) to reduce data sizes for the second step. The trimmed and clipped data will be loaded into a special database optimized in terms of write access speed (Fig 10 C). This step is one of the most time consuming, but guaranties maximum flexibility in the further analysis. To ensure reliable mappings, an export module creates several output files which can be used as input for different mapping algorithms. Currently FASTQ, SFF and standard FASTA formats are available (Fig. 10 D).

Afterwards read data is mapped with segemehl and / or the bowtie algorithm (Fig. 10 E). Both currently implemented mapping tools have their own pros and cons and their usage has to be assessed on a case by case basis. Data export, mapping and re-import of mapping results was implemented such that additional algorithms and result formats can be quickly included into the workflow (Fig. 10 D+E+F).

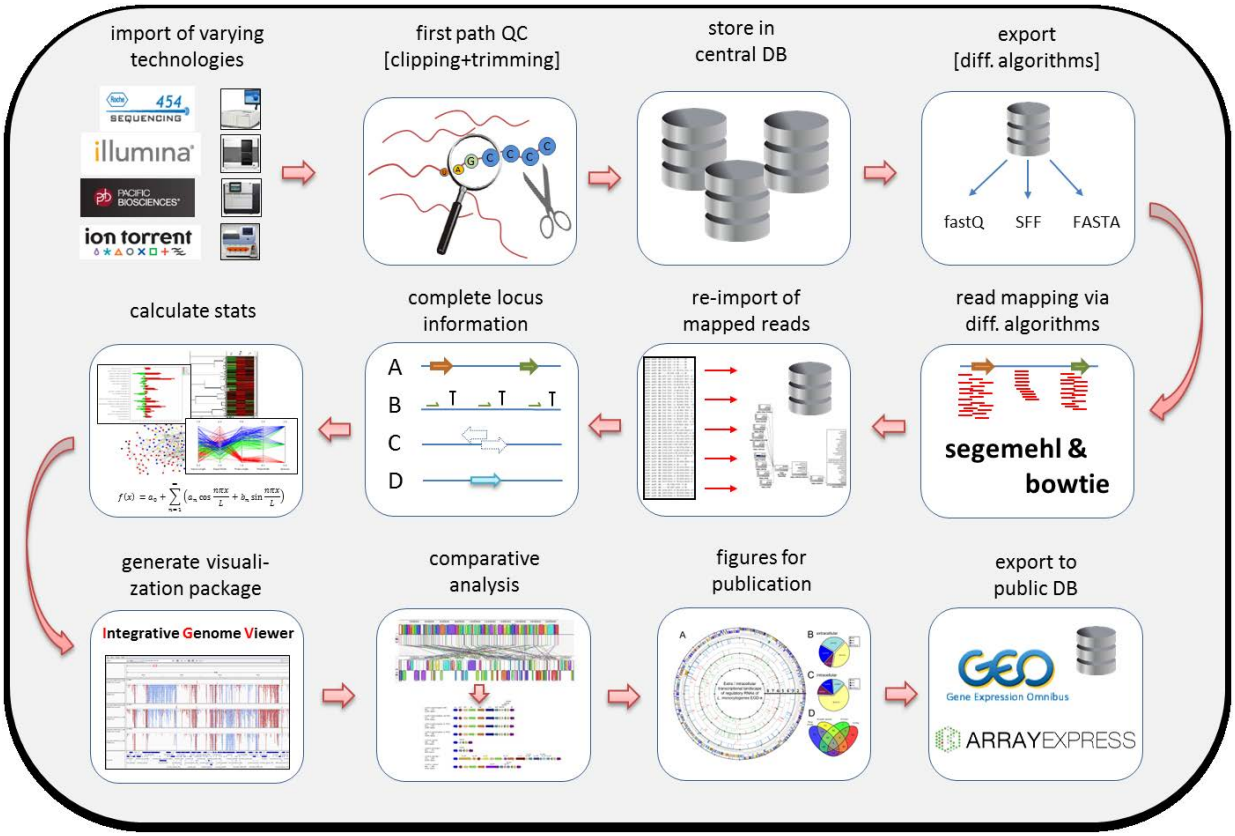


Fig. 10 RNA-Seq workflow

After re-importing positions, quality and several additional parameters for each mapped read, a complementation step is performed. Id est, reference genome information like CDS, RBS and terminators etc. will be extracted from provided input files (gbk, embl or gff3). Additional predictions for terminators via TransTermHP [16] and a promoter binding site prediction with the major known promoter motifs will be performed. Furthermore the output of several prediction tools such as sipht [17] or sRNAscanner [18] is also implemented in the database. Finally, already verified sRNAs from e.g. publications or databases will be extracted and incorporated (Fig. 10 G). All this additional information will be merged with the experimental data to guide prediction algorithms as well as the user to support understanding of particular circumstances.

Based on this additional information, sncRAS first calculates statistics about the analysed data set (Fig.10 H). Briefly, coverage of e.g. tRNAs, published sRNAs, CDS or known

features can be obtained to enable statements about quality and convenience of the executed experiment.

The next module (Fig. 10 I) will provide the possibility to interactively work with the whole transcriptome data. As a lot of expression constellations within a transcriptome experiment are yet unknown or cannot be rated by current software programs, manual inspection is of great importance. Therefore sncRAS produces all needed data files to use the integrated genomics viewer (IGV) from Broad Institute [19] for more detailed analysis.

Fig.11 Screenshot of the software IGV from Broad

IGV additionally allows a detailed comparative manual analysis of multiple transcriptomes of an organism differing in e.g. experimental conditions. This step also allows the comparison of the expression of different experiments to identify putative hot spots where e.g. an sRNA candidate shows highly differential expression between conditions (Fig. 10 J). sncRAS furthermore exports lists and figures in a publication ready format. Currently different types of Venn diagrams and pie charts can be calculated and list exports for circular representations (e.g. via GenomeViz) are available (Fig. 10 K). Finally, sncRAS offers the opportunity to export data in a particular format which can be submitted to a public database such as GEO or ArrayExpress to obtain an accession number for publication (Fig. 10 L).

The established workflow was used for the analysis of RNA-Seq data from four additional gram-positive pathogenic species within the sncRNAomis project.

Additional Tools

To improve analysis with regard to quality, speed and compatibility, several tools were created to simplify standard tasks as well as further analysis.

BlastIt

Due to the large amounts of data required for mapping and comparative analysis a small software tool called *BlastIt* was developed. *BlastIt* allows flexible analysis of sequences using the NCBI Blast Suite. The major benefit of this tool is that it enables the user to implement new parameters and variables to improve the value of output data.

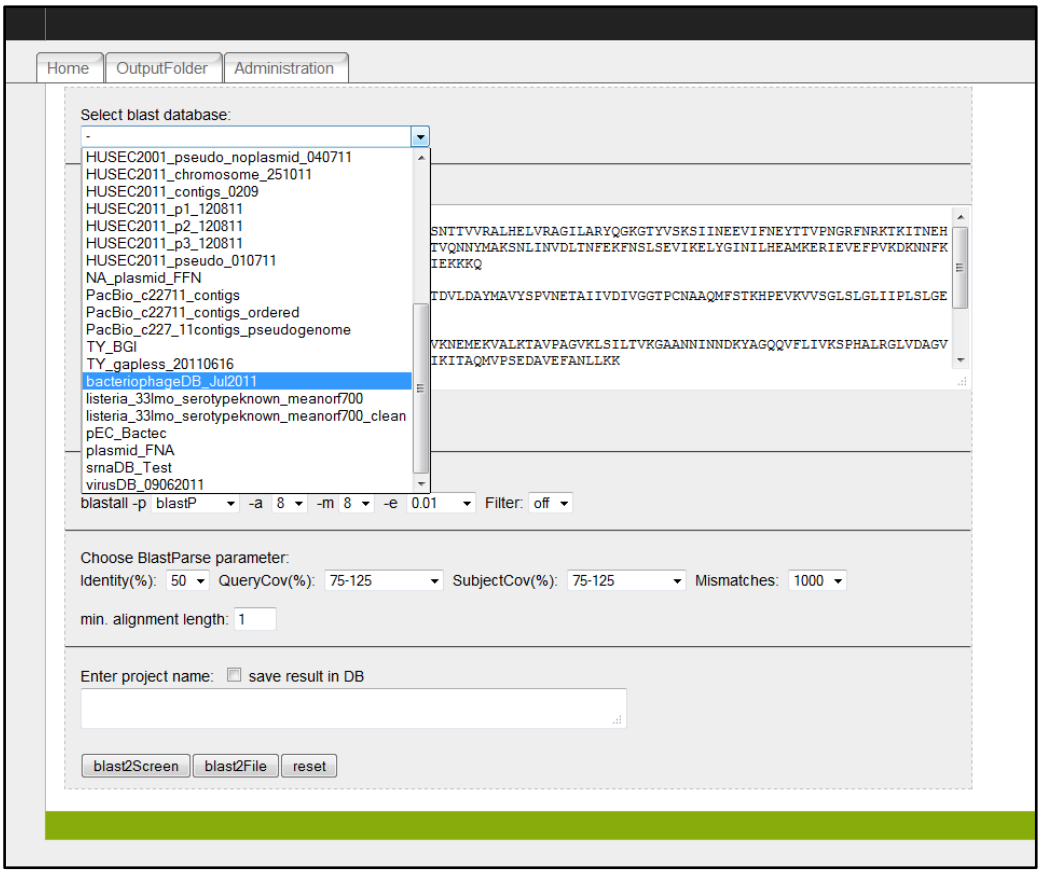


Fig. 12 Main interface of BlastIt

Results can be obtained in tabular format and additionally stored in *BlastIt* database for further investigation.

Nr.	queryID	subjectID	annotation	subject start	subject end	query start	query end	alignment len.	query len.	subject len.	mismatches	identity	query coverage	subject coverage
1.	sat	FFM_Contig_13		4646	2889	1	1758	1758	1758	121738	19	98.92	100	1.44408483793
2.	pic_Set1A_Set1B_true	FFM_Contig_75		5126	353	865	5639	4775	5639	5126	17	99.62	84.678134421	93.1525555989
3.	aafD	FFM_Contig_4		11631	11060	82	653	577	726	27899	155	71.40	79.476584022	2.06817448654
4.	aggR	FFM_Contig_78		1787	2584	1	798	798	798	3831	1	99.87	100	20.8300704777
5.	agg3C	FFM_Contig_4		10892	8431	88	2546	2481	2547	27899	649	72.19	97.4087161366	8.89279185634
6.	eatA_true	FFM_Contig_7A		2310	3867	143	1696	1566	4095	12237	488	67.56	38.2417582418	12.797254229
7.	sigA_plus_hypo_true	FFM_Contig_13		4646	789	1	3858	3858	3858	121738	0	100.00	100	3.16910085594

Table 3 Tabular output of BlastIt

KeggMapper

In order to map expression data onto individual pathways of metabolism, signal transduction and other cellular processes within any organism, a tool was created to project the transcriptome onto Kegg pathways [20]. KeggMapper allows the user to take any set of data and convert it to a format which can automatically linked to pathways available in database. The user simply enters gene identifier and fold changes or expression values and the software creates an overview based on all pathways in metabolism based on KEGG-ATLAS.

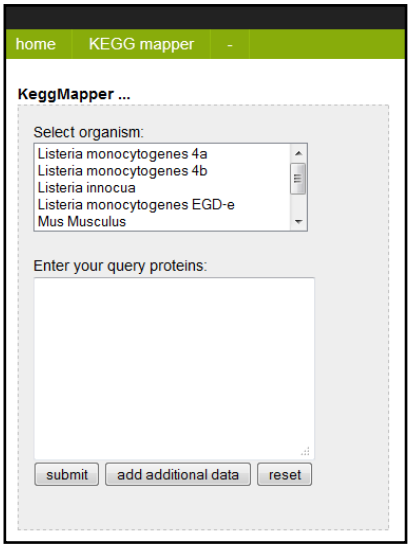


Fig. 13 General overview map composed of all classified genes mapped onto all metabolic pathways.

Fig. 15 Standard KEGG pathway visualization, which e.g. shows in detail all involved enzymes and metabolites (not shown)

Fig. 14 New schematic representation of more complex structures such as receptors, pores and proteins.

Fig. 13 Main input screen

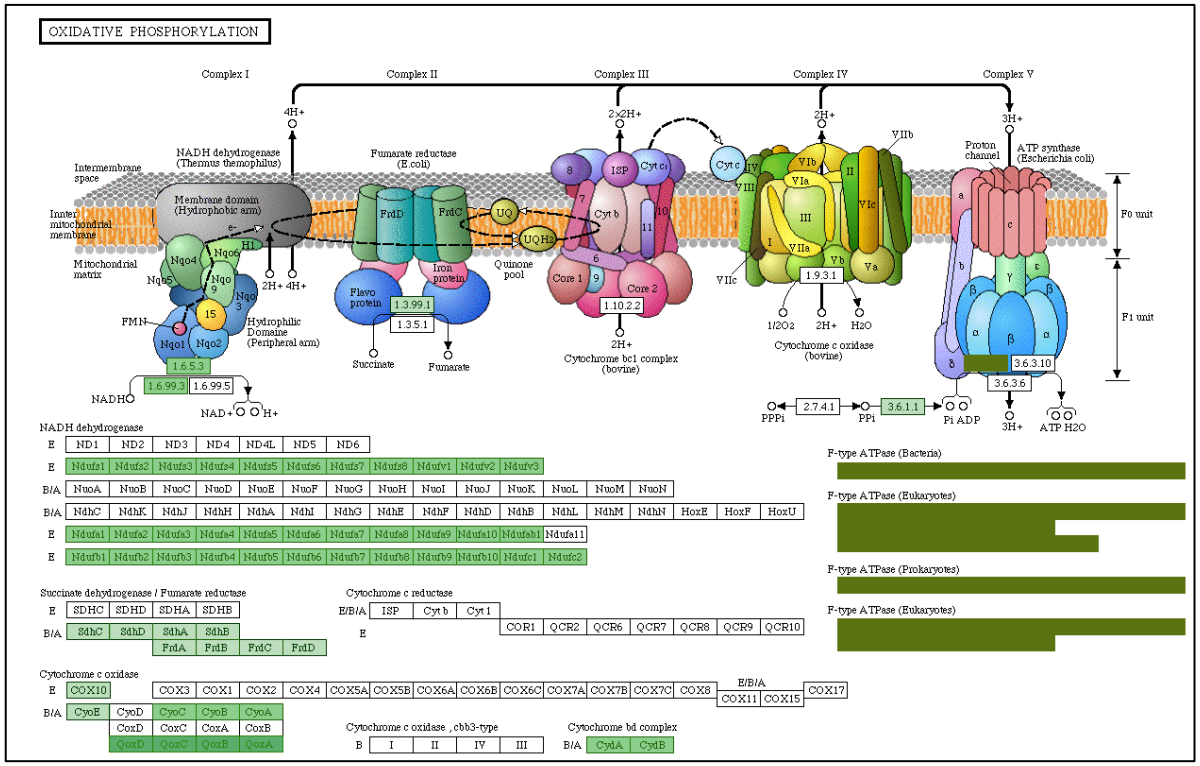


Fig. 14 KeggMapper; complex structure representation

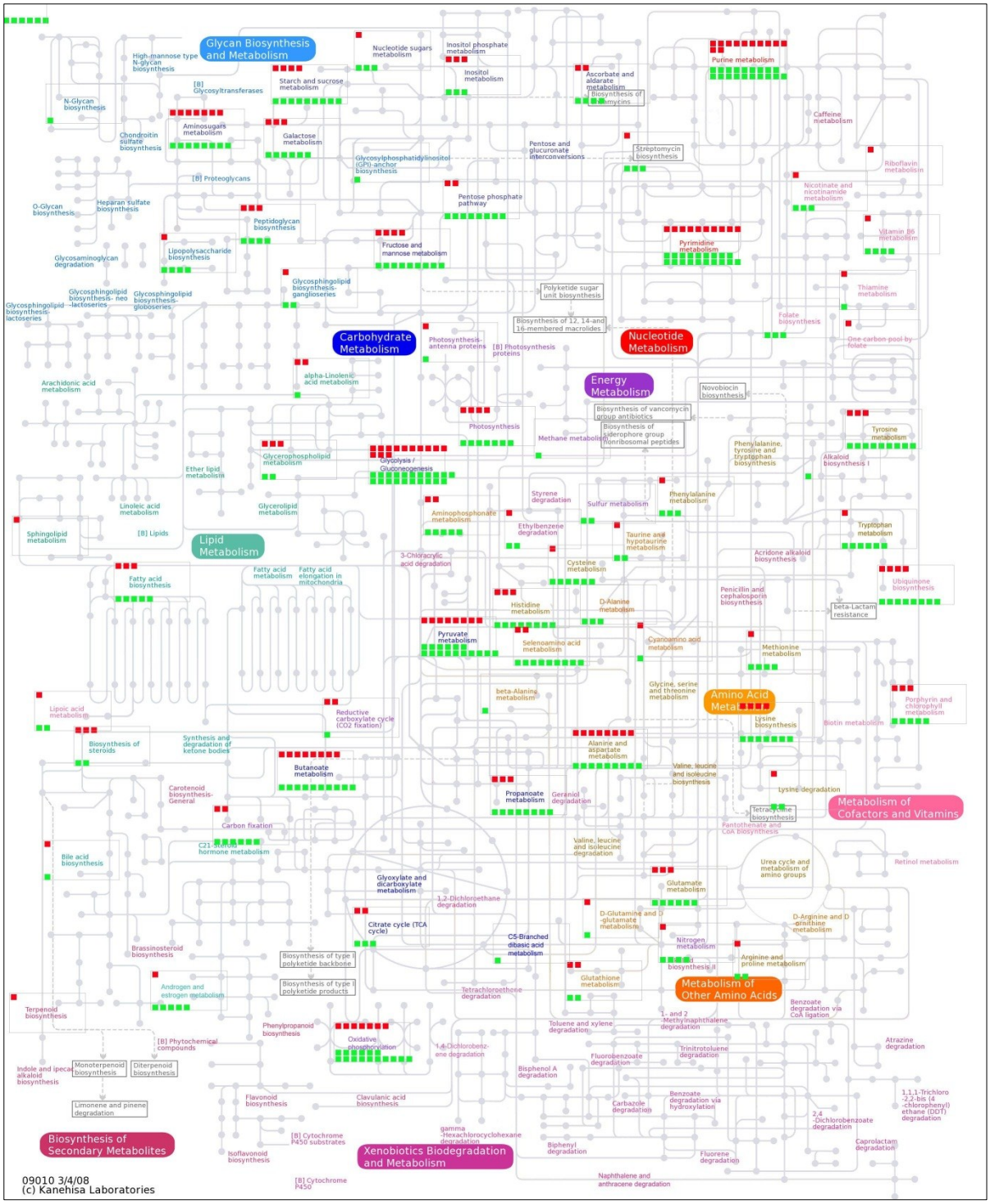


Fig.15 KeggMapper main overview

2 Chapter I

Comparative genomics and transcriptomics of lineages I, II, and III strains of *Listeria monocytogenes*

Comparative genomics and transcriptomics of lineages I, II, and III strains of *Listeria monocytogenes*

Torsten Hain*, Rohit Ghai*, André Billion*, Carsten Tobias Kuenne*, Christiane Steinweg, Benjamin Izar, Walid Mohamed, Mobarak Abu Mraheil, Eugen Domann, Silke Schaffrath, Uwe Kärst, Alexander Goesmann, Sebastian Oehm, Alfred Pühler, Rainer Merkl, Sonja Vorwerk, Philippe Glaser, Patricia Garrido, Christophe Rusniok, Carmen Buchrieser, Werner Goebel and Trinad Chakraborty.

*BMC Genomics 2012, **13**:144 doi:10.1186/1471-2164-13-144; Published: 24 April 2012*

2.1 Contribution

The author (A.B.) jointly conceived and assisted in writing the manuscript. He pre-processed, analysed and carried out statistics of transcriptome data in collaboration with R.G.. Microarray data submission to ArrayExpress was done by MARSII workflow, conducted by the author. Furthermore collection, comparative analysis and visualisation of ActA proteins as well as analysis of duplicated genes especially of PTS-Systems were done by the author. Additionally, collaborative evaluation and interpretation of SNPs and flagellin analysis as well as development of corresponding figures was carried out by A.B. Surface protein detection, comparative analysis and visualization was done with the AUGUR software, performed by A.B. which as well is responsible for CRISPR and CAS detection, identification, analysis, interpretation and visualization. The author contributed to the annotation of *L. monocytogenes* 4a L99.

2.2 Abstract and Introduction

Here we report the first comparative analysis of *Listeria* linage I, II, and III based on genomic sequences as well as transcriptomic evaluation. Comprehensive genome analysis showed extensive evidence of virulence gene decay, including loss of important surface

proteins. Linage I strains feature different compositions of invasion genes but share the lack of prophage genes. Comparative transcriptome analysis revealed distinct expression levels, especially in the ability of downregulation of flagellar genes. Furthermore, differences were found in metabolic flux control and phosphorylated sugar uptake. Deletion mutans of the *lma* operon showed severe attenuation of virulence which, combined with the strong induction of prophage genes, revealed the yet unknown importance of phage genes for intracellular survival of *Listeria monocytogenes*.

2.3 Material and Methods

Genome sequencing and annotation

Genome sequencing of *Listeria monocytogenes* L99 4a was performed on ABI PRISIM 3100 and 3730xl Genetic Analyser (Applied Biosystems). After assembly with the Phred/Phrap/Consed software [21, 22] whole genome shotgun sequencing revealed 27.637 sequences. Additionally 1684 fosmids and 671 gap closure PCRs were required to assemble a closed genome with about 6.7-fold coverage. Annotation was carried out by GenDB and manual inspection [23].

Sequencing of strain *L. monocytogenes* 4b CLIP80459 was conducted by colleagues of Institute Pasteur. Briefly, a library of 2-3 kb inserts were generated and cloned into pcDNA-2.1 vector. After sequencing on 3700 and 3730 DNA sequencers the assembly of 35.610 reads with Phred/Phrap/Consed software resulted in 361 contigs which were connected with 379 PCRs to a circular genome with approximately 7.8-fold coverage. Genome annotation was carried out as previously described [24].

ActA repeat analysis

All available ActA protein sequences of *Listeria monocytogenes* were collected and only sequences with at least 500 AA were used for further analysis. After filtering of duplicates etc., 218 sequences were aligned using ClustalW and manually examined.

Single nucleotide polymorphisms

Single nucleotide polymorphism (SNP) detection was done by the MUMmer software [25]. Resulting SNPs were mapped against CDS regions following a calculation of SNP-density and visualization in GenomeViz [26] (SNP-density per gene results from SNPs per gene, normalized by gene length).

CRISP repeat analysis

CRISPR repeats and spacers were identified by PILER-CR [27] and CRT [28] software. Spacer classification and visualization was performed using BLAST and ACT [29]. Comparative analysis and visualization of CRISPR loci including CAS genes were carried out by manual inspection and GeCo [30].

Microarrays

A whole genome microarray approach for all four strains was performed using Febit's Geniom One platform. Background correction of the data was based on several controls and normalization was done using quantile algorithm [31]. To assess reproducibility of at least two technical and three biological replicates we calculated Pearson's correlation coefficients (r2 >= 0.94) for all experiments. The statistical significance analysis was carried out by SAM software [32].

Nucleotide sequence and microarray accession number

The microarray data have been submitted via MARS2 software [unpublished] to ArrayExpress database with accession number E-MEXP-1947.

2.4 Results

General features of complete genomes of the three lineages of L. monocytogenes.

The general features of sequenced genomes are given in table four. Genomes of *L. monocytogenes* are highly synthentic in relation to genome size, GC content, composition of tRNA and rRNA sites, and characteristics of protein coding regions in considering length percentage coding.

Differences could be observed between 4a and 4b CLIP genome regarding lack of mobile genetic elements and plasmids. Furthermore we could identify four different prophage regions within 4a, whereas only one is present in the 4b CLIP genome. Interestingly, prophage loci I and III were found adjacent to tRNAs which represent genetic "anchoring elements" for the uptake of listerial prophage DNA [23]. Deeper analysis of prophage locus III revealed that in addition *L. welshimeri* as well as *L. ivanovii*, which both are apathogenic strains, harbour prophage or pathogenicity islands at this chromosomal location.

We could detect noticeably more pseudogenes within both 4b strains (42 for F2365 and 26 for CLIP80459) compared to EGD-e (9), 4a (1), and *L. innocua* (13).

	L. monocytogenes **4a L99**	*L. monocytogenes* **4b CLIP80459**	*L. monocytogenes* **4b F2365**	*L. monocytogenes* **1/2a EGD-e**	*L. innocua* **6a CLIP11262**
Size of chromosome [bp]	2979198	2912690	2905187	2944528	3011208
G+C content [%]	38.2	38.1	38.0	38.0	37.4
G+C content of protein-coding genes [%]	38.7	38.5	38.5	38.4	37.8
Protein-coding genes (pseudogenes)	2925 (1)	2790 (24)	2821 (26)	2855 (9)	2981 (13)
Average length of protein-coding genes [aa]	301	311	303	306	300
Number of rRNA operons (16S-23S-5S)	6	6	6	6	6
Number of tRNA genes	67	67	67	67	66
Percentage coding	88.9	89.4	88.4	89.2	89.2
Number of prophages (genes)	4 (191)	1 (16)	1 (16)	2 (79)	6 (322)
Plasmid	0	0	0	0	1
Number of strain-specific genes*	111	49	105	120	89
Number of orthologous genes*	2623	2725	2699	2656	2570
Number of transposons	0	0	0	1	0

*Prophage genes excepted.

Table 4 General features of sequenced genomes in this study

Comparison of the two 4b genomes revealed 115 specific genes for strain CLIP80459 which encode functions for regulation, sugar metabolism, and surface characteristics. Specific genes (146) for F2365 show similar functionalities, however most of these are annotated as hypothetical.

Comparing genomes of different lineages on nucleotide level, surface proteins show the highest SNP number of all CDS classes. No recognizable difference in surface SNPs could be found when comparing two closely related genomes or two with a larger evolutionary divergence thus SNPs of *Listeria* show a considerable evolutionary adaption among surface-associated genes.

Comparison of the virulence gene cluster of linage I, II, and III

A sequence alignment of the VGC of the four studied strains showed a homogeneous pattern. Closer analysis revealed a truncation of the *actA* sequence in 4b and 4a genomes. Additionally, a small truncation upstream of the *mpl* gene and a truncation of a short repeat region distal to the PrfA box of *mpl* was observed in the 4a cluster but shows not adverse effect for the PrfA binding site. The most divergent gene within the VGC is a lipoprotein (*lmo0207*), however it is unknown if this plays a role in virulence attenuation in 4a L99.
Interestingly, the 4a and both 4b strains display an identical deletion of *actA* repeats, which results in lower speed of movement of intracellular bacteria [33].

Loss of surface proteins in linage III

Several genes encoding internalins or internalin-like proteins are absent in the *L. monocytogenes* 4a L99 genome including the entire *inlGHE* cluster, InlJ as a PrfA-independent

internalin, InlC which is involved in cell-to-cell spread, and internalin F. Additionally, internalin-like genes such as homolog to lmo1666, lmo2470, and lmo2821 are absent in the 4a L99 genome. We also analysed the surface repertoire of three additional 4a genomes and confirmed the lack of a similar number of proteins. To validate the findings we performed further PCR analysis to confirm the absence of surface associated proteins.

Comparisons of the two 4b genomes with *L. monocytogenes* EGD-e as reference showed no noticeable differences in surface protein composition.

In contrast, 4a L99 revealed differences in non-internalin cell-wall associated proteins like the GW-motif containing *auto*, which is a PrfA-independent autolysin responsible for the entry into non-phagocytic eukaryotic cells [34]. Furthermore the PrfA-dependent *vip* gene which is involved in the late stage of infection [35] is also absent in 4a L99. Other variations could be found in truncation of e.g. InlI.

Surprisingly, comparative analysis of remaining classes of surface proteins like lipoproteins, LysM and NLPC/P60-motif containing proteins showed minor differences between the four genomes.

Altogether 4a L99 exhibits that loss of a number of crucial functionalities of *listerial* invasion contributes to the decreased invasion ability and the attenuated nature of this strain.

Decay of phage genes in the L. monocytogenes L99 4a strain

The two prophage loci of 1/2a EGD-e contain 79 genes, whereas L99 4a exhibits 193 genes organized in four different loci. The two 4b genomes feature only one locus with about 16 genes designated as monocin. This monocin locus is mostly conserved in all *L. monocytogenes* strains except for 4a, which only have one-third of the genes left. Interestingly, the monocin locus of the apathogenic *L. welshimeri* represents the most similar structure when compared to linage III genomes.

Gene duplications in Listeria genomes expand metabolic systems

Gene duplications in *Listeria* seem to be an ancient event, because most of the duplications (ranging from 231 to 296 in numbers) are conserved across all species. Functional classification of duplicated genes revealed strong evidence for the importance of metabolic pathways, since the majority could be identified to be involved in e.g. pentose phosphate pathway, fructose and mannose metabolism, carbon fixation, glycolysis and pyruvate metabolism.

However, gene duplication events can also occur from HGT-events instead of e.g. an error in homologous recombination, because some PTS-system genes are specific to1/2a EGD-e.

The CRISPR system of Listeria

The L99 4a genome contains two regions with CRISPR-like structures, while only locus II comprises required CAS genes. Both exhibit spacers of 35bp linked with repeats of 29bp. While locus I is highly conserved across several lineages and strains, locus II is only present in 4a genomes and additionally exhibits more than 5 times more spacers than locus I (~6 for locus I and 27 for locus II).

Spacer identification of L99 4a revealed a 100% hit against the PSA-prophage, which is known to infect serotype 4 strains, suggesting resistance to PSA. Future studies have to elaborate whether the 4a CRISPR system is functional or not.

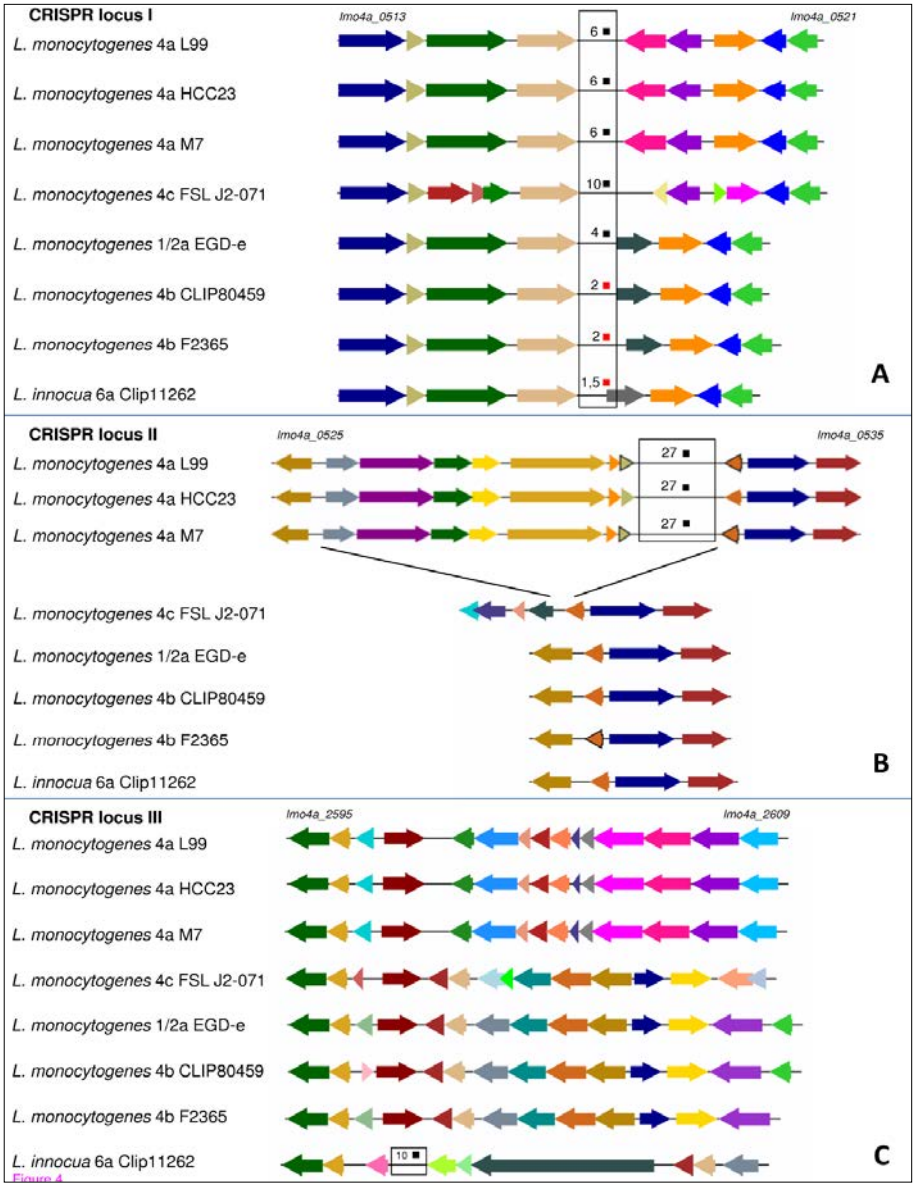


Fig.16 CRISPR loci in *Listeria monocytogenes*

Comparative intracellular transcriptomics of four L. monocytogenes strains of the three major linages

Major results found by comparative transcriptome analysis of two lineages revealed differences in virulence, cell wall, and stress response [36]. We performed a whole genome microarray experiment to investigate differences of intracellular gene expression of strains related to the three major lineages.

The intracellular core response is composed of upregulation of the entire virulence gene cluster which was highly induced within infected host cells. Furthermore, known genes (*hpt*, *clpE*, *bilEA*) important for survival are upregulated in all strains. Interestingly, 11 PTS systems of five different sugar transporters are upregulated, suggesting these sugars as the most frequent substrates in the cytosol. Unexpectedly, all strains showed strong induction of the *lma* operon and their surrounding phage genes including a conserved holin. The most conserved locus is the *lma* operon (excepting in L99 4a), whose gene products provoke a delayed type hypersensitivity reaction in hosts immune to *L. monocytogenes*. In summary, it can be stated that the common intracellular prophage gene expression in all *Listeria* lineages is the most striking observation, however functions of several genes are still unknown. Additionally, induction of the *eut* operon was observed, suggesting the use of ethanolamine as a carbon and nitrogen source. The common upregulation of Zinc transporters as well as the spermidine/putrescine ABC transporters indicate a role in intracellular survival. Another conspicuous gene induction was observed considering the non-oxidative branch of the pentose phosphate pathway. It's upregulation is important for several functionalities, such as the production of NADPH for countering oxidative stress, generation of erythrose-4-phosphat for aromatic AA biosynthesis, and also for the generation of pentose sugars. Correspondingly, several genes of the pyrimidine and purine biosynthesis, using pentose sugars, are downregulated.

Downregulation was observed comprising the *agr* locus, tryptophan biosynthesis, and some tRNA synthetase genes. Also decreased expression of the cytochrome genes caused less energy production.

Differences in flagellin expression are the most prominent differences among strains

When only comparing both 4b strains, and because 4b CLIP grows more efficiently intracellularly, most important differences were found regarding regulation of flagellar genes. While 4b F2365 shows induction of 10 flagellar genes, 4b CLIP only upregulates *fliR*. Additionally 4a L99 also exhibits a strong intracellular induction of nearly all flagellar genes

(21 genes) in contrast to 1/2a EGD-e and both 4b strains. This is counterproductive, because flagellar expression enables the host to detect the bacterial presence supporting the formation of an inflammasome.

Differential growth of the three lineages and ΔlmaB and ΔlmaD isogenic mutants in a mouse and cell infection model

Due to the lack of several internalin genes and intensive upregulation of flagellar genes, we observed severe deficiency of 4a L99 in entry of HeLa and CaCo-2 cells as well as fragile cell-to-cell transmission which finally leads to a rapid clearance in *in vivo* experiments. Additionally, the upregulation of several DNA repair genes suggests heavy genomic damage during intracellular growth.

Mouse infection experiments revealed the complete clearance of 4a L99 after five days post infection, which is in contrast to the 4b CLIP strain that shows even stronger growth than 1/2a and 4b F2365 genomes. Isogenic mutants (*ΔlmaB, ΔlmaD*) have shown less growth than the wild type of 1/2a EGD-e.

2.5 Discussion

This study attempted to unravel required abilities for intracellular survival of *Listeria*. Therefore we sequenced new genomes, compared their gene content, and produced intracellular expression profiles of the three major lineages of *L. monocytogenes* to disclose the different adaptation strategies for survival in the host.

The 4a L99 genome showed loss of important functionalities for entry and survival inside the host. A large number internalins, internalin-like, and surface proteins were deleted. Additionally, a truncation within the ActA gene, which could be also observed in many other lineage I strains, tends to result in a lowered motility in the cytosol. Furthermore half of the VGC is missing compared to other lineages which has a severe effect for the survival in macrophages. Analysis of the induction of the flagellar apparatus showed the inability of all strains (except for 4b CLIP) to downregulate most of the genes important for evading host detection and thus intracellular survival, because cell wall components and flagellin may be sensed by TLRs and NCRs. Deletion mutants of the f*laA* gene showed increased intracellular survival, which confirmed transcriptional findings. While all strains in this study were able to induce genes of the VGC, we have to assume that there are different

strategies in genes encoding proteins involved in carbohydrate transport, surface- and membrane proteins for intracellular adaptation.

Gene duplication

Gene duplication including HGT and gene loss may influence the evolution of bacteria in a very strong manner. In the past, it could be shown that gene duplication improved several functionalities such as temperature adaption, toxin exposure, antibiotic treatment, and even parasitic and symbiotic lifestyle. Classification of duplicated genes revealed that systems which are involved in energy generation, modulation, and transfer are extremely important for adaption to hostile conditions like intracellular growth. Furthermore we could observe that marginal differences in number of PTS systems may have dramatic effects on survival of listeria inside the host. Transposon insertion mutants showed that presence/absence of a single pentitol PTS system significantly influenced growth in epithelial cells [37]. Ethanolamine may be used by intracellular listeria as an additional carbon source, because homologous genes of this locus from *Salmonella* are intracellularly upregulated. While this locus is not specific for pathogenic *Listeria*, it may indicate a more general feature to compete for nutrients in particular situations. Another important aspect is the ability of adapting e.g. HGT genes in considering regulation and expression. It could be shown that 1/2 EGD-e is quicker and more successful in recruiting and adapting HGT genes than all other strains compared in this study. This feature may contribute to the better survival rate of intracellular 1/2 EGD-e.

Beside improvements in carbohydrate uptake and processing, it has been shown [38] that rerouting the carbohydrate flux from glycolysis to pentose phosphate pathway is a common mechanism in counteracting oxidative stress. Interestingly, 4b F2365 was the only one with the inability to reroute these pathways, which certainly contributes to the poor intracellular survival of this strain.

The CRISPR system in Listeria

Today, CRISPR systems are described for a wide range of bacteria and nearly all archaea. In general, CRISPR represents an innate phage-resistance mechanism composed of cas genes followed by an array of specific spacer sequences separated by direct repeats. Among the genomes compared in this study, only one locus associated with several cas genes was identified in the 4a L99 genome. Detailed analysis of the spacer locus revealed a perfect DNA sequence of the prophage PSA which possibly enables resistance to this

phage. The position of the spacer in the array suggests recent acquisition. Interestingly, a small sRNA *rliB* was located within an ancient repeat region of 1/2 EGD-e which is associated with virulence in mice [70].

The role of phage genes in virulence of listeria

Genome analysis revealed different numbers of phage genes within genomes in this study, however a massive induction of gene expression regardless of location and number could be observed across all strains. The only conserved locus in all three lineages is *lma*. *Lma* is induced under intracellular condition and a deletion mutant in 1/2 EGD-e showed reduced intracellular growth. The role of prophage genes have not examined in detail, but seems to be an essential mechanism for short-term evolution [39, 40].

2.6 Conclusion

Comparative genomics and whole-genome based transcriptomic analysis of representatives of all lineages revealed (i) reductive evolution is responsible for attenuated phenotypes (ii) recruitment and adaption of prophage genes and metabolic systems reveal new virulence-associated factors (iii) downregulation of surface proteins may serve to avoid detection by host cell receptors.

3 Chapter II

Adaptation of Listeria monocytogenes to oxidative and nitrosative stress in IFN-γ-activated macrophages

Adaptation of Listeria monocytogenes to oxidative and nitrosative stress in IFN-γ-activated macrophages

Mobarak Abu Mraheil*, André Billion, Walid Mohamed, Deepak Rawool, Torsten Hain, Trinad Chakraborty

International Journal of Medical Microbiology 301 (2011) 547– 555

3.1 Contributions

The author (A.B.) contributed to the results and provided input into the discussion section of the manuscript. Furthermore A.B. carried out all bioinformatics steps such as the design, analysis and publication-ready versions of microarray experiments. Adaptation of existing methods to the specific requirements of experiment as well as the construction of several figures was done by the author.

3.2 Abstract

While phagocytes represent a nearly host defence mechanism, this study investigates the transcriptome of *Listeria monocytogenes* to assess specific genes contributing to the intracellular survival in IFN-γ-activated macrophages. A whole genome-based microarray revealed 21 differentially expressed genes involved in overcoming oxidative and nitrosative stress. Especially two categories of genes with enhanced induction involved in metabolic pathways were identified, *trpE* encoding an anthranilate synthase and genes of the kynurenine pathway.

3.3 Introduction

Listeria monocytogenes is an important pathogen that has been involved in numerous outbreaks [41] and causes severe disease in pregnant women, neonates, the elderly and immu-

nosuppressed individuals [42]. Innate immune response during infection with *L. monocytogenes* is based on activation of macrophages which play a key role in clearance of bacteria. Infected macrophages cause, amongst others, the production of IFN-y which in turn leads to generation of nitric oxide radicals, superoxide radicals, and depletion of tryptophan to counter bacterial infection.

Recent studies [7] demonstrated how the bacteria adapts its intracellular growth by remodelling transcription of metabolic, stress, and virulence genes. However to date, there is no clear indication to the role of INF-y in relation to infections with facultative intracellular bacteria [43-45]. Because *Listeria* represents an established model for evaluation of crucial cellular interactions, we used whole genome based microarrays of *L. monocytogenes* within IFN-γ activated macrophages to gain a better understanding of the mechanisms involved in adaptation to innate immune response.

3.4 Material and Methods

IFN-y activation of macrophages

Activation of P388D1 macrophage cells were carried out 18h before and during infection by incubating a cell monolayer with recombinant murine IFN-y (100U/ml; Sigma-Aldrich).

RNA isolation

Total RNA extraction has been performed as described by Chatterjee. *L. monocytogenes* EGD-e were incubated until optical density of 1.0 at 600nm. RNA isolation was done from 0,5 ml bacterial aliquots pre-incubated with 1 ml RNAprotect (Qiagen) for 5 min. Intracellular RNA was extracted by lysing macrophages and centrifugation over 10 min (8000 x *g*). Final extraction was performed with the RNeasyMini Kit (Qiagen) and quantity was determined by absorbance at 260 and 280 nm. Quality check was done using Agilent 2100 Bioanalyser. After extraction, cDNA was generated and labelled with CyDyes. Eukaryotic RNA from P388D1 macrophages was again extracted using RNeasy Mini Kit.

Transcriptome analysis

The whole genome microarray was performed as previously described by Chatterjee. Two different experimental layouts have been carried out. The first was composed of bacteria grown in BHI versus intracellularly grown bacteria 4h post infection with in non-activated

macrophages, and second bacteria (BHI) versus intracellularly grown bacteria 4h post infection from IFN-y activated macrophages. Each condition was surveyed with three independent biological replicates, then each replicate was hybridized as dye swap. Signal intensities were quantified via Spotfinder software and identification of differentially expressed genes was done using a standardized workflow consisting of (i) a quality control step (outlier detection, filtering, and imputation) (ii) normalization and (iii) expression analysis by SAM software [32]. Final results were filtered by a false discovery rate of < 1% and a fold change of > 2.

3.5 Results and Discussion

IFN-y stimulation of P388D1 macrophages

Macrophage activation is mainly driven by secreted IFN-y of T-lymphocytes and natural killer cells. To investigate adaption of *Listeria monocytogenes* to activated macrophages, two celllines (J774 and P388D1) were analysed concerning basal expression of iNOS (inducible nitric oxide synthetase) which act as a marker for IFN-y activation. It turned out, that cellline P388D1 induced higher levels of IFN-y and was less sensitive in killing LMO during infection. Additional tests found an optimal dose of IFN-y (100 U/ml) for macrophage activation.

To additionally proof activation, another detection system was used. It has been shown that 3'-p-hydroxyphenyl fluorescein (HPF) is oxidized by ROS and NOS [46]. Therefore we compared untreated and IFN-y activated cells concerning production of ROS and NOS by using flow cytometry. It turn out that IFN-y activated cells induce a significant increase of oxygen and nitrogen radicals. Confirmation of increased levels of NOS and ROS was done by addition of catalase and L-NMMA respectively. Oxidative and nitrosative stress was reduced as expected. All these data confirmed that P388D1 cells undergo IFN-y dependent activation.

Intracellular expression profiling

To investigate the transcription profile of LMO regarding IFN-y activated macrophages we performed a whole genome two-color microarray. Following conditions were compared (i) bacteria from exponential phase grown in BHI (ii) IFN-y activated macrophages and (iii) non activated macrophages. Results from each condition consisted of three independent biological replicates. Comparative transcriptomic analysis of condition (i) and (ii) revealed

a total of 524 (363 up and 161 down) differentially regulated genes. To identify the effect of IFN-y activation we compared condition (ii) and (iii). Finally only 21 (6 up and 15 down) genes could be classified as specifically regulated in response to IFN-γ stimulation.

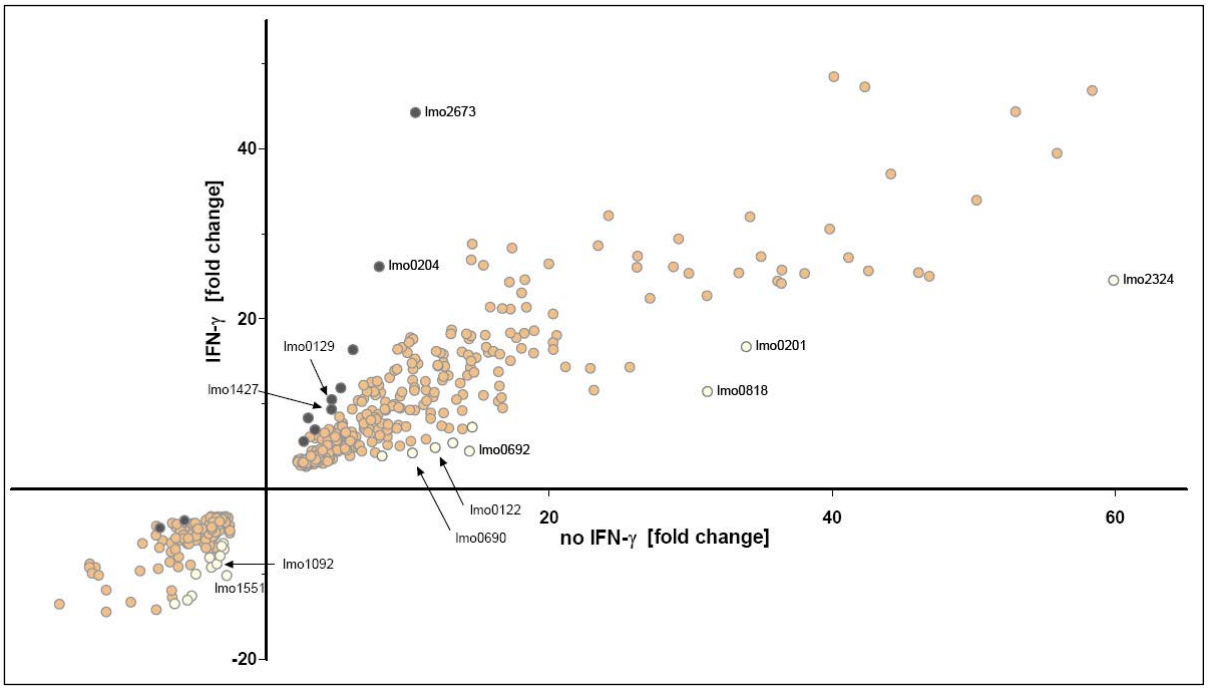


Fig.17 Scatter diagram of intracellular expressed genes

To validate these findings, regulation of 12 (6 up and 6 down) genes were confirmed via real-time PCR analysis.

Evaluation of genes which are involved in adaption to macrophage activation revealed essential requirements for intracellular survival. To overcome oxidative stress distinct functionalities like polyamide associated ABC-Transporters, multidrug efflux pumps, or SAM-methyltransferases could be identified within the induced gene repertoire. Lmo0807 is very similar to an ABC-Transporter which is connected to polyamine transport. Polyamins have been demonstrated to be involved in SOS-response and H_2O_2 protection [47]. A Na^+-driven multidrug efflux pump (lmo0990) which may act as an antiporter for a wide range of drugs (e.g. cationic dyes or amino glycosides) could also be found among the highly induced genes. Lmo1485 encodes a SAM-methyltransferase which supports the countering of oxidative stress induced by ethanol [48]. Investigation of downregulated genes revealed transporters involved in selenite uptake, auroporphyrinogen-III-methyl-transferase involved in hemebiosynthesis, a 5-enolpyruvylshikinate-3-phosphate synthase which is responsible for aromatic amino acid biosynthesis. All these functionalities are involved in reducing oxidative stress such as selenite directly promoting oxidative stress through a reaction with cellular reduced thiol-containing compounds. Down regulation of the

hemebiosynthesis reduces again oxidative stress, because cytocromes (most abundant heme-containing proteins) produce stress in reducing oxygen in aerobic respiration. Finally suppressing expression of the aromatic amino acid biosynthesis serves to reduce production of internal aggressive oxygen, since a lack of menaquinone leads to defective oxidative respiration.

Further experiments showed that lmo0807, which codes for an ABC-Transporter, plays an important role in mediating stress in general and especially considering nitrosative stress in Listeria.

Other genes like *lmo1445* (*zurM*) or *lmo447* (*zurR*) are responsible for zinc uptake, which is essential in protection versus oxidative stress [49, 50]. Another highly induced gene *actA*, responsible for recruitment of the host cytoskeleton, might help Listeria to escape from autophagy.

3.6 Conclusion

This study proved already known mechanisms and revealed new aspects of transcriptional consequences of IFN-y stimulation on intracellular growth of bacteria.

Major effects were increased oxidative and nitrosative stress levels. Thus our data revealed, that 21 genes occupy an important role in intracellular survival of *Listeria monocytogenes*.

4 Chapter III

The intracellular sRNA transcriptome of *Listeria monocytogenes* during growth in macrophages

The intracellular sRNA transcriptome of *Listeria monocytogenes* during growth in macrophages

Mobarak A. Mraheil*, Andre´ Billion*, Walid Mohamed, Krishnendu Mukherjee, Carsten Kuenne, Jordan Pischimarov, Christian Krawitz, Julia Retey, Thomas Hartsch, Trinad Chakraborty, and Torsten Hain.

Nucleic Acids Research, 2011, 1–14 doi:10.1093/nar/gkr033

4.1 Contributions

The author (A.B.) conceived and designed the study together with M.A.M. and T.H.. Furthermore he was responsible for the design of the RNA-Seq experiment itself. A.B. implemented analytical workflows for the analysis of microarray and RNA-Seq data generated by this project. The author performed the identification of small non-coding RNA candidates by using developed workflows combined with manual inspection. Furthermore, A.B. conducted all bioinformatics analysis such as transcription factor binding site detection, terminator prediction and identification of small ORFs.

4.2 Abstract

To investigate the regulatory role of small non-coding RNAs (sRNAs) contributing to intracellular survival, we used 454 pyro sequencing from size fractioned RNA of extra- and intracellular growing *Listeria monocytogenes*. Of the 150 reported sRNAs, of which 71 have never been described before, 29 are specifically expressed intracellularly. For verification, several northern blots were conducted as well as three isogenic mutants for specific intracellularly expressed sRNAs (rli31, rli33-1, rli50*). All three mutants showed attenuated growth in mice and insect infection models. This study demonstrates the importance of sRNAs during intracellular survival of *L. monocytogenes*.

4.3 Introduction

The association of novel experimental approaches and computational advances have led to an increasing number of complete bacterial genome sequences and to the identification of numerous small, non-coding RNAs. These small molecules are present in a broad range of organisms extending from eubacteria to humans. sRNAs are known to act as a vital regulatory instrument to e.g. coordinate adaptation to environmental changes, but they also play an important role in controlling virulence and pathogenesis. The first well-described sRNA was *Hfq* from E.coli. *Hfq* acts as a chaperone to promote sRNA-mRNA duplex formation [51, 52]. Until today various studies described several sRNAs under different growth conditions. While *Listeria monocytogenes* is able to invade and survive inside vertebrate host cells, this study attempts to reveal the role of sRNAs during intracellular survival. Therefore RNA from extra- and intracellularly growing bacteria was collected and transcribed to cDNA as described [53, 54]. cDNA was size-fractionated and deep sequenced with 454-Technology to provide a comprehensive view of listerial sRNA candidates preferentially induced following infection in murine macrophages.
Here we report on the discovery of 150 putative sRNAs, of which 29 are specifically intracellular expressed and highly conserved among other pathogenic listeria strains. Isogenic mutans showed attenuated virulence in *in vivo* experiments.

4.4 Material and Methods

RNA sequencing and data analysis
RNA isolation from extra- and intracellular grown *Listeria monocytogenes* was performed as described previously [7].
Quantity of isolated total RNA was determined by absorbance at 260 and 280 nm, whereas quality of quantified RNA was verified using Nano-chips for Agilents 2100 Bioanalyser. Afterward a small RNA-chip was used to detect and estimate the presence of small RNAs ranging from 20 to 500 nt.
RNA-sequencing (RNA-Seq) was performed for the total RNA fraction determined to be smaller than 500 nt. RNA were first treated with tobacco acid pyrophosphatase to discriminate between degraded and primary 5'-ends RNAs.
All primary transcripts were poly-A tailed and first-strand cDNA synthesis was performed. Resulting cDNA were PCR-amplified by high fidelity polymerase and afterwards size-

fractionated to sizes of 20-500 nt by using a 6% PAA-gel. Roche 454 pyrosequencing using GSFlx Titanium series chemistry was carried out by Eurofins MWG Operon.

As a result of RNA-seq we obtained raw data files for each condition comprising approximately 190K reads (31 mio. bases). A combination of automated and manual clipping of 5'-linker, poly-A tail and short reads was used to enhance sequence and mapping quality. Based on this adjusted data set, we performed reference mapping against *Listeria monocytogenes* using NCBI Blast with specifically adjusted parameters. The e-value was set to 0.001 to avoid accidental resemblance. The wordsize was adjusted to two to increase specificity. In order to avoid false positive hits, we established a cutoff combination of identity and coverage to ensure correct identification. Minimum nucleotide identity of the alignment between read and reference had to exceed 60 percent. Implementation of a "coverage" parameter is necessary, because it ensures that the nucleotide identity refers to the complete query- and reference sequence which reduce the number of errors especially considering long sequences. The nucleotide identity within the coverage (greater 80%) has to be over 60 percent. Only if these requirements were fulfilled, a read was regarded as mapped. The final set of mapped reads, which represent the starting point for sRNA candidate analysis, ranged from 21 to 521nt in size with an average length of 74 nt.

sRNA detection

To allow a more precise characterization, we establish new definitions to group identified sRNA candidates in three different categories: (I) the first category is named *regular sRNAs* and comprises all candidates which are found within an intergenic region (IGR), with no overlap to known annotations or structures. (II) *antisense sRNAs* identified on the complementary strand of annotated coding regions. (III) the third class contains *riboswitches including cis-regulatory sRNAs*.

Several steps were performed by sncRAS supported by manual inspection to detect putative sRNAs. The second step of analysis identified candidates located within an IGR with a minimal length of 50 nt, no overlap to adjacent genes, and covered by a minimum of 10 reads in at least one condition.

Resulting candidates were manually examined by means of the Integrative Genome Viewer (IGV). Valid candidates were searched against the Rfam database using a minimal score of 50 for detection of cis-regulatory sRNAs.

Transcription factor binding site and terminator prediction

In order to concretize identification of sRNA candidates, we searched 50 nt upstream of each sRNA for possible promoter binding sites with the FUZZNUC and HMMER3 softwares. Terminator prediction was extracted from a pre-calculated set by TransTermHP with a confidence score greater than 30. Motifs within a sliding window of +/- 10 nt around the 3'-end were identified as transcription termination site.

Small open reading frame detection

To exclude the possibility that one of the candidates describes no sRNA but a small ORF which was not predicted before, we conducted a reanalysis of small open reading frame detection by GenDB [55] applying a lowered ORF detection cut-off (> 10 amino acids). The resulting predictions of small peptides were mapped on the final sRNA candidate list.

Comparative genomics

To investigate the distribution of regulatory sRNAs within the genus *Listeria* we performed a comparative analysis of the 150 candidates against seven Listeria strains (4 pathogenic / 3 apathogenic). Comparison was carried out by the sRNAdb software which employs NCBI BLASTN to identify sequence similarities between candidates and genome sequences of compared Listeria strains using a minimum nucleotide identity of 60 % with at least 80% coverage of each sRNA.

4.5 Results

RNA sequencing

To investigate the intracellular transcriptome landscape of *Listeria monocytogenes*, an experiment based on 454 pyro-sequencing of extracellularly and intracellularly isolated total RNA was conducted as described in M&M earlier. The resulting two libraries comprised a total of 189,381 reads of which 73,223 (39%) were derived from extracellularly and 116,158 (61%) from intracellularly isolated RNA. Table five gives an overview of the data.

Table S2. cDNA sequencing read statistics

raw reads							
	raw reads	100% linker	100% 5' linker	100% 3' linker	100% 3 ' linker + PolyA	raw reads after clipping	average read length
EC[a]	73223	69066	68611	499	274	73128	89
IC[b]	116158	108750	108482	299	161	115867	59

after mapping and removal of duplicates					reads after mapping, including duplicates		
	mapped reads	cds reads	tRNA reads	rRNA reads	class I sRNA reads[c]	class II sRNA reads[c]	class III sRNA reads[c]
EC[a]	59218	24932	5467	14923	10838	159	4885
IC[b]	55241	31400	2958	4616	11889	475	2096

[a] EC = extracellular condition; [b] IC = intracellular condition; [c] candidate prediction includes multiple hits

Table 5 cDNA sequencing read statistics

To identify new small regulatory RNAs, sequenced reads were mapped against the reference genome using NCBI BLAST. The maximum e-value was set to 0.001 and rewards for a nucleotide match to two. With a combination of at least 60% identity and 80% coverage, the number of false positive hits was reduced to a minimum. Reads which did not fulfil these requirements were removed from the data set. After processing data through several steps of quality control and finally mapping, 114,459 unique reads were used for further analysis. We observed 100% perfect matching reads for about 50% of IC reads, but only 28% for the EC condition. The virtually computed *intergenome* of *Listeria monocytogenes*, which describes the intergenic regions between predicted ORFs, comprises approximately 10% (300 000 bases) of the entire genome. Under intracellular condition, nearly one-third of the *intergenome* is expressed which emphasizes the importance of investigation of e.g. small regulatory RNAs (Fig. 18 A). There is evidence to suggest that depletion of rRNA and tRNA has to be optimized, since approximately 60% of all mapped reads still belong to these RNA classes (Fig. 18 B and C), implying a waste of possibly interesting sequencing data.

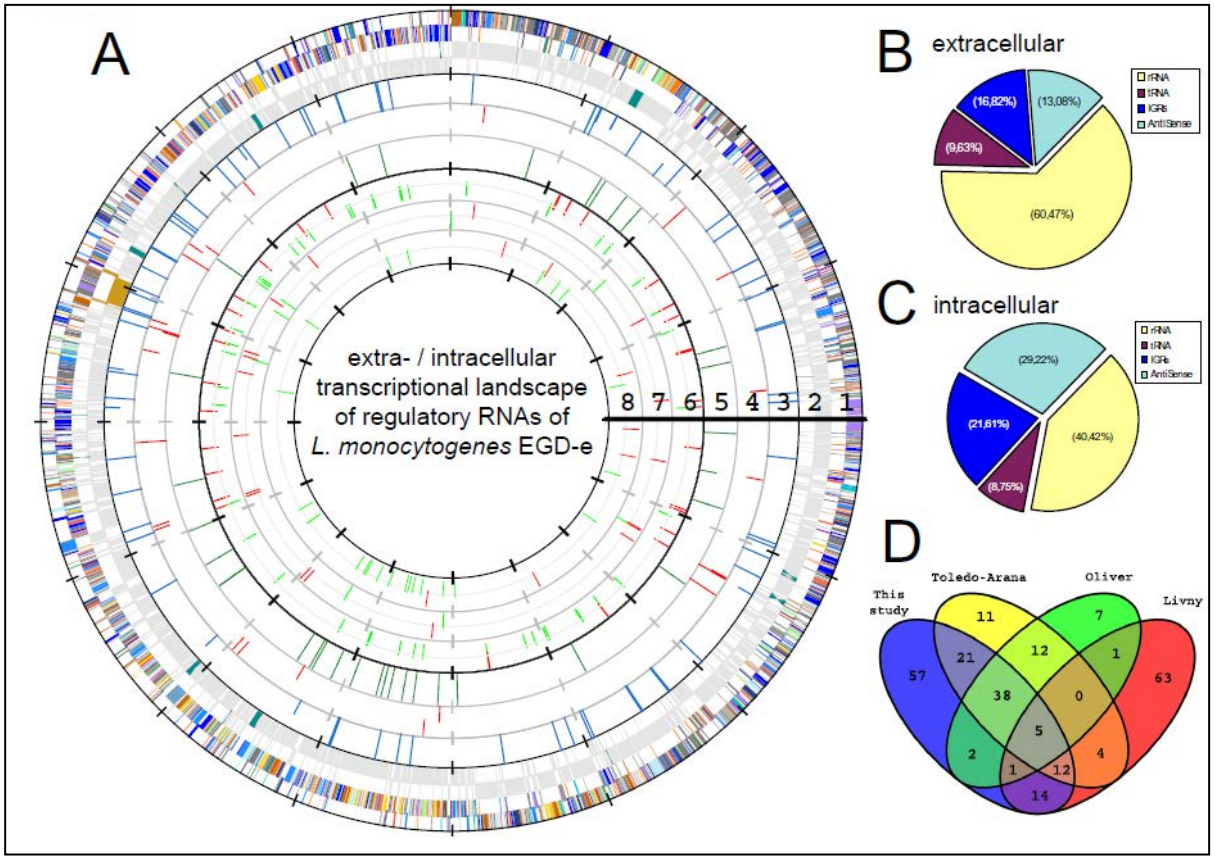


Fig.18 Discovery of the intracellular sRNAome of *L. monocytogenes* using RNA-Seq

Identification of sRNAs, asRNAs and cis-regulatory RNAs including riboswitches

With the help of extensive bioinformatics we were able to identify 150 putative regulatory sRNAs of *Listeria monocytogenes* expressed under either one or both growth conditions (Fig. 18A).

To ensure a better insight and distinction, candidates were grouped into three different classes: **class I** comprises sRNAs which are located between predicted ORFs with no overlap of currently known structures, **class II** RNAs include antisense RNAs, located antisense to coding sequences and **class III** describes *cis*-regulatory RNAs including riboswitches identified by Rfam and manually inspection. We identified 88 class I and 33 class III sRNAs which together represent approximately one-third of the regulatory *intergenome*. In addition, we classified 29 putative antisense sRNAs into class II (Fig. 18 A).

Of the 150 candidates determined, 121 were expressed extracellularly. A small share of ~7% (8 sRNAs) were specifically expressed, while ~49% (60) were more than two fold induced under extracellular growth. In contrast, 142 candidates were expressed under intracellular condition. Around ~20% (29) were expressed specifically and ~35% (50) showed enhanced induction under intracellular growth conditions. A common set of 113 candidates were observed to show induction under both conditions and therefore represent the first core "small regulatory RNAome".

Sizes of putative regulatory RNAs ranged from 50 to 517 nt with a minimal read count of 10. The highest expression was detected for the intracellular sRNA *rli28* with 1603 reads.
Of the 88 in class I, 77 showed expression under extracellular condition whereas 85 showed induction under intracellular stress. We observed eleven candidates to be specific for intracellular adaption whereby only three sRNAs were specific for extracellular lifecycle. It is noticeable that the eleven intracellular specific candidates were at least 25% higher induced than the 113 small regulatory RNAs which showed up under both conditions.
Class II sRNAs showed an even more explicit hint that specific regulation is needed intracellularly. About 86% (25) of total asRNAs (29) were intracellular induced whereas only 38% (11) were found under extracellular condition. Not surprisingly, the distribution of specifically regulated asRNAs is also shifted in intracellular direction. We observed 18 specific asRNAs for intra- and only 4 for extracellular growth.
Compared with class I and II, class III showed no major differences in sRNA expression (32 IC versus 33 EC). Only the fact that all extracellular cis-regulatory RNAs are, based on the geometric mean, at least twice as highly induced as the intracellular riboswitches is conspicuous.
Transcription factor binding site analysis showed no extraordinary results. For only 24% of the 150 candidates, a putative promoter binding site could be identified. A similar situation is observed for terminator prediction. For only half of the candidates a rho-independent terminator could be found.

Chromosomal distribution of regulatory RNA

In order to investigate the distribution of sRNAs, we created a circular representation of the genome and marked all putative candidates (Fig. 18 A). We wanted to determine if the distribution hinted towards a specific clustering or functionality, but found that sRNAs are neither located within specific chromosomal locations, nor do they exhibit a specific strand prevalence or cluster in noticeable patterns.

Comparative analysis with whole genome tiling array data, RNA-Seq and in silico prediction

To evaluate the findings of this study, we compared our candidates to previously identified sRNAs of *Listeria monocytogenes*. Toledo-Arana published 103 sRNAs extracted from five different conditions and the comparison showed an overlap of 75% to our data. The missed 26 candidates were composed of 12 sRNAs without any expression, 12 failed to

reach at least one of our predefined cutoffs (e.g. length, read count etc.), one candidate was expressed on the other strand and one was located in the same IGR but showed no overlap. Because conditions are quite diverse in this analysis, we performed an additional comparison of growth in BHI vs. BHI and intracellular vs. growth in blood. BHI growth comparison showed an overlap of about 68% whereas IC vs. blood indicated an even higher agreement of 72%. In addition, we compared an Illumina sequencing approach of Oliver and colleagues [54] and detected an overlap of nearly 70%. Finally, we wanted to evaluate the power of software prediction within small regulatory RNA investigation. For this comparison we included sRNAs from three different research groups, produced with three different technologies.

Overall, candidates were retrieved from different datasets of 12 (8 cossart, 2 ugmi, 2 oliver) different conditions which cover most conditions ever examined in genus *Listeria*. A comparison of 100 sRNAs predicted by the SIPHT software versus 150 of the current study, revealed an overlap of only 33%, which illustrates the inaccuracy of current software prediction.

We performed an additional small ORF prediction to minimize the possibility of small peptides erroneously identified as sRNAs. For 14 regulatory RNAs we identified an overlap to a putative small ORF. The smallest putative sRNAs for *L. monocytogenes* were identified within the *rliB* locus, which is involved in virulence [53] and was already noticed by Mandin [56]. We used a CRISPR detection tool to identify five repeats (29 nt each) and five spacers (36 nt each) and verified this finding with our transcriptome data. Interestingly, no homologs of CAS genes could be detected within this region which, may warrant the conclusion that this RNA is no longer used for phage defence, but its expression levels suggest a remaining functionality.

Comparative analysis of putative regulatory RNAs among members of the genus Listeria

To investigate the regulation of virulence through sRNAs we compared four pathogenic and three apathogenic *Listeria* strains with our pathogenic reference *L. monocytogenes* EGD-e. We detected a highly similar sRNA content within linage I and II. When compared to phylogenetically distant strains, the correlation decreased rapidly.

4.6 Discussion

We sequenced and analysed the transcriptome of *Listeria monocytogenes* EGD-e based on extracellular and intracellular growth conditions. We identified 150 putative regulatory RNAs from which 71 have never been described before. All candidates have been grouped into three different classes: class I comprises 88 sRNAs, class II includes 29 asRNAs and class III is composed of 33 *cis*-regulatory RNAs including riboswitches. To confirm our findings we performed a comparative analysis to a recent study of Toledo-Arana, despite differences of both studies considering condition and technology. Interestingly, we achieved a high overlap of 75% between both studies which suggest that dissimilar technologies as well as different labs and scientists can yield the same results.

Our findings show that diversity of regulation at the post-transcriptional level is an important component of adaption to specific circumstances.

For a more detailed functional analysis we focused on three intracellular strongly upregulated candidates, namely *rli31*, *rli33-1*, and *rli50**. Construction of isogenic mutants resulted in reproducible growth properties following infection of macrophages and in virulence attenuation in both insect and mouse models. Attention should be paid to the fact, that sRNA mutants are required in vertebrate and in invertebrate hosts, since *G. mellonella* is a model for innate immunity.

A possible target of *rli31* could be the downstream gene *lmo0559* which is involved in magnesium and cobalt uptake. Previous studies show a down-regulation of these transporters due to high ion concentration in the host cytosol. Furthermore, studies of *Samonellaenterica* (57) revealed a transcriptional termination of *mgtA* at high MG^{+} concentration via a stem-loop structure.

The candidate with the strongest effects in insect and mice model is *rli50**. Our analysis revealed two partially overlapping sRNAs (*rli50* and *rli112*) on different strands in the IGR of *lmo2709* and *lmo2710,* which was previously identified as *rli50*. The construction of a mutant of *rli112* tends to result in a double mutant which may be responsible for the impaired virulence observed. Interestingly, a homolog with 94% similarity to *rli112* exists (*rli78*) but for some reason it is not able to compensate for the loss of *rli50*. This may demonstrate a highly specific adaption to environmental changes.

Another example for sRNAs present in multiple variants is the already described *LhrC*. We detected five copies at two different locations which are intracellularly upregulated, however their role has still to be uncovered.

The fact that three sRNAs (*rli28*, *rli29* and *rli78*) exhibited a lower GC-content than the rest of the chromosome suggest horizontal gene transfer (HGT) events. Further analysis of the surrounding area indicated a gene of a possible IS3 transposase which additionally strengthens the hypothesis that sRNAs could be spread via different chromosomes by HGT transposition. To shape their microbial host genomes, they additionally use HGT to e.g. insert sRNAs, which was described for *ipeX* of E. coli [57]. Hence we analysed all phage or phage-related regions and found four sRNAs (*rli48*, *rli62*, *rli89* and *rli99*) within the A118 prophage locus. It is common knowledge that phage-related genes play a role in intracellular survival [7, 37,58] so it may be assumed, that sRNAs are also involved in the infection process as well. While CRISPR-systems are related to phages or rather to phage defence, we confirmed that *rliB*, as noticed by Mandin, is related to CRISPR elements.

A emerging class of small RNAs are the antisense sRNAs which have been described in several organisms and plasmids [59-63] to e.g. regulate expression at the translational level, support RNA stability or adjust transcription. Antisense RNAs could be found all over the listerial genome in various lengths. We were able to demonstrate, for the first time, the existence of two intracellularly upregulated asRNAs using northern blots. As the majority of our 29 identified candidates were intracellularly upregulated, there was unfortunately no correlation to their possible target genes in case of induction.

The third class of sRNAs are represented by *cis*-acting RNAs and riboswitches, respectively. Our study identified 33 candidates of 42 known listerial riboswitches incorporated in the RFAM database, which shows a slight difference to *B. subtillis* who uses class III sRNA regulation for at least 2% of its genes [64]. Interestingly, the majority of our 33 candidates was downregulated intracellularly suggesting an economy mode due to the fact that T-boxes represent the largest group in class III, which are involved in sensing uncharged tRNAs in the bacterial cell. Additionally co-workers found out, that tRNA synthetase genes are downregulated in the host cytosol [7], indicating an involvement of T-box regulation due to infection. There is evidence for the involvement of riboswitches in pathogenicity, because Loh and colleagues found two (in our data downregulated) *in trans* acting SAM riboswitches (SreA and SreB) allowing a intracellular induction of the PrfA regulator [65]. Comparative analysis disclosed several differences in the sRNA repertoires of *Listeria* strains and serotypes enabling the use of small RNAs as diagnostic markers e.g. *rli112* which is specific for linage II.

4.7 Conclusion

In this study we were able to show the important participation of sRNAs during intracellular survival. We identified 71 previously unknown putative small regulatory RNAs and therefore increased the size of the "sRNAome" of *Listeria* to 180. Furthermore, experiments revealed 29 candidates specifically expressed during intracellular growth which confirms the conclusion that sRNAs are required for virulence of the bacterium. Future development should aim at the establishment of methods and techniques to enable the identification of targets of sRNAs to enable us to understand the regulatory response triggered by the organism when shifting from extra- to intracellular environment.

5 Chapter IV

Comparative genome-wide analysis of small RNAs of major Gram-positive pathogens: from identification to application

Comparative genome-wide analysis of small RNAs of major Gram-positive pathogens: from identification to application

Mobarak A. Mraheil*, André Billion*, Carsten Kuenne, Jordan Pischimarov, Bernd Kreikemeyer, Susanne Engelmann, Axel Hartke, Jean-Christophe Giard, Maja Rupnik,Sonja Vorwerk, Markus Beier, Julia Retey, Thomas Hartsch, Anette Jacob, Franz Cemic, Jürgen Hemberger, Trinad Chakraborty and Torsten Hain

Microbial Biotechnology (2010) doi:10.1111/j.1751-7915.2010.00171.x

5.1 Contribution

The author conceived the structure of the review and drafted the manuscript in collaboration with M.A.M.. In addition, the author contributed figures and tables and was involved in comparative analysis of sRNAs from five pathogens.

5.2 Abstract

The number of drug- and multiresistant bacteria has increased substantially in recent years and therefore the demand in health care of new approaches for identification of putative targets and development of new anti-infectives gains a new as well as important significance. Small regulatory RNAs promise to become new drug targets, since their explicit role in gene expression is still mostly unknown for wide-spread gram-positive pathogens. This review focuses on current knowledge about sRNAs of the five major socioeconomically relevant gram-positive pathogens *Listeria monocytogenes*, *Staphylococcus aureus*, *Streptococcus pyogenes*, *Enterococcus faecalis*, and *Clostridium difficile*. It will summarise the current state of next generation sequencing techniques and closely related bioinformatics considering identification and classification of sRNAs. Finally, an outlook about the future use of modified peptide nucleic acids (PNAs) for inactivation of sRNAs, as well as the rapid and specific detection of pathogens is offered.

5.3 Introduction

Small non-coding RNAs as well as micro RNAs (miRNAs) are a current subject of study for eukaryotes and could be identified to act as key regulators of several cellular processes [66]. Since these types of RNAs represent an extra layer of gene regulation, it is not surprising that e.g. expression of miRNAs could be detected in various types of tumour cells [66, 67].

In contrast to eukaryotes, the function of sRNAs in prokaryotes is almost completely unknown. In general, sRNAs act as additional regulators of transcription, translation and affect mRNA stability [1]. Major interaction comprises pairing with RNA, adopting structures and binding of RNA-protein complexes [68]. As already introduced for other areas of science, *E. coli* acts as a model organism for which several identification techniques were established. These include microarray, shotgun cloning, co-purification with proteins and the development of *in silico* tools for the prediction of putative sRNA candidates [69]. Further investigation of e.g. *Salmonella typhimurium* revealed a clear relation of sRNAs and virulence [70].

A survey of the presence of sRNAs in five major gram-positive pathogens revealed 103 sRNAs for *L. monocytogenes* [53, 56, 71,72]. 12 sRNAs were identified for *S. aureus*, some of which showing differential patterns of expression in pathogenic strains, suggesting an involvement in virulence [73]. Only two sRNAs could be identified and further studied for S. pyogenes [74, 75], whereas for S. pneumoniae five candidates were verified [76].

Currently, sRNAs are mostly associated with stress response, iron homeostasis, outer membrane proteins, various metabolic systems (especially sugar metabolism) and quorum sensing. sRNA involvement in important functions as well as various experiments suggest an essential role for regulatory RNAs for the pathogenicity of many bacteria.

To further examine these hypotheses, a European consortium has been established which uses current technologies such as novel high throughput sequencing, state of the art bioinformatics, whole genome transcriptomics and proteomics combined with standardized molecular characterisation methods to investigate production, regulation and impact on pathogenicity of sRNAs in these high-risk pathogens.

Finally, results should enable the consortium to develop novel therapeutics on the basis of PNAs and to use this knowledge to create very fast and highly specific diagnostic test systems.

Listeria monocytogenes

L. monocytogenes is a facultatively anaerobic intracellular pathogen that is tolerant to low temperature, low pH and high salt concentrations. Seven species are currently known of which *L. monocytogenes* can cause fatal infections with a mortality of up to 30% in humans and animals, with disease patterns ranging from meningitis, abortion, prenatal infections to gastroenteritis [77]. The most striking feature is the ability to invade, survive and replicate inside vertebrate host cells. Several abilities are required for successful intracellular survival, however the most important genes are organized within the extensively studied VGC cluster [78]. Based on their mode of action, sRNAs can be divided into two groups: (i) *cis*-regulatory RNAs (UTR) and *trans*-acting RNAs. *Cis*-regulatory RNAs e.g. modulate their structure through the binding of metabolites or temperature changes. Both types of sRNAs often act by fine-tuning expression to react to changes in environmental conditions. The first reported *cis*-regulatory RNA is the well-known virulence regulator PrfA [79]. PrfA is located at the very start of the VGC and acts as an RNA thermometer, which enables the transcription of the VGC at higher temperatures (~37°C). Other genes which are essential for intracellular survival are also controlled by 5'-UTRs such as actA, hly, and inlA. Further structures like the Hfq chaperone support the effect of sRNAs in stress tolerance and virulence [80].

In the past, a number of transcriptomic studies have been performed and beside findings of several specific adaptational strategies, we found that approximately 17% of the entire chromosome is regulated to ensure intracellular survival [7]. The first whole genome sRNA study was performed by Toledo-Arana and colleagues in 2009 using Affymetrix tiling arrays. They studied sRNA expression under five different conditions e.g. exponential and stationary phase in BHI, low oxygen, high temperature, and human blood. This experiment identified 103 putative regulatory RNAs, and the first comprehensive operon map of *L. monocytogenes*. Nevertheless, further detailed studies are necessary to understand the exact role of sRNAs for the adaptation to specific situations.

Staphylococcus aureus

This bacterium is a major cause of nosocomial infections worldwide with a still increasing number of multiple resistant strains. Currently, only two fully sequenced genomes are available including 32 annotated sRNAs. The well studies virulence factors are under transcriptional control of an sRNA called RNAIII. RNAIII was the first sRNA for which an involvement in virulence was described [81]. In the year 2005, 12 sRNAs were identified

by a comparative approach using strain N315 [73]. Geismann and colleagues identified 11 additionals RNAs through a transcriptional analysis mainly based on growth phase experiments using three strains of *S. aureus* (RN6390, Newman and COL). Since knowledge of sRNAs and their role in virulence is still limited, Geismann performed further transcriptomic and proteomic analyses and revealed an important regulatory RNA (RsaE). Since the current number of identified sRNAs still do not represent the complete catalogue, new studies using whole genome tiling array and next generation sequencing are supposed to produce a more comprehensive picture of sRNA content under different conditions. Identification and characterization of important regulatory elements may produce new targets for therapeutics in the future.

		SIPHT[a]		Experimentally verified	
Genus	Species	Chromosome	Plasmid	Chromosome	Plasmid
Staphylococcus	*aureus*	32–79 (12)	0–5 (9)	23[b]	1[c]
	epidermidis	116–127 (2)	0–4 (7)		
	haemolyticus	74 (1)	–		
	saprophyticus	38 (1)	1 (2)		
Streptococcus	*agalactiae*	29–34 (3)	–		
	mutans	18 (1)	–		
	pneumoniae	28–66 (3)	–	5[d]	
	pyogenes	18–29 (12)	–	3[e]	
	thermophilus	31–36 (3)	0 (2)		
	sanguinis	34 (1)	–		
	suis	23–24 (2)	–		
Enterococcus Clostridium	*faecalis*	14 (1)	0–2 (3)		2[f]
	acetobutylicum	18 (1)	2 (1)	1[g]	
	beijerinckii	31 (1)	–		
	botulinum	54–68 (4)	0 (2)		
	difficile	2 (1)	0 (1)		
	kluyveri	46 (1)	0 (1)		
	novyi	26 (1)	–		
	perfringens	14–18 (3)	0–2 (3)	1[h]	
	tetani	45 (1)	0 (1)		
	thermocellum	6 (1)	–		
Listeria	*innocua*	115 (1)	2 (1)		
	monocytogenes	94–124 (2)	–	27[i]	
	welshimeri	100 (1)	–		

a. Livny *et al.* (2008).
b. Novick *et al.* (1993); Morfeldt *et al.* (1995); Pichon and Felden (2005); Geissmann *et al.* (2009).
c. Kwong *et al.* (2004; 2006).
d. Halfmann *et al.* (2007).
e. Kreikemeyer *et al.* (2001); Mangold *et al.* (2004); Roberts and Scott (2007).
f. Weaver (2007).
g. Fierro-Monti *et al.* (1992).
h. Shimizu *et al.* (2002).
i. Barry *et al.* (1999); Christiansen *et al.* (2006); Mandin *et al.* (2007); Nielsen *et al.* (2008); Toledo-Arana *et al.* (2009).
The SIPHT columns show the minimum and maximum number of annotated sRNAs of each species.
The quantity of analysed strains is depicted in brackets.

Table 6 Overview of published gram-positive sRNAs of the genera *Staphylococcus, Streptococcus, Enterococcus, Clostridium* and *Listeria*

Streptococcus pyogenes

This exclusively human pathogen is responsible for over 500,000 deaths yearly through severe invasive diseases revealing a major impact of these bacteria for global mortality [82]. As in several other bacterial species, research on the role of sRNAs in e.g. virulence has become a major focus. Two independent bioinformatics approaches (Sipht and MO-

SES) identified multiple putative sRNAs, however later verifications showed a relatively low overlap between prediction and experiment. This may be a result of inconsistent conditional parameters and a limited amount of basic information to develop robust prediction algorithms.

Later experiments revealed and characterized two important sRNAs (*sagA* and *fasX*) which are at least partially involved in adherence and internalization of GAS [83] and thus mediate virulence gene regulation. Whole genome based tiling array analysis revealed 40 additional candidates, which reveal contrary results to further studies [84].

In summary, 40 candidates were identified by tiling array, three were identified and characterized experimentally, whereas 39 were predicted by two bioinformatics approaches.

Enterococcus faecalis

E. faecalis is a further human commensal strain and also known to cause nosocomial infections affecting mainly young and immunocompromised patients. Despite the ability to initiate endocarditis, meningitis, pneumonias, peritonitis, visceral abscesses, and urinary infections [85], *E. faecalis* is also used in the food industry e.g. as probiotics [86]. Thus, it is very important to increase efforts in understanding the mechanisms which turn commensal living bacteria into dangerous pathogens.

Currently, only few transcriptional regulations are studied and mechanisms of virulence are poorly understood. Only one regulation based on two sRNAs (RNA I and RNA II) has been studied in more detail. Beyond that, only predictions of 17 putative candidates via Sipht software are available. Compared to other studies of gram-positive bacteria, the number of potential candidates of sRNAs in this germ remains high.

Clostridium difficile

C. difficile belongs to the group of nosocomial pathogens and can cause severe forms of colitis, pseudomembranous colitis and bowel perforation [87]. These are mainly caused by the well-studied toxins A and B, but mechanisms of gene regulation remain unclear. Furthermore, the roles of sRNAs are completely unknown for pathogenesis of *C. difficile*. For *C. perfringens,* only one study exists which reveals the first insights into the regulation of toxins by sRNAs in Clostridia. It is thus very important to fill this knowledge gap by the identification and characterization of novel regulatory small RNAs of *C. difficile*.

General usage of RNA-binding proteins remains unclear

Half of all currently sequenced bacteria encode a Hfq homologue which improves the pairing between sRNA and target mRNA, and supports the stability and translation of many mRNAs. Surprisingly, several strains in this study seem not to use Hfq for sRNA-mRNA interactions [88] or do not encode a Hfq protein at all.

It is therefore essential to establish the mechanisms that allow these bacteria to bypass the use of such RNA-binding proteins.

sRNA identification methods

In the past, small regulatory RNAs were hard to detect as only few techniques were available to verify short transcripts. The first sRNA were accidently identified by radiolabeling of total RNA four decades ago [89-91]. The technical evolution over the last decade went from single labelling and sequencing to whole genome approaches such as the SELEX approach [92], whole genome shotgun sequencing [93], microarrays [94] and finally to NGS-technologies to complete the catalogue of sRNAs for many bacteria.

Description of the whole genome sRNA expression of bacteria became possible in a specific growth condition after introduction of the tiling microarray approach. This requires the development of an appropriate array design to cover e.g. the whole genome or only intergenic regions. While most known sRNAs are encoded within intergenic regions, an array design of the *intergenome* is more cost effective and offers the opportunity to optimize parameters such as overlap and length of probes, melting temperatures, and control probes to improve detection quality. To further optimize sRNA identification, a two phase experiment can be performed, were the first round covers the whole bacterial genome and a second round focuses on promising regions with different parameter settings in terms of e.g. length and overlap of probes to achieve a higher resolution.

To further increase detection quality, an accurate size selection reduces the complexity of the RNA sample. Furthermore, direct labelling is preferred because amplification induces inaccuracies. Next-generation sequencing offers a new level of sRNA analysis, because the amount of generated information enables an unsurpassed accuracy in terms of e.g. start and stop position of transcripts, no signals-errors through cross-hybridizations, and detection of other transcriptional activities, which support the identification of sRNA candidates. However, the growth of data creates new challenges in information processing and interpretation.

Prediction and target identification of sRNAs

Prediction of small regulatory RNAs, especially *cis*-acting sRNAs, remains an open problem. The majority of available software uses nucleotide sequence, motifs, and homology based alignments as input. In the past, algorithms tried to find the lowest free energy of structured RNAs [95], but Rivas and Eddy [96] showed that minimum free energy (MFE) cannot be used as a distinction as it is not specific enough. Another approach was based on the observation, that sRNAs have a mean GC-content of 50% [97]. However, this method is only applicable to AT- or GC-rich organisms.

Comparative analysis arguably represents the optimal strategy for the *in silico* identification of sRNAs. Several groups tried to identify novel sRNAs by comparing related genome sequences and searching within highly orthologous regions for either thermodynamic information or covariant evidence.

The most widely accepted *de novo* prediction method is the HMM algorithm used by the Rfam database. Rfam identifies individual characteristics of a set of sRNAs and creates a flexible pattern of this sequence. Similar transcripts can then be screened for this pattern. The majority of HMMs were constructed for trans-acting RNAs including riboswitches, since automatic classification is till date the most reliable *in silico* prediction for type of RNA. Several other software programs such as SIPHT [17], sRNAscanner [18], and MOSES [98] attempt to identify new sRNAs on the basis of e.g. phylogenetic analysis, presence of Rho-independent terminators, comparative analysis, and presence of transcription factor binding sites. Current software evaluation showed a rather small overlap between prediction and experimental identification, but with the discovery and verification of further sRNAs the algorithms will become more precise. The next level of sRNA research focuses on the identification of targets, which requires computational predictions as well as experimental verification. Several tools are available to predict putative targets for a given nucleotide sequence. Software like TargetRNA [99] or IntaRNA [100] compute, based on base-pair-binding, the minimum free energy of a e.g. hybridized RNA-DNA complex to identify its possible targets.

Antisense RNAs as a putative treatment of infections

Increasing knowledge about sRNAs and their role in regulation in e.g. pathogenesis, makes this class of RNAs interesting for drug development. To date, antisense RNAs are tested to treat e.g. diabetes, cancer, asthma, and arthritis. While the first and second generation of

those drugs struggled with parameters such as stability, delivery, specificity, low binding affinity to complementary sequences and toxic side effects, a new class of analogues was needed. Third generation of antisense agents comprisedlocked nucleic acids(LNA), phosphorodiamidate morpholino oligomers (PMO) and peptide nucleic acids(PNA). LNAs have still difficulties with self-binding due to self-complementary sequences, a negatively charged backbone and stability. In contrast, PMOs and PNAs which own a neutral backbone, offer a good specificity and increased duplex stability in complexes with RNA and DNA. PMOs and PNAs primarily differ in their solubility in water whereas thermodynamic properties of PMOs are not well defined and thus not comparable to PNAs.

Since PNAs are well defined in terms of chemical and biological stability, solubility, specificity, and duplex stability, they are excellent candidates for drug development. To overcome cell-wall permeability, PNAs will be coupled to cell penetrating peptides (CPPs) mediating the non-invasive transport into the cell. The most commonly used CPP is (KFF_3) K, but Mellbye and colleagues recently showed improved cell delivery by modifying amino acid composition[101]. The development of efficient CPPs is an emerging field with an enormous potential for the future. The first proof of principle was published by Tan and colleagues [102] demonstrating a drug (coupled CPP with PNA) against a bacterial infection. A combination of CPP and PNA as an antisense drug is a promising alternative to actual antibiotic treatments in the future.

PNAs for diagnostics

Putative drugs comprising PNAs are attractive anti-infective agents because of their specificity and stability, and their ability to be used for detecting small regulatory RNAs in diagnostics systems. Current standard diagnostic tests are based on nucleic acid interactions

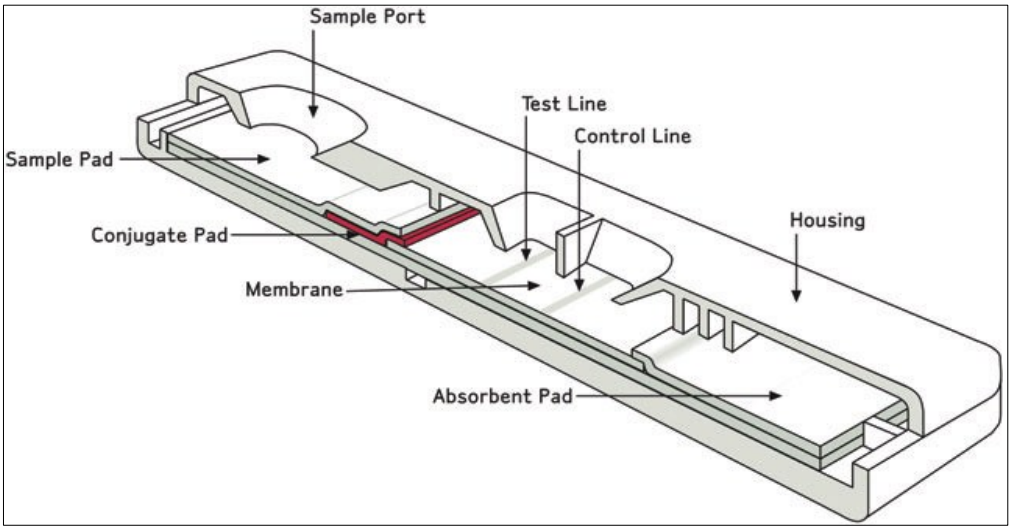


Fig.19 Standard lateral-flow assay device

(PCR) which require trained personnel and time to complete these test are not suitable for point-of-care-testing (POC).Standard LFAs use immobilized oligonucleotides on a nitrocellulose membrane for the chromatographic detection step. To adapt this principle for the identification of nucleic acids, several interim steps have to be performed, which reduce the usability in POC testing.

Therefore bead technology was successful applied to LFAs. Gold nanoparticles are used, because they are small, easily visible and robust in manufacturing [103]. Mao and colleagues describes typical detection limits with around 2.5 mg DNA ml^{-1}[104]. To further increase sensitivity, several modifications such as multiple hapten moieties, fluorescence lables, and quantum dots have been tested. All modifications improved sensitivity to a certain degree but background noise and costs prevent commercial application so far. To overcome these obstacles, members of this consortium used superparamagnetic nanoparticles which were intensively studied in the past and exhibit advantages such as commercial availability and surface chemistry, as well as stable paramagnetism to prevent clustering. The major advantage stems from the very low magnetic background of most samples, which permits detection of very low signals when compared to fluorescence or optical detection.

5.4 Conclusion

The actual state of affairs about identification and characterisation of small regulatory RNAs in the major five gram-positive pathogens only describes a subsection and will be extended with the increasing number of additionally sequenced genomes of these five pathogens. This information provides the base for further research within the area of regulatory networks, the systematic understanding of pathogenic features and the development of novel therapeutic drugs. As described previously, sRNAs form potential markers for diagnostics and will become increasingly relevant for POC application.

The most interesting questions for the future will be: (i) what is the mechanism behind the action of sRNAs (ii) to which extent are sRNAs involved in pathogenesis, (iii) can sRNAs be used as novel targets for therapy e.g. with PNAs, and (iv) how successful is the use of sRNAs for diagnostics in e.g. lateral flow assays.

6 General Discussion

This thesis demonstrates, in a chronological manner, the rapid development of the last few years in the area of high-throughput technologies and their invaluable contribution towards understanding the biological phenotype. The dramatic progress of these technologies is documented here as it describes the progression from microarrays to RNA-Seq using NGS technologies within the lifetime of this thesis.

The major focus of bioinformatics when dealing with microarray technology deals with the statistical analysis of expression data. Other possible issues such as the volume of data, complexity and computing power are negligible. Unlike the analysis of microarray based data which involves work on data encompassing several megabytes, NGS primary data ranges up to terabytes. Whereas a standard microarray analysis was manageable on a regular desktop computer, analyses of data of NGS technologies require high performance computing (HPC). Even today primary datasets are already too large to be transported via the internet. This massive expansion necessitates new investments in hardware, software and experienced personnel to be able to keep up with the progress of next generation technologies.

Therefore, the primary job of a bioinformatician does not only include the use of prepackage software for data analysis, but also involves creation of a hardware environment for data aggregation and processing, development of specific software for quality control, visualization tools, modelling and analysis for interpreting data for studies and publications.

At the beginning of this thesis, the primary job was focused on the analysis of data derived from microarrays. Even though standardized workflows for the manufacturing and analysis of microarrays were available at that time, these softwares revealed various drawbacks. To create a seamless platform that could analyse such data in a generalized way, new software solutions were implemented with specific idiosyncrasies of a particular technology with an emphasis on data quality.

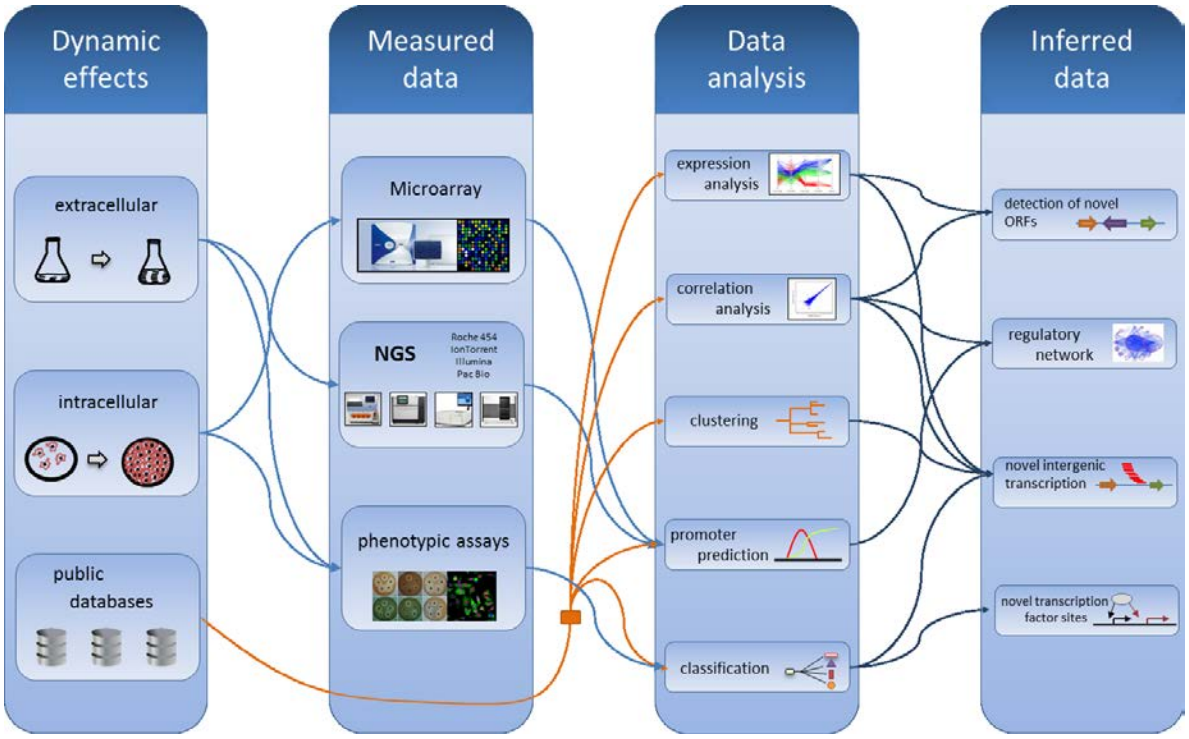


Fig. 20 Overview of the current work- and analysis workflow

Comparative whole genome transcriptome analysis reveals common and divergent expression profiles for intracellular survival

The availability of genome sequences from representative *Listeria monocytogenes* strains of different lineages permitted the design of consensus microarrays covering commonly occurring genes as well as strain-specific ORFs to examine gene expression of these strains under extra- and intracellular growth conditions. This allowed the correlation of genome content and expression analysis for individual strains. In addition, comparative bioinformatic analysis was used to characterise new pathogenic traits as well as to detect strain specific adaptive genes for intracellular survival.

The greatest challenge of this section of this thesis was the detailed comparison of four highly syntenic genomes and the evaluation of rather large expression data sets. Furthermore, genome content was linked to transcriptome data in order to obtain comparative information on *Listeria's* adaptation to the intracellular milieu. To enable structured analysis of this quantity of data, a new microarray workflow entitled MARS II were implemented. In MARS II modules from the existing program was extended by developing software packages that allowed better quality control, a scripting interface, storage of results in a central database, and the possibility for the generation of different kinds of exports.

In the first section of the work, it was necessary to close the existing gaps in the genome sequences of pathogenic lineages II and III, represented by th e*L monocytogenes* 4b CLIP 80459 and 4a L99 strains. For further in depth analysis, manual curation was used to annotate individual genomes. Thus the first complete comparative analysis of the gene content could be performed, indicating that the reduction and loss of virulence factors and surface proteins dramatically affected the intracellular survival rate. Deletions of internal repeats of the ActA gene were detected in some strains. Loss of a single repeat resulted in decreased intracellular bacterial motility which may affect survival. In contrast, duplications of transport genes such as those involving PTS-systems, or the presence of certain bacteriophage genes had a positive impact on intracellular growth.

We combined whole transcriptome analysis of these four strains under conditions of extracellular- and intracellular growth using the Febit Geniom platform. This platform allowed the generation of highly reproducible arrays comprising all ORFs of all strains.

The experiment was characterized by the high number of technical as well as biological replicates. Overall, 144 individual experiments were performed which allowed very accurate (low error rate) observations of differentially regulated genes. Unsurprisingly, the analysis revealed increased intracellular induction of virulence-associated genes as well as of many PTS systems. A remarkable observation was a change in metabolic flux whereby bacteria shifted from the use of glycolysis pathways to those involving the pentose phosphate pathway to produce pentose sugars for biosynthesis. As use of the PPP is associated with the production of reducing equivalents of NADPH, this is expected to counteract oxidative stress experienced by the bacteria during intracellular growth. Interestingly the most highly expressed intracellular genes included those within the *lma* operon and its adjacent genes. The function of prophage-related genes for intracellular survival is poorly understood. However, their importance in intracellular bacterial growth is demonstrated by the fact that mutants lacking these genes are attenuated for infection in the mouse model.

Downregulation of genes involved in nucleic acid biosynthesis, i.e. for DNA and RNA, expression of tRNA synthetase, as well as cytochrome synthesis indicate that bacteria growing intracellularly reduce replication rates to lower their energy consumption. This is supported by the observation that bacteria growing extracellularly in BHI medium double once every 30 minutes as compared to > 70 minutes for intracellular growth. We observed lowered expression of flagellar genes during intracellular growth, which supports the notion that intracellular motility is ActA based while extracellular motility is governed by

flagella. In addition, it has been shown that flagella are recognised by intracellular pattern recognition receptors (PRRs) which activate host immune responses.

Infection of IFN-γ activated macrophages by L. monocytogenes requires bacterial adaptation to oxidative and nitrosative stress

During infection, *L. monocytogenes* are often confronted with activated macrophages whose killing abilities are enhanced by the release of cytokines. Activated macrophages remove and kill bacterial pathogens by the induction of a large array of microbicidal factors including reactive oxygen and nitrogen oxide as well as antimicrobial peptides. Further microarray analysis, based on the in-house developed platform, revealed listerial regulation of numerous genes to counter oxidative stress. Additional bioinformatics analysis showed that listeria has an increased demand for tryptophan during replication in activated macrophages. This is achieved by the induction of enzymes which are involved e.g. in the anthranilat synthesis or the kynurenine pathway.

The primary challenge of this section was the wet lab experiment itself, which involved working with nanogram quantities of RNA as compared to six µg of RNA required for a regular microarray experiment. Generally, MOI's should be very small to ensure a natural cellular response. However, when using small MOI's, the resulting amount of RNA is correspondingly low. The difficulty of the experimental setup also resulted in relatively small signal to noise ratios from hybridization. Bioinformatics analysis was employed to determine optimal cutoffs and filters to distinguish between real signals and technical artefacts while the entire assay operated at the limit of detection. Final parameter settings could only be obtained through an iteration via several settings. Since the whole experiment was carried out near the limit of detection, we performed quantitative real-time PCR to independently confirm over 50% of the genes identified by bioinformatics analysis.

The transcriptional landscape of intracellular gene expression in Listeria monocytogenes reveals a central role for small non-coding RNAs

With the development of NGS, it was possible for the first time to analyse with great accuracy the frequency of expression of genes in bacteria growing either extracellularly or following infection. In order to analyse the results of next generation sequencing, development of new software was necessary, because the small number of existing tools did meet the requirements considering flexibility and transparency. The resulting platform called *sncRAS* allows processing of a wide range of NGS technologies such as 454 Roche, Illu-

mina or Ion Torrent. Sequence data can be, depending on the condition, analysed via different algorithms and internal pipelines. The platform is based on a central database where interim stages of the analysis are stored to enable fast and straightforward re-analysis using different parameters. This platform was conceived as a basis for future sequencing projects and was designed to keep up with the rapid developments in the area of DNA and RNA sequencing.

Implementation of new hardware with enhanced processing abilities was required for analysis of data from NGS based sequencing technologies. To enable this, a server with 50 GBs of memory, a TESLA C2050 graphics processing unit, 24 cores and a large raid array with around 12 TB of disk space was installed. Based on the development of the platform *sncRAS* it was possible for the first time to explore the intracellular transcriptome of *Listeria monocytogenes*. The result of the study was the identification of 71 as hitherto unknown small regulatory RNAs of which 29 are specifically expressed during intracellular growth. Several sRNA candidates were confirmed by northern blots and additionally, 12 deletion mutants were produced of which three showed an attenuated phenotype during intracellular growth. These deletion mutants are the first documented specific small regulatory RNA mutants produced that have an important role during intracellular growth. The major challenge in this study was the development of analysis tools to analyse high throughput NGS data to enable data interpretation in a reproducible and timely way. The gained advantage in terms of quality, accuracy and novelty created new challenges for interpretation of the data. Because the intracellular transcriptome can now be mapped at the individual nucleotide level, detailed characterization of promoter and termination structures became feasible, e.g. in determining binding sites for transcriptional factors or repressors that are commonly regulated. Further higher order information can now be generated by linking commonly regulated genes to pathways in the context of bacterial physiology.

A common platform for the analysis of whole genome transcriptomes of important gram-positive pathogens

The bioinformatics platforms developed for *Listeria monocytogenes* were used to analyse whole genome transcriptomes of major gram-positive pathogens such as *Staphylococcus aureus*, *Streptococcus pyogenes*, *Enterococcus faecalis*, and *Clostridium difficile* under different growth conditions. The major focus of this on-going study is to catalogue and

describe the role of small non-coding RNAs in these bacteria during stress and infection and to develop novel diagnostic systems for the detection of these major pathogens.

Even though the bioinformatics pipelines developed in this study were focused on examining and understanding the lifestyle of *L. monocytogenes* under different growth conditions, programs developed during this study can be utilized as general tools for whole genome and transcriptome analysis of any microorganism. They provide the base for further development of bioinformatics pipelines that examine higher order expression and interactions of genes and their products e.g. by defining common regulons or interacting proteins within a signal transduction pathway. By extension, these programs can be further developed to examine complex communities by correlating metagenomes and metatranscriptomes.

7 Summary

The rapid technical development in recent years within the scope of sequence detection ranging from microarray via tilingarray to direct RNA sequencing enables new insights into gene expression as well as gene regulation of hitherto unparalleled accuracy and quality.

This thesis chronologically describes the use of currently available technologies to analyse the intracellular transcriptome of pathogenic gram-positive germs, especially *Listeria monocytogenes* EGD-e. Necessary adaptations and recent developments of the bioinformatics workflows facilitated, among other things, comparative analysis of extra- and intracellular transcription profiles to identify specific adaptations for intracellular survival of bacteria.

For this purpose, a sequence of operations composed of existing and new software has been developed to ensure a standardized procedure for microarray analysis. Concerning this, the MARS Suite was extended, with the result that MARS II, in combination with MARSlab, is capable of uptake and processing of raw data, statistics, analysis, archiving and publication. Several studies were published using this framework.

In search for specific adaptations to the survival under intracellular conditions, transcription of intergenic regions was investigated by whole genome tiling arrays. A further development of the existing microarray workflow permitted insights into the regulation of intergenic regions of *L. monocytogenes*. For the first time, small non-coding RNAs, large antisense transcripts as well as long untranslated regions were detected under intracellular conditions based on this technology.

Due to the technically limited resolution and the uncertainty of results, the first intracellular transcriptome of the pathogen *L. monocytogenes* has been sequenced with 454 Life Sciences technology. A completely new workflow had to be developed for the analysis of this new technology. The resulting software, sncRAS (small non-coding RNA analysis suite) is able to process information from several sequencing technologies, performs quality controls and enhances sequence quality through e.g. sequence trimming. Furthermore sncRAS implements a variety of algorithms for mapping reads against reference genomes as well as for the execution of different analyses and statistics. These supports the preparation of lists of sRNA candidates assisted by further data such as promoters, terminators, sRNA predictions, and experimentally verified candidates. Additionally, sncRAS allows the generation of several listings and graphics to assist the interpretation of the data, as well as the export of experimental data to public databases such as Array Express to

facilitate the publication of the data. This workflow was successfully deployed in the publication of intracellular small RNAs.

Using genomic as well as transcriptomic analyses of four intracellular *Listeria* strains representing all three lineages by employing the MARS II pipeline, the major differences between extra- and intracellular growth could be obtained. Initially, complete sequencing and annotation of strain *L. monocytogenes* 4a L99 were conducted. Bioinformatic analysis found that a reduction or loss of both, virulence genes as well as surface proteins, has an attenuating effect on the intracellular survival. In addition, the loss of several repeats within the *actA* gene leads to a deterioration of mobility which in turn results in a reduced intracellular survival rate. In contrast, duplication of PTS transporters and presence of different prophages had a positive impact on intracellular growth. Furthermore, a switch of the metabolic flow from regular glycolysis towards the pentose phosphate metabolism may serve multiple purposes including the production of NADPH countering oxidative stress.

Another challenge was presented by the investigation of gene regulation of bacteria, which were extracted from IFN-γ activated macrophages. Due to the experimental setup, resulting signals were recognized near the detection limit. The analysis revealed a significant shift on the transcriptional level to counter oxidative as well as nitrosative stress in combination with an increased demand for tryptophan during replication.

In the following course of the dissertation, the intracellular intergenic transcriptome of *L. monocytogenes* was examined in detail for the first time using modern RNA-Seq technology. As a result, 71 previously unknown sRNAs, of which 29 were expressed specifically intracellular, could be added to the collection. Based on phenotypic studies of 12 deletion mutants a strong influence of sRNAs upon intracellular survival could be demonstrated.

In conclusion, bioinformatic workflows developed in this work can be used as a general tool for the analysis of genomes and transcriptomes as well as for the special application in search of new structures such as small non-coding RNAs.

8 Zusammenfassung

Die rasante technische Entwicklung der letzten Jahre im Bereich der Sequenzdetektion mit Hilfe von Mikroarrays über Tilingarrays bis hin zur RNA Sequenzierung ermöglicht neue Einblicke in die Genexpression sowie Genregulation von bislang unerreichter Sensitivität. Diese Dissertation beschreibt chronologisch die Nutzung der jeweils aktuellen Technologie zur Analyse des intrazellulären Transkriptomes von gram-positiven Keimen, im speziellen von *Listeria monocytogenes* EGD-e. Die hierfür notwendigen Anpassungen und Neuentwicklungen der bioinformatischen Workflows ermöglichten u.a. komparative Analysen von extrazellulären und intrazellulären Transkriptionsprofilen, um spezifische Anpassungen an das intrazelluläre Überleben von Bakterien zu identifizieren.

Zu diesem Zweck wurde zunächst ein Arbeitsablauf, bestehend aus vorhandener und eigens programmierter Software erstellt, um eine standardisierte Analyse von Microarray Experimenten zu gewährleisten. Hierzu wurde weitere Software auf der Basis der MARS Suite entwickelt, so dass MARS II in Kombination mit dem MARSlab alle Funktionen von der Aufnahme der Rohdaten über Bearbeitung, Statistik und Analyse bis hin zur Archivierung und Publikation der Daten zur Verfügung stellen kann. In diesem Rahmen wurden bis heute mehrere Experimente erfolgreich publiziert.

Auf der Suche nach spezifischen Anpassungen an das Überleben unter intrazellulären Bedingungen wurde als nächster Schritt die Transkription intergenischer Bereiche mit Hilfe von Tilingarrays untersucht. Eine Neu- bzw. Weiterentwicklung des vorhandenen Microarray-Arbeitsablaufs ermöglichte erste Einblicke in die Regulation der intergenischen Bereiche von *Listeria monocytogenes*. Mit Hilfe dieser Technologie konnten erstmals kleine nicht-codierende RNAs, große antisense Transkripte sowie lange UTRs unter intrazellulären Bedingungen nachgewiesen werden.

Bedingt durch die technische Unschärfe der Microarray-Ergebnisse wurde kurze Zeit später das erste intrazelluläre Transkriptom von *L. monocytogenes* mit Hilfe der 454 Life Science Technologie sequenziert. Für die Analyse dieser innovativen Technologie bedurfte es einer neuen Software (sncRAS), welche den Anforderungen in Bezug auf Informationsdichte und Interpretation gerecht werden konnte. SncRAS ist in der Lage, Ergebnisse unterschiedlichster Sequenziertechnologien aufzunehmen, Qualitätskontrollen durchzuführen und die Sequenzen mit Hilfe verschiedener Algorithmen gegen das Referenzgenom auszurichten. Zusätzlich erlaubt sncRAS die Durchführung von Analysen und Statistiken, sowie Erstellung von Kandidatenlisten unter Zuhilfenahme weiterer Daten

(Promotoren, Terminatoren, Vorhersagen, verifizierte Ergebnisse etc.). Weiterhin ist die Ausgabe von Listen und Grafiken möglich, sowie der Export von Experimenten in öffentliche Datenbanken wie z.B. ArrayExpress um die Publikation der Daten zu vereinfachen. Dieser Workflow wurde erfolgreich in der Publikation intrazellulärer kleiner RNAs eingesetzt. Mit Hilfe von genomischen und transkriptomischen Analysen von vier intrazellulär wachsenden *Listerien* aller drei Abstammungslinien, konnten die wesentlichen Unterschiede zwischen extrazellulärem und intrazellulärem Wachstum ermittelt werden. Zunächst wurde hierzu das Genom des Stammes *L. monocytogenes* 4a L99 vollständig sequenziert und annotiert. Eine komparative bioinformatische Analyse zeigte, dass eine Reduktion bzw. Verlust von sowohl Virulenz-Genen als auch Oberflächenprotein-kodierenden Genen einen großen Einfluss auf das intrazelluläre Überleben hat. Des Weiteren konnte der Verlust von Sequenzwiederholungen innerhalb des *actA*-Gens und eine somit einhergehende Verschlechterung der Mobilität auf eine verminderte intrazelluläre Überlebensrate gezeigt werden. Im Gegensatz dazu haben sich Vervielfältigungen von PTS-Transportern und die Anwesenheit unterschiedlicher Reste von Bakteriophagen positiv auf die intrazelluläre Überlebensrate ausgewirkt. Darüber hinaus konnte ein Umschalten des metabolischen Flusses von regulärer Glykolyse hin zum Pentosephosphat Metabolismus beobachtet werden, was zusätzlich, durch Herstellung von NADPH oxidativem Stress entgegen wirkt.

Eine Herausforderung bestand in der weiteren Untersuchung der Genregulation von Bakterien, welche aus IFN-γ aktivierten Makrophagen extrahiert wurden. Die resultierenden Signale bewegten sich, bedingt durch den Aufbau des Experiments, nahe der Nachweisgrenze. Die Analyse zeigte eine deutliche Reaktion auf Transkriptionsebene um dem aufkommenden oxidativen Stress entgegenzuwirken, sowie einen gesteigerten Bedarf an Tryptophan während der Replikation.

Im weiteren Verlauf der Arbeit wurde erstmals mit Hilfe moderner RNA-Seq Technologie das intrazelluläre intergenische Transkriptom von *L. monocytogenes* EGD-e näher untersucht. Als Ergebnis konnten 71 bis dato unbekannte sRNAs, von denen 29 spezifisch intrazellulär exprimiert wurden, identifiziert werden. Anhand von 12 hergestellten Deletionsmutanten konnte, nach phänotypischen Untersuchungen, ein starker Einfluss von sRNAs auf das intrazelluläre Überleben nachgewiesen werden.

Zusammenfassend lässt sich also sagen, dass die in dieser Arbeit entwickelten Bioinformatischen Programme und Abläufe sowohl zur Analyse von Genomen und Transkriptomen sämtlicher Bakterien, als auch speziell für die Suche von neuen Strukturen wie z.B. sRNAs verwendet werden können.

9 Publications (2010 - 2012)

Thesis:

1.

Torsten Hain*, Rohit Ghai*, André Billion*, Carsten Tobias Kuenne*, Christiane Steinweg, Benjamin Izar, Walid Mohamed, Mobarak Abu Mraheil, EugenDomann, Silke Schaffrath, Uwe Kärst, Alexander Goesmann, Sebastian Oehm, Alfred Pühler, Rainer Merkl, Sonja Vorwerk, Philippe Glaser, Patricia Garrido, Christophe Rusniok, Carmen Buchrieser, Werner Goebel and Trinad Chakraborty

Comparative genomics and transcriptomics of lineages I, II, and III strains of *Listeria monocytogenes*. *BMC Genomics*. 2012 Apr 24;13:144.PMID: 22530965.

2.

Mraheil MA*, Billion A, Mohamed W, Rawool D, Hain T, Chakraborty T. Adaptation of Listeria monocytogenes to oxidative and nitrosative stress in IFN-y-activated macrophages. *Int J Med Microbiol.* 2011 Nov;301(7):547-55. doi: 10.1016/j.ijmm.2011.05.001. Epub 2011 Jun 21. PubMed PMID: 21697010.

3.

Mraheil MA*, Billion A*, Mohamed W, Mukherjee K, Kuenne C, Pischimarov J, Krawitz C, Retey J, Hartsch T, Chakraborty T, Hain T. The intracellular sRNA transcriptome of Listeria monocytogenes during growth in macrophages. *Nucleic Acids Res.* 2011 May;39(10):4235-48. Epub 2011 Jan 29. PubMed PMID: 21278422;PubMed Central PMCID: PMC3105390.

4.

Mraheil MA*, Billion A*, Kuenne C, Pischimarov J, Kreikemeyer B, Engelmann S, Hartke A, Giard JC, Rupnik M, Vorwerk S, Beier M, Retey J, Hartsch T, Jacob A, Cemiz F, Hemberger J, Chakraborty T, Hain T. Comparative genome-wide analysis of small RNAs of major Gram-positive pathogens: from identification to application. *MicrobBiotechnol.* 2010 Nov;3(6):658-76. doi: 10.1111/j.1751-7915.2010.00171.x. Review. PubMed PMID: 21255362.

Additional Publications: (Appendix)

5.

Kuenne C*, Billion A*, Mshana SE, Schmiedel J, Domann E, Hossain H, Hain T, Imirzalioglu C, Chakraborty T. Complete Sequences of Plasmids from the Hemolytic-Uremic Syndrome-Associated Escherichia coli Strain HUSEC41. *JBacteriol.* 2012 Jan;194(2):532-3. PubMed PMID: 22207742.

Contribution:

The author (A.B.) gave conceptual advice in the sequencing of the four plasmids and supervised the project. Furthermore A.B. was, together with C.K., involved in gap closure, annotation and comparative analysis of outlined plasmids. A.B. additionally assisted in writing the manuscript.

6.

Imirzalioglu C*, Dahmen H, Hain T, Billion A, Kuenne C, Chakraborty T, Domann E. Highly specific and quick detection of Mycobacterium avium subsp. paratuberculosis in feces and gut tissue of cattle and humans by multiple real-time PCR assays. *J ClinMicrobiol.* 2011 May;49(5):1843-52. Epub 2011 Mar 23. PubMed PMID: 21430100; PubMed Central PMCID: PMC3122678.

Contribution:

The author (A.B.) conducted all bioinformatics analyses such as searching and testing of specific probes, comparative analysis and preparation of representative figures for publication.

7.

Seifart Gomes C*, Izar B, Pazan F, Mohamed W, Mraheil MA, Mukherjee K, Billion A, Aharonowitz Y, Chakraborty T, Hain T. Universal stress proteins are important for oxidative and acid stress resistance and growth of *Listeria monocytogenes* EGD-e in vitro and in vivo. *PLoS One.* 2011;6(9):e24965. Epub 2011 Sep 30. PubMed PMID: 21980369; PubMed Central PMCID: PMC3184099.

Contribution:

The author (A.B.) conducted bioinformatics analyses and provide technical IT support for the project. Furthermore he discussed the results and implications and commented on the manuscript.

8.

Steinweg C*, Kuenne CT*, Billion A, Mraheil MA, Domann E, Ghai R, Barbuddhe SB, Kärst U, Goesmann A, Pühler A, Weisshaar B, Wehland J, Lampidis R, KreftJ, Goebel W, Chakraborty T, Hain T. Complete genome sequence of Listeria seeligeri, a nonpathogenic member of the genus Listeria. *J Bacteriol.* 2010Mar;192(5):1473-4. Epub 2010 Jan 8. PubMed PMID: 20061480; PubMed Central PMCID: PMC2820865.

Contribution:

The author (A.B.) performed bioinformatics analysis such as comparative surface protein prediction using AUGUR, preparation of AUGUR for COG classification and assists with comparative analysis on related species using GeCo server. Furthermore A.B. was involved in annotation of genome sequence, discussed results and assists writing the manuscript.

9.

Pischimarov J*, Kuenne C*, Billion A, Hemberger J, Cemic F, Chakraborty T and Hain T. sRNAdb: A small non-coding RNA database for gram-positive bacteria.*BMC Genomics* 2012, **13**:384 doi:10.1186/1471-2164-13-384 Published: 10 August 2012

Contribution:

The author (A.B.) discussed the results and implications and commented on the manuscript at all stages. A.B. provided extensive information about sRNAs such as lists of experimentally verified candidates, software for candidate as well as target prediction.

10.

Kuenne C*, Billion A, Mraheil M, Strittmatter A, Daniel R, Goesmann A, Barbuddhe S, Hain T, Chakraborty T. Dynamic integration hotspots and mobile genetic elements shape the genome structure of the species *Listeria monocytogenes*.
(Submitted to BMC Genomics)

Contribution:

The author (A.B.) prepared and executed AUGUR software to identify all surface-associated proteins and performed a first pass comparative analysis. A.B. contributed in CRISPR/Cas analysis while identifying repeat/spacer arrays via PILER-CR and CRT software and further identification of phage sequences using BlastIt. A.B. discussed the results and implications and commented on the manuscript at all stages.

10 References

1. Waters LS, Storz G: **Regulatory RNAs in bacteria**. *Cell* 2009, **136**(4):615-628.
2. Sorek R, Cossart P: **Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity**. *Nat Rev Genet* 2010, **11**(1):9-16.
3. Chatterjee SS, Otten S, Hain T, Lingnau A, Carl UD, Wehland J, Domann E, Chakraborty T: **Invasiveness is a variable and heterogeneous phenotype in Listeria monocytogenes serotype strains**. *Int J Med Microbiol* 2006, **296**(4-5):277-286.
4. Mraheil MA, Billion A, Mohamed W, Rawool D, Hain T, Chakraborty T: **Adaptation of Listeria monocytogenes to oxidative and nitrosative stress in IFN-gamma-activated macrophages**. *Int J Med Microbiol* 2011, **301**(7):547-555.
5. van der Veen S, Hain T, Wouters JA, Hossain H, de Vos WM, Abee T, Chakraborty T, Wells-Bennik MH: **The heat-shock response of Listeria monocytogenes comprises genes involved in heat shock, cell division, cell wall synthesis, and the SOS response**. *Microbiology* 2007, **153**(Pt 10):3593-3607.
6. Mukherjee K, Altincicek B, Hain T, Domann E, Vilcinskas A, Chakraborty T: **Galleria mellonella as a model system for studying Listeria pathogenesis**. *Appl Environ Microbiol* 2010, **76**(1):310-317.
7. Chatterjee SS, Hossain H, Otten S, Kuenne C, Kuchmina K, Machata S, Domann E, Chakraborty T, Hain T: **Intracellular gene expression profile of Listeria monocytogenes**. *Infect Immun* 2006, **74**(2):1323-1338.
8. Hain T, Hossain H, Chatterjee SS, Machata S, Volk U, Wagner S, Brors B, Haas S, Kuenne CT, Billion A *et al*: **Temporal transcriptomic analysis of the Listeria monocytogenes EGD-e sigmaB regulon**. *BMC Microbiol* 2008, **8**:20.
9. Guell M, van Noort V, Yus E, Chen WH, Leigh-Bell J, Michalodimitrakis K, Yamada T, Arumugam M, Doerks T, Kuhner S *et al*: **Transcriptome complexity in a genome-reduced bacterium**. *Science* 2009, **326**(5957):1268-1271.
10. Horvath P, Coute-Monvoisin AC, Romero DA, Boyaval P, Fremaux C, Barrangou R: **Comparative analysis of CRISPR loci in lactic acid bacteria genomes**. *Int J Food Microbiol* 2009, **131**(1):62-70.
11. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome**. *Genome Biol* 2009, **10**(3):R25.
12. Hoffmann S, Otto C, Kurtz S, Sharma CM, Khaitovich P, Vogel J, Stadler PF, Hackermuller J: **Fast mapping of short sequences with mismatches, insertions and deletions using index structures**. *PLoS Comput Biol* 2009, **5**(9):e1000502.
13. Vazquez-Boland JA, Kuhn M, Berche P, Chakraborty T, Dominguez-Bernal G, Goebel W, Gonzalez-Zorn B, Wehland J, Kreft J: **Listeria pathogenesis and molecular virulence determinants**. *Clin Microbiol Rev* 2001, **14**(3):584-640.
14. Duffy LL, Vanderlinde PB, Grau FH: **Growth of Listeria monocytogenes on vacuum-packed cooked meats: effects of pH, aw, nitrite and ascorbate**. *Int J Food Microbiol* 1994, **23**(3-4):377-390.
15. Maurer RH, Roth DR, Kinnison JD, Goldsten JO, Gold RE, Fainchtein R: **MArs Neutron Energy Spectrometer (MANES): an instrument for the Mars 2003 Lander**. *Acta Astronaut* 2003, **52**(2-6):405-410.
16. Kingsford CL, Ayanbule K, Salzberg SL: **Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake**. *Genome Biol* 2007, **8**(2):R22.

17. Livny J, Teonadi H, Livny M, Waldor MK: **High-throughput, kingdom-wide prediction and annotation of bacterial non-coding RNAs**. *PLoS One* 2008, **3**(9):e3197.
18. Sridhar J, Sambaturu N, Sabarinathan R, Ou HY, Deng Z, Sekar K, Rafi ZA, Rajakumar K: **sRNAscanner: a computational tool for intergenic small RNA detection in bacterial genomes**. *PLoS One* 2010, **5**(8):e11970.
19. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP: **Integrative genomics viewer**. *Nat Biotechnol* 2011, **29**(1):24-26.
20. Okuda S, Yamada T, Hamajima M, Itoh M, Katayama T, Bork P, Goto S, Kanehisa M: **KEGG Atlas mapping for global analysis of metabolic pathways**. *Nucleic Acids Res* 2008, **36**(Web Server issue):W423-426.
21. Ewing B, Hillier L, Wendl MC, Green P: **Base-calling of automated sequencer traces using phred. I. Accuracy assessment**. *Genome Res* 1998, **8**(3):175-185.
22. Gordon D, Abajian C, Green P: **Consed: a graphical tool for sequence finishing**. *Genome Res* 1998, **8**(3):195-202.
23. Hain T, Steinweg C, Kuenne CT, Billion A, Ghai R, Chatterjee SS, Domann E, Karst U, Goesmann A, Bekel T *et al*: **Whole-genome sequence of Listeria welshimeri reveals common steps in genome reduction with Listeria innocua as compared to Listeria monocytogenes**. *J Bacteriol* 2006, **188**(21):7405-7415.
24. Glaser P, Frangeul L, Buchrieser C, Rusniok C, Amend A, Baquero F, Berche P, Bloecker H, Brandt P, Chakraborty T *et al*: **Comparative genomics of Listeria species**. *Science* 2001, **294**(5543):849-852.
25. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL: **Versatile and open software for comparing large genomes**. *Genome Biol* 2004, **5**(2):R12.
26. Ghai R, Hain T, Chakraborty T: **GenomeViz: visualizing microbial genomes**. *BMC Bioinformatics* 2004, **5**:198.
27. Edgar RC: **PILER-CR: fast and accurate identification of CRISPR repeats**. *BMC Bioinformatics* 2007, **8**:18.
28. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, Hugenholtz P: **CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats**. *BMC Bioinformatics* 2007, **8**:209.
29. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J: **ACT: the Artemis Comparison Tool**. *Bioinformatics* 2005, **21**(16):3422-3423.
30. Kuenne CT, Ghai R, Chakraborty T, Hain T: **GECO--linear visualization for comparative genomics**. *Bioinformatics* 2007, **23**(1):125-126.
31. Bolstad BM, Irizarry RA, Astrand M, Speed TP: **A comparison of normalization methods for high density oligonucleotide array data based on variance and bias**. *Bioinformatics* 2003, **19**(2):185-193.
32. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response**. *Proc Natl Acad Sci U S A* 2001, **98**(9):5116-5121.
33. Sokolovic Z, Schuller S, Bohne J, Baur A, Rdest U, Dickneite C, Nichterlein T, Goebel W: **Differences in virulence and in expression of PrfA and PrfA-regulated virulence genes of Listeria monocytogenes strains belonging to serogroup 4**. *Infect Immun* 1996, **64**(10):4008-4019.
34. Cabanes D, Dussurget O, Dehoux P, Cossart P: **Auto, a surface associated autolysin of Listeria monocytogenes required for entry into eukaryotic cells and virulence**. *Mol Microbiol* 2004, **51**(6):1601-1614.

35. Cabanes D, Sousa S, Cebria A, Lecuit M, Garcia-del Portillo F, Cossart P: **Gp96 is a receptor for a novel Listeria monocytogenes virulence factor, Vip, a surface protein**. *EMBO J* 2005, **24**(15):2827-2838.
36. Severino P, Dussurget O, Vencio RZ, Dumas E, Garrido P, Padilla G, Piveteau P, Lemaitre JP, Kunst F, Glaser P *et al*: **Comparative transcriptome analysis of Listeria monocytogenes strains of the two major lineages reveals differences in virulence, cell wall, and stress response**. *Appl Environ Microbiol* 2007, **73**(19):6078-6088.
37. Joseph B, Przybilla K, Stuhler C, Schauer K, Slaghuis J, Fuchs TM, Goebel W: **Identification of Listeria monocytogenes genes contributing to intracellular replication by expression profiling and mutant screening**. *J Bacteriol* 2006, **188**(2):556-568.
38. Ralser M, Wamelink MM, Kowald A, Gerisch B, Heeren G, Struys EA, Klipp E, Jakobs C, Breitenbach M, Lehrach H *et al*: **Dynamic rerouting of the carbohydrate flux is key to counteracting oxidative stress**. *J Biol* 2007, **6**(4):10.
39. Orsi RH, Borowsky ML, Lauer P, Young SK, Nusbaum C, Galagan JE, Birren BW, Ivy RA, Sun Q, Graves LM *et al*: **Short-term genome evolution of Listeria monocytogenes in a non-controlled environment**. *BMC Genomics* 2008, **9**:539.
40. Verghese B, Lok M, Wen J, Alessandria V, Chen Y, Kathariou S, Knabel S: **comK prophage junction fragments as markers for Listeria monocytogenes genotypes unique to individual meat and poultry processing plants and a model for rapid niche-specific adaptation, biofilm formation, and persistence**. *Appl Environ Microbiol* 2011, **77**(10):3279-3292.
41. Denny J, McLauchlin J: **Human Listeria monocytogenes infections in Europe--an opportunity for improved European surveillance**. *Euro Surveill* 2008, **13**(13).
42. Liu D, Lawrence ML, Wiedmann M, Gorski L, Mandrell RE, Ainsworth AJ, Austin FW: **Listeria monocytogenes subgroups IIIA, IIIB, and IIIC delineate genetically distinct populations with varied pathogenic potential**. *J Clin Microbiol* 2006, **44**(11):4229-4233.
43. van Dissel JT, Stikkelbroeck JJ, Michel BC, van den Barselaar MT, Leijh PC, van Furth R: **Inability of recombinant interferon-gamma to activate the antibacterial activity of mouse peritoneal macrophages against Listeria monocytogenes and Salmonella typhimurium**. *J Immunol* 1987, **139**(5):1673-1678.
44. Langermans JA, van der Hulst ME, Nibbering PH, van der Meide PH, van Furth R: **Intravenous injection of interferon-gamma inhibits the proliferation of Listeria monocytogenes in the liver but not in the spleen and peritoneal cavity**. *Immunology* 1992, **77**(3):354-361.
45. Herskovits AA, Auerbuch V, Portnoy DA: **Bacterial ligands generated in a phagosome are targets of the cytosolic innate immune system**. *PLoS Pathog* 2007, **3**(3):e51.
46. Setsukinai K, Urano Y, Kakinuma K, Majima HJ, Nagano T: **Development of novel fluorescence probes that can reliably detect reactive oxygen species and distinguish specific species**. *J Biol Chem* 2003, **278**(5):3170-3175.
47. Oh TJ, Kim IG: **The expression of Escherichia coli SOS genes recA and uvrA is inducible by polyamines**. *Biochem Biophys Res Commun* 1999, **264**(2):584-589.
48. Rea RB, Gahan CG, Hill C: **Disruption of putative regulatory loci in Listeria monocytogenes demonstrates a significant role for Fur and PerR in virulence**. *Infect Immun* 2004, **72**(2):717-727.

49. Gaballa A, Helmann JD: **A peroxide-induced zinc uptake system plays an important role in protection against oxidative stress in Bacillus subtilis**. *Mol Microbiol* 2002, **45**(4):997-1005.
50. Schapiro JM, Libby SJ, Fang FC: **Inhibition of bacterial DNA replication by zinc mobilization during nitrosative stress**. *Proc Natl Acad Sci U S A* 2003, **100**(14):8496-8501.
51. Valentin-Hansen P, Eriksen M, Udesen C: **The bacterial Sm-like protein Hfq: a key player in RNA transactions**. *Mol Microbiol* 2004, **51**(6):1525-1533.
52. Aiba H: **Mechanism of RNA silencing by Hfq-binding small RNAs**. *Curr Opin Microbiol* 2007, **10**(2):134-139.
53. Toledo-Arana A, Dussurget O, Nikitas G, Sesto N, Guet-Revillet H, Balestrino D, Loh E, Gripenland J, Tiensuu T, Vaitkevicius K *et al*: **The Listeria transcriptional landscape from saprophytism to virulence**. *Nature* 2009, **459**(7249):950-956.
54. Oliver HF, Orsi RH, Ponnala L, Keich U, Wang W, Sun Q, Cartinhour SW, Filiatrault MJ, Wiedmann M, Boor KJ: **Deep RNA sequencing of L. monocytogenes reveals overlapping and extensive stationary phase and sigma B-dependent transcriptomes, including multiple highly transcribed noncoding RNAs**. *BMC Genomics* 2009, **10**:641.
55. Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J, Kalinowski J, Linke B, Rupp O, Giegerich R *et al*: **GenDB--an open source genome annotation system for prokaryote genomes**. *Nucleic Acids Res* 2003, **31**(8):2187-2195.
56. Mandin P, Repoila F, Vergassola M, Geissmann T, Cossart P: **Identification of new noncoding RNAs in Listeria monocytogenes and prediction of mRNA targets**. *Nucleic Acids Res* 2007, **35**(3):962-974.
57. Castillo-Keller M, Vuong P, Misra R: **Novel mechanism of Escherichia coli porin regulation**. *J Bacteriol* 2006, **188**(2):576-586.
58. Camejo A, Buchrieser C, Couve E, Carvalho F, Reis O, Ferreira P, Sousa S, Cossart P, Cabanes D: **In vivo transcriptional profiling of Listeria monocytogenes and mutagenesis identify new virulence factors involved in infection**. *PLoS Pathog* 2009, **5**(5):e1000449.
59. Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermuller J, Reinhardt R *et al*: **The primary transcriptome of the major human pathogen Helicobacter pylori**. *Nature* 2010, **464**(7286):250-255.
60. Georg J, Voss B, Scholz I, Mitschke J, Wilde A, Hess WR: **Evidence for a major role of antisense RNAs in cyanobacterial gene regulation**. *Mol Syst Biol* 2009, **5**:305.
61. Liu JM, Livny J, Lawrence MS, Kimball MD, Waldor MK, Camilli A: **Experimental discovery of sRNAs in Vibrio cholerae by direct cloning, 5S/tRNA depletion and parallel sequencing**. *Nucleic Acids Res* 2009, **37**(6):e46.
62. Romby P, Charpentier E: **An overview of RNAs with regulatory functions in gram-positive bacteria**. *Cell Mol Life Sci* 2010, **67**(2):217-237.
63. Wurtzel O, Sapra R, Chen F, Zhu Y, Simmons BA, Sorek R: **A single-base resolution map of an archaeal transcriptome**. *Genome Res* 2010, **20**(1):133-141.
64. Phan TT, Schumann W: **Transcriptional analysis of the lysine-responsive and riboswitch-regulated lysC gene of Bacillus subtilis**. *Curr Microbiol* 2009, **59**(4):463-468.
65. Loh E, Dussurget O, Gripenland J, Vaitkevicius K, Tiensuu T, Mandin P, Repoila F, Buchrieser C, Cossart P, Johansson J: **A trans-acting riboswitch controls expression of the virulence regulator PrfA in Listeria monocytogenes**. *Cell* 2009, **139**(4):770-779.

66. Garzon R, Calin GA, Croce CM: **MicroRNAs in Cancer**. *Annu Rev Med* 2009, **60**:167-179.
67. Negrini M, Nicoloso MS, Calin GA: **MicroRNAs and cancer--new paradigms in molecular oncology**. *Curr Opin Cell Biol* 2009, **21**(3):470-479.
68. Storz G, Opdyke JA, Zhang A: **Controlling mRNA stability and translation with small, noncoding RNAs**. *Curr Opin Microbiol* 2004, **7**(2):140-144.
69. Sharma CM, Vogel J: **Experimental approaches for the discovery and characterization of regulatory small RNA**. *Curr Opin Microbiol* 2009, **12**(5):536-546.
70. Padalon-Brauch G, Hershberg R, Elgrably-Weiss M, Baruch K, Rosenshine I, Margalit H, Altuvia S: **Small RNAs encoded within genetic islands of Salmonella typhimurium show host-induced expression and role in virulence**. *Nucleic Acids Res* 2008, **36**(6):1913-1927.
71. Christiansen JK, Nielsen JS, Ebersbach T, Valentin-Hansen P, Sogaard-Andersen L, Kallipolitis BH: **Identification of small Hfq-binding RNAs in Listeria monocytogenes**. *RNA* 2006, **12**(7):1383-1396.
72. Nielsen JS, Olsen AS, Bonde M, Valentin-Hansen P, Kallipolitis BH: **Identification of a sigma B-dependent small noncoding RNA in Listeria monocytogenes**. *J Bacteriol* 2008, **190**(18):6264-6270.
73. Pichon C, Felden B: **Small RNA genes expressed from Staphylococcus aureus genomic and pathogenicity islands with specific expression among pathogenic strains**. *Proc Natl Acad Sci U S A* 2005, **102**(40):14249-14254.
74. Kreikemeyer B, Boyle MD, Buttaro BA, Heinemann M, Podbielski A: **Group A streptococcal growth phase-associated virulence factor regulation by a novel operon (Fas) with homologies to two-component-type regulators requires a small RNA molecule**. *Mol Microbiol* 2001, **39**(2):392-406.
75. Mangold M, Siller M, Roppenser B, Vlaminckx BJ, Penfound TA, Klein R, Novak R, Novick RP, Charpentier E: **Synthesis of group A streptococcal virulence factors is controlled by a regulatory RNA molecule**. *Mol Microbiol* 2004, **53**(5):1515-1527.
76. Halfmann A, Kovacs M, Hakenbeck R, Bruckner R: **Identification of the genes directly controlled by the response regulator CiaR in Streptococcus pneumoniae: five out of 15 promoters drive expression of small non-coding RNAs**. *Mol Microbiol* 2007, **66**(1):110-126.
77. Hof H, Szabo K, Becker B: **[Epidemiology of listeriosis in Germany: a changing but ignored pattern]**. *Dtsch Med Wochenschr* 2007, **132**(24):1343-1348.
78. Hain T, Chatterjee SS, Ghai R, Kuenne CT, Billion A, Steinweg C, Domann E, Karst U, Jansch L, Wehland J *et al*: **Pathogenomics of Listeria spp**. *Int J Med Microbiol* 2007, **297**(7-8):541-557.
79. Johansson J, Mandin P, Renzoni A, Chiaruttini C, Springer M, Cossart P: **An RNA thermosensor controls expression of virulence genes in Listeria monocytogenes**. *Cell* 2002, **110**(5):551-561.
80. Christiansen JK, Larsen MH, Ingmer H, Sogaard-Andersen L, Kallipolitis BH: **The RNA-binding protein Hfq of Listeria monocytogenes: role in stress tolerance and virulence**. *J Bacteriol* 2004, **186**(11):3355-3362.
81. Novick RP, Ross HF, Projan SJ, Kornblum J, Kreiswirth B, Moghazeh S: **Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule**. *EMBO J* 1993, **12**(10):3967-3975.
82. Carapetis JR, Steer AC, Mulholland EK, Weber M: **The global burden of group A streptococcal diseases**. *Lancet Infect Dis* 2005, **5**(11):685-694.

83. Klenk M, Koczan D, Guthke R, Nakata M, Thiesen HJ, Podbielski A, Kreikemeyer B: **Global epithelial cell transcriptional responses reveal Streptococcus pyogenes Fas regulator activity association with bacterial aggressiveness**. *Cell Microbiol* 2005, **7**(9):1237-1250.
84. Perez N, Trevino J, Liu Z, Ho SC, Babitzke P, Sumby P: **A genome-wide analysis of small regulatory RNAs in the human pathogen group A Streptococcus**. *PLoS One* 2009, **4**(11):e7668.
85. Gilmore MS, Ferretti JJ: **Microbiology. The thin line between gut commensal and pathogen**. *Science* 2003, **299**(5615):1999-2002.
86. Domann E, Hain T, Ghai R, Billion A, Kuenne C, Zimmermann K, Chakraborty T: **Comparative genomic analysis for the presence of potential enterococcal virulence factors in the probiotic Enterococcus faecalis strain Symbioflor 1**. *Int J Med Microbiol* 2007, **297**(7-8):533-539.
87. Rupnik M, Wilcox MH, Gerding DN: **Clostridium difficile infection: new developments in epidemiology and pathogenesis**. *Nat Rev Microbiol* 2009, **7**(7):526-536.
88. Bohn C, Rigoulay C, Bouloc P: **No detectable effect of RNA-binding protein Hfq absence in Staphylococcus aureus**. *BMC Microbiol* 2007, **7**:10.
89. Hindley J: **Fractionation of 32P-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting**. *J Mol Biol* 1967, **30**(1):125-136.
90. Griffin BE: **Separation of 32P-labelled ribonucleic acid components. The use of polyethylenimine-cellulose (TLC) as a second dimension in separating oligoribonucleotides of '4.5 S' and 5 S from E. coli**. *FEBS Lett* 1971, **15**(3):165-168.
91. Ikemura T, Dahlberg JE: **Small ribonucleic acids of Escherichia coli. I. Characterization by polyacrylamide gel electrophoresis and fingerprint analysis**. *J Biol Chem* 1973, **248**(14):5024-5032.
92. Lorenz C, von Pelchrzim F, Schroeder R: **Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels**. *Nat Protoc* 2006, **1**(5):2204-2212.
93. Vogel J, Bartels V, Tang TH, Churakov G, Slagter-Jager JG, Huttenhofer A, Wagner EG: **RNomics in Escherichia coli detects new sRNA species and indicates parallel transcriptional output in bacteria**. *Nucleic Acids Res* 2003, **31**(22):6435-6443.
94. Selinger DW, Cheung KJ, Mei R, Johansson EM, Richmond CS, Blattner FR, Lockhart DJ, Church GM: **RNA expression analysis using a 30 base pair resolution Escherichia coli genome array**. *Nat Biotechnol* 2000, **18**(12):1262-1268.
95. Le SV, Chen JH, Currey KM, Maizel JV, Jr.: **A program for predicting significant RNA secondary structures**. *Comput Appl Biosci* 1988, **4**(1):153-159.
96. Rivas E, Eddy SR: **Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs**. *Bioinformatics* 2000, **16**(7):583-605.
97. Rivas E, Eddy SR: **Noncoding RNA gene detection using comparative sequence analysis**. *BMC Bioinformatics* 2001, **2**:8.
98. Raasch P, Schmitz U, Patenge N, Vera J, Kreikemeyer B, Wolkenhauer O: **Non-coding RNA detection methods combined to improve usability, reproducibility and precision**. *BMC Bioinformatics* 2010, **11**:491.

99. Tjaden B: **TargetRNA: a tool for predicting targets of small RNA action in bacteria**. *Nucleic Acids Res* 2008, **36**(Web Server issue):W109-113.
100. Busch A, Richter AS, Backofen R: **IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions**. *Bioinformatics* 2008, **24**(24):2849-2856.
101. Mellbye BL, Puckett SE, Tilley LD, Iversen PL, Geller BL: **Variations in amino acid composition of antisense peptide-phosphorodiamidate morpholino oligomer affect potency against Escherichia coli in vitro and in vivo**. *Antimicrob Agents Chemother* 2009, **53**(2):525-530.
102. Tan XX, Actor JK, Chen Y: **Peptide nucleic acid antisense oligomer as a therapeutic strategy against bacterial infection: proof of principle using mouse intraperitoneal infection**. *Antimicrob Agents Chemother* 2005, **49**(8):3203-3207.
103. Carney J, Braven, H., Seal, J., and Whitworth, E.: **Present and future applications of gold in rapid assays.***IVD Technol* 2006, **12**(41–50).
104. Mao X, Ma Y, Zhang A, Zhang L, Zeng L, Liu G: **Disposable nucleic acid biosensors based on gold nanoparticle probes and lateral flow strip**. *Anal Chem* 2009, **81**(4):1660-1668.

11 Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne unzulässige Hilfe oder Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nichtveröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten sowie ethische, datenschutzrechtliche und tierschutzrechtliche Grundsätze befolgt. Ich versichere, dass Dritte von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen, oder habe diese nachstehend spezifiziert. Die vorgelegte Arbeit wurde weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde zum Zweck einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt. Alles aus anderen Quellen und von anderen Personen übernommene Material, das in der Arbeit verwendet wurde oder auf das direkt Bezug genommen wird, wurde als solches kenntlich gemacht. Insbesondere wurden alle Personen genannt, die direkt und indirekt an der Entstehung der vorliegenden Arbeit beteiligt waren. Mit der Überprüfung meiner Arbeit durch eine Plagiatserkennungssoftware bzw. ein internetbasiertes Softwareprogramm erkläre ich mich einverstanden.

Gießen, am <Datum>

12 Acknowledgements

I am indebted to many people for their long-lasting support which was of great value for me and without this encouragement this thesis wouldn't be what it is today.

Firstly, I must thank Prof. Dr. Trinad Chakraborty who initiated this thesis and gave me the opportunity to use the excellent research facilities in his institute. Furthermore, I want to express my most profound thanks to him for countless ideas and discussions and never ending assistance in nearly all affairs during this thesis.

Further on, I have to thank Dr. Torsten Hain who gave me the opportunity to participate in his project *sncRNAomics* whose results represent a large portion of this thesis. Furthermore I want to thank him for mostly fruitful discussions ☺, successful publications and all the support over the years.

My sincere thanks goes to many friends and colleagues for the consistently good support and scientific discussions always so greatly appreciated, among them I want to thank Dr. Mobarak Abu Mraheil for excellent teamwork during plenty of scientific projects, Carsten Kuenne and Anita Höhland for any number of discussions and support, and Alexandra Amend which done an inexpressible good job in the lab to provide data for bioinformatics analysis.

Last but not least, I like to thank the best (my) family on this wonderful planet for never-ending love and encouragement. Without their love, trust and support none of this could have been possible.

13 Appendix

13.1 Bioinformatics workflows technical manuals

The following paragraphs include descriptions and technical instructions regarding the bioinformatics workflows briefly introduced in the beginning of this thesis.

13.1.1 Microarray workflow

There is a great amount of management and analysis software for microarray experiments and resulting data. Finding software which complies with all requirement, and is easy to use in a formidable task. With increasing number of differently formatted experiments from several sources and the necessity of uniform formats, the importance of flexible software rises continually. Only on the basis of uniform data formats is comparative analysis from different sources possible. This data analysis has to be transparent and easy to understand because it is intended that it be used not only by computer scientists and bioinformatic personnel but also from lab personnel without any extended knowledge about computers and analysis algorithms. The MARSlab program group was developed to save microarray experiments from several sources into an uniform format at the basis of the MARS database (TU Graz) and to make transparent analysis possible where all single steps of processing are separate and all interim results can be used any desired. Experiments of several chip technologies (Affymetrix, cDNA, Oligo) from different databases (ArrayExpress, GEO – Gene expression omnibus) and local production serve as sources. The use of individual scripts in analysis preparation permits flexibility in their modification because they are only binded at runtime. At present MARSlab offers the following pre-processing functionality: Merging of two or more raw data sets, normalization, quality control, imputation, export into several formats and graphical output as scatter-, MA-, density plots and heatmaps.

Essentials:

Database: PostgreSQL

The used database management system is based on an open source PostgreSQL project. PostgreSQL comes with a broad protocol for transaction control as well as several interfaces and procedures for different programming languages.

Statistics: R and Bioconducter

The programming language R is only used by scripts executed from MARSlab software. R is an open-source scripting language mainly used for statistical analysis. Compared to other sources such as "S" or "S-Plus", R has advantages in speed and flexibility. R possess a central archive CRAN (comprehensive R archive network) where plenty of additional modules can be obtained.

Furthermore, Bioconductor provide an extensive resource of modules for specific bioinformatics tasks e.g. for the analysis of Affymetrix data or for various statistical tests.

Programming language: Java

The basis for MARS and MARSlab is the object oriented and cross-platform programming language Java developed by SUN Microsystems (taken over by Oracle 2008). Advantages of Java are the cross-platform and webstart ability which makes developed software flexible and through Java webstart secure. In recent years Java community developed a wide range of additional modules such as JavaBeans, Applets, Portlets, SOAP and RMI which makes this language even more attractive.

Interface

Microarray retrieval and storage system (MARS)

The underlying software MARS has been completely developed in Java combined with a standard interface to PostgreSQL and is organized as two independent modules. The main database of MARS consist of 191 tables and serves for the main module *retrieval and*

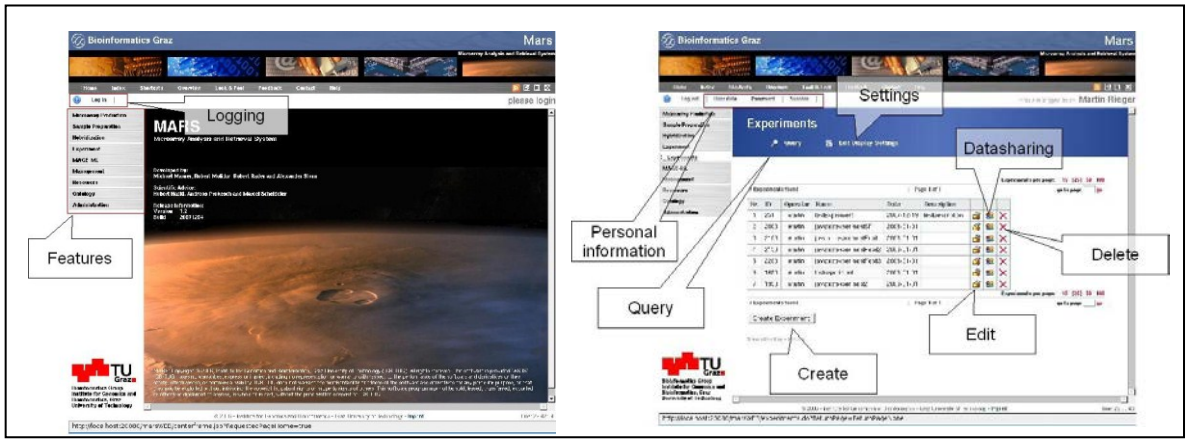


Fig.21 MARS interface

storage system as well as for the second module which is responsible for the *user management*. Both modules run independently as server processes served by a JBOSS server. They enable the user to view, organise, complete and delete stored microarray data. The

user management module makes sure that only authorized people can view, modify or delete data.

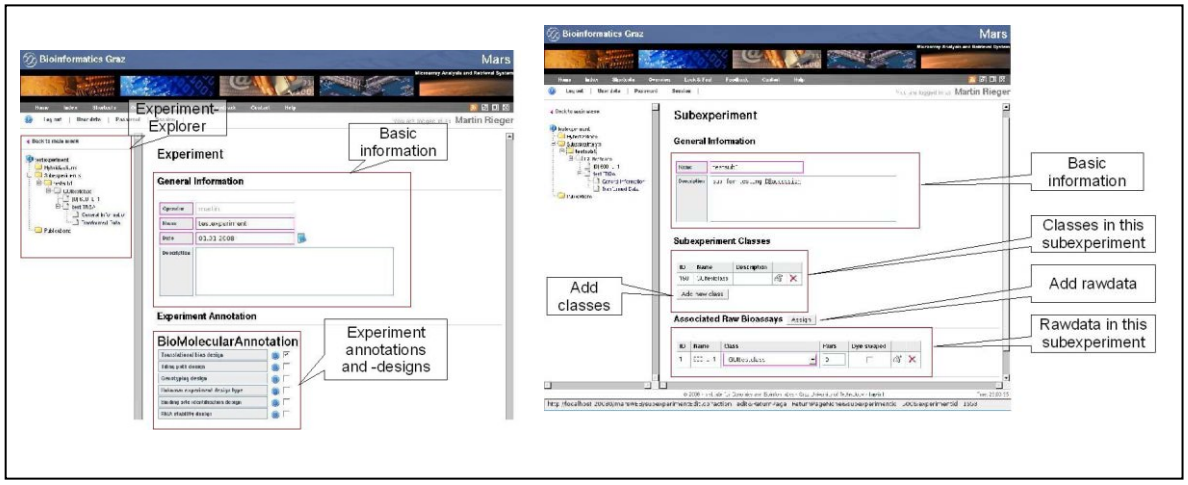
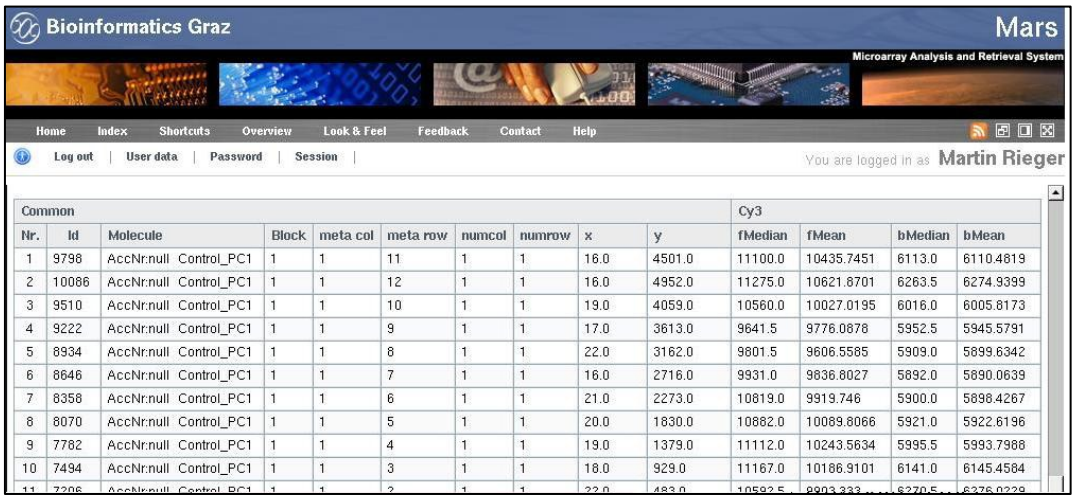


Fig.22 MARS experiment view

Each Experiment has to be created in the central experiment explorer. When creating a new experiment, several information's such as study design, protocols, quality controls, experiment parameters etc. are necessary. Once a main experiment is created, several types of data can be imported via the import function of the additional developed MARSlab tool (described later). Imported or entered data can be easily inspected and modified via the structured web interface of MARS. Even microarray raw data can be verified over the web-interface, because all information's are stored in the central database.



Common										Cy3			
Nr.	Id	Molecule	Block	meta col	meta row	numcol	numrow	x	y	fMedian	fMean	bMedian	bMean
1	9798	AccNr:null Control_PC1	1	1	11	1	1	16.0	4501.0	11100.0	10435.7451	6113.0	6110.4819
2	10086	AccNr:null Control_PC1	1	1	12	1	1	16.0	4952.0	11275.0	10621.8701	6263.5	6274.9399
3	9510	AccNr:null Control_PC1	1	1	10	1	1	19.0	4059.0	10560.0	10027.0195	6016.0	6005.8173
4	9222	AccNr:null Control_PC1	1	1	9	1	1	17.0	3613.0	9641.5	9776.0878	5952.5	5945.5791
5	8934	AccNr:null Control_PC1	1	1	8	1	1	22.0	3162.0	9801.5	9606.5585	5909.0	5899.6342
6	8646	AccNr:null Control_PC1	1	1	7	1	1	16.0	2716.0	9931.0	9836.8027	5892.0	5890.0639
7	8358	AccNr:null Control_PC1	1	1	6	1	1	21.0	2273.0	10819.0	9919.746	5900.0	5898.4267
8	8070	AccNr:null Control_PC1	1	1	5	1	1	20.0	1830.0	10882.0	10089.8066	5921.0	5922.6196
9	7782	AccNr:null Control_PC1	1	1	4	1	1	19.0	1379.0	11112.0	10243.5634	5995.5	5993.7988
10	7494	AccNr:null Control_PC1	1	1	3	1	1	18.0	929.0	11167.0	10186.9101	6141.0	6145.4584

Fig.23 Raw data

Because the basic version of MARS supports no batch or automatic import of microarray data and provide no opportunity for statistical analysis, we developed an additional tool MARSlab.

MARSlab

MARSlab consist of a combination of R-scripts and Java modules that compensate for not available functions in the original software. Java code (version JDK1.6.0_04) was used to develop the central backbone of MARSlab which employs Java Webstart from SUN and is served via the APACHE web server of version 6_0_16. An advantage of this construction is the ease of use for the user, because the software itself will be automatically downloaded and started from the server. No local installation or configuration is needed. To enable webstart functionality a JNLP file has to be created within the installation directory. This file administers access privileges, procedures, and different parameters for specific programs like R.

```
<?xml version="1.0" encoding="utf-8"?>
<jnlp>
        <codebase="file:///apache-tomcat-6.0.16/webapps/t2m" href="t2m.jnlp" />
        <information>
                <title>Transmission2MARS</title>
                <vendor>Andre Billion / Martin Rieger</vendor>
        </information>
        <resources>
                <j2se version="1.6+" />
                <jar href="file:///apache-tomcat-6.0.16/webapps/T2M/T2M.jar" />
                <jar href="file:///apache-tomcat-6.0.16/webapps/T2M/myBib.jar" />
                <jar href="file:///apache-tomcat-6.0.16/webapps/
                        T2M/postgresql-8.3-603.jdbc4.jar" />
        </resources>
        <security><all-permissions/></security>
        <offline-allowed/>
        <application-desc main-class="transmission2mars_00">
        <argument>apache-tomcat-6.0.16/webapps/T2M/MARSlab.conf.WIN</argument>
        </application-desc>
</jnlp>
```

Fig.24 Example of a configuration file

Concept

The storage and management of all data as well as the export of e.g. MAGE-ML formatted experiments is organized and implemented in the basic MARS software. Additional features such as the import of several experiment- and data formats of different sources and technologies into the central MARS database and the analysis part with functions for merging, normalization, imputation, quality control (QC), export and generation of graphics and plots was developed with in the MARSlab module.

A standard procedure of a microanalysis can be described in six steps:

1. Input of a new Experiment via the MARS interface. Data of MAGE-ML, MINiML, Febit, Imagene and Spotfinder formats are supported and a guided process ensures a uniform database and data structure.
2. Pre-processing of input data with tools of the MARSlab suite such as merging, QC, normalization and imputation.

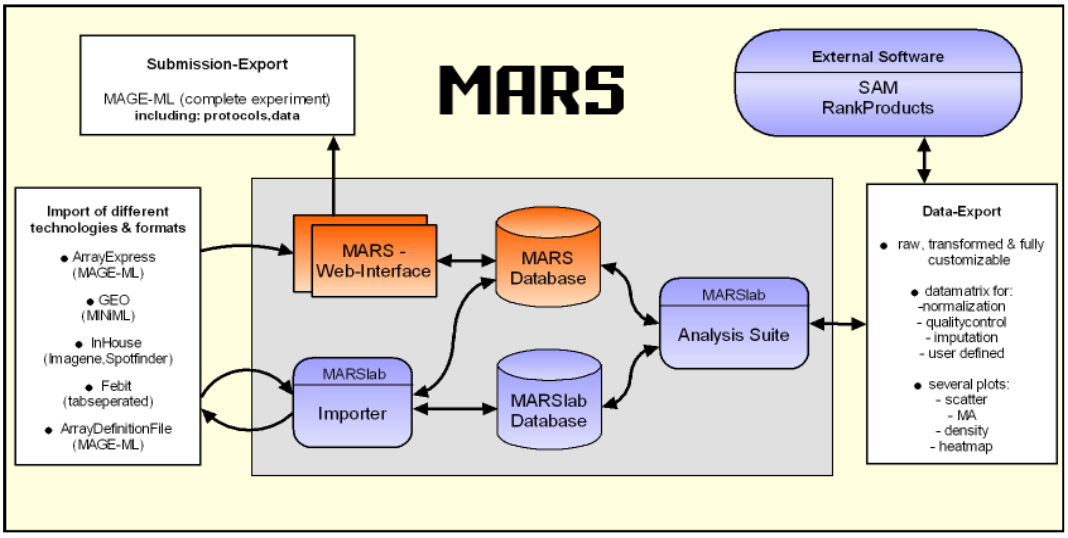


Fig.25 MARS structure

3. Data export for statistical analysis
4. Analysis of data via external software (SAM &RankProducts)
5. Import and graphical illustration of results
6. Export in MAGE-ML format for submission to public databases such as ArrayExpress of GEO

Configuration

In order to keep configuration flexible, a central configuration file contains all parameters for environment- and server typical setting such as paths and needed files and programs. The file (MARSlab.conf) is used in both modules and is tabulator separated!

parameter	firstvalue	secondvalue	description
RSCRIPT	Path tolocalrscript	version	serveral lines (for each script) are possible
MARSDUMMY	Username: marsdummy	Password: marsdummy	essential for database access (change before first use)
USERMANAGEMENT	Username: usermanagement	Password: usermanagament	
INFO	Path to info.gif		
BANNERLEFT	Path to left banner element		
BANNERRIGHT	Path to right banner element		
LOG	Path to log files		Contains all program stages as well as errors

DBDRIVER	postgreSQLdriver		e.g. org.postgresql.driver
DBURL	URL to used MARS DB		e.g. jdbc:postgresql://localhost:5432/mars
SCRIPTS	Path toscripfolder		This folder contains scripts for normalization
QC	Path to qc.jar		Quality control
OS	WINDOWS or LINUX		OS
TEMP	Path fortemporaryfiles		

Table 7 MARS configuration script

The following listing represents the default configuration file which we include in our initial version:

```
RSCRIPT      /MARS/R-2.7.0/bin/Rscript          2.7.0
RSCRIPT      /MARS/R-2.8.0/bin/Rscript          2.8.0
MARSDUMMY        marsdummy   marsdummy
USERMANAGEMENT       usermanagement      sermanagement
INFO   /MARS/T2M-and-MARSlab/images/Info.gif
BANNERLEFT       /MARS/T2M-and-MARSlab/images/jlu.png
BANNERRIGHT      /MARS/T2M-and-MARSlab/images/fh.png
LOG    /MARS/T2M-and-MARSlab/log
DBDRIVER     org.postgresql.Driver
DBURL        jdbc:postgresql://127.0.0.1:5432/mars
SCRIPTS      /MARS/T2M-and-MARSlab/scripts
QC     /MARS/T2M-and-MARSlab/qc.jar
OS     LINUX
TEMP   /MARS/T2M-and-MARSlab/temp
```

Entries like INFO, BANNERLEFT, and BANNERRIGHT refer to files which are provided with the basic program. To change icons, colors, style, graphics etc. customize corresponding default files. The option RSCRIPT can be used multiple times. This is needed in cases where several modules of e.g. bioconducter are used which require different R versions. Selection of the respective version is managed via a pop-up window.
The options MARSDUMMY, USERMANAGEMENT, DBURL, and DBDRIVER are essential for the use of the central database. Usernames are fixed whereas associated passwords should be change after installation. Additional users and groups can be created via the user management module of MARS. Parameters such as DBURL and DBDRIVER are specific to each host system. Please contact your local administrator if you need assistance.

Construction of an Array design file (ADF)

The array design describes the structure of the respective chip technology used. Each spot is represented by several informations such as position, expression intensity, and identifiers. For an import of experimental data into the MARS database and for the export to a

public database like ArrayExpress, an array design is an essential condition. The following paragraph describes the structure and gives support for the preparation of such a file:

Spalte	Beschreibung	Kommentar
MetaColumn	Array column	This combination describes a unique spot on the slide
MetaRow	Array row	
Column	column	
Row	row	
Reporter Identifier	ID	Identifier of the gene / molecule
Reporter Name	Name	Name of the gene / molecule
Reporter Comment	Description	optional
Reporter BioSequence Database Entry [datenbank]	Database entry	optional
Reporter BioSequence Type	Molecule type	e.g. 'PCR_amplicon' or 'ss_oligo'
Reporter BioSequencePolymerType	Polymer type	e.g. 'DNA' or 'RNA' or 'PROTEIN'
Reporter BioSequence [ActualSequence]	Oligosequenz	e.g. 'atgcctgacctagtgatcg', optional
Reporter Group [role]	Data category	e.g. 'Control' or 'Experimental'
Reporter ControlType	Control type	e.g. 'control_lable'

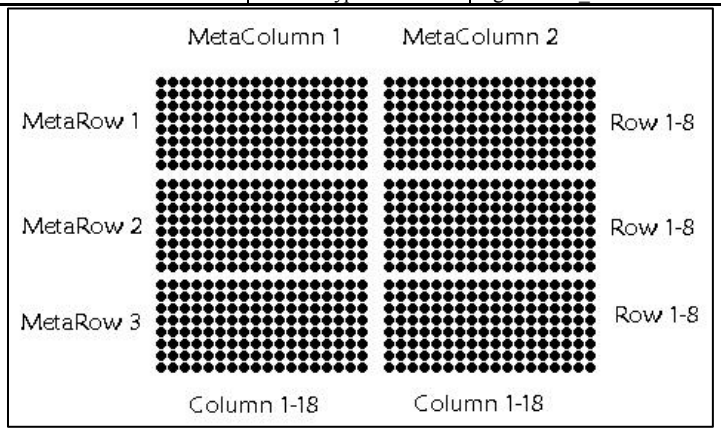


Fig.26 Description of a chip position

In order to extend an existing MAGE-ML file to a valid ADF, an additional FeatureReporterMap (FRM) is needed. The FRM links experimental with description data via an additional matrix. This procedure is faster and even more efficient than building a completely new ADF. The following table describes the structure and entry of a FRM:

Reporter ID (ADF)	Reporter ID (Experiment)	MetaColumn	MetaRow	Column	Row
lmo1741	ebi.ac.uk:MIAMExpress:Feature:A-MEXP-752.1.1.26.6	1	1	26	6
lmo1768	ebi.ac.uk:MIAMExpress:Feature:A-MEXP-752.1.2.1.3	1	2	1	3
24909	bugs.sgul.ac.uk:BuG@Sbase/Feature:19_1.1.17.1	1	1	1	17
24649	bugs.sgul.ac.uk:BuG@Sbase/Feature:19_1.1.19.1	1	1	1	19
TE	ebi.ac.uk:MIAMExpress:Feature:A-MEXP-310.1	1	12	12	11
YOL074C	ebi.ac.uk:MIAMExpress:Feature:A-MEXP-310.3.1.12.4	3	1	12	4
2	F:A-MEXP-66:3.2.8.6	3	2	8	6

1	F:A-MEXP-66:3.2.12.6	3	2	12	6
2	F:A-MEXP-66:3.2.13.6	3	2	13	6

Table 8 Description of a FRM matrix

The official MARS and MARSlab manuals contain extensive information on the creation off all needed files and formats.

Import of public microarray data

In order to extend the locale collection of expression data e.g. to enhance comparative transcriptomic analysis, an importer was implemented to e.g. enables the user to download data from the public database ArrayExpress from EBI and store them in the MARS database. Both databases have a similar structure which supports data import.

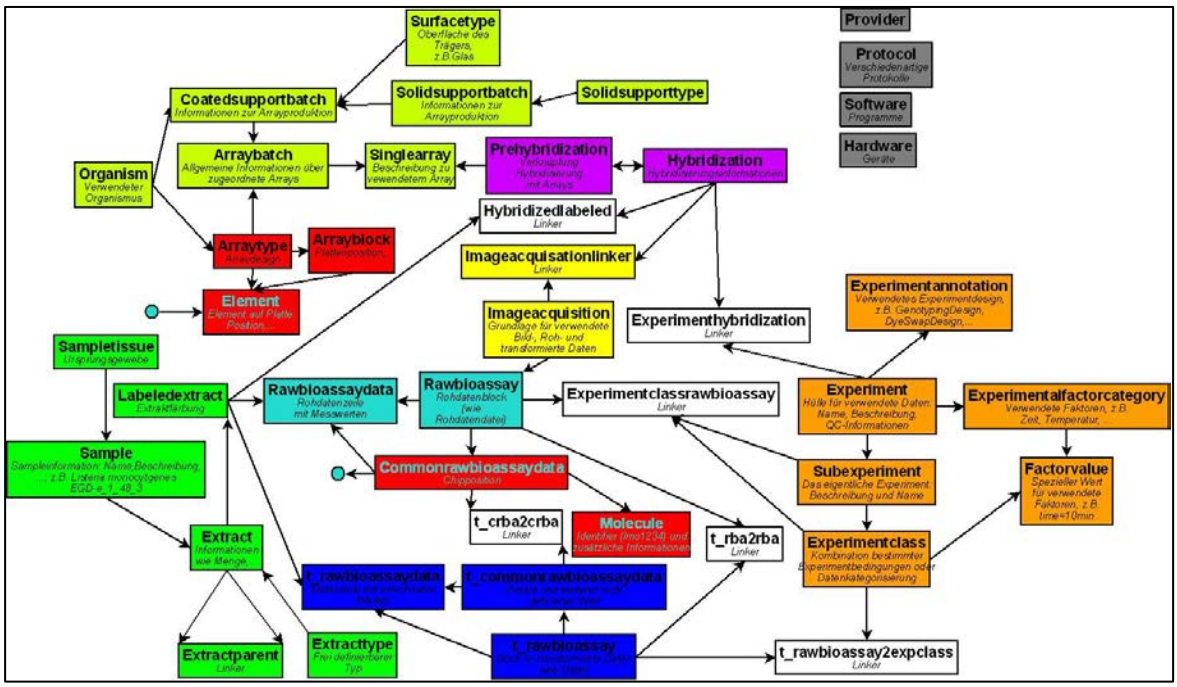


Fig.27 Used tables for ArrayExpress upload to central MARS database

Import steps of the guided process of MARSlab:

1. Download complete experiment archive from ArrayExpress
2. Copy raw data into *tmp* folder on the server (can be moved via the webinterface)
3. Create ADF as described
4. Define position and label corresponding to columns within MARS and ADF
    a. MARS: raw bio assay data ADF: fmedian,fmean,fstd,bmedian,bmean,bstd
    b. MARS: common raw bio assay data ADF: x,y,diameter
5. Upload data (may take a while … depends on the number of experiments)

The user will be informed if the import was successful. It is highly recommended to re-check the imported data, because a rollback function was implemented for special cases. A subsequent manual rollback is difficult and time consuming, because of the complexity of the database and corresponding entries in the data tree structure.

MARSlab - Analyse

The MARSlab tool enables the user to perform different tasks on the data stored in the central MARS database:

Fig.28 Main menu of MARSlab

- merge data of replicates
- normalize data through different normalization algorithms
- impute missing values through the SKNN algorithm
- execute a three step quality control
- visualisation of data
- export data into several formats

To execute a statistical significance analysis with the SAM-Excel-Plugin, MARSlab solely provide a special export in the specific format used by SAM.

Merging

The merging step was developed to enable MARS to import experimental data from dye-swap experiments. Dye-swap data contains two signals per oligo, one with the Cyanine3-Lable (Cy3) and one with Cyanine5-Label (Cy5). Both signals are merged and the calculated mean value will be stored in the database.

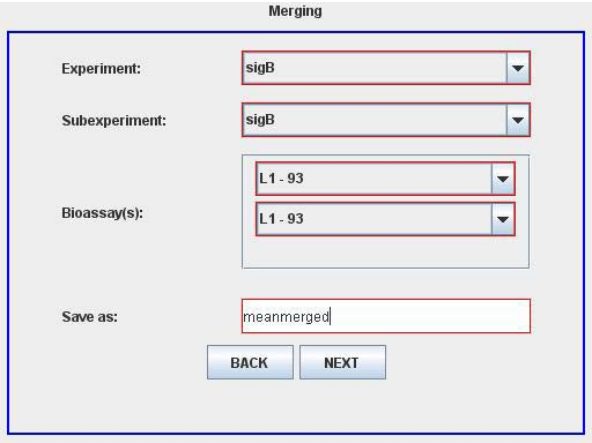


Fig.29 Merging of e.g. two color data

Normalisation

The normalization uses an external R-script which has advantages in flexibility and speed. MARSlab just provide the input files and is responsible for the uptake of normalized results.

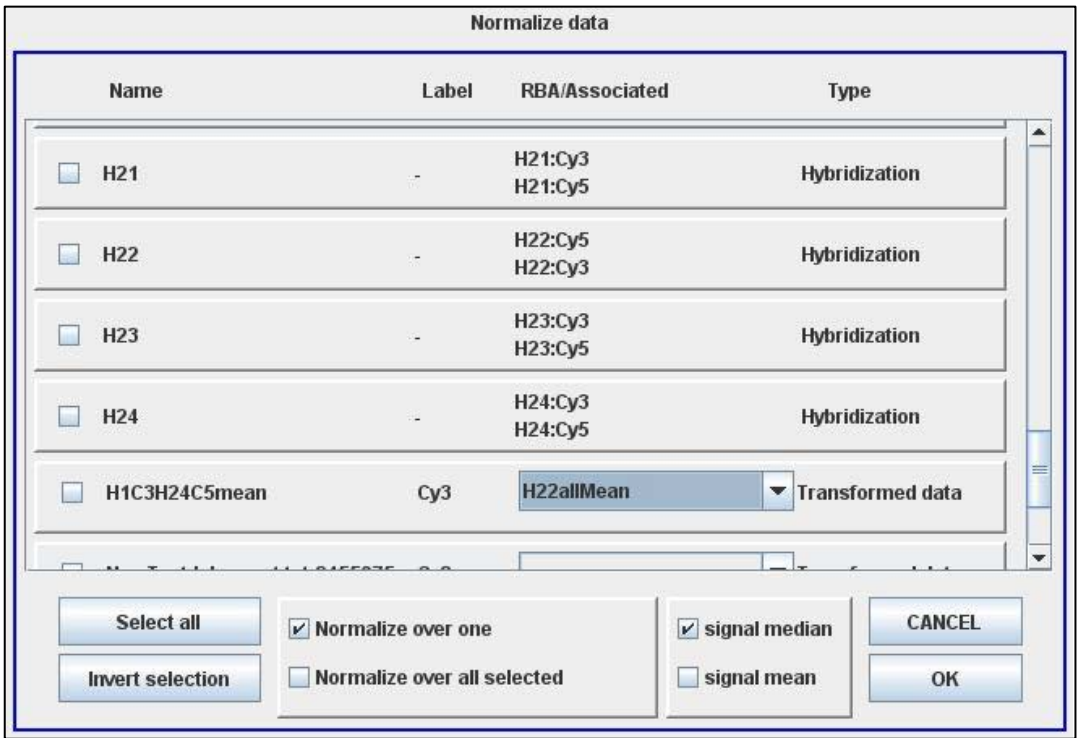
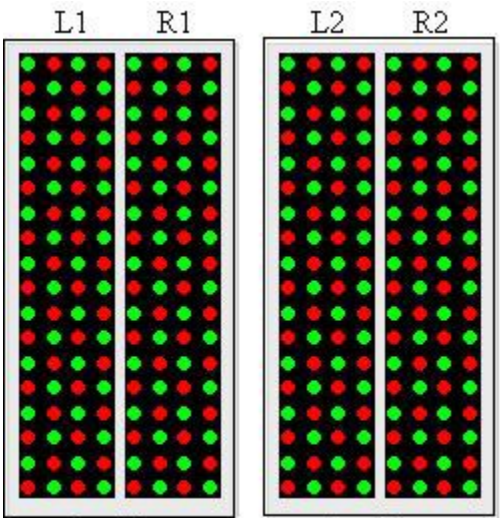


Fig.30 Chooser for normalizing hybridizations

Normalization is limited to a single experiment. The user has to choose all bioassays of a hybridization. Raw as well as further transformed data can be used as input. If the user wanted to normalize Hybridization (raw) data, all used raw data blocks (e.g. H21:Cy3) are displayed. In the case of transformed data, the user has to choose the options Cy3 or Cy5 from a dropdown menu. Chosen bioassay pairs combined with a unique identifier will be exported and automatically normalized based on the defined normalization script.



The resulting data matrix is available for download and can then be imported into the MARS database.

Imputation

In case of missing expression values caused by e.g. faulty measurements, MARSlab implements the SKNN algorithm [Kim et al 2004] to impute missing values. It is important to execute the imputation before the quality control step,otherwise it could happen that the corresponding gene will be eliminated through a high percentage of missing expression values. Following parameters are needed for the execution of the imputation script:

Nr	Parameter	Description
1	Input file	Path to the input matrix
2	Header	Header included? (TRUE / FALSE)
3	k	Count of usedneighbours
4	Outputfile	Path toresultmatrix

Table 9 Parameters for imputation script

Special attention should be paid to parameter three (number of checked neighbours), because this value needs to be adapted to each dataset specifically.

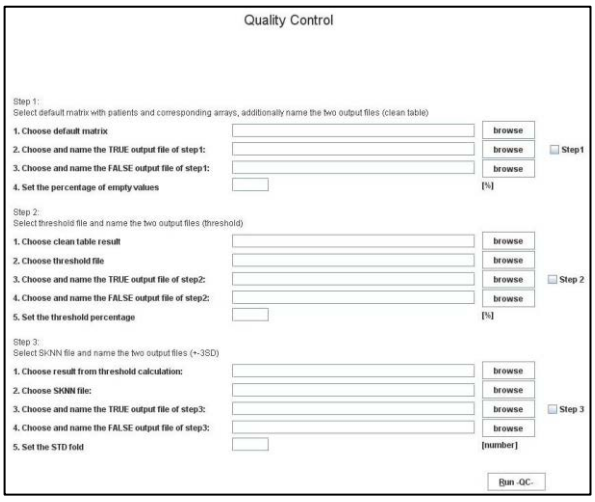
The program itself is relatively small, since it implements *MASS* and *SeqKnn* libraries from the R package. Implementation is structured as follows:

```
library(MASS)
library(SeqKnn)
param<-commandArgs(TRUE)
infile<- read.table(file=param[1],sep="\t",header=param[2])
sknn5 <- SeqKNN(infile,k=param[3])
write.matrix(sknn5,file=param[4])
```

The result of the sknn-imputation step can be stored, tagged as ‘transformed’, in the central MARS database.

Quality control (QC)

The QC step is intended to ensure that analysed data correspond to defined quality values.



It tags and eliminates signals which are identified as outliers to make sure that later statistics and analysis are not affected through false measured expressions. Genes will also not taken into account, if more than 50% (default value) of measured signals over all conditions are missing. Furthermore, it must be ensured that all expression values are greater than a given

threshold (e.g. background noise or control probes).

This QC module will be published as part of publication **11.**Hilgendorf et al., and is implemented in MARSlab as a separate Java application. MARSlab provides several methods to export data for QC input as well as it takes care of writing the results to database.

Visualisation

For the purpose of analysis and further quality control, MARSlab implements a visualisation module. This module is able to produce different plots such as scatterplots, MA-plots, histograms, densityplots, and heatmaps.

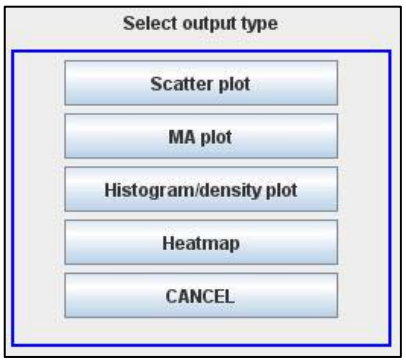


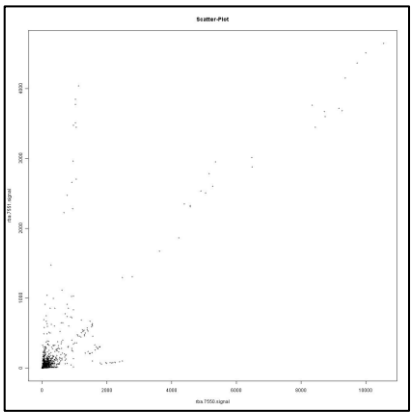
Fig.31 Visualization menu

Usage of this module is very similar to the merging step described before. The user has to choose from a listing of the database which experiment and data blocks should be visualized.

Scatterplot

The scatterplot is used to discriminate between two given set of values and calculate the degree of similarity. If the pattern of dots slopes from lower left to upper right, it suggests a positive correlation between the variables being studied. Following script is used to execute a R-script based function:

```
par<-commandArgs(TRUE)
mat.1<-read.csv(par[1],header=TRUE, sep="\t", na.strings="NA", dec=".")
jpeg(par[2],width=1024,height=1024)
plot.new()
plot(mat.1[,1:2],main="Scatter-Plot", cex=0.6)
```

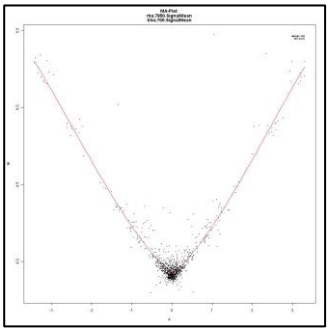


As a result, MARSlap construct a plotfile which can be stored by pressing the space key.

MA-Plot

The MA-Plot is very similar to the scatterplot, but it correlates A and M values. A values are calculated based on the major spot intensity whereas M values are logged fold changes.

```
par<-commandArgs(TRUE)
library("affy")
dat<-read.csv(par[1],header=TRUE, sep="\t",
na.strings="NA", dec=".")
amat<-log(dat[,1]+dat[,2])/2
mmat<-log(dat[,1]/dat[,2])
jpeg(par[2],width=1024,height=1024)
plot.new();
ma.plot(mmat,amat,pch=20,cex=.6,main=c("MA-Plot
```



Histogramm/Densityplot

Densityplots show the frequency of measured data of one or more bioassays. With this kind of plot, outlier and other irregularities can be identified. The function *plotDensity* from the R-library *affy* is used to create the main plot. For better differentiation it is possible to choose between a regular and a logarithmic plot.

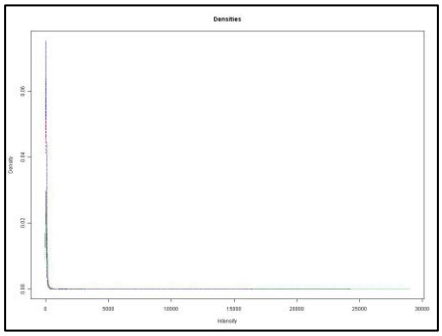


Fig.32 Regular density

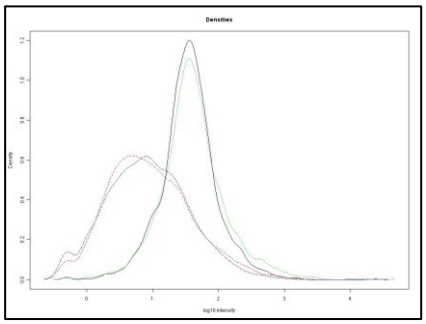
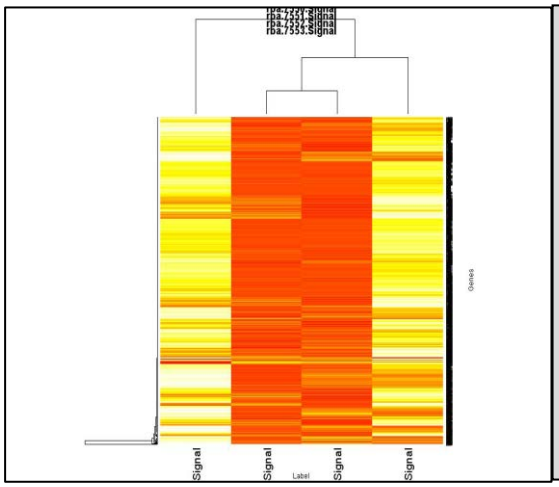


Fig.33 log10 density

```
par<-commandArgs(TRUE)

library("affy")

dateimatrix<-read.csv(
par[1],header=TRUE,sep="\t",row.names=1,dec=".",
na.strings="NA")

if(length(par)==3 && par[3]=="l10")
{
        datenmatrix<-log10(as.matrix(dateimatrix))
        xl="log10 Intensity"
}
if(!(length(par)==3 && par[3]=="l10"))
{
        datenmatrix<-as.matrix(dateimatrix)
        xl="Intensity"
}
jpeg(par[2],width=1024,height=768)

plotDensiy(mat=datenmatrix,xlab=xl,ylab="Density",
main="Densities",
na.rm=TRUE)
```

Heatmap

MARSlab can create heatmaps based on the regulation of genes from different bioassays. Genes can be clustered e.g. according to their relation or expression.




```
par<-commandArgs(TRUE)
dat<-read.csv(par[1],header=TRUE, sep="\t",
na.strings="NA", dec=".")
dmrows<-length(dat[,1])
dmcols<-length(dat[1,])-1
xl<-colnames(dat)
yl<-rownames(dat)
dat<-dat[,2:length(xl)]
dat<-as.matrix(dat)
jpeg(par[2],width=1024,height=768)
heatmap(dat,xlab="Label",ylab="Genes",labRo
w=yl,labCol=xl[2:length(xl)],
col=heat.colors(dmrows),main=c("Heatmap
",xl),na.rm=TRUE)
```

13.1.2 RNA sequencing workflow

In order to cope with the massive amount of data emerge from NGS technologies, a new software called small noncoding RNA analysis suite (sncRAS) has been developed. This program connects several analysis steps such as import, quality control, mapping, statistics, analysis, visualization, and export for publication. In this project, the emphasis was placed on quality and flexibility, since the field of NGS undergoes major changes caused by the continually development of new technologies.

Structure overview

Since this workflow is under continuous development, no GUI based interface has been programmed yet. All interactions will be carried out via command line and config files.

The workflow is currently grouped into twelve basically stand-alone modules, from which each module can be easily modified or extended since no dependences are made among themselves.

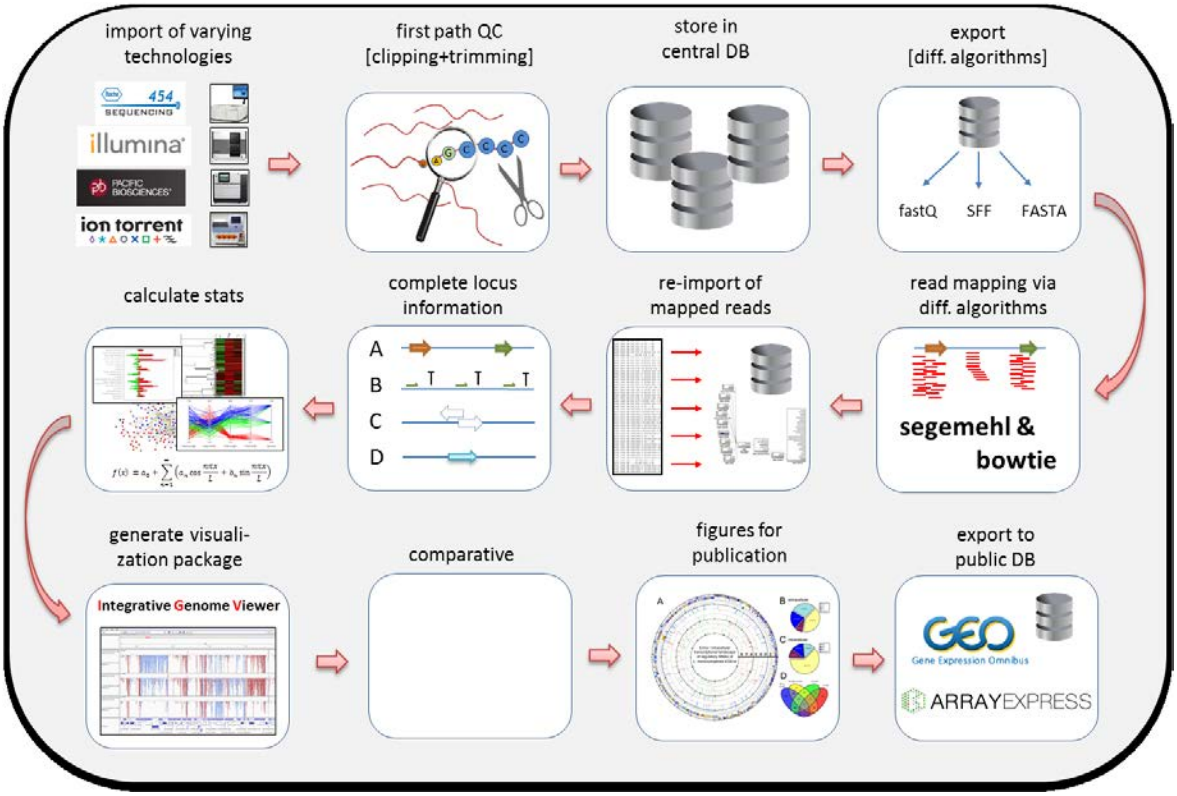


Fig.34 sncRAS workflow overview

Import of raw sequence data

Currently three different input formats are accepted. Regular FASTA format which comprises a description line and the sequence of the read itself. It should be noted, that this input format contains no quality information which may have a negative effect regarding mapping quality.

FASTA format

```
>FTLMN9J02D5EI0 length=279 xy=1584_2618 region=2 run=R_2009_04_02_16_27_35_
TGGATCTCTTGTAAATGCGCAATGTATAGTTAAGTGACCATGTCTGTGTGTACTTATTAA
ATATGAATTGGATGGGCTCCCTAGCCTAGTGTCCAGGCGTGGGGGTAGCTTGTCTGTGTG
CAATTTTCCAACCATTTTGCCATTGAAGTAATCTATGTATAGATATTAGAACTGTAATTT
TCCTCTTCACTTGATGTTTATTTATATTAAGATAATTTGTCATACTTGAAATATGATCTG
TTTCTCAATTTCTCCCCACTTCGCAGTGAGTGACAGGCC
```

The standard input used in *sncRAS* is the FASTQ format which uses four lines to describe the sequence of a single read. The first line has to begin with a @ character and is used as a unique identifier for each read, because by default it describes run number and position of this specific sequence (format depends on used technology). Line two is the sequence of the described read. Line four start with a "+" character and is optional and is currently not used. The last line of the block encodes the quality scores (in ASCII code) of the sequence in line two.

FASTQ format

```
@16_3_1_17374_944:0
NNNNNNTCTCTGCGCCAATGGTAGTTGGGGGCTTCCCCCTGCGAGAGTAGGTCGCTGCCGGGCAAAAAAAAAAAAA
+16_3_1_17374_944:0
BBBBBBBHKFHTQTQTYYYYYQHORL[[[[[_________YYYYYYVYYYY_____BBBBBBBBBBBBBBBBBBBB
```

The SFF format is the standard sequence format from 454 Life Sciences and contains sequencing results with additional flowgrams. While this format is a binary format and most widely specific to 454 sequencing, sncRAS transforms SFF files with the help of *sff2fastq* to FASTQ files.

Central database

For performance reasons, the central database is stored on a RAID 0 / SSD hardware combination. This setup allows very rapid read-and-write rates of about 500 MB/s and almost instant access time which is crucial for searching in tables with plenty of entries.

Since number of entries and size of tables is greatly depending on the particular experiment, the database can easily have around 60 to 100 GB with 600 million to one billion entries.

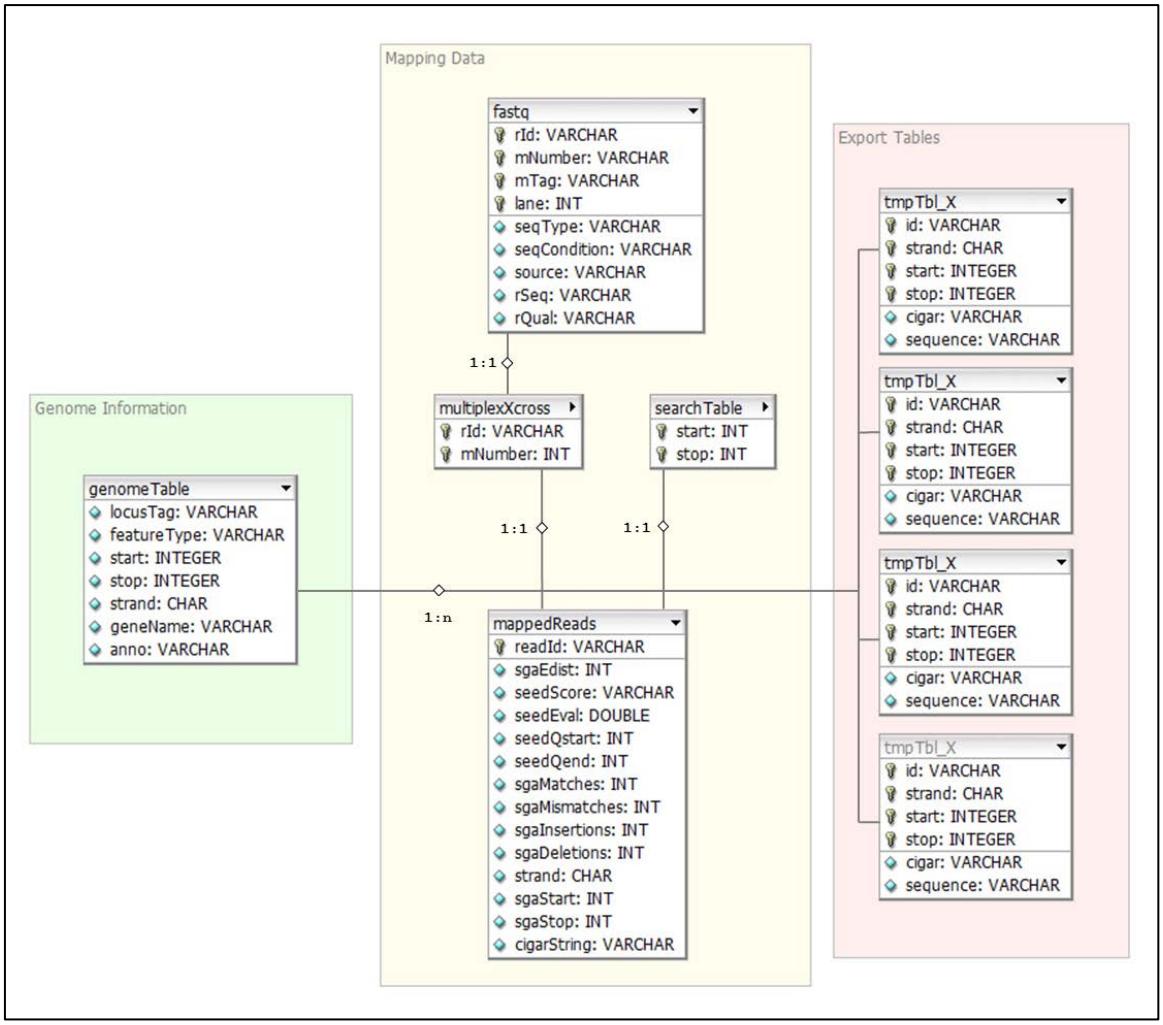


Fig.35 Database structure of sncRAS

Several temporary tables will be created and deleted during a complete analysis run. These temporary tables have been designed to increase operation speed. Furthermore, the sole purpose of the concept of sncRAS database is justified to enhance operation speed. Disadvantages of this concept are usability, redundancy, and less complexity (e.g. it does not allow complex linkages to be depicted).

Mapping of sequence reads

Currently, two different mapping methods are implemented in this workflow. The major differences between both algorithms are run time, number of tolerated mismatches and behavior in cases of insertions and deletions.

The segemehl algorithm is based on suffix arrays to compute seeds for a mapping starting point. After an update, this method is more robust to errors and 5' and 3'-contaminations. Furthermore, the algorithm was improved in its matching performance for data sets with a higher number of insertions and deletions (as found 454 reads). Best results can be obtained with short and medium reads, however it is not limited to a specific technology or read length.

The main disadvantage of the segemehl algorithm is its run time which results of a higher sensitivity. In cases of very large datasets, sncRAS implements the bowtie algorithm. Its layout is similar to segemehl and focuses on mismatches while the performance in cases of insertions and deletions is not comparable to segemehl. However bowtie is at least ten times faster than running segemehl (k=1) on the same dataset.

Since sncRAS was primarily developed for mapping transcriptomes of prokaryotes, the segemehl algorithm is favored for most of the processed mappings.

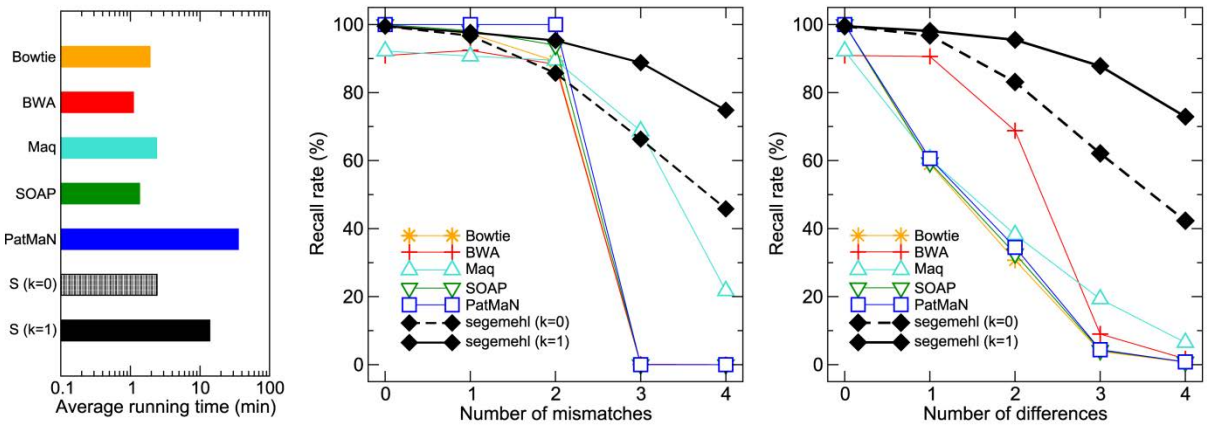


Fig.36 Performance of several mapping algorithms demonstrating that segemehl is superior in terms of accuracy [12]

Mapping sequenced reads against a reference genome with the segemehl algorithm can be performed in three steps. First, the input sequence file (reads) in FASTA format must be prepared. sncRAS assigns short but unique sequence identifiers and add the nucleotide sequence of each read. The next step is the build of an index structure for the reference sequence.

```
./segemehl.x -x chr1.idx -d chr1.fa
```

Options:

-x name of the returned index file

- d path and name of the reference genome for indexing

For indexing multiple FAST files, simply list them after parameter ‘d’

```
./segemehl.x -x chr1_2_3.idx -d chr1.fa chr2.fa chr3.fa
```

To map sequenced reads against desired reference, execute following command:

```
./segemehl.x -i chr1.idx -d chr1.fa -q myreads.fa --threads 64 -k 1 > mymap.sam
```

Options:

-i index of reference sequence

-d e.g. sequence of reference genome

-q sequenced reads

--threads number of threads used (number of cores used)

-k (0 or 1) defines sensitivity

> write output into result file

Result:

Following table describes 2 lines of results.

#"descr"	semi global alignment edist	seed score	seed Evalue	seed qstart	seed qend	semi global alignment matches	semi global alignment mismatches
>1_71_13864_19178:4:GGCG	0	71	0.0000	1	71	71	0
>1_10_5045_5967:4:GGCG	0	71	0.0000	1	71	71	0
semi global alignment insertions	semi global alignment deletions	strand	start of semi global alignment in subject sequence	end of semi global alignment in subject sequence	sequenceidx	cigar	
0	0	-	1713118	1713188	>gi|57650036	M71;	
0	0	-	1713084	1713154	>gi|57650036	M71;	

Table 10 segemehl output

Add locus information

Adding additional informations such as promoters, terminators, predictions etc. is the basis for a comprehensive analysis. We tried to incorporate data from as many sources as possible. As a result of this, a regular data record contains up to 50.000 entries.

```
EGD-e;Giessen;CDS;318;1673;.;+;.;"locus_tag=lmo0001;product=replication protein;gene=CDS_lmo0001"
…
EGD-e;Giessen;terminator;3030;3062;.;+;.;"nSeq=TCCGTCTTGGTTTCAAGGCGGATTCTTTTTTTG;gene=HP_Term"
…
EGD-e;Giessen;rbs;305;310;.;+;.;"nSeq=AGGGGG;gene=RBS"
…
EGD-e;Giessen;trna;82705;82777;.;+;.;"locus_tag=lmot01;product=tRNA Lys;gene=trna_lmot01"
…
EGD-e;Giessen;rrna;237466;239020;.;+;.;"locus_tag=lmor01;product=16S ribo RNA;gene=rrna_lmor01"
…
EGD-e;Giessen;sRNA;876921;877173;.;+;.;gene=sRNA_sipht_Candidate_9
…
EGD-e;Giessen;promoter;108945;108972;.;+;.;gene=promoter_sigmaA_16bp
```

A) CDS

Coding sequence information were obtained from GenBank flat files from NCBI FTP server. (*www.ncbi.nlm.nih.gov/Ftp/*)

B) Terminators

Transcription terminator predictions were obtained from TransTermHP version 2.09. This roh-independent terminator prediction was performed with a confidence score >= 30. (*http://transterm.cbcb.umd.edu/*)

C) rbs

Ribosomal binding sites were obtained from GenBank flat files from NCBI FTP server. (*www.ncbi.nlm.nih.gov/Ftp/*)

D) trna

Trnas were obtained from GenBank flat files from NCBI FTP server. (*www.ncbi.nlm.nih.gov/Ftp/*)

E) rrna

Rrnas were obtained from GenBank flat files from NCBI FTP server. (*www.ncbi.nlm.nih.gov/Ftp/*)

F) sRNA

Small noncoding RNAs were obtain from several prediction tools such as sipht [17] or sRNAscanner [18] as well as experimentally verified candidates from publications or cooperation partners.

G) Promoters

Promoter prediction was carried out with a HMM search from PPP server (http://bioinformatics.biol.rug.nl/websoftware/ppp)

Visualisation

Visualization has been realized via the integrated genomics viewer (IGV) developed by Broad Institute[19]. Advantage of this software is it's possibility to visualize multiple tracks at once to enable a direct comparative analysis. Furthermore, IGV can store plenty of information within the main memory, making it to the fastest viewer for high reads visualizations. Following input files are needed and will be generated by sncRAS:

A) Genome fasta file

   This sequence file is used to display the reference sequence in the browser and forms the backbone of the visualization.

B) GFF annotation file

   The gff file contains all informations added in the *locus information* step. Given data will be linked to the genome sequence and is displayed on the bottom of the viewer and is used for information and orientation purpose.

C) SAM file

   The SAM (Sequence Alignment/Map) file contains the actual read / sequence information.

After importing the reference genome, annotation and sequence files, IGV can be used for detailed manual inspection. Since IGV can incorporate all kind of information using gff formatted files, this viewer represents a comprehensive solution in analysing all kind of sequencing data.

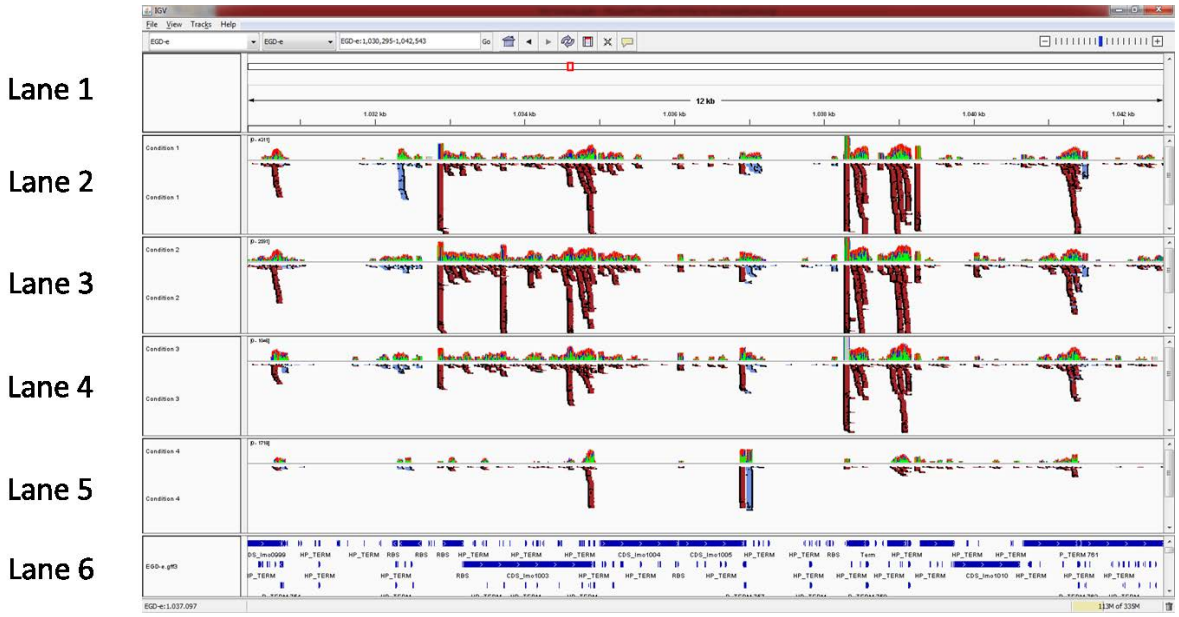


Fig.37 IGV screenshot with six lanes

Figure 20 shows a typical IFV view with four experiments loaded. Lane 1 indicates the actual position in the genome. Lane two to five represent the loaded experiments, divided in a Histogram and a read view.

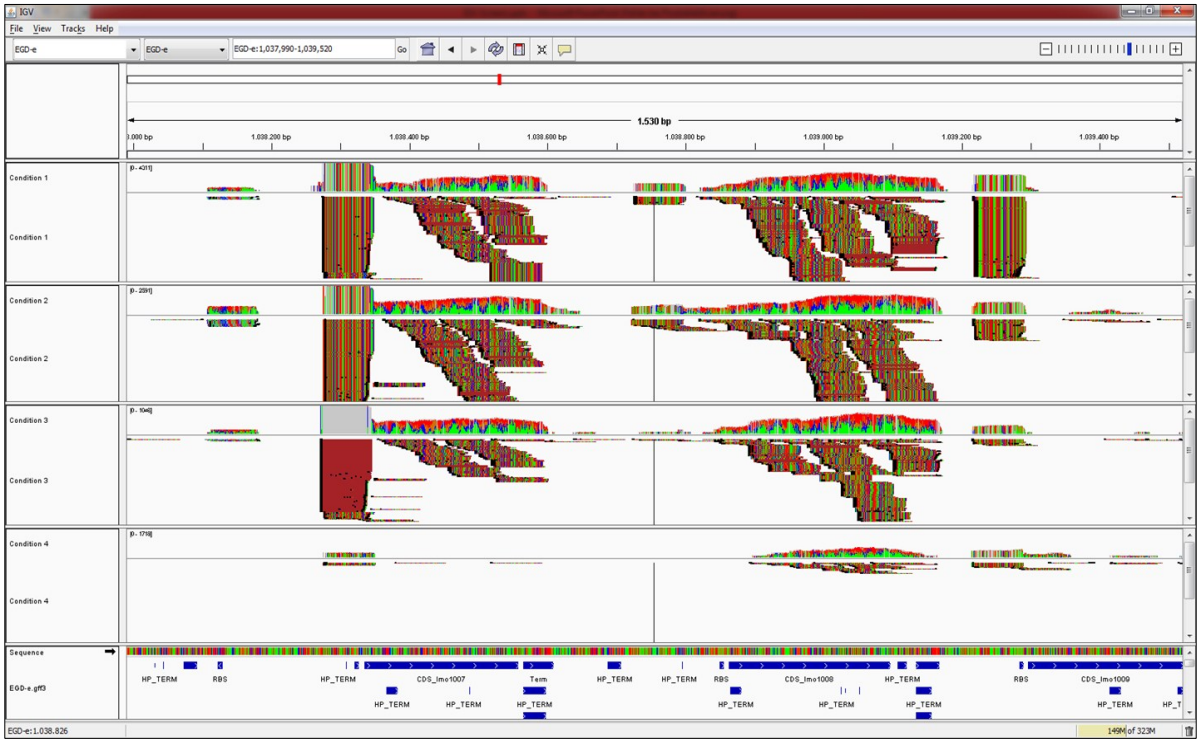


Fig.38 IGV screenshot with minimum zoom

A very helpful feature of IGV is its progressive zoom. With minimum zoom, the user has the possibility to inspect a region of at least 50kb, whereas with the maximum zoom it goes down to the base pair level (Fig. 22).

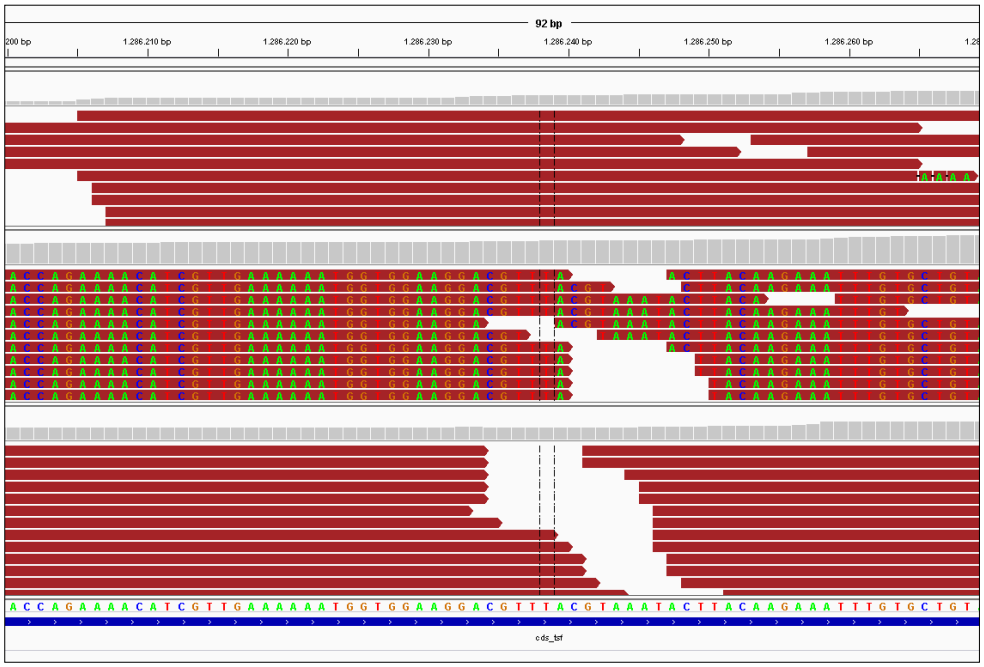


Fig.39 IGV screenshot with maximum zoom

Figures and tables

Currently sncRAS provide data export for three different types figures. Additionally several lists can be created with informations about distributions and coverage of reads. sncRAS can provide data files for the import into GenomeViz [26] to enable the user to generate a circular representation of the respective genome with several features.

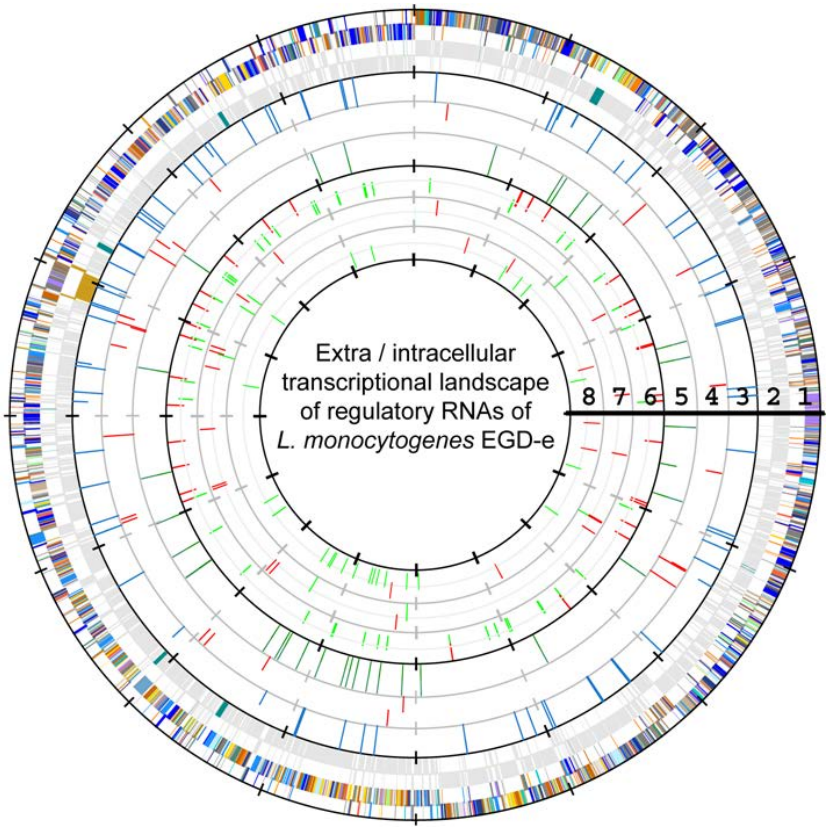


Fig.40 Circular representation of transcriptomic informations using GenomeViz

Furthermore, sncRAS provide data exports of calculated distributions and statistical comparisons. For example the user wants to compare a set of predicted small noncoding RNAs against three other e.g. published candidate lists. Therefore he has to provide the lists of candidates to compare with (List must include: identifier, start, stop, strand) and has to define an overlap cut off.

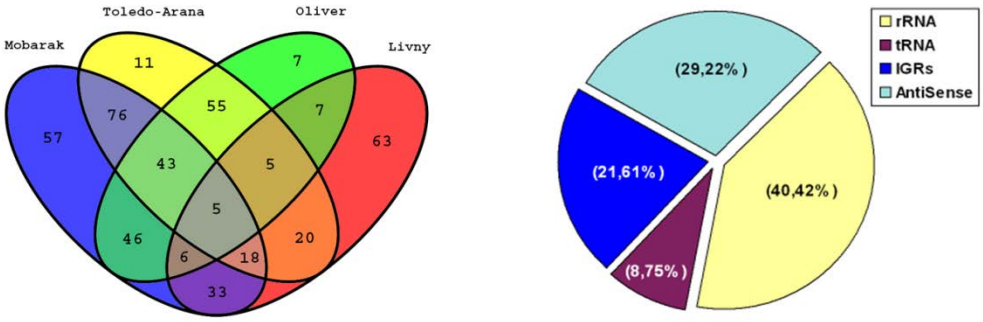


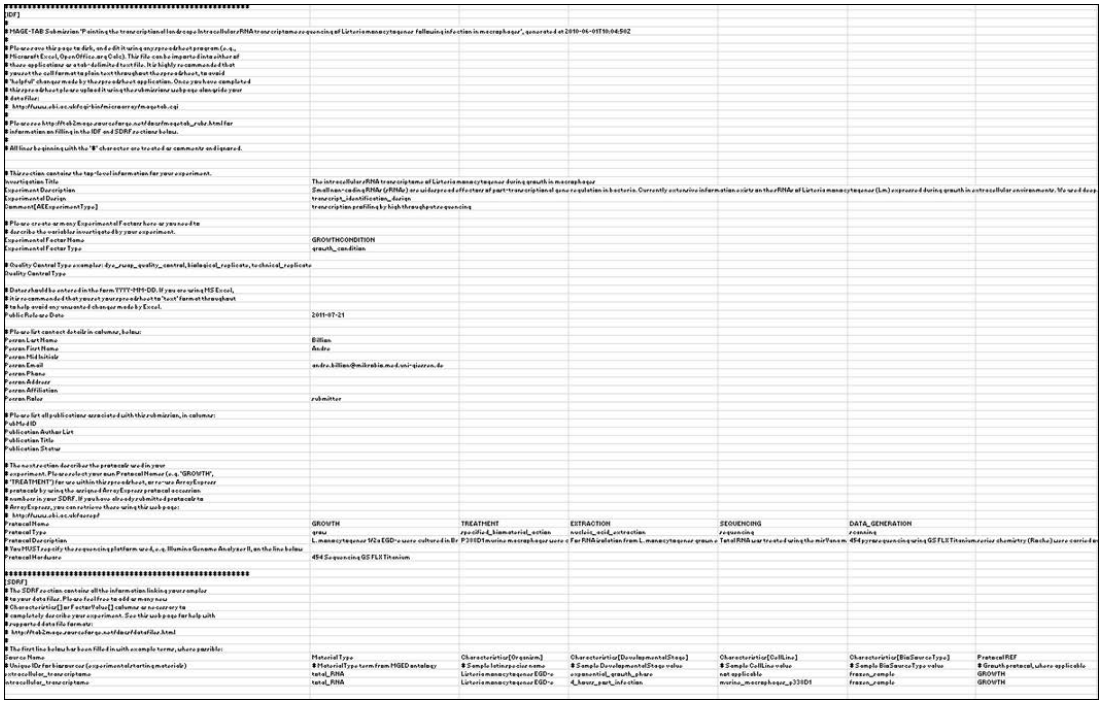
Fig.41 sncRAS plots

Export to public databases

sncRAS provides several exports for the submission of experiment data to a public database such as ArrayExpress or GEO. Four different categories of informations are needed:

A) Mage-tab file

The mage-tag file contains a summary about the whole experiment such as study investigation title, experiment description, growth conditions, contact informations, and a detailed description about submitted data sets.



B) Normalized data

The submitter has to provide all normalized sequence data sequence data in the following format (readId, mapping start, mapping stop, stand, sequence):

C) Protocols

Similar to the section Material & Methods of a regular publication, one has to submit the corresponding protocols of created data. Following protocol types are recommended: growth, extraction, sample treatment, and sequencing.

D) Raw data

Finally, the submitter hast to provide all raw data information of the current experiment. File types depends on used technology (SFF, FASTq etc.).

13.2 Appendix publication 1

BMC Genomics

BioMed Central
The Open Access Publisher

This Provisional PDF corresponds to the article as it appeared upon acceptance. Fully formatted PDF and full text (HTML) versions will be made available soon.

Comparative genomics and transcriptomics of lineages I, II, and III strains of Listeria monocytogenes

BMC Genomics 2012, **13**:144 doi:10.1186/1471-2164-13-144

Torsten Hain (Torsten.Hain@mikrobio.med.uni-giessen.de)
Rohit Ghai (rghai@umh.es)
Andre Billion (andre@erdna.net)
Carsten Tobias Kuenne (Carsten.Kuenne@mikrobio.med.uni-giessen.de)
Christiane Steinweg (steinweg.christiane@gmx.de)
Benjamin Izar (benjamin.izar@med.uni-giessen.de)
Walid Mohamed (Walid.Mohamed@mikrobio.med.uni-giessen.de)
Mobarak Mraheil (Mobarak.Mraheil@mikrobio.med.uni-giessen.de)
Eugen Domann (eugen.domann@mikrobio.med.uni-giessen.de)
Silke Schaffrath (silke.schaffrath@roche.com)
Uwe Kärst (Uwe.Kaerst@helmholtz-hzi.de)
Alexander Goesmann (agoesman@CeBiTec.Uni-Bielefeld.DE)
Sebastian Oehm (Oehm@CeBiTec.Uni-Bielefeld.DE)
Alfred Pühler (puehler@CeBiTec.Uni-Bielefeld.DE)
Rainer Merkl (rainer.merkl@biologie.uni-regensburg.de)
Sonja Vorwerk (sv@ascendispharma.com)
Philippe Glaser (pglaser@pasteur.fr)
Patricia Garrido (pgarrido@pasteur.fr)
Christophe Rusniok (rusniok@pasteur.fr)
Carmen Buchrieser (cbuch@pasteur.fr)
Werner Goebel (goebel@mvp.uni-muenchen.de)
Trinad Chakraborty (Trinad.Chakraborty@mikrobio.med.uni-giessen.de)

ISSN 1471-2164

Article type Research article

Submission date 25 August 2011

Acceptance date 24 April 2012

Publication date 24 April 2012

Article URL http://www.biomedcentral.com/1471-2164/13/144

© 2012 Hain *et al.* ; licensee BioMed Central Ltd.
This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

BMC Genomics

BioMed Central
The Open Access Publisher

Like all articles in BMC journals, this peer-reviewed article was published immediately upon acceptance. It can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

Articles in BMC journals are listed in PubMed and archived at PubMed Central.

For information about publishing your research in BMC journals or any BioMed Central journal, go to

http://www.biomedcentral.com/info/authors/

© 2012 Hain *et al.* ; licensee BioMed Central Ltd.
This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Comparative genomics and transcriptomics of lineages I, II, and III strains of *Listeria monocytogenes*

Torsten Hain[1†]
Email: Torsten.Hain@mikrobio.med.uni-giessen.de

Rohit Ghai[1†]
Email: rghai@umh.es

André Billion[1†]
Email: Andre.Billion@mikrobio.med.uni-giessen.de

Carsten Tobias Kuenne[1†]
Email: Carsten.Kuenne@mikrobio.med.uni-giessen.de

Christiane Steinweg[1]
Email: steinweg.christiane@gmx.de

Benjamin Izar[1]
Email: benjamin.izar@med.uni-giessen.de

Walid Mohamed[1]
Email: Walid.Mohamed@mikrobio.med.uni-giessen.de

Mobarak Abu Mraheil[1]
Email: Mobarak.Mraheil@mikrobio.med.uni-giessen.de

Eugen Domann[1]
Email: eugen.domann@mikrobio.med.uni-giessen.de

Silke Schaffrath[1]
Email: silke.schaffrath@roche.com

Uwe Kärst[2]
Email: Uwe.Kaerst@helmholtz-hzi.de

Alexander Goesmann[3]
Email: agoesman@CeBiTec.Uni-BIelefeld.DE

Sebastian Oehm[3]
Email: Oehm@CeBiTec.Uni-Bielefeld.DE

Alfred Pühler[3]
Email: puehler@CeBiTec.Uni-Bielefeld.DE

Rainer Merkl[4]
Email: rainer.merkl@biologie.uni-regensburg.de

Sonja Vorwerk[5]
Email: sv@ascendispharma.com

Philippe Glaser[6]
Email: pglaser@pasteur.fr

Patricia Garrido[7]
Email: pgarrido@pasteur.fr

Christophe Rusniok[7]
Email: rusniok@pasteur.fr

Carmen Buchrieser[7]
Email: cbuch@pasteur.fr

Werner Goebel[8]
Email: goebel@mvp.uni-muenchen.de

Trinad Chakraborty[1*]
[*] Corresponding author
Email: Trinad.Chakraborty@mikrobio.med.uni-giessen.de

[1] Institute of Medical Microbiology, Justus-Liebig-University, Schubertstrasse 81, Giessen D-35392, Germany

[2] Helmholtz-Zentrum für Infektionsforschung GmbH, Inhoffenstraße 7, Braunschweig 38124, Germany

[3] Center for Biotechnology, University of Bielefeld, Bielefeld D-33594, Germany

[4] Institut für Biophysik und physikalische Biochemie, Universität Regensburg, Universitätstrasse 31, Regensburg D-93053, Germany

[5] Febit Biomed GmbH, Im Neuenheimer Feld 519, Heidelberg D-69120, Germany

[6] Institut Pasteur, Laboratoire Evolution et Génomique Bactériennes and CNRS URA 2171, Paris 75724, France

[7] Institut Pasteur, Biologie des Bactéries Intracellulaires and CNRS URA 2171, Paris 75724, France

[8] Max von Pettenkofer-Institut for Hygiene and Medical Microbiology, Ludwig Maximilians-University München, Pettenkoferstrasse 9a, Munich D-80336, Germany

[†] Equal contributors.

Abstract

Background

Listeria monocytogenes is a food-borne pathogen that causes infections with a high-mortality rate and has served as an invaluable model for intracellular parasitism. Here, we report complete genome sequences for two *L. monocytogenes* strains belonging to serotype 4a (L99) and 4b (CLIP80459), and transcriptomes of representative strains from lineages I, II, and III, thereby permitting in-depth comparison of genome- and transcriptome -based data from three lineages of *L. monocytogenes*. Lineage III, represented by the 4a L99 genome is known to contain strains less virulent for humans.

Results

The genome analysis of the weakly pathogenic L99 serotype 4a provides extensive evidence of virulence gene decay, including loss of several important surface proteins. The 4b CLIP80459 genome, unlike the previously sequenced 4b F2365 genome harbours an intact *inlB* invasion gene. These lineage I strains are characterized by the lack of prophage genes, as they share only a single prophage locus with other *L. monocytogenes* genomes 1/2a EGD-e and 4a L99. Comparative transcriptome analysis during intracellular growth uncovered adaptive expression level differences in lineages I, II and III of *Listeria*, notable amongst which was a strong intracellular induction of flagellar genes in strain 4a L99 compared to the other lineages. Furthermore, extensive differences between strains are manifest at levels of metabolic flux control and phosphorylated sugar uptake. Intriguingly, prophage gene expression was found to be a hallmark of intracellular gene expression. Deletion mutants in the single shared prophage locus of lineage II strain EGD-e 1/2a, the *lma* operon, revealed severe attenuation of virulence in a murine infection model.

Conclusion

Comparative genomics and transcriptome analysis of *L. monocytogenes* strains from three lineages implicate prophage genes in intracellular adaptation and indicate that gene loss and decay may have led to the emergence of attenuated lineages.

Keywords

Listeria monocytogenes, Lineage, Comparative genomics, Gene decay, Comparative transcriptomics, Flagella, Prophage, Monocin, Isogenic deletion mutants, Murine infection

Background

Listeria monocytogenes is a Gram-positive, motile, non-sporulating, rod shaped bacterium. It is the causative agent of listeriosis, a food-borne disease, which afflicts both humans and animals. There are only eight species in the entire genus, *L. monocytogenes*, *L. marthii*, *L. innocua*, *L. seeligeri*, *L. welshimeri*, *L. ivanovii*, *L. grayi* and *L. rocourtiae*. *L. monocytogenes* and *L. ivanovii* are the pathogenic species while the others are apathogenic [1,2]. In the genus *Listeria*, non-pathogenic species have been hypothesized to have evolved through genome reduction from pathogenic progenitor strains [3]. *L. monocytogenes* is able to invade and

replicate in both phagocytic and non-phagocytic cells. The infectious life cycle has been elucidated in detail, and several virulence factors, essential for each stage of infection have been identified [4,5]. Pathogenic listeriae encode several virulence factors that are localized in a virulence gene cluster (*vgc*) or *Listeria* pathogenicity island-1 (LIPI-1) in the genome. However, a number of genes required for virulence are not localized in this cluster, including the two internalins *inlA* and *inlB*. These encode proteins that are expressed on the surface of the bacterium and facilitate the entry of the bacterium into the eukaryotic cell and their incorporation into a membrane-bound vacuole [6,7]. Further pathogenicity islands present in the genus *Listeria* code for multiple internalins and additional hemolysin genes in species *L. ivanovii* (LIPI-2) [8] and a subset of strains of lineage I (LIPI-3) [9].

Within the four lineages of *L. monocytogenes*, strains are generally classified by serotyping or MLST [10,11], of which 1/2a, 1/2b and 4b are most commonly associated with human listerial infections [2,12]. The first outbreak of *L. monocytogenes* was described for the strain EGD-e, a serotype 1/2a strain of lineage II, following an epidemic in rabbits and guinea pigs in 1926 by E.G.D. Murray [13]. This strain has become a model *Listeria* strain, and was the first listerial strain to be completely sequenced, along with the non-pathogenic *Listeria innocua* 6a CLIP11262 [14]. Subsequently, the first genome of a 4b serotype strain (F2365) of lineage I was completely sequenced [14,15]. It was isolated from Jalisco cheese during a listeriosis outbreak in California in 1985 and mainly associated with pregnancy-related cases. However, it has been recently shown that this strain contains nonsense and frameshift mutations in several genes. Owing to a frameshift in *inlB*, F2365 is severely compromised in Caco-2 invasion assays [16].

Here we report thus the genome sequence of a clinical isolate of the 4b serotype of lineage I, the *L. monocytogenes* 4b strain CLIP80459 that was isolated in a clinical outbreak of listeriosis in France affecting 42 persons [17]. We also present the complete genome sequencing of *L. monocytogenes* strain 4a L99 of lineage III. L99 was originally isolated from food by Kampelmacher in 1950s in the Netherlands. This strain is attenuated in its virulence properties and exhibits a restricted ability to grow within the liver and spleen of infected mice [18]. The availability of the complete genome of *L. monocytogenes* EGD-e serotype 1/2a has permitted analysis of the intracellular gene expression profile of this strain [19-21].

The genome sequences of strains 4a L99 and 4b CLIP80459 presented in this work provide a unique opportunity to delineate specific adaptations of these lineage representives both at the genomic and at the transcriptional level.

Results

General features of complete genomes of three lineages of *L. Monocytogenes*

The overall features of the completely sequenced circular genomes of *L. monocytogenes* 4a L99, *L. monocytogenes* 4b CLIP80459, *L. monocytogenes* 1/2a EGD-e, *L. monocytogenes* 4b F2365 and *L. innocua* 6a CLIP11262 are given in Table 1. Computational multi-virulence-locus sequence typing (MVLST) [22] analysis showed that strain 4b CLIP80459 belongs to epidemic clone ECII and strain 4b F2365 to epidemic clone ECI as previously reported by Nelson and colleagues [15], respectively. The *L. monocytogenes* genomes are remarkably syntenic: genome size, G+C content, percentage coding and average length of protein-

coding genes are similar among all four strains (which was previously reported for other listerial genomes) [14,15]. All four *L. monocytogenes* genomes harbour 67 tRNA genes and contain six complete copies of rRNA operons (16 S-23 S-5 S), of which two are located on the right and four on the left replichore. The chromosomes of 4a L99 and 4b CLIP80459 are devoid of mobile genetic elements and harbour no plasmid.

Table 1 General features of *L. monocytogenes* 1/2a EGD-e, *L. monocytogenes* 4a L99, *L. monocytogenes* 4b CLIP80459, *L. monocytogenes* 4b F2365 and *L. innocua* 6a CLIP11262

	L. monocytogenes 4a L99	*L. monocytogenes* 4b CLIP80459	*L. monocytogenes* 4b F2365	*L. monocytogenes* 1/2a EGD-e	*L. innocua*6a CLIP11262
Size of chromosome [bp]	2979198	2912690	2905187	2944528	3011208
G+C content [%]	38.2	38.1	38.0	38.0	37.4
G+C content of protein-coding genes [%]	38.7	38.5	38.5	38.4	37.8
Protein-coding genes (pseudogenes)	2925 (1)	2790 (24)	2821 (26)	2855 (9)	2981 (13)
Average length of protein-coding genes [aa]	301	311	303	306	300
Number of rRNA operons (16 S-23 S-5 S)	6	6	6	6	6
Number of tRNA genes	67	67	67	67	66
Percentage coding	88.9	89.4	88.4	89.2	89.2
Number of prophages (genes)	4 (191)	1 (16)	1 (16)	2 (79)	6 (322)
Plasmid	0	0	0	0	1
Number of strain-specific genes*	111	49	105	120	89
Number of orthologous genes*	2623	2725	2699	2656	2570
Number of transposons	0	0	0	1	0

***Prophage genes excepted.**

Core and specific genes were analyzed using orthologous pairs excluding prophage genes as described previously [3]

We observed four different prophage regions in the genome of the 4a L99 and only one in the 4b CLIP80459 strain (see prophage region II). *L. monocytogenes* 4a L99 prophage I is located at position 71438 bp (*lmo4a_0064-lmo4a_0115*), prophage II at (*lmo4a_0148-lmo4a 0153,* prophage-remnant: *lmaDC*; 4b ClIP80459 Lm4b_00117b-Lm4b00134 or monocin region), prophage III at 1224779 bp (*lmo4a_1221-lmo4a_1293*) and prophage IV at 2668913 bp (*lmo4a_2599-lmo4a_2658*). Two prophage regions, I and III, are located adjacent to tRNAs. Prophage region I is flanked by $tRNA^{Lys}$ and prophage region III is inserted within the region between the gene for $tRNA^{Arg}$ and *ydeI* compared to *L. monocytogenes* 1/2a EGD-e. At this very chromosomal location in *L. welshimeri* 6b SLCC5334 there is an insertion of a prophage [3,23,24], while *L. ivanovii* harbours the species-specific *Listeria* pathogenicity island 2 (LIPI-2), which contains a sphingomyelinase C (SmcL) and also a cluster of internalin genes [8]. These findings confirm previous observations [3] indicating that tRNAs represent genetic "anchoring elements" for the uptake of listerial prophage DNA by transduction processes and thus contributing to evolutionary genome diversity of listeriae. Pseudogenes were detected for both 4b F2365 (24 pseudogenes) and 4b CLIP80459 (26 pseudogenes) genomes respectively, which is a higher number compared to that seen in *L. monocytogenes* 1/2a EGD-e (9 pseudogenes), *L. monocytogenes* 4a L99 (one pseudogene) and *L. innocua* (13 pseudogenes).

When comparing the two *L. monocytogenes* 4b genomes (CLIP80459 and F2365) 115 genes are specific for strain 4b CLIP80459 with respect to strain 4b F2365. The dominant functions

encoded by these genes are related to sugar metabolism as they comprise five PTS systems and five sugar permeases or sugar transporters. Furthermore, four transcriptional regulators and four surface anchored proteins are specific to 4b CLIP80459 indicating differences in regulation, sugar metabolism and surface characteristics between the two strains. Of the 146 genes found to be specific for strain 4b F2365, the majority were of unknown function, apart from a PTS system and a specific surface protein. Most interestingly, *inlB* although it is reported to be important for virulence of *L. monocytogenes* has a frameshift mutation in this strain [15].

When comparing the genomes of different lineages at the nucleotide sequence level a number of genomic differences were revealed (Figure 1). Surface proteins showed the highest number of single nucleotide polymorphisms (SNPs). Even in the comparison of the two closely related 4b genomes, two LPXTG-motif containing proteins were identified as encoding a large number of SNPs. One of these, *lm4b_01142* shares substantial similarity to internalins. Comparison of the 4a L99 and the 1/2a EGD-e genomes reflected larger evolutionary divergence, but once again involved surface proteins, such as the LPXTG-motif containing protein *lmo1799*, internalin *lmo0409* (*inlF*), autolysin *lmo1215*, as well as proteins involved in surface antigen biosynthesis like *lmo2552* (*murZ*) and *lmo2549* (*gtcA*). Further analysis identified genes that are most divergent in the three lineages and classification of the most divergent orthologous gene groups was performed (Additional file 1: Table S1). Thus, distribution of SNPs in *Listeria* suggests considerable evolutionary adaptation among surface-associated genes.

Figure 1 Comparative SNP analysis of five listerial strains From outside to inside: genome of *L. monocytogenes* 1/2a EGD-e colored according to COG categories (two strands shown separately). Number of SNPs normalized by gene length in the comparison of 1/2a EGD-e and *L. innocua* 6a CLIP11262, 1/2a EGD-e and 4a L99, 1/2a EGD-e and 4b CLIP80459, 1/2a EGD-e and 4b F2365, and the two 4b strains (4b F2365 and 4b CLIP80459). The innermost circle shows the location of phage genes (blue) and virulence genes (black) in the 1/2a EGD-e genome. Line graphs indicate the number of SNPs/gene length reflecting loci in the genome having a disproportionate number of SNPs. However, if a gene is specific to a certain genome, this will also be shown as a peak indicating a region of divergence within the two genomes under comparison. This analysis was performed using the MUMmer package [25] and SNPs were mapped to coding regions using PERL scripts. Data were visualized by GenomeViz [26]. For each pairwise comparison of strains, percentage of SNPs per gene length of surface- and non-surface-associated genes, as well as the ratio of these values is given in the table. The latter was named "nucleotide divergence ratio" and denotes the relative amount of difference between those two classes of genes, in order to identify more (positive value) or less (negative value) abundant mutation in surface-associated than in non-surface-associated genes

Comparison of the virulence genes cluster of lineage I, II and III

All genes of the virulence gene cluster are present in the four studied strains [27]. We performed a nucleotide sequence alignment of the entire virulence genes cluster, using the EGD-e sequence as a reference. As shown in Figure 2 we identified a truncation in the *actA* sequence of the 4b and the 4a genomes. In addition, a small truncation upstream the *mpl* gene and a truncation of a short repeat region distal to the PrfA binding box of *mpl* was present in the 4a genome. However, the PrfA binding site was not affected. Moreover, the alignment identity decreased slightly in the latter half of the cluster, with differences most prominently

visible in the regions containing *lmo0207* and *lmo0209*. *lmo0207* encodes a lipoprotein and was identified as one of the most divergent genes of the LIPI-1 when comparing three lineages.

Figure 2 Alignment of the virulence gene cluster of representatives of three *L. monocytogenes* lineages *L. monocytogenes* 1/2a EGD-e was used as reference genome. Nucleotide sequence identity of compared genomes is visualized. The top panel indicates location and direction of virulence genes

Interestingly, both the *L. monocytogenes* 4b strains (CLIP80459 and F2365), and the *L. monocytogenes* 4a L99 strain, have an identical repeat truncation in the ActA protein compared to ActA of the 1/2a EGD-e (Additional file 2: Table S2 Additional file 3: Table S3). Such truncations in *actA* have been reported previously for strain 4a L99 and affect the speed of movement of intracellular bacteria [28]. We surveyed sequenced *act*A alleles present in GenBank and discovered that the truncation in the ActA protein is far more frequent in 1/2b and 4b strains (77% and 51% respectively) than in 1/2a strains (7.5%).

Loss of surface proteins in lineage III

Several genes encoding internalin-like proteins are absent in the *L. monocytogenes* 4a L99 genome in comparison to the 1/2a EGD-e and the 4b strains (Additional file 4: Table S4) as previously reported for lineage III strains [27,29]. The entire *inlGHE* cluster [30] is absent in the 4a L99 genome (Additional file 5: Table S5) [27,30]. The corresponding loci in both 4b genomes are identical to each other, but different to strain 1/2a EGD-e. Another PrfA-independent internalin (InlJ) that has been shown to be specifically expressed only in vivo [31] is also absent from the 4a L99 genome. Similarly, Internalin C [27], involved in cell-to-cell spread and innate immune response in the vertebrate host [32-35], is absent in 4a L99 but is conserved in both 4b strains and 1/2a EGD-e. A comparable situation was identified for internalin F [27], however deletion mutants have not been shown to be reduced in invasion into non-phagocytic cells [36]. Apart from the absence of these characterized internalin genes, several other internalin-like genes (*lmo1666*, *lmo2470* and *lmo2821*, Additional file 4: Table S4) are present in the 1/2a EGD-e and 4b genomes, but are absent from the 4a L99 genome. In addition, we analysed the repertoire of genes encoding surface proteins for recently published 4a genomes of strain HCC23 [37] and M7 [38] as well as 4c FSL J2-071 (*Listeria monocytogenes* Sequencing Project, Broad Institute of Harvard and MIT; http://www.broad.mit.edu) (Additional files 4: Table S4, Additional file 6: Table S6 Additional file 7: Table S7 Additional file 8: Table S8 Additional file 9: Table S9 Additional file 10: Table S10 and Additional file 11: Figure S1). We confirmed by comparative genomics that these 4a genomes lack a similar number of surface proteins (Additional files 4: Table S4, Additional file 6: Table S6 Additional file 7: Table S7 Additional file 8: Table S8 Additional file 9: Table S9 Additional file 10: Table S10 and Additional file 11: Figure S1). These findings were independently verified by additional PCR analysis to confirm the absence of genes encoding surface proteins for four 4a strains and three 4c strains, respectively. Half of the inspected chromosomal loci differed by PCR analysis among 4a and 4c genomes (Additional file 11: Figure S1). Some non-internalin like cell-wall proteins that have been shown to be important for invasion are also absent, e.g. *auto* a GW-motif containing (Additional file 6: Table S6), PrfA-independent, surface autolysin. Previous studies revealed an essential role for *auto* in the entry into non-phagocytic eukaryotic cells [39]. The *vip* gene product, a PrfA-dependent LPXTG protein (Additional file 7: Table S7), described as a receptor for the eukaryotic Gp96 surface protein and important for late stages

of infection [40], is also absent from the 4a L99 genome. In addition to these missing genes, InlI is slightly truncated. However Ami (Additional file 6: Table S6), an important listerial adhesion protein seems to be present in a shorter version in both 4b strains [41,42], whereas the number of lipoproteins (Additional file 8: Table S8), LysM- and (Additional file 9: Table S9) NLPC/P60-motif containing proteins (Additional file 10: Table S10) was comparable among the four strains under study.

Overall, in comparison to 1/2a EGD-e and the two 4b genomes, 4a L99 strain has lost a number of crucial determinants required for listerial invasion. The selective loss of genes primarily responsible for the first steps of infection may contribute to the poor invasion ability and the attenuated nature of the 4a L99 strain.

Decay of phage genes in the *L. Monocytogenes* 4a L99 strain

The 1/2a EGD-e genome contains 79 prophage genes in two different loci, the 4a L99 genome includes 193 phage genes at four loci, while the 4b genomes encode with 16, for the smallest number of prophage genes limited to a single locus (also called the monocin-locus) at the same position in the chromosomes.

This monocin locus, a cryptic prophage region, is conserved in all *L. monocytogenes* lineages and includes the *lma* genes [43]. Although previously thought to be specific to *L. monocytogenes*, it was shown that *lmaDCBA* is also present in several apathogenic *L. innocua* strains. However, not all genes of the operon are present in all *L. monocytogenes* strains. The 4a L99 genome lacks *lmaA* and *lmaB* (Additional file 12: Figure S2). The entire locus in 1/2a EGD-e and the two 4b genomes has 16 genes, but only five of these genes are present in the 4a L99 genome. *lmaA* and *lmaB* are absent in *L. welshimeri*. Interestingly, the structure of this prophage locus in strain 4a L99 and other lineage III strains is more similar to *L. welshimeri* than to other pathogenic listeriae (Additional file 12: Figure S2).

The CRISPR system of *Listeria*

The *L. monocytogenes* 4a L99 genome was found to contain two adjacent CRISPR loci (I and II) with CRISPR repeats (Figure 3A and 3B). Both loci contain sequences of length 35 bp separated by repeat sequences of length 29 bp. However, they differ considerably in the number of repeat copies (6 in locus I, and 29 in locus II, respectively). While locus I is highly conserved in the 4b strains, 1/2a EGD-e and *L. innocua*, locus II was exclusively present in 4a genomes of L99, HCC23, M7, but not in another lineage III genome of 4c FSL J2-071 (Figure 3 A-C). It is not known whether the CRISPR system is functional in the 4a L99 genome. However, by sequence similarity searches using the spacers to detect possible prophage DNA traces, we were able to identify the PSA prophage that is known to infect serotype 4 strains. Assuming a functional CRISPR system in 4a L99 suggests a resistance to the PSA bacteriophage (Additional file 13: Figure S3).

Figure 3 Overview of CRISPR (clustered regularly interspaced short palindromic repeats) loci in *L monocytogenes* 1/2a EGD-e, *L. monocytogenes* 4a L99, *L. monocytogenes* 4a HCC23, *L. monocytogenes* 4a M7, *L. monocytogenes* 4c FSL J2-071, *L. monocytogenes* 4b CLIP80459, *L. monocytogenes* 4b F2365 and *L. innocua* 6a CLIP11262. (A): CRISPR locus I is shown for all five listeriae, black boxes indicate complete CRISPR repeats, red boxes represent incomplete or truncated (*) CRISPR repeats. No *cas* genes were found to be associated with this locus. Flanking genes are conserved in 1/2a EGD-e and both 4b

genomes. Comparison of the intergenic sequences with the 4a L99 genome revealed a sequence footprint of decaying repeat elements (2 repeat copies in both 4b genomes, and 1 copy in *L. innocua 6a* CLIP11262), indicating loss of the CRISPR repeats. (B): Locus II shows 29 copies of repeats and is associated with several *cas* genes (*cas2*, *cas3*, *cas5* and *cas6*. *cas1* is partially detectable, but seems to be truncated. (C): *L. innocua* 6a CLIP11262 harbours the CRISPR locus III at position 2.77 Mb in the genome, which is neighboured by a single *cas2* gene. No other CRISPR repeats nor any *cas* gene homologs were found in the 4b genomes

Gene duplications in the *Listeria* genomes expand metabolic systems

We found substantial evidence for a minimum of 231 to a maximum of 296 gene duplications in the *Listeria* genomes (Additional file 14: Figure S4 and Additional file 15: Figure S5). It is evident that the majority of these duplications are ancient events as they are shared among all species and the number of gene pairs with a very high percentage identity is very low (1-12% per strain). Functional classification of the duplicated genes revealed that many of these have important implications in metabolic pathways, like the pentose phosphate pathway, fructose and mannose metabolism, carbon fixation, glycolysis and pyruvate metabolism.

While several duplicated genes could be mapped to central metabolic pathways from the KEGG database, this was not possible for horizontally transferred genes (Additional file 16 Figure S6 and Additional file 17: Figure S7). However, not all duplicated genes seem to have arisen from true duplications, but some may have been transferred horizontally, like some PTS system genes that are *L. monocytogenes* EGD-e strain-specific genes. The number of genes classified into known metabolic pathways or systems was significantly higher for duplicated genes, while several horizontally transferred genes could not be mapped.

Comparative intracellular transcriptomics of four *L. Monocytogenes* strains of the three major lineages

Comparative transcriptome analysis of *Listeria monocytogenes* strains of the two major lineages revealed differences in virulence, cell wall, and stress response [44]. Here we performed intracellular gene expression analyses using whole genome microarrays between four *L. monocytogenes* strains belonging to the three major lineages to investigate eventual differences. P388D1 murine macrophages were infected and total RNA was isolated four hours post infection and hybridized to bioarrays.

In order to determine the core intracellular response of *L. monocytogenes* we created a dataset of core-syntenic homologous genes for all four genomes and the expression data for these genes were compared. We found that in all strains studied the entire virulence genes cluster, (*prfA*, *plcA*, *hly*, *mpl*, *actA*, *plcB* and *orfX)* was highly induced within the infected host cells. Furthermore genes known to be important for bacterial survival, such as *hpt*, *clpE*, *bilEA* and two LRR domain-containing proteins (*lmo0514* and *lmo2445*) were upregulated in all strains.

Interestingly, three mannose transporting PTS systems (*lmo0021-lmo0024*, *lmo0781-lmo0784*, *lmo1997-lmo2002*), two fructose specific systems (*lmo2335* and *lmo2733*), two galacitol specific systems (*lmo0503*, *lmo0507*, *lmo0508* and *lmo2665-lmo2667*), two beta-glucoside systems (the partial system *lmo0373-lmo0374* and *lmo0874-lmo0876*), and two cellobiose specific systems (the partial system *lmo0901* and *lmo0914-lmo0916*) were commonly upregulated in all strains. These possibly represent the most frequently used

substrates of listeriae in the cytosol. Only one mannose specific PTS system, (*lmo0096-lmo0098*) is downregulated by all studied strains (Additional file 18 Figure S8 and Additional file 19: Text S1).

Most surprisingly, all *Listeria* strains studied expressed the genes of the *lma* operon and surrounding prophage genes of the monocin locus, including a conserved holin (*lmo0112, lmo0113, lmo0115, lmo0116, lmo0128*) during intracellular growth. However, the functions of several of these genes are not defined. The only locus that is conserved in all three lineages (albeit with some deletions in 4a L99) is the monocin *lma* locus. The *lmaA* gene product has been shown to provoke a delayed type hypersensitivity reaction in mice immune to *L. monocytogenes*. It is also secreted at 20°C but much less [45] at 37°C. The *lma* operon produces two transcripts, a 2100 bp *lmaDCBA* transcript expressed both at 20°C and 37°C, and a 1050 bp *lmaBA* transcript induced at lower temperatures [43]. Additional prophage genes were highly expressed in the individual strains (Figure 4). Taken together, high intracellular prophage gene expression, despite several differences in prophage gene content, is one of the most striking observations across all *Listeria* lineages.

Figure 4 Comparative transcriptomics of four *L. monocytogenes* genomes: *L. monocytogenes* 1/2a EGD-e, *L. monocytogenes* 4a L99, *L. monocytogenes* 4b CLIP80459, *L. monocytogenes* 4b F2365 (from outside to inside). There are two tracks per strain: the first one shows the coding sequences (gray), phage genes (blue) and virulence genes (black). The second one visualizes increase (red) or decrease (green) of intracellular gene expression (log fold changes). Phage and virulence genes are clearly upregulated intracellularly. Data were illustrated using GenomeViz [26]

All strains showed induction of the *eut* operon suggesting that ethanolamine may be used as a carbon and nitrogen source in intracellular conditions. The zinc transporters were also commonly upregulated indicating a role of zinc in intracellular survival as well as the spermidine/putrescine ABC transporters (*potB, potC* and *potD*). Furthermore, the non-oxidative branch of the pentose phosphate pathway was utilized by all listeriae, possibly to generate NADPH for countering oxidative stress in intracellular conditions. The upregulation of genes of the pentose phosphate pathway has been shown previously [19,20,46] and it has been speculated that it is important for generation of erythrose-4-phosphate for aromatic amino acid biosynthesis or for generation of pentose sugars. Accordingly; we observed a downregulation of several genes involved in pyrimidine and purine biosynthesis from pentose sugars (e.g. *lmo1463, lmo1497, lmo1565, lmo1832, lmo1836, lmo1856, lmo1929, lmo2154, lmo2155, lmo2390* and *lmo2559*).

Downregulated genes included the *agr* locus (*lmo0048-lmo0051*) as demonstrated previously [20,46] and several genes of the tryptophan biosynthesis operon (*trpA, trpB, trpF* and *trpD*), and some tRNA synthetase genes (*ileS, valS, glyS* and *glyQ*). Diminished energy generation was indicated by decreased expression of the cytochrome genes cluster *cytABCD*. With respect to the pentose phosphate pathway, we detected downregulation of the phosphoribosyl pyrophosphate synthetase (*prs, lmo0199*) gene, which is required for the production of PRPP (phosphoribosyl pyrophosphate) that links the pentose phosphate pathway to the biosynthesis of purines and pyrimidines. While several genes of the glycolytic operon, and several individual genes were downregulated by 1/2a EGD-e, 4b CLIP80459 strain or 4a L99, the 4b F2365 strain showed increased expression (Additional file 20: Text S2).

Differences in flagellin expression are the most prominent differences among strains

To address the observation that strain 4b CLIP80459 grows more efficiently inside the host than strain 4b F2365, we performed a direct comparison of the transcriptome data derived from these two strains. Most important differences were found in the regulation of flagellar genes. While intracellular bacteria of strain 4b F2365 upregulated a substantial number of flagellar genes, including *fliS*, *fliI*, *flhA*, *fliF*, *filE*, *flgB*, *flgC*, *flgG*, *fliD* as well as the transcriptional regulator *degU* (*lmo2515*), in the 4b strain CLIP80459 only *fliR* was upregulated. When comparing the intracellular transcriptome of strain 4a L99 to the 1/2a and 4b strains the most striking difference was again the expression of the flagellar operon. We observed a strong induction of nearly all flagellar genes in the operon, including flagellin (Additional file 21: Text S3) (homologues of *lmo0675*, *lmo0676*, *lmo0681*, *lmo0685*, *lmo0686*, *lmo0690*-*lmo0696*, *lmo0698*-*lmo0701*, *lmo0703*-*lmo0706*, *lmo0708*, *lmo0709*, *lmo0712*, *lmo0714* and *lmo0715*) in strain 4a L99. Strong expression of these genes is counterproductive within infected cells, because it probably enables the host to efficiently detect bacterial presence and the formation of an inflammasome.

Apart from genes that are important for pathogen recognition mechanisms by the host, a concerted expression profile (Additional file 22: Figure S9) involving genes of cell wall synthesis, host cell invasion, response to oxidative stress, utilization of host carbohydrates and propanediol, which are crucial for intracellular survival as well as virulence and surface proteins were identified.

Differential growth of the three lineages and Δ*lmaB* and Δ*lmaD* isogenic mutants in a mouse infection and cell infection models

We observed a severe deficiency in entry of strain L99 in HeLa and Caco-2 cells as well as poor cell-to-cell transmission with macrophages and L929 fibroblasts when compared to 1/2a EGD-e (data not shown). Impaired invasion ability of host cells may be due to lack of several internalin genes in the genome of strain 4a L99. It is likely that both, decreased invasive ability and strong intracellular expression of flagellar genes contribute towards the rapid clearance of the 4a L99 strain in in vivo experiments in mice. Upregulation of several DNA repair genes was also seen in strain 4a L99 compared to the other strains, e.g. (*recF*, *recN*, *radA* and *mutL*), suggesting genomic damage during the infection process.

To further assess the virulence potential of the three lineages, we performed mouse infection experiments with each of the four strains (1600 cfu/mouse), and measured bacterial loads in spleens and livers at different time points (Figure 5A and 5B). The 4a L99 strain was cleared rapidly from the mice and was not detectable after five days of infection, in accordance with previous results [18], indicating that the 4a L99 strain is attenuated in its pathogenicity. However, the other three strains were able to survive in both spleens and livers of infected mice. Interestingly, while they could comparably replicate in the spleen, the 1/2a EGD-e and the 4b F2365 bacterial loads in liver were significantly lower than the 4b CLIP80459 strain whose counts remained significantly higher even on days five and eight post-infection. Isogenic mutants of Δ*lmaB* and Δ*lmaD* showed similar counts in mice spleens and livers. However, both mutants have shown a significantly lower level of growth than 1/2a EGD-e on days 3 and 5 post-infection (Figure 5C and 5D).

Figure 5 Murine infection studies with three different *Listeria* serotypes and two chromosomal deletion mutants of Δ*lmaB* and Δ*lmaD* of *L. monocytogenes* 1/2a EGD-e
Mice were infected i.v. with 2000 cfu of *L. monocytogenes* serotypes 1/2a EGD-e (filled circles), 4b F2365 (open circles), 4b CLIP80459 (filled triangles), and 4a L99 (open triangles). On days 1, 3, 5, and 8 after infection, the numbers of viable bacteria in spleens (A) and livers (B) of three animals per group were determined ($P \leq 0{,}05$ and $P \leq 0{,}001$ of 4b CLIP80459 vs. 1/2a EGD-e and 4b F2365 vs. 1/2a EGD-e in spleen and liver respectively). Bacterial load in mice organs were also determined following i.v. infection with 2000 cfu of *L. monocytogenes* 1/2a EGD-e wild type strain (filled circles) as well as its isogenic mutants Δ*lmaB* (open circles), and Δ*lmaD* (filled triangles). On days 1, 3, and 5 after infection, the numbers of viable bacteria in spleens (C) and livers (D) of three animals per group were determined ($P \leq 0{,}05$ and $P \leq 0{,}01$ of 1/2a EGD-e versus Δ*lmaB* and Δ*lmaD* in spleen and liver respectively). Data presented are representative of three independent experiments. An asterisk indicates means that are significantly different from the wild type. Significance analysis was performed with student *t*-test

Discussion

We sequenced and analysed the genomes of representatives of three major lineages of species *L. monocytogenes* to correlate gene content with (i) its wide spectrum of pathogenic abilities, (ii) its differing properties for survival in the hosts, and (iii) its adaptive properties during growth under extracellular conditions.

Decay of surface proteins in the virulence attenuated *L. Monocytogenes* 4a strain

Analysis of the 4a L99 genome revealed extensive loss of a large number of internalins, internalin-like proteins and other surface proteins important for invasive ability. For strain 4a L99, which was isolated from contaminated food in the 1950's, it might be possible that mutations have taken place over this lengthy time of storage under in vitro conditions. Surprisingly, a previously known *act*A truncation in the 4a genomes of L99, HCC23 and M7, was also found in a higher number of lineages I strains compared to lineage II, but not in the *act*A gene of another lineage III strain of 4c FSL J2-071 indicating a serotype-specific heterogeneity of ActA sequences within the genus *Listeria*. The loss of this proline-repeat in ActA is correlated with lowered actin-based motility in the cytosol. In addition, comparative nucleotide analysis indicated that the latter half of the LIPI-I pathogenicity island in strain 4a L99 has diverged significantly from that of the 4b and 1/2a strain leading to a loss of the open reading frames *lmo0206* to *lmo0209*. Loss of *lmo0206* (*orfX*) has been shown to confer a severe growth effect on survival in macrophages, [20] while loss of *lmo0207* has a small effect on growth in macrophages and no data are presently available for *lmo0208* and *lmo0209* and their role in virulence.

Differential regulation of intracellular flagella gene expression by strains of different lineages

Highly sensitive and widely distributed host microbe-associated microbial pattern receptors (TLRs and NLRs) continuously patrol the cell surface, endosomes and the cytosol for signs of microbial presence by sensing cell wall components, bacterial DNA, lipoproteins and flagellin. Ligands may be shared between the surface and the cytosolic receptors, e.g. cell

wall components and flagellin may be sensed both by TLRs and also by cytosolic receptors. We detected the intracellular expression of the flagellin gene in 1/2a EGD-e [20]. Recently, it has been shown that cytosolic flagellin, expressed by *L. monocytogenes* strain 10403 S (serotype 1/2a) is detected by multiple Nod-like receptors, including IPAF and NALP3, and also by a pathway involving the adaptor protein ASC and the cytosolic DNA sensor AIM2, which is required for the formation of the inflammasome [47-49]. Detection of flagellin in the cytosol via these pathways leads to caspase-1 mediated cleavage of pro-IL-1B and release of active IL-1B. Mice lacking caspase-1 or ASC are unable to mount active IL-1B response to intracellular pathogens such as *Shigella flexneri* and *Francisella tularensis* [50,51]. All strains investigated in this study were found to express flagellar genes in the cytosol, except for strain 4b CLIP80459. The ability to successfully downregulate flagellar (*flaA*) gene expression is probably critical for evading host detection and promoting bacterial intracellular growth. In line with this observation, a 1/2a EGD-e chromosomal deletion mutant of the gene displayed increased survival in mouse infection assays [52].

In keeping with this finding, both strains 4b F2365 and 4a L99 displayed strong induction of several flagellar genes during intracellular growth and were more readily cleared from the host. This suggests strain-specific differences in the ability to avoid host recognition can lead to large differences in virulence manifestation, despite several commonalities in the adaptations of the lineages to the intracellular lifestyle. Although all the strains investigated in this study were able to induce all genes of the virulence genes cluster intracellularly, it is likely that there are a multitude of effects including differences in virulence gene expression, uptake of carbohydrates, membrane protein expression and flagellar biosynthesis, all of which contribute to the observed phenotypic properties.

Effects of gene duplication events on metabolic adaptation and survival within the host

The processes of gene duplications, horizontal gene transfer and gene loss influence the short- and long-term evolution of prokaryotic genomes. The benefits of gene duplications in the short term can be seen clearly in conditions of antibiotic treatment [53,54], toxin exposure [55], heavy metal stress [56,57], extreme temperatures [58], nutrient limitation [59,60] and even parasitic and symbiotic lifestyles [54,61]. Duplications found in all *Listeria* genomes seem to have been ancient i.e. precede species differentiation, with only the exception of the recent prophage duplication in *L. innocua* 6a CLIP11262. Classification of duplicated genes revealed several paralogous genes in metabolic pathways, while very few horizontally transferred genes could be classified at all.

The highest numbers of gene duplications were identified in the following categories: ABC transporters, PTS systems, pentose phosphate pathway, starch and sucrose metabolism, fructose and mannose metabolism, and carbon fixation. Surprisingly, we found a high number of duplicated gene paralogues involved in the regulation of the non-oxidative branch of the pentose phosphate pathway and in the generation of ribose-5-phospate from ribulose-5-phosphate. Under conditions of intracellular growth, we observed differences in the ability of the lineages to express horizontally transferred genes. 1/2a EGD-e was most successful in this regard (17 genes), followed by 4a L99 (10 genes), 4b F2365 (6 genes) and 4b CLIP80459 (2 genes). Apart from the horizontally transferred genes, differences in the expression of strain-specific genes in the cytosol were apparent (1/2a EGD-e: 45; 4a L99: 49; 4b F2365 11; 4b CLIP80459: 3).

PTS systems enable listeriae to utilize host carbohydrates, a mechanism that is essential for the intracellular survival. PTS systems (EII) for the utilization of fructose and beta-glucosides, mannose and cellobiose were most frequently observed in the investigated *Listeria* genomes. Although the numbers of PTS systems are comparable among the investigated genomes (Additional file 18: Figure S8), even a slight difference in presence/absence of a PTS system available as an additional carbohydrate utilization mechanism may have dramatic effects on listerial survival inside the host cytosol [61-63], specifically on the master regulator PrfA [61,62,64,65]. For instance, the pentitol PTS system in 1/2a EGD-e is not present in either the 4b or the 4a L99 genomes. A transposon insertion mutant of this system (*lmo1971*) has been shown to have significantly attenuated growth in epithelial cells [46]. Several partial PTS systems are also present in the genome (Additional file 19: Text S1). These are independently expressed intracellularly, and represent broadly shared and commonly regulated systems. In accordance, the pathogenic strain 4b CLIP80459 was found to upregulate more PTS systems than strain 4b F2365, which may contribute to better intracellular survival of 4b CLIP80459.

In addition to phosphorylated sugars, there are other nitrogen and carbon sources available to intracellular bacteria, such as ethanolamine. Ethanolamine is used as substrate and an energy supply by *Salmonella enterica* grown under anaerobic conditions and is suggested to be used by other bacteria [66]. A locus homologous to that of the ethanolamine operon of *S. enterica* has also been described in *Listeria* [67]. The gene organization of the locus is not identical to the *Salmonella* cluster, but all the genes of the cluster have homologous sequences in *Listeria* (Additional file 23: Figure S10). Previous studies identified genes of the locus to be upregulated intracellularly during infection and were shown to play a critical role for intracellular survival [46]. Our data support this observation and further demonstrate upregulation of several genes of this locus across all three pathogenic lineages of *Listeria*, suggesting that the functions of the locus are conserved. However, since the locus is also present in the apathogenic *L. innocua* strain 6a CLIP11262, it may exemplify a general requirement of *Listeria* to cope with nutrient rather than a specific virulence adaptation. Furthermore, degradation of the phagosomal membrane that traps intracellular listeriae, results in the release of ethanolamine as a byproduct and may serve an energy source in the host cytosol.

Not only the efficient recruitment of carbohydrate substrates, but also the differential channeling through different pathways represents an important adaption within the host cytosol. It has been shown that an essential mechanism to counteract oxidative stress is to reroute carbohydrate flux via the pentose phosphate pathway, which is required for the biosynthesis of reductive substrates rather than through glycolysis pathway [68]. Indeed, we observed that all lineages prefer to channel carbohydrate flux via the pentose phosphate pathway, rather than glycolysis. In contrast to the other strains, only strain 4b F2365 was unable to downregulate glycolysis, suggesting that the inability to route sugars efficiently via pentose phosphate contributes to the poor intracellular growth of this strain.

The CRISPR system in *Listeria* reveals expansion and atrophy

A CRISPR (Clustered, regularly interspaced short palindromic repeats) locus, associated with several *cas* genes was identified in the 4a L99 genome. CRISPRs are highly divergent loci found in genomes of all archaea and several bacteria [69]. A CRISPR system is composed of the *cas* (CRISPR-associated) genes, a leader sequence and arrays of direct repeats separated by non-repetitive spacer sequences resulting in a RNA-interference like innate phage-

resistance mechanism [70]. A recent study in *Streptococcus thermophilus* demonstrated how bacteria are able to integrate new spacer sequences derived from infecting phages, directly into the CRISPR arrays, and that this ability confers phage-resistance [71]. The mechanism of resistance has also been elucidated [70]. Among the genomes compared in this study, only the 4a L99 genomes of L99, HCC23 and M7 possesses *cas* genes and several CRISPR repeats. There are only two repeats in each 4b genome, five in 1/2a EGD-e a single one in *L. innocua* 6a CLIP11262, but none of these strains harbour identifiable *cas* genes. In addition, a small sRNA *rliB* is located in the repeat region of 1/2a EGD-e and contributes to virulence in mice [72]. We were also able to detect a DNA sequence of a potential prophage (PSA) using the spacers from the 4a genome. As prophages evolve quite rapidly, it is likely that this acquisition is a recent event.

Distinct role of intracellularly upregulated phage genes in virulence of listerial strains

The four *L. monocytogenes* strains have different numbers of prophage genes (1/2a EGD-e: 79; 4a L99: 191; 4b CLIP80459: 16 and 4b F2365: 16) distributed in different loci. Regardless of location and lineage, all strains expressed several prophage genes within the infected host cell. However, only a single locus, the *lma* locus is conserved across the three lineages and is also induced during infection. The role of prophage genes in the virulence of *Listeria* has not been examined in detail. We show that chromosomal deletion mutants of two genes in this locus (*lmaB* and *lmaD*) resulted in growth reduction of 1/2a EGD-e in a murine infection model. Although the underlying mechanisms leading to the attenuated phenotypes remain unclear, a recent study revealed that prophage diversification represents an essential mechanism for short-term genome evolution within the species *L. monocytogenes* [73,74] and is subject of further investigation.

Conclusion

Listeria monocytogenes is arguably one of the best characterized pathogens and has been established as an unparalleled model microorganism in infection biology. Detailed understanding of differences in virulence of the three major lineages of *Listeria* provides us with invaluable information about evolutionary adaptation of this pathogen. Here we used comparative genomics and whole-genome based transcriptome analysis of strains from all lineages to obtain a comprehensive view as to how these strains have evolutionarily diverged. This approach suggests that (i) reductive evolution of strains of serotype 4a such as L99, HCC23 and M7 is the major force driving the attenuated phenotype, (ii) acquisition and adaptation of prophage genes and metabolic systems, respectively, identify novel virulence-associated factors of listeriae and (iii) listeriae avoid detection and subsequent immune response of the host via downregulation of surface structures and by differences in intracellular expression of flagellar genes.

Methods

Strains and growth conditions

Four *L. monocytogenes* strains were used in the study, *L. monocytogenes* 1/2a EGD-e [14], *L. monocytogenes* 4a L99 [18], *L. monocytogenes* 4b CLIP80459 [17], *L. monocytogenes* 4b F2365 [15] and chromosomal deletion mutants of *L. monocytogenes* 1/2a EGD-e Δ*lmaB* and

Δ*lmaD*. Bacteria were grown in brain heart infusion (BHI) broth (Difco) at 37°C with shaking. For further comparative genomic analysis *L. monocytogenes* 4a HCC23 [37] *L. monocytogenes* 4a M7 [38] and *L. monocytogenes* 4c FSL J2-071) (*Listeria monocytogenes* Sequencing Project, Broad Institute of Harvard and MIT; http://www.broad.mit.edu) was used.

Genome sequencing and annotation

In brief, genome sequencing *L. monocytogenes* 4a L99 was performed on ABI PRISM 3100 or 3730xl Genetic Analyzers (Applied Biosystems). Whole genome shotgun sequencing was performed by LGC (Berlin, Germany). Sequence data were analysed and assembled using Phred/Phrap/Consed [75,76]. A total number of 27,637 sequences of shotgun libraries, 1684 fosmid and 671 PCR gap closure sequences were assembled by the Phrap software resulting in a ~6.7-fold coverage. Genome annotation was performed as previously described [3].

Genome sequencing of *L. monocytogenes* 4b CLIP80459 was performed using the conventional whole genome shotgun strategy [77,78]. One library (2–3 kb inserts) was generated by random mechanical shearing of genomic DNA and cloning into pcDNA-2.1 (Life technologies) and recombinant plasmids were used as templates for cycle sequencing reactions. Samples were loaded on capillary automatic 3700 and 3730 DNA sequencers (Applied Biosystems). In an initial step 35,610 sequences were assembled into 361 contigs using the Phred/Phrap/Consed software [75,76]. CAAT-Box [79] was used to predict links between contigs. 379 PCR products amplified from *L. monocytogenes* CLIP80459 chromosomal DNA as template were used to fill gaps and to re-sequence low quality regions. Final assembly resulted in a ~7.8-fold coverage. Genome annotation was performed as previously described [14].

Alignment of the virulence gene cluster

The alignment was performed using MAVID [80] after extracting the virulence gene cluster of all genomes. The plot was created using VISTA [81].

ActA repeat analysis

Available ActA protein sequences for all *L. monocytogenes* strains were retrieved from GenBank (http://www.ncbi.nlm.nih.gov/Genbank/). Only sequences that contained at least 500 amino acids (reference strain 1/2a EGD-e ActA: 639 amino acids) were downloaded (774 sequences). It was possible to assign a lineage to only 386 ActA sequences. Duplicates with identical length, strain and sequence were also removed, leaving a total of 218 sequences for the analysis. These were aligned using ClustalW and the alignment of repeat regions was examined manually.

Single nucleotide polymorphisms

Single nucleotide Polymorphisms (SNPs) were detected by the MUMmer [25] and SNPs were mapped to coding regions using PERL scripts. The SNP-density per gene normalized by gene length was calculated and the data were visualized in GenomeViz [26].

CRISPR repeats analysis

Comparative visualization of the CRISPR related genome loci was performed by GECO [82]. CRISPR repeats were identified using the PILER-CR software [83]. Subsequent analysis and visualization of repeat footprints was performed using BLAST and ACT [84].

Horizontal gene transfer and gene duplications

Horizontally transferred genes were detected using SIGI [85] and SIGI-HMM [86]. Duplicated genes were identified using BLAST cut-offs of at least 40% identity and 80% coverage considering both sequences.

Cell culture and infection model

All cell culture experiments were performed as described by Chatterjee and colleges [20].

Microarrays

For each of the four strains of the study, a genome-wide custom microarray chip was designed and implemented using the Geniom One platform from Febit Biomed GmbH, Germany. All transcriptome studies were performed with this platform. Complete details of the protocols are provided in the ArrayExpress database (http://www.ebi.ac.uk/microarray-as/ae/). Data were background corrected and then normalized using quantile normalization [87]. Pearson's correlation coefficients were used to assess reproducibility within at least two technical and three biological replicates ($r^2 >= 0.94$ in all cases). The significance analysis of microarrays (SAM) program was used to analyze the data [88] as an unpaired response.

Construction of the deletion mutants Δ*lmaB* and Δ*lmaD*

Chromosomal in frame deletion mutants of *L. monocytogenes* 1/2a EGD-e Δ*lmaB* and Δ*lmaD* were constructed by generating the 5′ (with primers P1 and P2) and the 3′ (with primers P3 and P4) flanking region of the gene concerned. Primers used to generate the flanking regions are shown (Additional file 24: Table S11). The purified PCR fragments of 5′ and 3′ flanking regions were amplified using primer P1 and P4, ligated into pCRII (Life technologies) and transformed into *E. coli* InvαF' electrocompetent cells (Life technologies). Subsequently, the vector was digested with restriction enzyme *Eco*RI and ligated into the temperature sensitive suicide vector pAUL-A which was digested with the same enzymes and transformed into *E. coli* InvαF' electrocompetent cells. Plasmid DNA of pAUL-A bearing the fragment was isolated from the recombinants and used to transform *L. monocytogenes* EGD-e to generate the chromosomal deletion mutants as described in detail by Schaeferkordt et al. [89]. The deletion in the gene concerned was identified by PCR and confirmed by sequencing the PCR fragment using primers P1 and P4.

Murine infection assay

Primary infection with *L. monocytogenes* serotypes and mutants was performed by intravenous injection of viable bacteria in a volume of 0.2 ml of PBS. Bacterial growth in spleens and livers was determined by plating 10-fold serial dilutions of organ homogenates on BHI after several days. The detection limit of this procedure was 10^2 CFU per organ.

Colonies were counted after 24 h of incubation at 37°C. Six- to eight-week-old female BALB/c mice, purchased from Harlan Winkelmann (Borchen, Germany), were used in all experiments.

Ethics statement

This study was carried out in strict accordance with the regulation of the National Protection Animal Act (§7-9a Tierschutzgesetz). The protocol was approved by the local Committee on the Ethics of Animal Experiments (Regierungsbezirk Mittelhessen) and permission was given by the local authority (Regierungspraesidium Giessen, Permit Number: GI 15/5-Nr.63/2007).

Statistical data analysis of infection experiments

All infection experiments were performed a minimum of three times. Significant differences between two values were compared with a paired Student's *t*-test. Values were considered significantly different when the p value was less than 0.05 ($p < 0.05$).

Nucleotide sequence and microarray accession number

The genome sequences have been deposited in the EMBL database with accession numbers FM211688 for *L. monocytogenes* 4a L99 and FM242711 for *L. monocytogenes* 4b CLIP80459 respectively. The microarray data have been submitted to ArrayExpress with the accession number E-MEXP-1947.

Competing interests

The authors declare that they have no competing interests.

Author' contributions

TH and TC designed the study. CB, PG, CS and TH performed genome sequencing, MAM isolated total RNA. SV performed microarray samples preparation and hybridization. TH, AB, CTK UK, ED, BI, RG, SS, PG, CR, CB, AG, SO and WM performed the genome annotation and analysis work. RG, AB, CTK, BI, TH and TC drafted and wrote the manuscript. All authors contributed to and approved the final manuscript.

Acknowledgments

We thank Alexandra Amend, Claudia Zörb and Nelli Schklarenko for excellent technical assistance. This work was supported by the fund obtained from the BMBF through the Competence Network PathoGenoMik (031U213B) to T.H. and T.C and through the ERA-NET Pathogenomics Network (62080061) to T.H.

References

1. Hain T, Chatterjee SS, Ghai R, Kuenne CT, Billion A, Steinweg C, Domann E, Karst U, Jansch L, Wehland J, Eisenreich W, Bacher A, Joseph B, Schar J, Kreft J, Klumpp J,

Loessner MJ, Dorscht J, Neuhaus K, Fuchs TM, Scherer S, Doumith M, Jacquet C, Martin P, Cossart P, Rusniock C, Glaser P, Buchrieser C, Goebel W, Chakraborty T: **Pathogenomics of *Listeria* spp.** *Int J Med Microbiol* 2007, **297:**541–557.

2. Vazquez-Boland JA, Kuhn M, Berche P, Chakraborty T, Dominguez-Bernal G, Goebel W, Gonzalez-Zorn B, Wehland J, Kreft J: ***Listeria* pathogenesis and molecular virulence determinants.** *Clin Microbiol Rev* 2001, **14:**584–640.

3. Hain T, Steinweg C, Kuenne CT, Billion A, Ghai R, Chatterjee SS, Domann E, Karst U, Goesmann A, Bekel T, Bartels D, Kaiser O, Meyer F, Puhler A, Weisshaar B, Wehland J, Liang C, Dandekar T, Lampidis R, Kreft J, Goebel W, Chakraborty T: **Whole-genome sequence of *Listeria welshimeri* reveals common steps in genome reduction with *Listeria innocua* as compared to *Listeria monocytogenes*.** *J Bacteriol* 2006, **188:**7405–7415.

4. Cossart P, Vicente MF, Mengaud J, Baquero F, Perez-Diaz JC, Berche P: **Listeriolysin O is essential for virulence of *Listeria monocytogenes*: direct evidence obtained by gene complementation.** *Infect Immun* 1989, **57:**3629–3636.

5. Cossart P, Toledo-Arana A: ***Listeria monocytogenes*, a unique model in infection biology: an overview.** *Microbes Infect* 2008, **10:**1041–1050.

6. Gaillard JL, Berche P, Frehel C, Gouin E, Cossart P: **Entry of *L. monocytogenes* into cells is mediated by internalin, a repeat protein reminiscent of surface antigens from gram-positive cocci.** *Cell* 1991, **65:**1127–1141.

7. Lecuit M, Ohayon H, Braun L, Mengaud J, Cossart P: **Internalin of *Listeria monocytogenes* with an intact leucine-rich repeat region is sufficient to promote internalization.** *Infect Immun* 1997, **65:**5309–5319.

8. Dominguez-Bernal G, Muller-Altrock S, Gonzalez-Zorn B, Scortti M, Herrmann P, Monzo HJ, Lacharme L, Kreft J, Vazquez-Boland JA: **A spontaneous genomic deletion in *Listeria ivanovii* identifies LIPI-2, a species-specific pathogenicity island encoding sphingomyelinase and numerous internalins.** *Mol Microbiol* 2006, **59:**415–432.

9. Cotter PD, Draper LA, Lawton EM, Daly KM, Groeger DS, Casey PG, Ross RP, Hill C: **Listeriolysin S, a novel peptide haemolysin associated with a subset of lineage I Listeria monocytogenes.** *PLoS Pathog* 2008, **4:**e1000144.

10. Orsi RH, den Bakker HC, Wiedmann M: ***Listeria monocytogenes* lineages: Genomics, evolution, ecology, and phenotypic characteristics.** *Int J Med Microbiol* 2011, **301:**79–96.

11. Ward TJ, Ducey TF, Usgaard T, Dunn KA, Bielawski JP: **Multilocus genotyping assays for single nucleotide polymorphism-based subtyping of *Listeria monocytogenes* isolates.** *Appl Environ Microbiol* 2008, **74:**7629–7642.

12. Jacquet C, Doumith M, Gordon JI, Martin PM, Cossart P, Lecuit M: **A molecular marker for evaluating the pathogenic potential of foodborne *Listeria monocytogenes*.** *J Infect Dis* 2004, **189:**2094–2100.

13. Murray EGD, Webb RA, Swann HBR: **A disease of rabbits characterized by a large mononuclear leucocytosis caused by a hitherto undescribed bacillus *Bacterium monocytogenes* (n.sp.).** *J Pathol Bacteriol* 1926, **29:**407–439.

14. Glaser P, Frangeul L, Buchrieser C, Rusniok C, Amend A, Baquero F, Berche P, Bloecker H, Brandt P, Chakraborty T, Charbit A, Chetouani F, Couve E, de Daruvar A, Dehoux P, Domann E, Dominguez-Bernal G, Duchaud E, Durant L, Dussurget O, Entian KD, Fsihi H, Garcia-del Portillo F, Garrido P, Gautier L, Goebel W, Gomez-Lopez N, Hain T, Hauf J, Jackson D: **Comparative genomics of *Listeria* species.** *Science* 2001, **294:**849–852.

15. Nelson KE, Fouts DE, Mongodin EF, Ravel J, DeBoy RT, Kolonay JF, Rasko DA, Angiuoli SV, Gill SR, Paulsen IT, Peterson J, White O, Nelson WC, Nierman W, Beanan MJ, Brinkac LM, Daugherty SC, Dodson RJ, Durkin AS, Madupu R, Haft DH, Selengut J, Van Aken S, Khouri H, Fedorova N, Forberger H, Tran B, Kathariou S, Wonderling LD, Uhlich GA, Bayles DO, Luchansky JB, Fraser CM: **Whole genome comparisons of serotype 4b and 1/2a strains of the food-borne pathogen *Listeria monocytogenes* reveal new insights into the core genome components of this species.** *Nucleic Acids Res* 2004, **32:**2386–2395.

16. Nightingale KK, Milillo SR, Ivy RA, Ho AJ, Oliver HF, Wiedmann M: ***Listeria monocytogenes* F2365 carries several authentic mutations potentially leading to truncated gene products, including *inlB*, and demonstrates atypical phenotypic characteristics.** *J Food Prot* 2007, **70:**482–488.

17. de Valk H, Vaillant V, Jacquet C, Rocourt J, Le Querrec F, Stainer F, Quelquejeu N, Pierre O, Pierre V, Desenclos JC, Goulet V: **Two consecutive nationwide outbreaks of Listeriosis in France, October 1999-February 2000.** *Am J Epidemiol* 2001, **154:**944–950.

18. Chakraborty T, Ebel F, Wehland J, Dufrenne J, Notermans S: **Naturally occurring virulence-attenuated isolates of *Listeria monocytogenes* capable of inducing long term protection against infection by virulent strains of homologous and heterologous serotypes.** *FEMS Immunol Med Microbiol* 1994, **10:**1–9.

19. Camejo A, Buchrieser C, Couve E, Carvalho F, Reis O, Ferreira P, Sousa S, Cossart P, Cabanes D: **In vivo transcriptional profiling of *Listeria monocytogenes* and mutagenesis identify new virulence factors involved in infection.** *PLoS Pathog* 2009, **5:**e1000449.

20. Chatterjee SS, Hossain H, Otten S, Kuenne C, Kuchmina K, Machata S, Domann E, Chakraborty T, Hain T: **Intracellular gene expression profile of *Listeria monocytogenes*.** *Infect Immun* 2006, **74:**1323–1338.

21. Joseph B, Goebel W: **Life of *Listeria monocytogenes* in the host cells' cytosol.** *Microbes Infect* 2007, **9:**1188–1195.

22. Chen Y, Zhang W, Knabel SJ: **Multi-virulence-locus sequence typing identifies single nucleotide polymorphisms which differentiate epidemic clones and outbreak strains of *Listeria monocytogenes*.** *J Clin Microbiol* 2007, **45:**835–846.

23. Canchaya C, Fournous G, Brussow H: **The impact of prophages on bacterial chromosomes.** *Mol Microbiol* 2004, **53:**9–18.

24. Dorscht J, Klumpp J, Bielmann R, Schmelcher M, Born Y, Zimmer M, Calendar R, Loessner MJ: **Comparative genome analysis of *Listeria* bacteriophages reveals extensive mosaicism, programmed translational frameshifting, and a novel prophage insertion site.** *J Bacteriol* 2009, **191:**7206–7215.

25. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL: **Versatile and open software for comparing large genomes.** *Genome Biol* 2004, **5:**R12.

26. Ghai R, Hain T, Chakraborty T: **GenomeViz: visualizing microbial genomes.** *BMC Bioinforma* 2004, **5:**198.

27. Doumith M, Cazalet C, Simoes N, Frangeul L, Jacquet C, Kunst F, Martin P, Cossart P, Glaser P, Buchrieser C: **New aspects regarding evolution and virulence of *Listeria monocytogenes* revealed by comparative genomics and DNA arrays.** *Infect Immun* 2004, **72:**1072–1083.

28. Sokolovic Z, Schuller S, Bohne J, Baur A, Rdest U, Dickneite C, Nichterlein T, Goebel W: **Differences in virulence and in expression of PrfA and PrfA-regulated virulence genes of *Listeria monocytogenes* strains belonging to serogroup 4.** *Infect Immun* 1996, **64:**4008–4019.

29. Jia Y, Nightingale KK, Boor KJ, Ho A, Wiedmann M, McGann P: **Distribution of internalin gene profiles of *Listeria monocytogenes* isolates from different sources associated with phylogenetic lineages.** *Foodborne Pathog Dis* 2007, **4:**222–232.

30. Raffelsbauer D, Bubert A, Engelbrecht F, Scheinpflug J, Simm A, Hess J, Kaufmann SH, Goebel W: **The gene cluster i*nlC2DE* of *Listeria monocytogenes* contains additional new internalin genes and is important for virulence in mice.** *Mol Gen Genet* 1998, **260:**144–158.

31. Linden SK, Bierne H, Sabet C, Png CW, Florin TH, McGuckin MA, Cossart P: ***Listeria monocytogenes* internalins bind to the human intestinal mucin MUC2.** *Arch Microbiol* 2008, **190:**101–104.

32. Domann E, Zechel S, Lingnau A, Hain T, Darji A, Nichterlein T, Wehland J, Chakraborty T: **Identification and characterization of a novel PrfA-regulated gene in *Listeria monocytogenes* whose product, IrpA, is highly homologous to internalin proteins, which contain leucine-rich repeats.** *Infect Immun* 1997, **65:**101–109.

33. Engelbrecht F, Chun SK, Ochs C, Hess J, Lottspeich F, Goebel W, Sokolovic Z: **A new PrfA-regulated gene of *Listeria monocytogenes* encoding a small, secreted protein which belongs to the family of internalins.** *Mol Microbiol* 1996, **21:**823–837.

34. Gouin E, dib-Conquy M, Balestrino D, Nahori MA, Villiers V, Colland F, Dramsi S, Dussurget O, Cossart P: **The *Listeria monocytogenes* InlC protein interferes with innate immune responses by targeting the I{kappa}B kinase subunit IKK{alpha}.** *Proc Natl Acad Sci U S A* 2010, **107:**17333–17338.

35. Rajabian T, Gavicherla B, Heisig M, Muller-Altrock S, Goebel W, Gray-Owen SD, Ireton K: **The bacterial virulence factor InlC perturbs apical cell junctions and promotes cell-to-cell spread of *Listeria*.** *Nat Cell Biol* 2009, **11:**1212–1218.

36. Dramsi S, Dehoux P, Lebrun M, Goossens PL, Cossart P: **Identification of four new members of the internalin multigene family of *Listeria monocytogenes* EGD.** *Infect Immun* 1997, **65:**1615–1625.

37. Steele CL, Donaldson JR, Paul D, Banes MM, Arick T, Bridges SM, Lawrence ML: **Genome sequence of lineage III *Listeria monocytogenes* strain HCC23.** *J Bacteriol* 2011, **193:**3679–3680.

38. Chen J, Xia Y, Cheng C, Fang C, Shan Y, Jin G, Fang W: **Genome sequence of the nonpathogenic *Listeria monocytogenes* serovar 4a strain M7.** *J Bacteriol* 2011, **193:**5019–5020.

39. Cabanes D, Dussurget O, Dehoux P, Cossart P: **Auto, a surface associated autolysin of *Listeria monocytogenes* required for entry into eukaryotic cells and virulence.** *Mol Microbiol* 2004, **51:**1601–1614.

40. Cabanes D, Sousa S, Cebria A, Lecuit M, Garcia-del Portillo F, Cossart P: **Gp96 is a receptor for a novel *Listeria monocytogenes* virulence factor, Vip, a surface protein.** *EMBO J* 2005, **24:**2827–2838.

41. Milohanic E, Jonquieres R, Glaser P, Dehoux P, Jacquet C, Berche P, Cossart P, Gaillard JL: **Sequence and binding activity of the autolysin-adhesin Ami from epidemic *Listeria monocytogenes* 4b.** *Infect Immun* 2004, **72:**4401–4409.

42. Milohanic E, Jonquieres R, Cossart P, Berche P, Gaillard JL: **The autolysin Ami contributes to the adhesion of *Listeria monocytogenes* to eukaryotic cells via its cell wall anchor.** *Mol Microbiol* 2001, **39:**1212–1224.

43. Schaferkordt S, Chakraborty T: **Identification, cloning, and characterization of the *Ima* operon, whose gene products are unique to *Listeria monocytogenes*.** *J Bacteriol* 1997, **179:**2707–2716.

44. Severino P, Dussurget O, Vencio RZ, Dumas E, Garrido P, Padilla G, Piveteau P, Lemaitre JP, Kunst F, Glaser P, Buchrieser C: **Comparative transcriptome analysis of *Listeria monocytogenes* strains of the two major lineages reveals differences in virulence, cell wall, and stress response.** *Appl Environ Microbiol* 2007, **73:**6078–6088.

45. Gohmann S, Leimeister-Wachter M, Schiltz E, Goebel W, Chakraborty T: **Characterization of a *Listeria monocytogenes*-specific protein capable of inducing delayed hypersensitivity in *Listeria*-immune mice.** *Mol Microbiol* 1990, **4:**1091–1099.

46. Joseph B, Przybilla K, Stuhler C, Schauer K, Slaghuis J, Fuchs TM, Goebel W: **Identification of *Listeria monocytogenes* genes contributing to intracellular replication by expression profiling and mutant screening.** *J Bacteriol* 2006, **188:**556–568.

47. Sauer JD, Witte CE, Zemansky J, Hanson B, Lauer P, Portnoy DA: ***Listeria monocytogenes* triggers AIM2-mediated pyroptosis upon infrequent bacteriolysis in the macrophage cytosol.** *Cell Host Microbe* 2010, **7:**412–419.

48. Warren SE, Mao DP, Rodriguez AE, Miao EA, Aderem A: **Multiple Nod-like receptors activate caspase 1 during *Listeria monocytogenes* infection.** *J Immunol* 2008, **180:**7558–7564.

49. Wu J, Fernandes-Alnemri T, Alnemri ES: **Involvement of the AIM2, NLRC4, and NLRP3 inflammasomes in caspase-1 activation by *Listeria monocytogenes.*** *J Clin Immunol* 2010, **30:**693–702.

50. Mariathasan S, Weiss DS, Dixit VM, Monack DM: **Innate immunity against *Francisella tularensis* is dependent on the ASC/caspase-1 axis.** *J Exp Med* 2005, **202:**1043–1049.

51. Sansonetti PJ, Phalipon A, Arondel J, Thirumalai K, Banerjee S, Akira S, Takeda K, Zychlinsky A: **Caspase-1 activation of IL-1beta and IL-18 are essential for *Shigella flexneri*-induced inflammation.** *Immunity* 2000, **12:**581–590.

52. Bigot A, Pagniez H, Botton E, Frehel C, Dubail I, Jacquet C, Charbit A, Raynaud C: **Role of FliF and FliI of *Listeria monocytogenes* in flagellar assembly and pathogenicity.** *Infect Immun* 2005, **73:**5530–5539.

53. Koch AL: **Evolution of antibiotic resistance gene function.** *Microbiol Rev* 1981, **45:**355–378.

54. Romero D, Palacios R: **Gene amplification and genomic plasticity in prokaryotes.** *Annu Rev Genet* 1997, **31:**91–111.

55. Reinbothe S, Ortel B, Parthier B: **Overproduction by gene amplification of the multifunctional arom protein confers glyphosate tolerance to a plastid-free mutant of Euglena gracilis.** *Mol Gen Genet* 1993, **239:**416–424.

56. Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV: **Selection in the evolution of gene duplications.** *Genome Biol* 2002, **3:**RESEARCH0008.

57. van Hoof NA, Hassinen VH, Hakvoort HW, Ballintijn KF, Schat H, Verkleij JA, Ernst WH, Karenlampi SO, Tervahauta AI: **Enhanced copper tolerance in *Silene vulgaris* (Moench) Garcke populations from copper mines is associated with increased transcript levels of a 2b-type metallothionein gene.** *Plant Physiol* 2001, **126:**1519–1526.

58. Riehle MM, Bennett AF, Long AD: **Genetic architecture of thermal adaptation in *Escherichia coli.*** *Proc Natl Acad Sci U S A* 2001, **98:**525–530.

59. Brown CJ, Todd KM, Rosenzweig RF: **Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment.** *Mol Biol Evol* 1998, **15:**931–942.

60. Sonti RV, Roth JR: **Role of gene duplications in the adaptation of *Salmonella typhimurium* to growth on limiting carbon sources.** *Genetics* 1989, **123:**19–28.

61. Lai CY, Baumann L, Baumann P: **Amplification of *trpEG*: adaptation of *Buchnera aphidicola* to an endosymbiotic association with aphids.** *Proc Natl Acad Sci U S A* 1994, **91:**3819–3823.

62. Stoll R, Mertins S, Joseph B, Muller-Altrock S, Goebel W: **Modulation of PrfA activity in *Listeria monocytogenes* upon growth in different culture media.** *Microbiology* 2008, **154:**3856–3876.

63. Stoll R, Goebel W: **The major PEP-phosphotransferase systems (PTSs) for glucose, mannose and cellobiose of *Listeria monocytogenes*, and their significance for extra- and intracellular growth.** *Microbiology* 2010, **156:**1069–1083.

64. Ake FM, Joyet P: *Deutscher J.* Milohanic E: **Mutational analysis of glucose transport regulation and glucose-mediated virulence gene repression in *Listeria monocytogenes*.** *Mol Microbiol* 2011.

65. Joseph B, Mertins S, Stoll R, Schar J, Umesha KR, Luo Q, Muller-Altrock S, Goebel W: **Glycerol metabolism and PrfA activity in *Listeria monocytogenes*.** *J Bacteriol* 2008, **190:**5412–5430.

66. Cannon GC, Bradburne CE, Aldrich HC, Baker SH, Heinhorst S, Shively JM: **Microcompartments in prokaryotes: carboxysomes and related polyhedra.** *Appl Environ Microbiol* 2001, **67:**5351–5361.

67. Buchrieser C, Rusniok C, Kunst F, Cossart P, Glaser P: **Comparison of the genome sequences of *Listeria monocytogenes* and *Listeria innocua*: clues for evolution and pathogenicity.** *FEMS Immunol Med Microbiol* 2003, **35:**207–213.

68. Ralser M, Wamelink MM, Kowald A, Gerisch B, Heeren G, Struys EA, Klipp E, Jakobs C, Breitenbach M, Lehrach H, Krobitsch S: **Dynamic rerouting of the carbohydrate flux is key to counteracting oxidative stress.** *J Biol* 2007, **6:**10.

69. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der OJ, Koonin EV: **Evolution and classification of the CRISPR-Cas systems.** *Nat Rev Microbiol* 2011, **9:**467–477.

70. Tang TH, Polacek N, Zywicki M, Huber H, Brugger K, Garrett R, Bachellerie JP, Huttenhofer A: **Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*.** *Mol Microbiol* 2005, **55:**469–481.

71. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P: **CRISPR provides acquired resistance against viruses in prokaryotes.** *Science* 2007, **315:**1709–1712.

72. Toledo-Arana A, Dussurget O, Nikitas G, Sesto N, Guet-Revillet H, Balestrino D, Loh E, Gripenland J, Tiensuu T, Vaitkevicius K, Barthelemy M, Vergassola M, Nahori MA,

Soubigou G, Regnault B, Coppee JY, Lecuit M, Johansson J, Cossart P: **The *Listeria* transcriptional landscape from saprophytism to virulence.** *Nature* 2009, **459:**950–956.

73. Orsi RH, Borowsky ML, Lauer P, Young SK, Nusbaum C, Galagan JE, Birren BW, Ivy RA, Sun Q, Graves LM, Swaminathan B, Wiedmann M: **Short-term genome evolution of *Listeria monocytogenes* in a non-controlled environment.** *BMC Genomics* 2008, **9:**539.

74. Verghese B, Lok M, Wen J, Alessandria V, Chen Y, Kathariou S, Knabel S: **comK prophage junction fragments as markers for *Listeria monocytogenes* genotypes unique to individual meat and poultry processing plants and a model for rapid niche-specific adaptation, biofilm formation, and persistence.** *Appl Environ Microbiol* 2011, 77:3279–3292.

75. Ewing B, Hillier L, Wendl MC, Green P: **Base-calling of automated sequencer traces using phred.** *I. Accuracy assessment. Genome Res* 1998, **8:**175–185.

76. Gordon D, Abajian C, Green P: **Consed: a graphical tool for sequence finishing.** *Genome Res* 1998, **8:**195–202.

77. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM: **Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.** *Science* 1995, **269:**496–512.

78. Frangeul L, Nelson KE, Buchrieser C, Danchin A, Glaser P, Kunst F: **Cloning and assembly strategies in microbial genome projects.** *Microbiology* 1999, **145(Pt 10):**2625–2634.

79. Frangeul L, Glaser P, Rusniok C, Buchrieser C, Duchaud E, Dehoux P, Kunst F: **CAAT-Box, Contigs-Assembly and Annotation Tool-Box for genome sequencing projects.** *Bioinformatics* 2004, **20:**790–797.

80. Bray N, Pachter L: **MAVID: constrained ancestral alignment of multiple sequences.** *Genome Res* 2004, **14:**693–699.

81. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I: **VISTA: computational tools for comparative genomics.** *Nucleic Acids Res* 2004, **32:**W273–W279.

82. Kuenne CT, Ghai R, Chakraborty T, Hain T: **GECO–linear visualization for comparative genomics.** *Bioinformatics* 2007, **23:**125–126.

83. Edgar RC: **PILER-CR: fast and accurate identification of CRISPR repeats.** *BMC Bioinforma* 2007, **8:**18.

84. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J: **ACT: the Artemis Comparison Tool.** *Bioinformatics* 2005, **21:**3422–3423.

85. Merkl R: **SIGI: score-based identification of genomic islands.** *BMC Bioinforma* 2004, **5:**22.

86. Waack S, Keller O, Asper R, Brodag T, Damm C, Fricke WF, Surovcik K, Meinicke P, Merkl R: **Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models.** *BMC Bioinforma* 2006, **7:**142.

87. Bolstad BM, Irizarry RA, Astrand M, Speed TP: **A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.** *Bioinformatics* 2003, **19:**185–193.

88. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci U S A* 2001, **98:**5116–5121.

89. Schaferkordt S, Chakraborty T: **Vector plasmid for insertional mutagenesis and directional cloning in *Listeria* spp.** *Biotechniques* 1995, **19:**720–725.

90. Billion A, Ghai R, Chakraborty T, Hain T: **Augur–a computational pipeline for whole genome microbial surface protein prediction and classification.** *Bioinformatics* 2006, **22:**2819–2820.

Additional files

Additional_file_1 as PDF
Additional file 1 Table S1. Nucleotide analysis of *actA* repeats of *Listeria*.

Additional_file_2 as XLS
Additionaf file 2 Table S2. Prediction of LRR region containing proteins by Augur [90].

Additional_file_3 as XLS
Additional file 3 Table S3. Prediction of proteins containing GW modules by Augur [90].

Additional_file_4 as XLS
Additional file 4 Table S4. Prediction of LPXTG motif harbouring proteins by Augur [90].

Additional_file_5 as XLS
Additional file 5 Table S5. Prediction of lipoproteins by Augur [90].

Additional_file_6 as XLS
Additional file 6 Table S6. Prediction of LysM domain containing proteins by Augur [90].

Additional_file_7 as XLS
Additional file 7 Table S7. Prediction of NLPC/P60 domain containing proteins by Augur [90].

Additional_file_8 as PDF
Additional file 8 Table S8. Comparative CRISPR analysis table.

Additional_file_9 as PDF
Additional file 9 Table S9. Metabolosome of *L. monocytogenes* 1/2a EGD-e.

Additional_file_10 as PDF
Additional file 10 Table S10. Primers used in this study.

Additional_file_11 as PNG
Additional file 11 Figure S1. Comparative analysis of *L. monocytogenes* ActA protein sequences.

Additional_file_12 as PDF
Additional file 12 Figure S2. Comparison of the *inlGHE* locus in the three listerial lineages. All three genes in this cluster have been absent in the *L. monocytogenes* 4a L99 genome.

Additional_file_13 as PDF
Additional file 13 Figure S3. Genome analysis of *lmaDCBA* region of six listeriae. Comparative analysis was performed using GECO [82] applying bidirectional pairs.

Additional_file_14 as PDF
Additional file 14 Figure S4. Frequency of distributions of the percentage identity between all duplicated gene pairs in the *Listeria* genomes.

Additional_file_15 as PDF
Additional file 15 Figure S5 Gene duplication and horizontal gene transfer in *Listeria* genomes.

Additional_file_16 as PDF
Additional file 16 Figure S6. Duplication vs. HGT classifiable genes in listeriae.

Additional_file_17 as PDF
Additional file 17 Figure S7. Complete PTS Systems in *L. monocytogenes* strains.

Additional_file_18 as PDF
Additional file 18 Figure S8. Partial PTS Systems in *L. monocytogenes* strains.

Additional_file_19 as PDF
Additional file 19 Text S1. SNP analysis of three listerial lineages.

Additional_file_20 as PDF
Additional file 20 Text S2. Differential regulation of glycolysis in *L. monocytogenes* 4b F2365.

Additional_file_21 as PDF
Additional file 21 Text S3. Comparison of two *L. monocytogenes* 4b strains CLIP80459 and F2365.

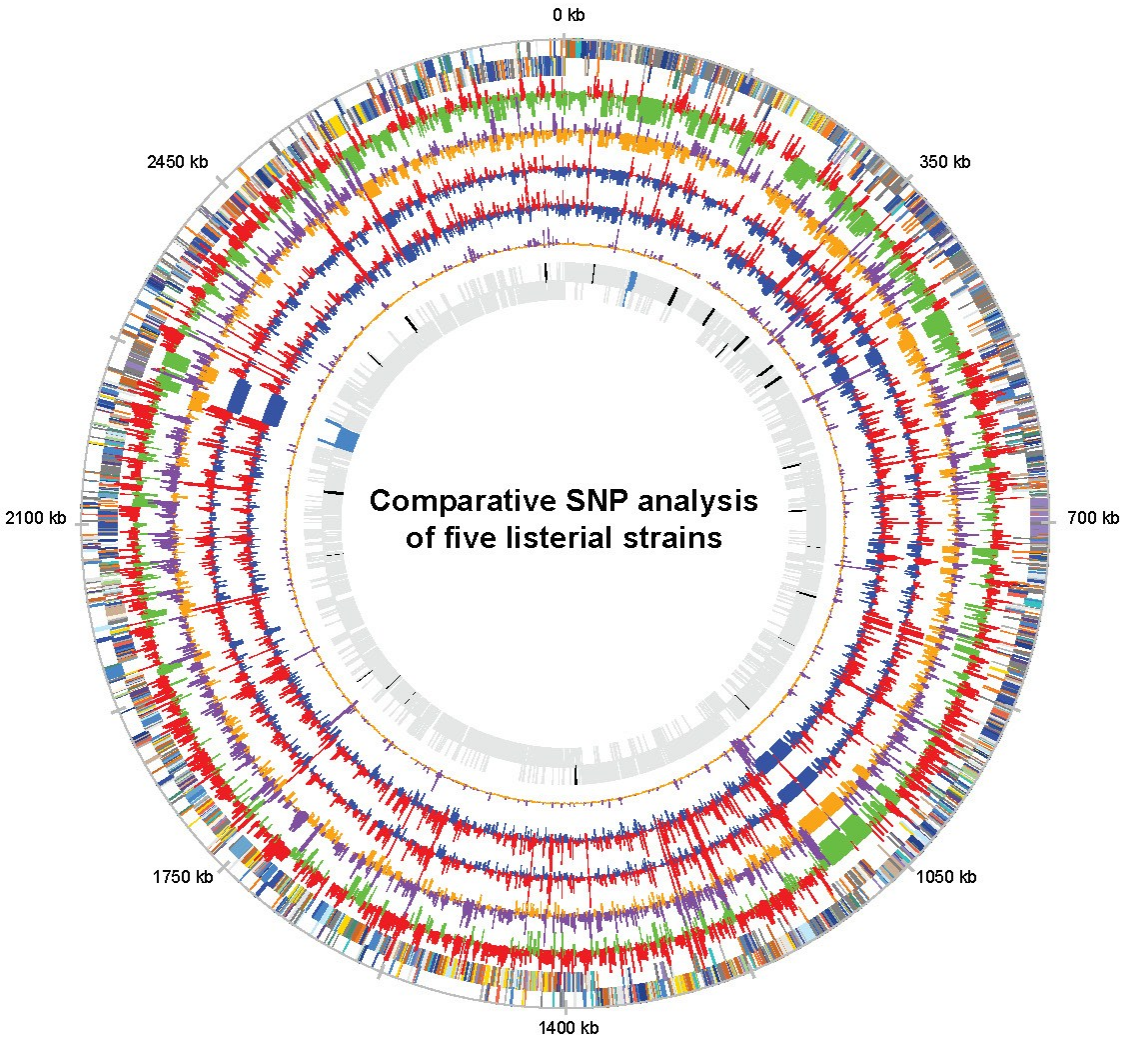
Additional_file_22 as PDF
Additional file 22 Figure S9. Confirmation of lacking genes encoding surface proteins in four *L. monocytogenes* 4a strains and three *L. monocytogenes* 4c strain generated by PCR analysis.

Additional_file_23 as PDF

Additional file 23 Figure S10. Intracellular flagellin expression data of *L. monocytogenes* 1/2a EGD-e, *L. monocytogenes* 4a L99, *L. monocytogenes* 4b CLIP80459 and *L. monocytogenes* 4b F2365 generated by qRT-PCR analysis.

Additional_file_24 as XLS

Additional file 24 Table S11. List of gene duplication in *Listeria* genomes.



	L. monocytogenes 1/2a EGD-e vs. *L. innocua* 6a CLIP11262	*L. monocytogenes* 1/2a EGD-e vs. *L. monocytogenes* 4a L99	*L. monocytogenes* 1/2a EGD-e vs. *L. monocytogenes* 4b CLIP80459	*L. monocytogenes* 1/2a EGD-e vs. *L. monocytogenes* 4b F2365	*L. monocytogenes* 4b F2365 vs. *L. monocytogenes* 4b CLIP80459
% nucleotide divergence (surface-associated CDS)	11.14	7.00	5.16	5.09	0.59
% nucleotide divergence (non-surface-associated CDS)	10.77	6.38	5.01	4.99	0.57
% nucleotide divergence ratio	3.46	9.79	2.96	2.02	4.55

Figure 1

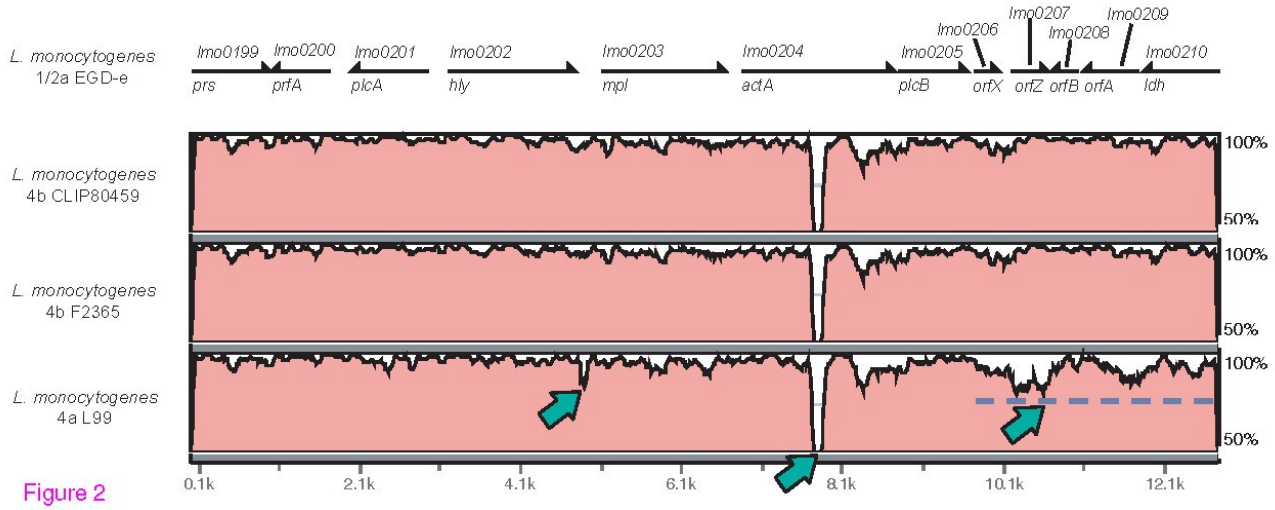
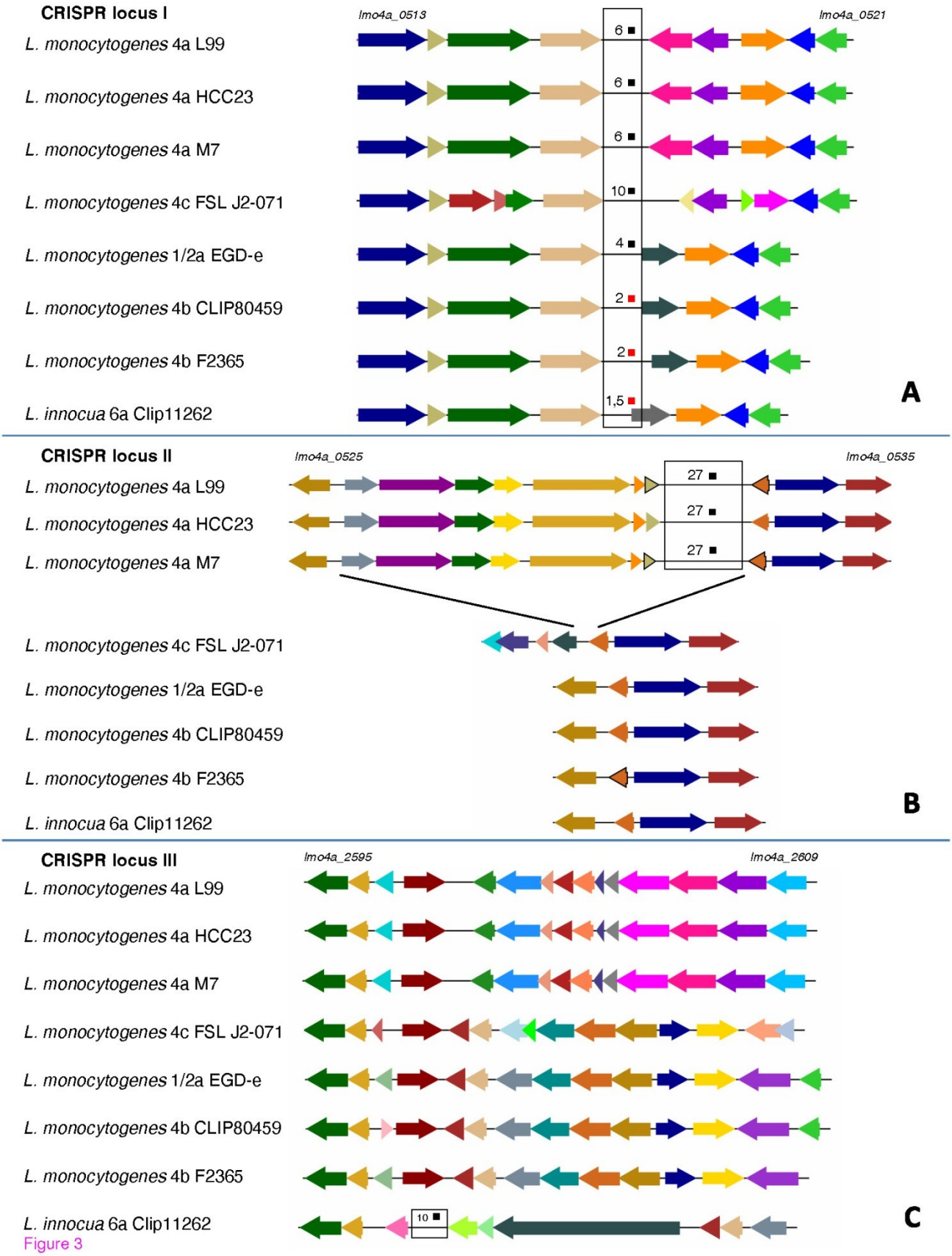

L. monocytogenes 1/2a EGD-e
lmo0199
lmo0200
lmo0201
lmo0202
lmo0203
lmo0204
lmo0205
lmo0206
lmo0207
lmo0208
lmo0209
lmo0210
prs
prfA
plcA
hly
mpl
actA
plcB
orfX
orfZ
orfB
orfA
ldh
L. monocytogenes 4b CLIP80459
L. monocytogenes 4b F2365
L. monocytogenes 4a L99
100%
50%
0.1k
2.1k
4.1k
6.1k
8.1k
10.1k
12.1k

Figure 2


CRISPR locus I
lmo4a_0513
lmo4a_0521
L. monocytogenes 4a L99
L. monocytogenes 4a HCC23
L. monocytogenes 4a M7
L. monocytogenes 4c FSL J2-071
L. monocytogenes 1/2a EGD-e
L. monocytogenes 4b CLIP80459
L. monocytogenes 4b F2365
L. innocua 6a Clip11262
6
6
6
10
4
2
2
1,5
A
CRISPR locus II
lmo4a_0525
lmo4a_0535
L. monocytogenes 4a L99
L. monocytogenes 4a HCC23
L. monocytogenes 4a M7
27
27
27
L. monocytogenes 4c FSL J2-071
L. monocytogenes 1/2a EGD-e
L. monocytogenes 4b CLIP80459
L. monocytogenes 4b F2365
L. innocua 6a Clip11262
B
CRISPR locus III
lmo4a_2595
lmo4a_2609
L. monocytogenes 4a L99
L. monocytogenes 4a HCC23
L. monocytogenes 4a M7
L. monocytogenes 4c FSL J2-071
L. monocytogenes 1/2a EGD-e
L. monocytogenes 4b CLIP80459
L. monocytogenes 4b F2365
L. innocua 6a Clip11262
10
C
Figure 3

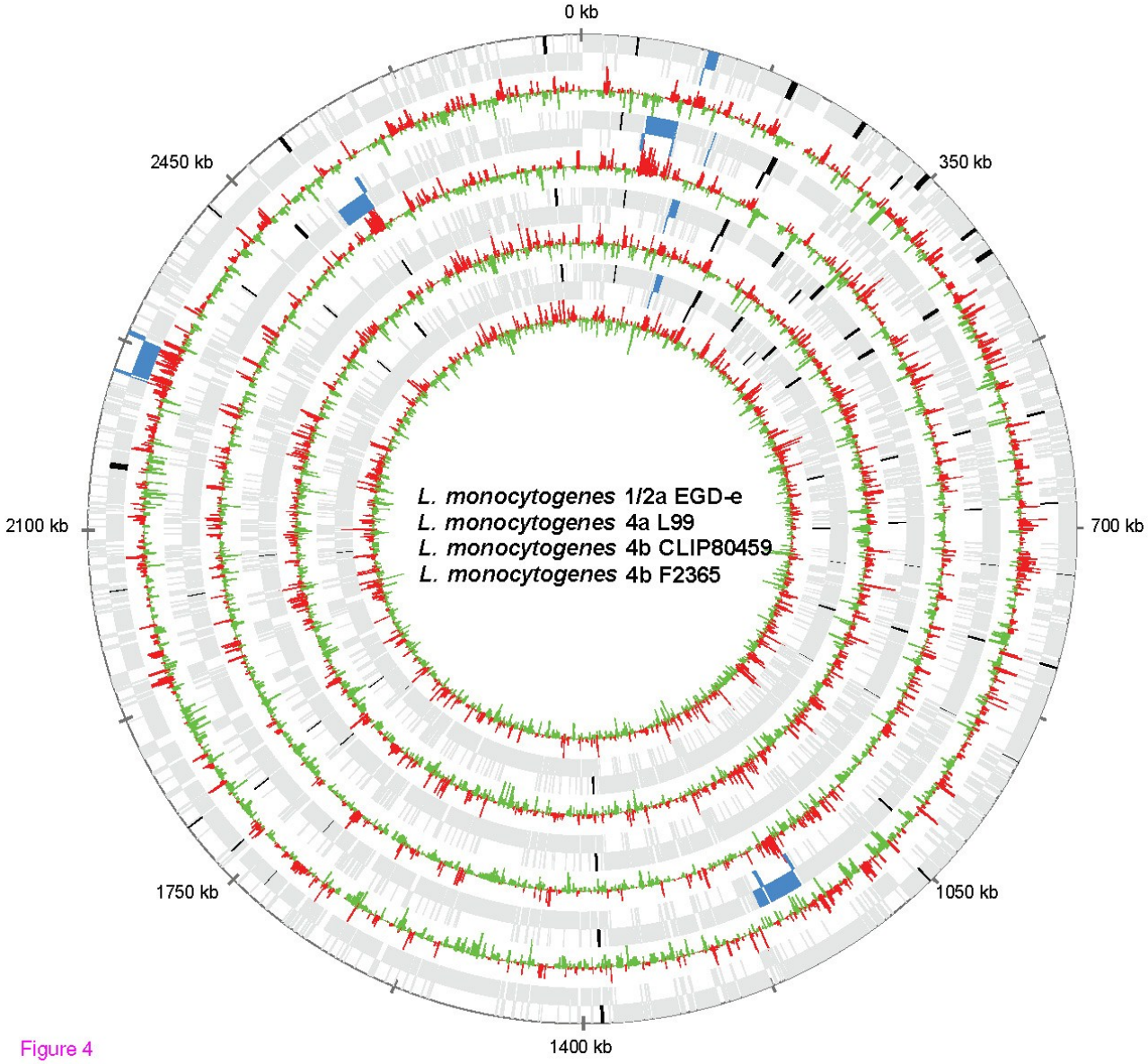

0 kb
350 kb
700 kb
1050 kb
1400 kb
1750 kb
2100 kb
2450 kb
L. monocytogenes 1/2a EGD-e
L. monocytogenes 4a L99
L. monocytogenes 4b CLIP80459
L. monocytogenes 4b F2365

Figure 4

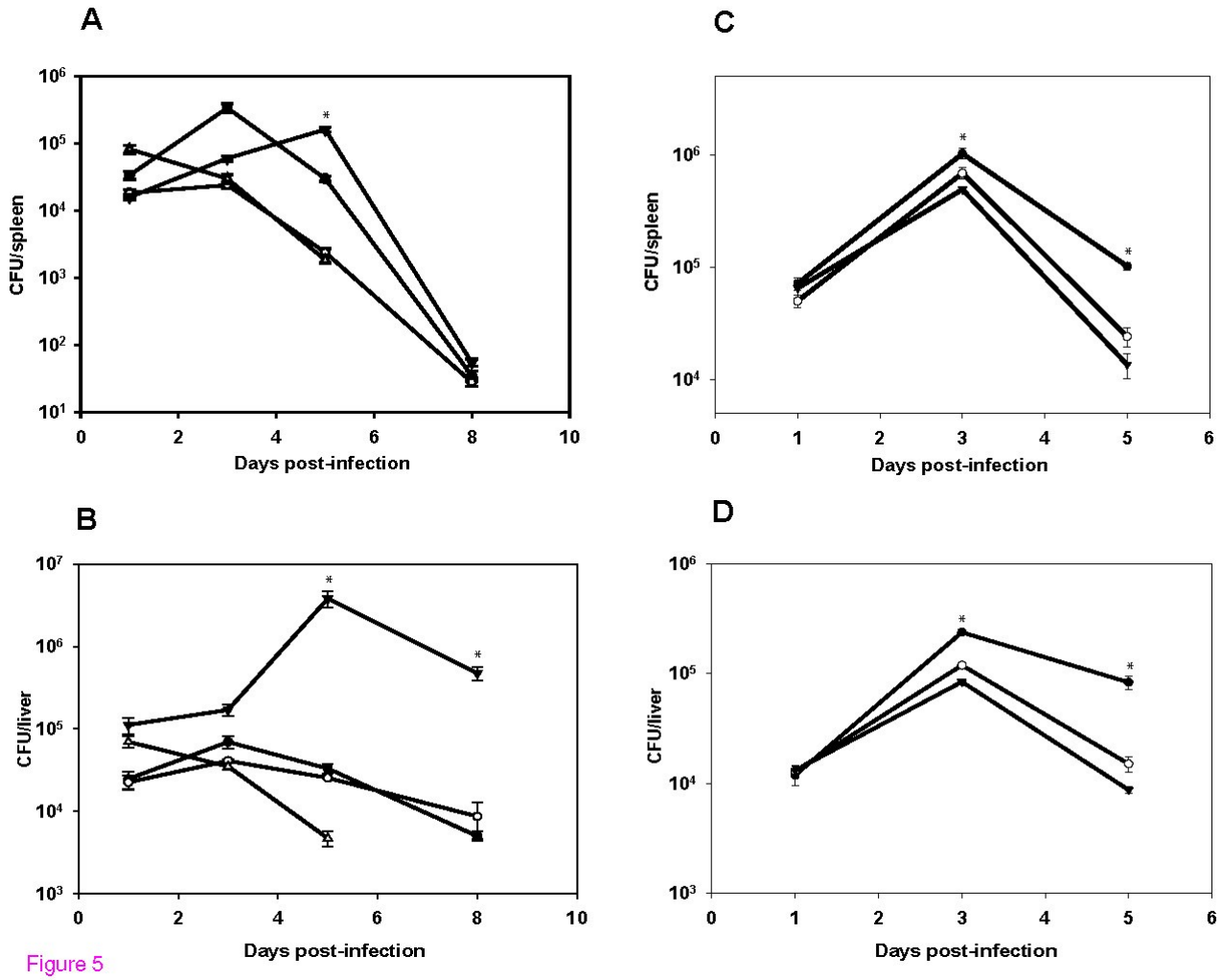

A
CFU/spleen
Days post-infection
B
CFU/liver
Days post-infection
C
CFU/spleen
Days post-infection
D
CFU/liver
Days post-infection

Figure 5

13.3 Appendix publication 2

International Journal of Medical Microbiology 301 (2011) 547–555

Contents lists available at ScienceDirect

International Journal of Medical Microbiology

journal homepage: www.elsevier.de/ijmm

Research paper

Adaptation of *Listeria monocytogenes* to oxidative and nitrosative stress in IFN-γ-activated macrophages

Mobarak Abu Mraheil *, André Billion, Walid Mohamed, Deepak Rawool, Torsten Hain, Trinad Chakraborty

Institute for Medical Microbiology, Justus-Liebig-University, Frankfurter Straße 107, 35392 Giessen, Germany

ARTICLE INFO

Article history:
Received 7 September 2010
Received in revised form 29 March 2011
Accepted 1 May 2011

Keywords:
Listeria monocytogenes
IFN-γ
Macrophages
ROS
HPF

ABSTRACT

IFN-γ-activated macrophages are considered to be the primary effector cells in host defense against *Listeria monocytogenes* infections. However despite the induction of the complex host defense mechanisms, survival of *L. monocytogenes* in activated macrophages is still observed. Here we used a whole genome-based transcriptome approach to examine for bacterial genes specifically induced in IFN-γ-activated macrophages. We demonstrated that cells activated by IFN-γ had elevated oxidative and nitrosative stress levels in both the activated macrophages as well as in the intracellular replicating bacteria isolated from these infected cells. We found that a subset of 21 transcripts were specifically differentially regulated in bacteria growing in cells pretreated with IFN-γ. Bioinformatics and functional analysis revealed that many of these genes have roles involved in overcoming oxidative stress and contribute to bacterial survival within activated macrophages. We detected increased transcription of the putative *trpE* gene of *L. monocytogenes*, encoding an anthranilate synthase, in bacteria growing in IFN-γ cells indicating host cell metabolic restriction of bacterial growth. Indeed we found enhanced activation of host cell genes involved in the kynurenine pathway indicating an increased need of *L. monocytogenes* for tryptophan during replication in IFN-γ-activated macrophages.

© 2011 Elsevier GmbH. All rights reserved.

Introduction

Listeria monocytogenes, the causative agent of listeriosis, is an important pathogen that has been involved in numerous outbreaks due to contaminated food (Schuchat et al., 1991; Swaminathan and Gerner-Smidt, 2007). The intracellular pathogen invades and multiplies in many cell types including epithelial cells and macrophages and can breach the intestinal, fetoplacental, and blood-brain barriers. As a result infection with *Listeria* is often associated with severe disease in pregnant women, neonates, the elderly, and immunosuppressed individuals (Vázquez-Boland et al., 2001).

L. monocytogenes is an established model for evaluation of crucial cellular interactions such as initiation of the host T-cell response which plays an important role in innate immunity (Pamer, 2004). Macrophages are the key actors for innate immune response during infection with *L. monocytogenes* as they are the major targets for intracellular replication and they are essential for mediating clearance of bacteria during infection. Infected macrophages secrete TNFα and IL-12 (Havell, 1987; Tripp et al., 1993; Hsieh et al., 1993) which in turn lead to the production of IFN-γ by natural killer cells and T-cells resulting in activation of macrophages and increase of their bactericidal activity. Cytokines such as IFN-γ act on macrophages and enhance their antimicrobial properties by targeting genes including those encoding inducible nitric oxide synthase (Bogdan et al., 2000), an enzyme that generates nitric oxide (NO) radicals. Other targets of IFN-γ are the subunits of the NADPH phagocyte oxidase, which generate superoxide (O_2^-) radicals, and gene products that starve intracellular pathogens by depleting the levels of tryptophan, such as indoleamin-2,3-dioxygenase (Decker et al., 2002).

Recently, we cataloged genes required for intracellular replication of *L. monocytogenes* in macrophages by comparing the expression profiles of bacteria grown in the cytosol of P388D1 macrophage cells with that of bacteria grown extracellularly in rich medium using whole genome microarrays (Chatterjee et al., 2006). Subsequent studies using bacterial mutants provided insight on how *L. monocytogenes* adapts to growth in the vacuolar and cytosolic compartments of infected macrophages by remodeling transcription through the coordinated expression of regulatory, metabolic, stress, and virulence genes. However to date, there is no information on those bacterial genes specifically regulated in IFN-γ-activated macrophages. Also, the role of IFN-γ in the resistance to infections with facultative intracellular bacteria still remains unclear as various studies have been shown to be contradictory

* Corresponding author. Tel.: +49 641 99 41251; fax: +49 641 99 41259.
E-mail address: Mobarak.Mraheil@mikrobio.med.uni-giessen.de (M.A. Mraheil).

1438-4221/$ – see front matter © 2011 Elsevier GmbH. All rights reserved.
doi:10.1016/j.ijmm.2011.05.001

Appendix

(van Dissel et al., 1987; Langermans et al., 1992; Herskovits et al., 2007).

In view of the above we used whole genome-based microarrays with *L. monocytogenes* within IFN-γ-activated macrophages to gain a better understanding of the genes involved in adaptation to the host innate immune response. We identified listerial genes specifically regulated in IFN-γ-activated cells and found many of them to be associated with counteracting increased oxidative and metabolic stress.

Materials and methods

Strains and growth conditions

L. monocytogenes EGD-e (Glaser et al., 2001) was grown in brain heart infusion (BHI) broth (Difco) at 37 °C with shaking at 180 rpm (Unitron, Infors). All experiments in the present study were performed with bacterial cultures growing to mid-exponential phase (optical density of 1.0 at 600 nm).

Cell culture infection assay

P388D1 murine macrophage cells were cultured and infected as described by Chatterjee et al. (2006). Tryptophan-free RPMI medium was purchased from PAN Biotech, Germany).

IFN-γ activation of macrophages

P388D1 macrophage cells were activated by incubating a cell monolayer with recombinant murine IFN-γ (100 U/ml; Sigma–Aldrich) for 18 h before and during infection.

Reagents

Catalase, N^G-Monomethyl-L-arginine acetate salt (L-NMMA), Spermine-Nitric oxide complex hydrate (Spermine NONOate), hydorogen peroxide (H_2O_2) and ethanol were obtained from Sigma–Aldrich. Catalase and L-NMMA were added to cultures at final concentrations of 100 μg/ml and 100 μM, respectively. Gentamicin solution (50 mg/ml) used in the study was procured from Gibco and 3′-(p-hydroxyphenyl) fluorescein (HPF) was purchased from Invitrogen.

RNA isolation

Total bacterial RNA was extracted as previously described by Chatterjee et al. (2006). Briefly, overnight cultures of *L. monocytogenes* EGD-e grown in BHI broth were diluted 1:50 in 20 ml fresh BHI broth in 100-ml Erlenmeyer flasks and were incubated at 37 °C with shaking at 180 rpm until the optical density of 1.0 at 600 nm was reached (Unitron, Infors). The RNA isolation from *L. monocytogenes* grown in media was performed as follows: Bacterial aliquots of 0.5 ml were preincubated with 1 ml RNAprotect (Qiagen) for 5 min and centrifuged for 10 min (8000 × *g*) at 20 °C. Pellets were stored at −80 °C until usage. RNA isolation from intracellularly grown bacteria at 4 h post infection was carried out as previously described (Chatterjee et al., 2006). Briefly, macrophages were lysed using a cold mix of 0.1% (w/v) sodium dodecyl sulfate, 1% (v/v) acidic phenol and 19% (v/v) ethanol in water. Bacteria were harvested by centrifugation for 10 min (8000 × *g*) and stored at −80 °C. RNA was extracted from bacteria using the RNeasy Mini kit (Qiagen) and stored at −80 °C. The quantity of the isolated total RNA was determined by absorbance at 260 nm and 280 nm, and the quality was assessed using the Agilent 2100 Bioanalyzer. cDNA was generated and labeled with CyDyes (both Cy3 and Cy5 for each probe) from 3 μg of total RNA using the CyScribe Post-Labelling Kit (GE Health Care). Quantification of cDNA was measured at the absorbance of 550 nm and 650 nm for Cy3 and Cy5 dyes, respectively, using an ND-1000 Spectrophotometer (NanoDrop Technologies, Inc., USA).

Eucaryotic RNA was isolated from macrophage cells as described here: RNA was extracted from P388D1 macrophages using the RNeasy Mini Kit (Qiagen). To synthesize cDNA, 1 μg of total RNA, SuperScript II Reverse transcriptase, QuantiTect Primer Assays, and RNAseOUT ribonuclease inhibitor (Invitrogen) were used.

Transcriptome analysis

The whole genome microarray for *L. monocytogenes* EGD-e was performed as previously described by Chatterjee et al. (2006). Briefly, hybridization of cDNA was carried out in ASP base Hybridizer (GE Health Care) for 12 h at 42 °C. The comparative experiment layout was as follows: (i) Bacteria grown in BHI versus intracellularly grown bacteria at 4 h postinfection from non-activated P388D1 cells and (ii) EGD-e grown in BHI versus EGD-e grown intracellularly for 4 h in IFN-γ-activated P388D1 cells. For each condition three independent biological experiments were performed. Also cDNA samples from each time-point were hybridized as dye swap experiments. The fluorescent signal intensities from each spot on the microarray were quantified using Spotfinder software from ImaGene version 7.5 (GE Health Care).

The identification of differentially expressed genes using Significance Analysis of Microarrays (SAM) was performed as described in the manual (Tusher et al., 2001). Briefly, data analysis was conducted with intra- and inter-slide normalization (Chatterjee et al., 2006; Beissbarth et al., 2000), $\log_2$-transformed data using the one class method in SAM 2.0 for the comparison of 2 experiments: Firstly, bacteria grown in BHI versus intracellular grown bacteria for 4 h in non-activated P388D1 cells and secondly, bacteria grown in BHI versus bacteria isolated from IFN-γ-activated P388D1 at 4 h postinfection. The resulting two gene lists were then filtered for genes with a false discovery rate (FDR) of ≤1.0% and a fold change of ≥2.

Microarray data have been deposited to ArrayExpress, accession number: E-TABM-663 (http://www.ebi.ac.uk/ArrayExpress/).

Quantitative PCR analysis

Bacterial genes: Quantitative real-time PCR was performed on the ABI Prism 7700 sequence detection system (PerkinElmer). Forward and reverse primers (purchased from eurofins MWG operon) (Table S3) were designed to produce an amplicon length of about 150 bp. Total RNA was isolated as described above with an additional treatment with the DNase free kit (Ambion). Isolation was followed by production of cDNA from 3 μg RNA, SuperScript II Reverse Transcriptase (Invitrogen), and the probes were subjected to quantitative real-time PCR in a final volume of 25 μl using QuantiTect SYBR Green PCR kit (Qiagen) according to the manufacturer's instruction.

A standard curve was generated for each primer pair using different copy numbers of genomic DNA from EGD-e. For each primer pair a negative control (water), RNA sample without reverse transcriptase (to determine genomic DNA contamination) and a sample with known amount of copy numbers (to test the efficiency of the reaction) were included as controls during cDNA quantification. All samples after real-time PCR were run on a 1.5% agarose gel to verify that only a single band was produced.

The expression level of each gene was calculated by normalizing its mRNA quantity to the quantity of the mRNA of 16S rRNA (Tasara and Stephan, 2007) for the same sample using a mathematical model for relative quantification in real-time PCR published by Pfaffl (2001).

Eukaryotic genes: Quantitative real-time PCR was performed in the same ABI Prism detection system used for bacterial real-time PCR. RNA was harvested from macrophages using RNeasy Mini kit (Qiagen). To synthesize cDNA 1.0 μg of total RNA, SuperScript (SS) II reverse transcriptase (Invitrogen), random primer, and RNaseOUT ribunuclease inhibitor (Invitrogen) were used. For PCR analysis, we used QuantiTect SYBR Green PCR kit (Qiagen) to amplify cDNA, after 1:10 dilution, in a final volume of 25 μl with specific primers. The used QuantiTect *Mus musculus* primers (Nos2, Indo, Kmo, Haao, Kynu, and Arbp) were obtained from Qiagen. For each indicated gene, as well as for the reference gene ArBP (acidic ribosomal phosphoprotein), a standard curve was generated to calculate the quantity of mRNA as a function of the Ct value. The expression level of each gene was calculated by normalizing its mRNA quantity to the quantity of the ArBP mRNA for the same sample using the same mathematical model mentioned above for bacterial genes.

ROS measurement by HPF

Measurement of ROS (reactive oxygen species) formation was carried out in IFN-γ-stimulated and untreated P388D1 macrophages using a fluorescent reporter dye, 3′-(p-hydroxyphenyl) fluorescein (HPF). For measurement of ROS, P388D1 macrophages were washed, centrifuged at 1200 rpm and resuspended in phosphate-buffered saline (PBS). Macrophages were incubated with HPF (10 μM) for 30 min at room temperature in the dark. Thereafter, macrophages were washed with PBS and analyzed on the FACSCalibatur flow cytometer (Becton Dickinson). Measurement of ROS formation in *L. monocytogenes* grown intracellularly in IFN-γ-activated and non-activated P388D1 cells was carried out after lysing the infected macrophages with ice cold water. The bacterial cells were centrifuged at 8000 rpm and the pellet including the bacteria was washed with PBS and used for ROS determination by flow cytometry. ROS formation was determined using a 488 nm argon laser and a 515–545 nm emission filter (FL1) at low flow rate. Data was analyzed using Cell-Quest Software (Becton Dickinson). At least 10,000 cells were collected for each sample.

Results and discussion

IFN-γ stimulation of P388D1 macrophages

Upon bacterial infection, macrophages of the host are activated when IFN-γ is secreted by cytokine-induced T lymphocytes and natural killer cells (NK). To mimic this host response, we analyzed two macrophage cell lines (P388D1 and J774) for basal expression of iNOS (inducible nitric oxide synthase), a marker for IFN-γ-mediated macrophage activation (Bogdan et al., 2000). Our experiments revealed that IFN-γ stimulation of the murine macrophage cell line P388D1 induced higher levels of iNOS mRNA than J774 cells (data not shown). We also found that J774 is more sensitive to killing by *L. monocytogenes* EGD-e infection. Based on these results we chose to use P388D1 cells for our further studies. The amount of IFN-γ necessary for optimal activation of P388D1 macrophages (measured by levels of iNOS mRNA) was assessed by performing a dose-dependent quantitative real-time PCR experiment. As shown in Fig. 1 the optimal dose of IFN-γ for activation of P388D1 cells was observed with 100 U/ml.

In addition to iNOS, IFN-γ-mediated oxidative stress exhibited by macrophages by targeting subunits of phagocyte NADPH oxidase, was detected using the fluorescent reporter dye 3′-(p-hydroxyphenyl) fluorescein (HPF). It has been shown that HPF is oxidized by ROS and NOS with high specificity (Setsukinai et al., 2003). Also superoxides generated due to oxidative damage of Fe–S

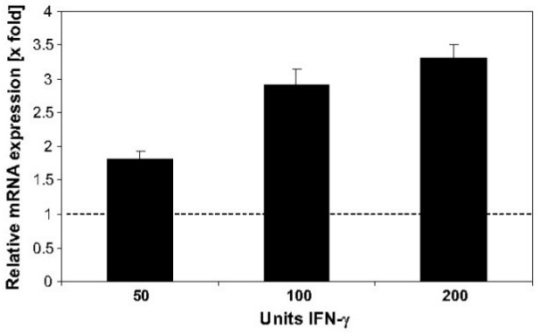


Fig. 1. Relative expression of iNOS (inducible nitric oxide synthase) in P388D1 macrophages after 18 h activation with different concentrations of IFN-γ (50, 100, and 200 U). The mRNA levels of iNOS were quantified by real-time RT-PCR and normalized to the quantity of the mRNA of ArBP (acidic ribosomal phosphoprotein) of the same sample. Dotted line denotes an expression ratio of 1 = no change.

clusters and iron leaching play a critical role in the generation of hydroxyl radicals by the Fenton's reaction (Imlay et al., 1988). Thus, using flow cytometry we compared untreated macrophage cells with IFN-γ-activated macrophages for production of ROS and NOS. A significant increase of oxygen and nitrogen radicals was exhibited by IFN-γ-activated macrophages, as we found elevated levels of HFP fluorescence that are generated by oxidization of HPF in comparison to untreated cells (Fig. 2). These findings were confirmed by addition of catalase (100 μg/ml) and L-NMMA (100 μM) which resulted in strong reduction of ROS and NOS levels. The effect of catalase which decreases oxidative stress by degradation of hydrogen peroxide to water and oxygen (Makino et al., 2005) was slightly stronger than that of L-NMMA. L-NMMA reduces the intracellular nitrosative stress by inhibiting NO generation from arginine. These data together with induction of iNOS indicate that P388D1 cells undergo IFN-γ dependent activation.

Intracellular expression profiling of Listeria monocytogenes in IFN-γ-activated macrophage cells

To study the whole-genome transcription profile of *L. monocytogenes*, total RNA was prepared from *L. monocytogenes* grown

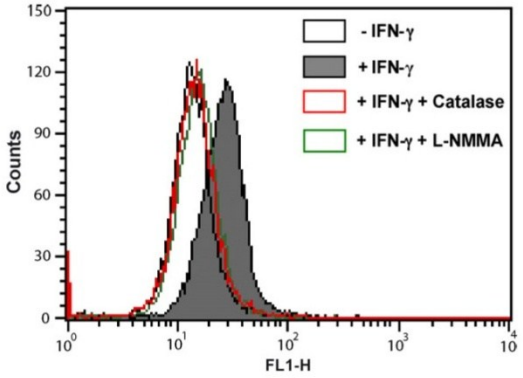


Fig. 2. Detection of reactive oxygen and nitrogen species (ROS and NOS) in non-activated and IFN-γ-activated P338D1 murine macrophages either without or following treatment with catalase (100 μg) and L-NMMA (100 μM). Measurement of ROS and NOS was carried out using a specific fluorescent dye 3′-(p-hydroxyphenyl) fluorescein (HPF). The intensity of fluorescence was determined by flow cytometry. Data are obtained from one experiment representative of three.

Appendix

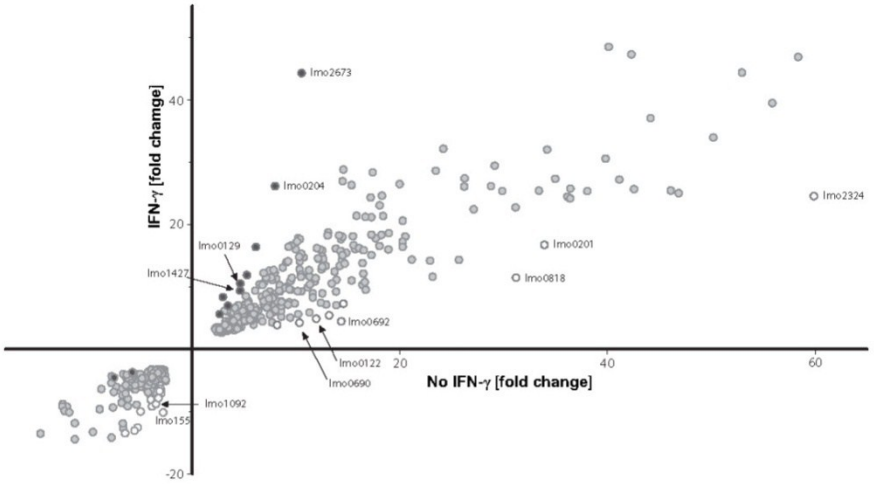


Fig. 3. Scatter diagram displaying genes whose intracellular expression changed ≥2-fold. Additionally, genes with a ratio (fc + IFN-γ/fc – IFN-γ) ≥2 were emphasized. Filled black circles indicate up regulated genes while empty circles represent down regulated genes with ratio ≥2-fold. The filled grey circles in the middle of the figure represent the rest of the genes with a ratio <2. Some representative genes with ratios >2 are named.

intracellularly for 4 h. A two-color microarray hybridization method was employed to compare the expression profile of bacteria cultivated in BHI broth at exponential phase with that of bacteria grown in either IFN-γ-activated or non-activated macrophages (4 h postinfection). Each data set represents the average result of three independent biological replicates.

Comparative transcriptomic analysis of EGD-e grown in BHI broth versus EGD-e isolated 4 h post infection (p.i.) from IFN-γ-activated P388D1 produced a total of 524 genes that were differentially regulated with a fold change ≥2 (Table S1). This accounts for ~18% of the whole genome representing the genes involved in intracellular survival of *L. monocytogenes* within macrophages. Among the genes 363 were induced and 161 were repressed. Fig. 3 displays a scatter diagram of all genes differentially regulated intracellularly with a fold change ≥2. Additionally the diagram also displays those genes with a difference in their ratio (fc + IFN-γ/fc – IFN-γ) of ≥2.

To reveal the effect of macrophage activation by IFN-γ on the intracellular transcriptome of the bacterium, we compared the expression profiles of bacteria residing in non-activated macrophages to those grown in IFN-γ-stimulated macrophages 4 h postinfection. Of the 524 identified genes in activated macrophages only 21 genes were found to exhibit an expression profile that was significantly regulated in response to IFN-γ-stimulation while the remaining 503 genes showed no difference between the activation states of the macrophages (Table 1). Out of these 21 IFN-γ-specifically regulated genes, six were induced and 15 were repressed.

Table 1
IFN-γ-specific intracellularly regulated genes of *L. monocytogenes* 4 h post infection.

Gene	fc	Gene name	Gene annotation
Specific up regulated genes			
lmo0655	3.11		similar to phosphoprotein phosphatases
lmo0807	2.95		similar to spermidine/putrescine ABC transporter, ATP-binding protein
lmo0990	4.41		conserved hypothetical protein
lmo1485	4.86		similar to unknown proteins
lmo1633	3.92	*trpE*	highly similar to anthranilate synthase alpha subunit
lmo1702	5.76	*fosX*	similar to glutathione transferase – fosfomycin resistance protein
Specific down regulated genes			
lmo0042	−4.7		similar to *E. coli* DedA protein
lmo0111	−4.8		lmo0111 protein
lmo0478	−4.7		putative secreted protein
lmo0490	−6.5	*aroE*	similar to shikimate 5-dehydrogenase
lmo0532	−3.7		lmo0532 protein
lmo0578	−8.5		putative conserved membrane protein
lmo0618	−5.1		similar to protein kinase
lmo0823	−4.6		similar to oxydoreductases
lmo0856	−3.5	*murF*	UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diamino pimelate-D-alanyl-D-alanyl ligase
lmo1201	−5.5		similar uroporphyrinogen-III methyltransferase/uroporphyrinogen-III synthase
lmo1363	−4.6		similar to geranyltranstransferase
lmo1669	−4.7		some similarity to hypothetical proteins
lmo1925	−4.1	*hisC*	similar to histidinol-phosphate aminotransferase and tyrosine/phenylalanine aminotransferase
lmo2164	−3.5	*lapB*	similar to transcriptional regulator (AraC/XylS family)
lmo2340	−3.8		similar to *Erwinia chrysanthemi* IndA protein

Appendix

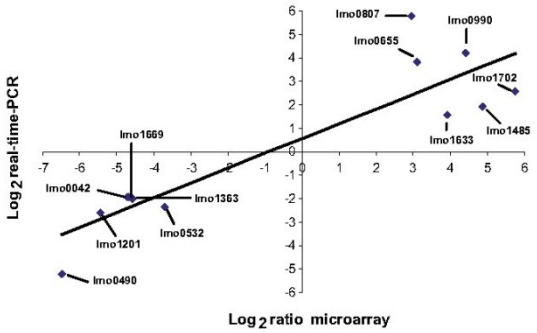


Fig. 4. Validation of microarray data with quantitative real-time PCR analysis. The fold changes of transcripts of 12 IFN-γ-specific regulated genes 4 h postinfection were log transformed, and the values were plotted against each other to evaluate their correlation. Each point represents an individual gene that was analyzed by microarray and real-time PCR respectively.

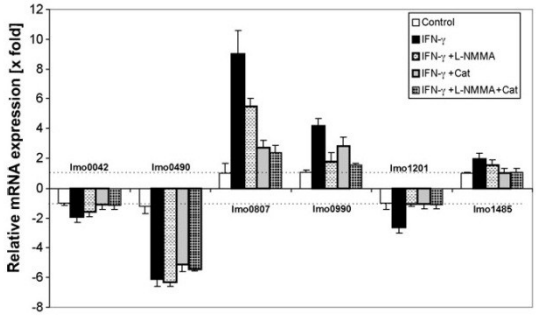


Fig. 6. Effect of L-NMMA (100 μM) and/or catalase (100 μg/ml) on the expression of IFN-γ specifically regulated *L. monocytogenes* genes predicted to be involved in stress response. The mRNA levels were quantified by real-time PCR. Quantified mRNA was normalized to the quantity of the 16S rRNA of the same sample. Dotted lines denote an expression ratio of 1 (no change). For the calculation of statistical significance we compared the values (IFN-γ + L-NMMA, IFN-γ + Cat and IFN-γ + L-NMMA + Cat) to the values obtained after IFN-γ activation. Values represent means ± SD of 5 different experiments ($P < 0.05$; paired Student's *t*-test).

Confirmation of microarray results using real-time quantitative PCR

Real-time PCR analysis of IFN-γ-specific regulated genes was performed to validate the microarray profiling data. The genes analyzed in this study include six up-regulated genes (*lmo0655*, *lmo0807*, *lmo0990*, *lmo1485*, *lmo1633*, and *lmo1702*) and six down-regulated genes (*lmo0042*, *lmo0490*, *lmo0532*, *lmo1201*, *lmo1363*, and *lmo1669*). Fig. 4 represents the results of real-time PCR which strongly correlates with microarray-derived expression profile data ($r = 0.80$). The primers used for these experiments are listed in Table S3.

Differentially regulated genes of L. monocytogenes grown in IFN-γ-activated macrophages

Evaluation of the 21 IFN-γ-specific differentially regulated genes for their presence or absence of a homolog in non-pathogenic *Listeria* strains such as *Listeria innocua* CLIP11262 (Glaser et al., 2001) and *Listeria welshimeri* (Hain et al., 2006), revealed that all of the genes were present in both strains. These data suggest that the genes might be involved in environmental niche-specific response and contribute to intracellular survival of the pathogenic *Listeria monocytogenes* under various conditions.

The 21 IFN-γ-specific regulated genes were distinct and were absent in the list of previously identified genes contributing to intracellular replication of *L. monocytogenes* grown in non-activated P388D1 macrophages (Chatterjee et al., 2006). Additionally, when we compared our data to another study performed with epithelial Caco-2 cells, only three of 21 specific regulated genes (lmo0578, lmo0618, and lmo1485) were observed (Joseph et al., 2006).

Genes involved in oxidative stress adaptation

We examined the function of the genes involved in the adaption to macrophage activation using bioinformatical analysis. Previous studies showed, that upon *in vitro* stimulation of macrophages

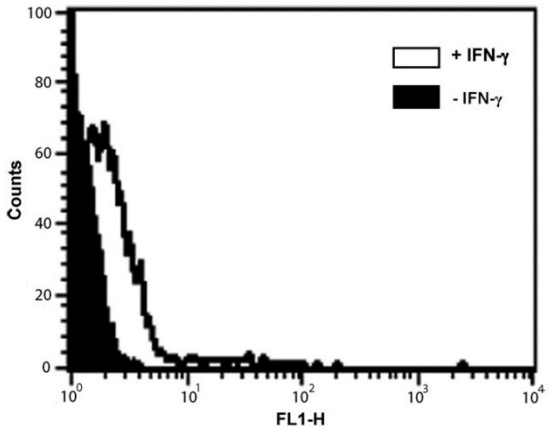


Fig. 5. Detection of ROS formation in *L. monocytogenes* EGD-e isolated 4 h postinfection from IFN-γ-activated and non-activated P338D1 murine macrophages (control). ROS concentration was measured using 3′-(p-hydroxyphenyl) fluorescein (HPF). Fluorescence intensity was determined by flow cytometry. Data are obtained from one experiment representative of three.

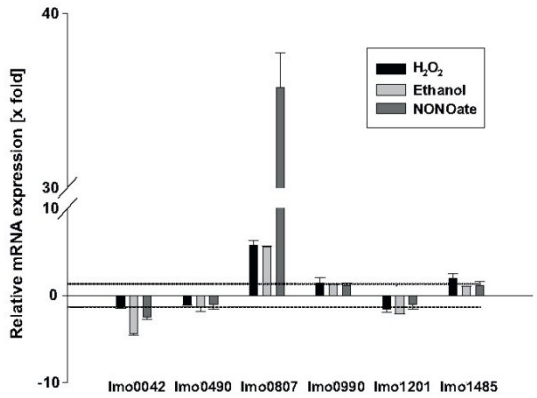


Fig. 7. Effect of H_2O_2, ethanol and NONOate on the expression of specific IFNγ-regulated genes of *L. monocytogenes* predicted to be involved in stress response was assessed. Bacterial cells were incubated for 2 h with H_2O_2 (0.05%), ethanol (5.0%) and NONOate (250 μM) separately during exponential growth in BHI broth. Values represent means of relative mRNA expression ± SD for 3 different experiments ($P < 0.05$; paired Student's *t*-test).

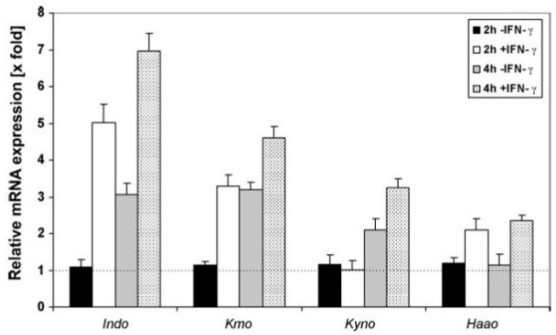


Fig. 8. Analysis of relative expression of the kynurenine pathway in non-activated and IFN-γ-activated murine macrophages P388D1 at 2 h and 4 h postinfection with *L. monocytogenes* EGD-e. Total RNA was isolated from infected macrophages and the mRNA levels of genes involved in the kynurenine pathway were quantified by real-time PCR. Quantified mRNA was normalized to the quantity of the ArBP (acidic ribosomal phosphoprotein) mRNA of the same sample. A dotted line denotes an expression ratio of 1 (no change). Data shown here are representative of three independent biological experiments. Values represent means ± SD of 4 different experiments ($P < 0.05$; paired Student's *t*-test).

with macrophage-activating factors such as IFN-γ the production of reactive oxygen intermediates is enhanced (Nathan et al., 1983; Martin and Edwards, 1994), nitric oxide is induced (Nathan and Hibbs, 1991) and phagosome fusion with lysosomes is facilitated (Xu et al., 1994). Interestingly, we found that a number of genes play a crucial role in overcoming the oxidative stress response during the intracellular life cycle of bacteria, including three up-regulated genes (*lmo0807*, *lmo0990*, and *lmo1485*) and three down-regulated genes (*lmo0042*, *lmo1201*, and *lmo0490*). The orf *lmo0807* is homologous to the ABC transporter complex *pot*ABCD involved in spermidine/putrescine import. A study with polyamines such as spermidine and putrescine revealed their participation in many cellular processes including modulation of gene expression, signal transduction, protein synthesis and regulation of cell growth (Igarashi and Kashiwagi, 2000). In addition, the abnormal growth of a polyamine-deficient *Escherichia coli* mutant has been attributed to oxidative stress-induced damage resulting in the requirement of the mutant strain for antioxidant or specific nutritional amino acids during growth (Jung et al., 2003). Furthermore, it has been documented that polyamines are the key factor in mediating resistance to nitrosative stress in uropathogenic *E. coli* (Bower and Mulvey, 2006). Putrescine and spermidine have been demonstrated to be involved in SOS induction and protection against H_2O_2 in *E. coli* (Oh and Kim, 1999a, b). Thus, induction of the *lmo0807* gene is probably required to overcome the oxidative stress bacteria are exposed to during stimulation of macrophages with IFN-γ.

Another important gene that was highly induced was *lmo0990* encoding a gene product that has similarity to a Na^+-driven multidrug efflux pump (NorM). This transporter system has been described to work as a drug/sodium antiport mediating resistance to a wide range of cationic dyes, fluroquinolones, aminoglycosides and other structurally diverse drugs (Morita et al., 2000; Huda et al., 2001). More recently, several other listerial multidrug resistance transporters (MDRs) were demonstrated to control the magnitude of a host cytosolic surveillance pathway, leading to the production of several cytokines, including type I IFN (Crimmins et al., 2008).

The third IFN-γ-specific up regulated gene *lmo1485*, encoded a putative protein with predicted SAM-methyltransferase function. Methyltransferases play a role in DNA methylation, protein-protein signaling, and biosynthesis of cellular components (Kolosova et al., 2001; Rando, 1996; Urnov, 2002). The involvement of SAM-methyltransferases with oxidative stress has been shown previously with the disruption of a gene, *lmo0581*, encoding a putative SAM-methyltransferase in *L. monocytogenes* EGD-e. The mutation resulted in extreme sensitivity to ethanol which is a potent inducer of oxidative stress (Rea et al., 2004).

Among the IFN-γ-specific down-regulated genes is *lmo0042* whose gene product shares homologies to the *E. coli* DedA integral membrane protein. Phenotypic characterization of the *ded*A mutants and selenite consumption experiments in *E. coli* strongly suggest that DedA is involved in the uptake of selenite, which has been demonstrated to directly promote oxidative stress through a reaction with cellular reduced thiol-containing compounds such as glutathione (Kramer and Ames, 1988; Ledgham et al., 2005). Another IFN-γ specific down-regulated gene is *lmo1201* encoding a protein with similar function to the uroporphyrinogen-III methyl-transferase/uroporphyrinogen-III synthase. This enzyme is believed to be involved in heme biosynthesis in bacteria. In prokaryotes, cytochromes are the most abundant heme-containing proteins and act as electron transfer proteins during the reduction of oxygen in aerobic respiration, which is one of the main sources of oxidative stress in bacterial cells. The intracellularly decreased expression of cytochrome d genes (*lmo2715–2718*) supports this finding (Table S1). An additional IFN-γ specifically down-regulated gene is *lmo0490* (AroE) encoding 5-enolpyruvylshikimate-3-phosphate synthase, an enzyme involved in aromatic amino acid biosynthesis which catalyzes the conversion of shikimate to 5-dehydroshikimate. The down-regulation of *lmo0490* gene is supported by findings of Stritzker et al. (2004) showing that an *aro*E mutant of *L. monocytogenes* is deficient in oxidative respiration due to lack of menaquinone.

In the same context we observed the down-regulation of *lmo0532*, a gene encoding a methylase involved in ubiquinone and menaquinone synthesis. Based on the functions of the IFN-γ-regulated genes discussed above, it is clear that overcoming oxidative stress is crucial for adaptation of the bacteria to survive within IFN-γ-activated macrophages.

To ascertain whether the differential regulation of genes after stimulation with IFN-γ was due to oxidative stress we isolated intracellular growing bacteria from non-activated macrophages and IFN-γ-stimulated macrophages and measured bacterial ROS levels using HPF and flow cytometry. The results revealed a significant increase of luminescence caused by higher ROS levels in bacteria grown in activated macrophages in comparison to those isolated from non-activated cells (Fig. 5).

Impact of reduced oxidative- and nitrosative stress conditions on intracellular expression levels

The differential expression of the six genes (*lmo0042*, *lmo0490*, *lmo0807*, *lmo0990*, *lmo1201*, and *lmo1485*), presumed to be involved in oxidative stress adaptation in IFN-γ-stimulated macrophages was further investigated. The mRNA levels of these genes in *L. monocytogenes* EGD-e were evaluated at 4 h postinfection in activated macrophages by quantitative real-time RT-PCR under the four following conditions: (i) IFN-γ activation; (ii) IFN-γ activation plus L-NMMA; (iii) IFN-γ activation plus catalase; (iv) IFN-γ activation plus L-NMMA and catalase. Analysis of the relative expression ratios of the six IFN-γ-specific regulated genes revealed that they are IFN-γ dependent as their mRNA levels were reduced ($P < 0.05$) in the presence of catalase and/or L-NMMA, with the exception of lmo0490 after addition of L-NMMA (Fig. 6).

Additionally, we addressed the relevance of these genes under oxidative and nitrosative stress conditions. Bacterial cells were incubated for 2 h with 0.05% H_2O_2, 5% ethanol or 250 μM spermine NONOate during exponential growth in BHI broth. The most significant results were obtained for the ABC transporter

Appendix

(*lmo0807*), which showed increased expression with all three reagents (Fig. 7). The strongest effect was achieved with NONOate. *Lmo0807* seems to play an important role in mediating resistance to stress in general and especially to nitrosative stress in *L. monocytogenes*. To our knowledge, this is so far the first evidence indicating the role of *lmo0807* in intracellular resistance of *L. monocytogenes* determined by IFN-γ activation. The genes *lmo0490*, *lmo0990*, *lmo1201* and *lmo1485* did not indicate significant changes in their expression levels. Ethanol treatment led to increased down regulation of *lmo0042* whose gene product shares homologies to the *E. coli* DedA integral membrane protein. These results confirm the impact of IFN-γ activation on both oxidative and nitrosative stress leading to expression of bacterial genes that help intracellular bacteria to overcome stress induced by activation of macrophages.

Enhanced tryptophan degradation by IFN-γ

We noted the specific induction of *lmo1633*, a gene encoding the anthranilate synthase (*trp*E) which is located in the operon responsible for tryptophan synthesis in *L. monocytogenes*. Generally, IFN-γ stimulation induces indoleamin-2.3-dioxygenase (IDO) (Bogdan et al., 2000), an enzyme that catalyzes the conversion of tryptophan to N-formylkynurenine which is the first and rate-limiting enzyme in the kynurenine pathway. IDO causes the depletion of tryptophan and thereby mediates starvation of intracellular bacteria (Popov et al., 2006). In order to investigate the impact of IFN-γ activation on tryptophan catabolism in the host cell, we analyzed the expression of the kynurenine pathway enzymes Indo (IDO), Kmo (kynurenine 3-monooxygenase), Kynu (kynureninase), and Haao (3-hydroxyanthranilate 3,4-dioxygenase) (Fig. S1) Quantitative real-time PCR was performed with isolated mRNA from IFN-γ-activated and non-activated P388D1 macrophage cells infected with *L. monocytogenes* for 2 h and 4 h. All four tested genes showed elevated mRNA levels during infection (Fig. 8). In non-activated cells the expression of the tested genes was increased at 4 h postinfection with the exception of the gene for Haao. However, the up-regulation of genes involved in the kynurenine pathway was even more pronounced in activated macrophages and supports data from other studies suggesting that tryptophan depletion may be a host cell defense response resulting in metabolic restriction of intracellular bacterial growth (Popov et al., 2006).

To assess the expression of *trp*E upon addition of an excess amount of tryptophan to activated cells during infection, we used tryptophan-free RPMI medium infection experiments with macrophages. Three different tryptophan concentrations were tested: at 5 μg/ml which is equal to the tryptophan concentration in "normal" tryptophan-containing RPMI 1640 medium, and with higher tryptophan concentrations i.e. at 10 and 20 μg/ml, respectively. The results presented in the supplementary Fig. S2 show decreased expression levels of *trp*E upon addition of different tryptophan concentrations. These results are in accordance with a previous study that showed that Indo is one of the most significantly up regulated genes in cells following infection with *L. monocytogenes* (Popov et al., 2006). They confirmed the correlation between increased production of host Indo concentrations and tryptophan depletion in infected cells.

In addition to those listerial genes involved in oxidative stress response and tryptophan synthesis, our data revealed other specifically regulated genes that are listed in Table 1. For many of these genes their putative gene products are uncharacterized with respect to their function. Further studies would be required to determine their role in bacterial life cycle during its adaptation in activated macrophages.

Additional genes of L. monocytogenes involved in adaptation to IFN-γ-activation

In addition to IFN-γ-specifically regulated genes we also looked for genes that were highly induced or repressed in IFN-γ-activated and non-activated macrophages by ≥2-fold change (Table S2). The orf *lmo2376* which was expressed at 4-fold higher in IFN-γ-stimulated cells was found to encode a protein similar to the universal stress protein (UspA). In *E. coli*, UspA is an autophosphorylating serine and threonine phosphoprotein (Freestone et al., 1997) that has been shown to be involved in oxidative stress response (Nyström and Neidhardt, 1994; Nachin et al., 2005).

In addition, a strong up-regulation was also observed for *lmo1446* (*zur*M) and *lmo1447* (*zur*A) encoding proteins involved in zinc (Zn) uptake and *lmo1445* which encodes the transcriptional regulator (zinc uptake regulator, *zur*R). The gene *zur*R is essential for zinc-specific repression of the operon *zur*RMA. Earlier studies have implicated the zinc uptake system in protection of *Bacillus subtilis* against peroxide-induced oxidative stress (Gaballa and Helmann, 2002). Furthermore, inhibition of bacterial DNA replication of *Salmonella enterica* during nitrosative stress was accompanied by zinc mobilization suggesting that DNA-binding zinc metalloproteins are critical targets of NO-related antimicrobial activity (Schapiro et al., 2003). We analyzed the expression levels of *lmo1445*, *lmo1446*, and *lmo1447* in the presence of NO-inhibitor L-NMMA to test if the increased expression of proteins of the zinc uptake system correlates with heightened NO concentrations in activated macrophages. We found no changes in the gene expression of *lmo1445*, *lmo1446*, and *lmo1447* despite reduced NO amounts measured by L-NMMA (data not shown) which suggests that elevated expression of zinc uptake proteins is not connected to NO formation.

Another well known virulence gene of *L. monocytogenes* that was highly up-regulated (3.3-fold) upon activation of macrophages is *act*A, encoding a surface protein required for recruitment of host cytoskeletal proteins (Kocks et al., 1992; Pistor et al., 1994). Results from recent studies have shown that elevated *act*A expression levels are required for bacterial evasion of autophagy in the cytosol at 4 h postinfection (Yoshikawa et al., 2009). Thus, an increase of *act*A expression in IFN-γ-activated macrophages would be expected to increase decoration of the bacterial cell surface with host cytoskeleton proteins thereby preventing it from being targeted for destruction by the host cell.

Finally, we note that the 503 commonly regulated genes of *L. monocytogenes* irrespective of activation of macrophages with IFN-γ represent genes that enable *L. monocytogenes* to survive and replicate within the eukaryotic host cells (Table S1; see also Chatterjee et al., 2006; Joseph et al., 2006).

Conclusion

This study aimed to reveal the transcriptional consequences of IFN-γ stimulation on intracellular growth of *L. monocytogenes*. The major effects triggered by IFN-γ were increased oxidative and nitrosative stress levels in both P388D1 cells as well as in *L. monocytogenes* replicating within the activated macrophages. Out of 21 IFN-γ-specifically regulated genes five (*lmo0042*, *lmo0807*, *lmo0990*, *lmo1201*, and *lmo1485*) were implicated in resistance against oxidative- and nitrosative stress. Regulation by IFN-γ was confirmed by quantitative real-time RT-PCR as these genes showed reduced mRNA levels in presence of catalase and/or L-NMMA. Additionally, we observed increased ROS levels in bacteria isolated from activated macrophages in comparison to those derived from non-activated cells. Here we show that increased expression of the putative listerial *trp*E gene is associated with concomitant catabolic

degradation of tryptophan by host cell enzymes including IDO. Thus our data provide specific insight to bacterial defenses invoked to cope with the host cell stress responses induced in IFN-γ-activated macrophages.

Acknowledgments

We thank Alexandra Amend for excellent technical assistance. This work was supported by funds from the Deutsche Forschungsgemeinschaft (DFG) via SFB535 TP 14, the Federal Ministry of Education and Research (BMBF) via ERA-NET PathoGenoMics 0313939A to T.H. and T.C. and ERA-NET PathoGenoMics to the sncRNAomics project (62080061) to T.H.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ijmm.2011.05.001.

References

Beissbarth, T., Fellenberg, K., Brors, B., Arribas-Prat, R., Boer, J., Hauser, N.C., Scheideler, M., Hoheisel, J.D., Schütz, G., Poustka, A., Vingron, M., 2000. Processing and quality control of DNA array hybridization data. Bioinformatics 16, 1014–1022.

Bogdan, C., Röllinghoff, M., Diefenbach, A., 2000. The role of nitric oxide in innate immunity. Immunol. Rev. 173, 17–26.

Bower, J.M., Mulvey, M.A., 2006. Polyamine-mediated resistance of uropathogenic *Escherichia coli* to nitrosative stress. J. Bacteriol. 188, 928–933.

Chatterjee, S.S., Hossain, H., Otten, S., Kuenne, C., Kuchmina, K., Machata, S., Domann, E., Chakraborty, T., Hain, T., 2006. Intracellular gene expression profile of *Listeria monocytogenes*. Infect. Immun. 74, 1323–1338.

Crimmins, G.T., Herskovits, A.A., Rehder, K., Sivick, K.E., Lauer, P., Dubensky Jr., T.W., Portnoy, D.A., 2008. *Listeria monocytogenes* multidrug resistance transporters activate a cytosolic surveillance pathway of innate immunity. Proc. Natl. Acad. Sci. U. S. A. 105, 10191–10196.

Decker, T., Stockinger, S., Karaghiosoff, M., Müller, M., Kovarik, P., 2002. IFNs and STATs in innate immunity to microorganisms. J. Clin. Invest. 109, 1271–1277.

Freestone, P., Nyström, T., Trinei, M., Norris, V., 1997. The universal stress protein, UspA, of Escherichia coli is phosphorylated in response to stasis. J. Mol. Biol. 274, 318–324.

Gaballa, A., Helmann, J.D., 2002. A peroxide-induced zinc uptake system plays an important role in protection against oxidative stress in *Bacillus subtilis*. Mol. Microbiol. 45, 997–1005.

Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A., Baquero, F., Berche, P., Bloecker, H., Brandt, P., Chakraborty, T., Charbit, A., Chetouani, F., Couvé, E., de Daruvar, A., Dehoux, P., Domann, E., Domínguez-Bernal, G., Duchaud, E., Durant, L., Dussurget, O., Entian, K.D., Fsihi, H., García-del Portillo, F., Garrido, P., Gautier, L., Goebel, W., Gómez-López, N., Hain, T., Hauf, J., Jackson, D., Jones, L.M., Kaerst, U., Kreft, J., Kuhn, M., Kunst, F., Kurapkat, G., Madueno, E., Maitournam, A., Vicente, J.M., Ng, E., Nedjari, H., Nordsiek, G., Novella, S., de Pablos, B., Pérez-Diaz, J.C., Purcell, R., Remmel, B., Rose, M., Schlueter, T., Simoes, N., Tierrez, A., Vázquez-Boland, J.A., Voss, H., Wehland, J., Cossart, P., 2001. Comparative genomics of Listeria species. Science 294, 849–852.

Hain, T., Steinweg, C., Kuenne, C.T., Billion, A., Ghai, R., Chatterjee, S.S., Domann, E., Kärst, U., Goesmann, A., Bekel, T., Bartels, D., Kaiser, O., Meyer, F., Pühler, A., Weisshaar, B., Wehland, J., Liang, C., Dandekar, T., Lampidis, R., Kreft, J., Goebel, W., Chakraborty, T., 2006. Whole-genome sequence of *Listeria welshimeri* reveals common steps in genome reduction with *Listeria innocua* as compared to Listeria monocytogenes. J. Bacteriol. 188, 7405–7415.

Havell, E.A., 1987. Production of tumor necrosis factor during murine listeriosis. J. Immunol. 139, 4225–42231.

Herskovits, A.A., Auerbuch, V., Portnoy, D.A., 2007. Bacterial ligands generated in a phagosome are targets of the cytosolic innate immune system. PLoS Pathog. 3, e51.

Hsieh, C.S., Macatonia, S.E., Tripp, C.S., Wolf, S.F., O'Garra, A., Murphy, K.M., 1993. Development of TH1 CD4+ T cells through IL-12 produced by Listeria-induced macrophages. Science 260, 547–549.

Huda, M.N., Morita, Y., Kuroda, T., Mizushima, T., Tsuchiya, T., 2001. Na^+-driven multidrug efflux pump VcmA from *Vibrio cholerae* non-O1, a non-halophilic bacterium. FEMS Microbiol. Lett. 203, 235–239.

Igarashi, K., Kashiwagi, K., 2000. Polyamines: mysterious modulators of cellular functions. Biochem. Biophys. Res. Commun. 271, 559–564.

Imlay, J.A., Chin, S.M., Linn, S., 1988. Toxic DNA damage by hydrogen peroxide through the Fenton reaction in vivo and in vitro. Science 240, 640–642.

Joseph, B., Przybilla, K., Stühler, C., Schauer, K., Slaghuis, J., Fuchs, T.M., Goebel, W., 2006. Identification of *Listeria monocytogenes* genes contributing to intracellular replication by expression profiling and mutant screening. J. Bacteriol. 188, 556–568.

Jung, I.L., Oh, T.J., Kim, I.G., 2003. Abnormal growth of polyamine-deficient *Escherichia coli* mutant is partially caused by oxidative stress-induced damage. Arch. Biochem. Biophys. 418, 125–132.

Kocks, C., Gouin, E., Tabouret, M., Berche, P., Ohayon, H., Cossart, P., 1992. *L. monocytogenes*-induced actin assembly requires the actA gene product, a surface protein. Cell 68, 521–531.

Kolosova, N., Sherman, D., Karlson, D., Dudareva, N., 2001. Cellular and subcellular localization of S-adenosyl-L-methionine:benzoic acid carboxyl methyltransferase, the enzyme responsible for biosynthesis of the volatile ester methylbenzoate in snapdragon flowers. Plant Physiol. 126, 956–964.

Kramer, G.F., Ames, B.N., 1988. Mechanisms of mutagenicity and toxicity of sodium selenite (Na2SeO3) in *Salmonella typhimurium*. Mutat. Res. 201, 169–180.

Langermans, J.A., van der Hulst, M.E., Nibbering, P.H., van der Meide, P.H., van Furth, R., 1992. Intravenous injection of interferon-gamma inhibits the proliferation of Listeria monocytogenes in the liver but not in the spleen and peritoneal cavity. Immunolog. 77, 354–361.

Ledgham, F., Quest, B., Vallaeys, T., Mergeay, M., Covès, J., 2005. A probable link between the DedA protein and resistance to selenite. Res. Microbiol. 156, 367–374.

Makino, M., Kawai, M., Kawamura, I., Fujita, M., Gejo, F., Mitsuyama, M., 2005. Involvement of reactive oxygen intermediate in the enhanced expression of virulence-associated genes of *Listeria monocytogenes* inside activated macrophages. Microbiol. Immunol. 49, 805–811.

Martin, J.H., Edwards, S.W., 1994. Interferon-gamma enhances monocyte cytotoxicity via enhanced reactive oxygen intermediate production. Absence of an effect on macrophage cytotoxicity is due to failure to enhance reactive nitrogen intermediate production. Immunolog. 81, 592–597.

Morita, Y., Kataoka, A., Shiota, S., Mizushima, T., Tsuchiya, T., 2000. NorM of Vibrio parahaemolyticus is an Na(+)-driven multidrug efflux pump. J. Bacteriol. 182, 6694–6697.

Nachin, L., Nannmark, U., Nyström, T., 2005. Differential roles of the universal stress proteins of *Escherichia coli* in oxidative stress resistance, adhesion, and motility. J. Bacteriol. 187, 6265–6272.

Nathan, C.F., Murray, H.W., Wiebe, M.E., Rubin, B.Y., 1983. Identification of interferon-gamma as the lymphokine that activates human macrophage oxidative metabolism and antimicrobial activity. J. Exp. Med. 158, 670–689.

Nathan, C.F., Hibbs Jr., J.B., 1991. Role of nitric oxide synthesis in macrophage antimicrobial activity. Curr. Opin. Immunol. 3, 65–70.

Nyström, T., Neidhardt, F.C., 1994. Expression and role of the universal stress protein, UspA, of Escherichia coli during growth arrest. Mol. Microbiol. 11, 537–544.

Oh, T.J., Kim, I.G., 1999a. The SOS induction of umu′-′lacZ fusion gene by oxidative damage is influenced by polyamines in *Escherichia coli*. Cell Biol. Toxicol. 15, 291–297.

Oh, T.J., Kim, I.G., 1999b. The expression of Escherichia coli SOS genes *recA* and *uvrA* is inducible by polyamines. Biochem. Biophys. Res. Commun. 264, 584–589.

Pamer, E.G., 2004. Immune responses to *Listeria monocytogenes*. Nat. Rev. Immunol. 4, 812–823.

Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 29, e45.

Pistor, S., Chakraborty, T., Niebuhr, K., Domann, E., Wehland, J., 1994. The ActA protein of *Listeria monocytogenes* acts as a nucleator inducing reorganization of the actin cytoskeleton. EMBO J. 13, 758–763.

Popov, A., Abdullah, Z., Wickenhauser, C., Saric, T., Driesen, J., Hanisch, F.G., Domann, E., Raven, E.L., Dehus, O., Hermann, C., Eggle, D., Debey, S., Chakraborty, T., Krönke, M., Utermöhlen, O., Schultze, J.L., 2006. Indoleamine 2,3-dioxygenase-expressing dendritic cells form suppurative granulomas following *Listeria monocytogenes* infection. J. Clin. Invest. 116, 3160–3170.

Rando, R.R., 1996. Chemical biology of isoprenylation/methylation. Biochem. Soc. Trans. 24, 682–687.

Rea, R.B., Gahan, C.G., Hill, C., 2004. Disruption of putative regulatory loci in *Listeria monocytogenes* demonstrates a significant role for Fur and PerR in virulence. Infect. Immun. 72, 717–727.

Schapiro, J.M., Libby, S.J., Fang, F.C., 2003. Inhibition of bacterial DNA replication by zinc mobilization during nitrosative stress. Proc. Natl. Acad. Sci. U. S. A. 100, 8496–8501.

Schuchat, A., Swaminathan, B., Broome, C.V., 1991. Epidemiology of human listeriosis. Clin. Microbiol. Rev. 4, 169–183.

Setsukinai, K., Urano, Y., Kakinuma, K., Majima, H.J., Nagano, T., 2003. Development of novel fluorescence probes that can reliably detect reactive oxygen species and distinguish specific species. J. Biol. Chem. 278, 3170–3175.

Stritzker, J., Janda, J., Schoen, C., Taupp, M., Pilgrim, S., Gentschev, I., Schreier, P., Geginat, G., Goebel, W., 2004. Growth, virulence, and immunogenicity of *Listeria monocytogenes* aro mutants. Infect. Immun. 72, 5622–5629.

Swaminathan, B., Gerner-Smidt, P., 2007. The epidemiology of human listeriosis. Microbes Infect. 9, 1236–1243.

Tasara, T., Stephan, R., 2007. Evaluation of housekeeping genes in *Listeria monocytogenes* as potential internal control references for normalizing mRNA expression levels in stress adaptation models using real-time PCR. FEMS Microbiol. Lett. 269, 265–272.

Tripp, C.S., Wolf, S.F., Unanue, E.R., 1993. Interleukin 12 and tumor necrosis factor alpha are costimulators of interferon gamma production by natural killer cells in severe combined immunodeficiency mice with listeriosis, and interleukin 10 is a physiologic antagonist. Proc. Natl. Acad. Sci. U. S. A. 90, 3725–3729.

Tusher, V.G., Tibshirani, R., Chu, G., 2001. Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. U. S. A. 98, 5116–5121.

13.4 Appendix publication 3

Nucleic Acids Research Advance Access published January 29, 2011

Nucleic Acids Research, 2011, 1–14
doi:10.1093/nar/gkr033

The intracellular sRNA transcriptome of *Listeria monocytogenes* during growth in macrophages

Mobarak A. Mraheil[1], André Billion[1], Walid Mohamed[1], Krishnendu Mukherjee[2], Carsten Kuenne[1], Jordan Pischimarov[1], Christian Krawitz[1], Julia Retey[3], Thomas Hartsch[3], Trinad Chakraborty[1,*] and Torsten Hain[1,*]

[1]Institute of Medical Microbiology, Justus-Liebig-University, Frankfurter Strasse 107, [2]Institute of Phytopathology and Applied Zoology, Justus-Liebig-University, Heinrich-Buff-Ring 26-32, 35392 Giessen and [3]Genedata Bioinformatik GmbH, Lena-Christ-Strasse 50, D-82152 Planegg-Martinsried, Germany

Received September 15, 2010; Revised January 12, 2011; Accepted January 13, 2011

Downloaded from nar.oxfordjournals.org at Universitaetsbibliothek on February 9, 2011

ABSTRACT

Small non-coding RNAs (sRNAs) are widespread effectors of post-transcriptional gene regulation in bacteria. Currently extensive information exists on the sRNAs of *Listeria monocytogenes* expressed during growth in extracellular environments. We used deep sequencing of cDNAs obtained from fractioned RNA (<500 nt) isolated from extracellularly growing bacteria and from *L. monocytogenes* infected macrophages to catalog the sRNA repertoire during intracellular bacterial growth. Here, we report on the discovery of 150 putative regulatory RNAs of which 71 have not been previously described. A total of 29 regulatory RNAs, including small non-coding antisense RNAs, are specifically expressed intracellularly. We validated highly expressed sRNAs by northern blotting and demonstrated by the construction and characterization of isogenic mutants of *rli31*, *rli33-1* and *rli50 for intracellular expressed sRNA candidates, that their expression is required for efficient growth of bacteria in macrophages. All three mutants were attenuated when assessed for growth in mouse and insect models of infection. Comparative genomic analysis revealed the presence of lineage specific sRNA candidates and the absence of sRNA loci in genomes of naturally occurring infection-attenuated bacteria, with additional loss in non-pathogenic listerial genomes. Our analyses reveal extensive sRNA expression as an important feature of bacterial regulation during intracellular growth.**

INTRODUCTION

The availability of an increasing number of complete bacterial genome sequences along with recent technical advances in DNA sequencing has led to an explosion in the identification of numerous small, non-coding RNAs (sRNAs) (1 3) and this number is constantly growing (2,4 6). A combination of both computational and novel experimental approaches have demonstrated the ubiquity of sRNAs and has led to the description of many functionally important sRNAs in organisms ranging from eubacteria to humans (7 12).

The majority of sRNA candidates identified to date are thought to regulate gene expression by hybridizing with target mRNA thus modulating its stability and/or translation activity. In addition, some sRNAs bind to proteins and modulate their activity or build functional complexes (13,14). In bacteria, the main function of sRNAs are in coordinating adaptation to environmental changes and signals by controlling target gene expression and includes, for example, responses to iron limitation, oxidative stress and low temperature (4,13 15).

The role of sRNA in controlling virulence and pathogenesis has been demonstrated for a number of Gram-negative bacteria including *Escherichia coli*, *Pseudomonas aeruginosa*, *Salmonella typhimurium*, *Vibrio cholerae*, *Chlamydia trachomatis* and *Helicobacter pylori*, as well Gram-positive bacteria such as *Staphylococcus aureus*, *Streptococcus pneumoniae* and *Clostridium perfringens* (16 18).

*To whom correspondence should be addressed. Tel: +49 641 99 46400; Fax: +49 641 99 46409; Email: torsten.hain@mikrobio.med.uni-giessen.de
Correspondence may also be addressed to Trinad Chakraborty. Tel: +49 641 99 41251; Fax: +49 641 99 41259;
Email: trinad.chakraborty@mikrobio.med.uni-giessen.de

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2011. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

An initial approach to sRNAs detection was the characterization of those RNAs which bind to Hfq, a protein originally identified as a host factor needed for Qβ bacteriophage replication in *E. coli* (19,20). In *E. coli*, Hfq modulates the activity of several sRNAs and acts as an RNA chaperone to promote sRNA mRNA duplex formation (21,22). Analysis of *hfq* mutants of *P. aeruginosa*, *V. cholerae*, *Legionella pneumophila* and *Brucella abortus* revealed attenuated growth in macrophages or mice (23 26).

In the Gram-positive, facultative intracellular pathogen bacterium *L. monocytogenes*, it was shown that Hfq contributes to pathogenesis in mice (21,27). Presently, three sRNAs have been identified by co-immunoprecipitation using Hfq (28). Additional studies have led to the description of further 12 sRNAs including the σ^B-dependent SrbA transcript (29,30). Recently a genome-based tiling array study provided a catalog of sRNA candidates expressed under different physiological growth conditions (31). These data have largely been confirmed by a study in which *L. monocytogenes* and its isogenic SigB mutant strain were grown under stationary growth conditions and the repertoires of RNA produced examined by deep RNA sequencing (32).

Listeria monocytogenes is the causative agent of listeriosis, a severe human infection with a high mortality rate. The bacterium inhabits numerous ecological niches, it can multiply at high salt concentrations (10% NaCl) and broad ranges of pH (4.5 9) and temperature (0 45°C) (33). A hallmark of this human pathogen is its ability to invade and survive inside vertebrate and invertebrate host cells, wherein the bacterium can freely multiply within the cytosol and can induce actin-based movement. Actin-based movement allows the bacterium to spread from cell-to-cell which leads to fatal outcomes of listerial infection. Prior to infection, internalin A and B induce the first step of the infection process in non-phagocytic cells by interacting with the eukaryotic host cell and promote the intracellular uptake of the pathogen after binding with the E-Cadherin and c-Met receptors in mammals. However, the main virulence genes, responsible for the intracellular life cycle of *L. monocytogenes* are clustered in a ~9 kb chromosomal region. This virulence cluster encodes the genes *viz.*, listeriolysin O (LLO) and phosphatidylinositol phospholipase A (PlcA), which are responsible for the escape of bacterium from the primary vacuole. Another phospholipase (PlcB) is required for bacterial cells to escape from the secondary vacuole and is processed with the assistance of the metalloprotease (Mpl). To facilitate actin recruitment for intracellular movement, the bacterium modulates components of the host cell machinery with the surface protein ActA, which is anchored within the cytoplasmic membrane. Finally, the regulation of the virulence gene cluster is dependent on the master regulator PrfA, a Crp/Fnr-like transcriptional regulator (34,35). In this context it is reasonable to assume that sRNAs may play a versatile role in the adaptation mechanisms of *L. monocytogenes* to these different environments.

Currently, there is extensive information available regarding the transcriptome of *L. monocytogenes* when grown under conditions that are external to the host cell (31,32). These include data for bacteria grown extracellularly in broth (exponential and stationary growth) under different conditions of stress, including low oxygen, low temperature (30°C), in blood and the lumen of the infected gut as well as analysis of several isogenic mutants such as Δ*prfA*, Δ*sigB* and Δ*hfq*. However, information as to whether specific sRNAs are expressed during intracellular growth and their respective roles are largely unknown (11).

Here we describe a genome-wide search for sRNAs expressed in the wild-type of *L. monocytogenes* during extracellular and intracellular growth. Extracellular cultures were grown until exponential phase like previous studies (31,32). Intracellular cultures were grown in P338D1 murine macrophages. RNA was collected 4 h post-infection and size-fractioned to <500 nt. cDNA generated from size-fractioned RNA was deep sequenced and provided a comprehensive view of listerial sRNA candidates preferentially induced following infection in murine macrophages. We report on the discovery of 150 putative regulatory RNAs of which 29 are specifically expressed intracellularly. Analysis of several sRNA candidates highly expressed during intracellular growth revealed that these loci are highly conserved in pathogenic *L. monocytogenes* strains. Isogenic mutants lacking these loci exhibit attenuated virulence in *in vivo* models of infection.

MATERIALS AND METHODS

Bacterial strains and growth conditions

The strain *L. monocytogenes* EGD-e (36) and its chromosomal deletion mutants (Supplementary Table S1A) were used in this study. Bacteria were grown in BHI broth (VWR) overnight at 37°C with shaking at 180 r.p.m. (Unitron, Infors) until mid-exponential phase ($OD_{600\ nm}$ 1.0). Overnight cultures were diluted 1:50 in 20 ml fresh BHI broth using a 100 ml Erlenmeyer flask and were incubated at the same conditions mentioned above until $OD_{600\ nm}$ 1.0.

RNA isolation

For RNA isolation from *L. monocytogenes* grown extracellularly in BHI broth (VWR) until mid-exponential phase, aliquots of 0.5 ml bacterial culture were treated with 1.0 ml RNA protect (Qiagen) for 5 min, the bacterial cells were collected by centrifugation for 10 min (8000*g*) and subsequently stored at −80°C until use. RNA extraction from intracellularly grown *L. monocytogenes* in macrophages 4 h post-infection was performed as described previously (37). Briefly, infected host cells were lysed using cold mix of 0.1% (w/v) sodium dodecyl sulfate, 1.0% (v/v) acidic phenol and 19% (v/v) ethanol in water. The bacterial pellets were collected by centrifugation for 3 min (16 000*g*).

Total RNA was extracted using miRNeasy kit (Qiagen) with some modifications. The collected pellets were washed with SET buffer [50 mM NaCl, 5 mM EDTA and 30 mM Tris HCl (pH 7.0)]. After centrifugation at 16000 *g* for 3 min pellets were resuspended into 0.1 ml

Downloaded from nar.oxfordjournals.org at Universitaetsbibliothek on February 9, 2011

Tris HCl (pH 6.5) containing 50 mg/ml lysozyme (Sigma), 25 U of mutanolysin (Sigma), 40 U of SUPERase (Ambion), 0.2 mg of proteinase K (Ambion) and incubated at 37°C for 30 min at 350 r.p.m. QIAzol (Qiagen) was added, mixed gently and incubated for 3 min at room temperature. An additional incubation at room temperature was done after adding 0.2 volume chloroform followed by centrifugation at 16 000*g* at 4°C for 15 min. The upper aqueous phase, containing RNA, was transferred to a new collection tube and 1.5 v of 100% ethanol was added and mixed thoroughly. The probes containing RNA and ethanol were transferred into columns supplied with the miRNeasy Kit (Qiagen) and treated according to the manual including an on-column DNase digestion (RNase-Free DNase, Qiagen). RNA was eluted by RNase-free water and stored at −80°C until needed. The quantity of the isolated total RNA was determined by absorbance at 260 and 280 nm, and the quality was assessed using Nano-chips for Agilent's 2100 Bioanalyzer. For detection and estimation of the small RNA fraction within the isolated total RNA, a small RNA-chip (Agilent) was used, which visualizes RNAs with sizes ranging from 20 to 150 nt.

RNA sequencing and data analysis

The fraction of the total RNA <500 bp was used for RNA-Seq as described below. The solubilized RNAs were first treated with tobacco acid pyrophosphatase (Epicentre) as recommended by the manufacturer. This treatment allows discriminating primary 5′-ends generated by transcription initiation from 5′-ends generated by RNA processing. Then the small RNAs were poly(A)-tailed using poly(A) polymerase followed by ligation of a RNA adapter to the 5′-phosphate of the RNA. First-strand cDNA synthesis was performed using an oligo(dT)-adapter primer and M-MLV reverse transcriptase (Promega) assay. The resulting cDNAs were PCR-amplified in 20 cycles to a concentration of 20 ng/µl using a high fidelity DNA polymerase. For size fractionation PCR-amplified cDNA was run on a preparative 6% PAA-gel. PCR products containing small RNA sequences of 20 500 nt were isolated from PAA-gel. The cDNAs were purified using the Macherey and Nagel NucleoSpin Extract II kit. 454 pyrosequencing using GS FLX Titanium series chemistry (Roche) were carried out by Eurofins MWG Operon (Germany).

A sequence file was created for each condition, respectively. In total 189 381 reads with approximately 31 million bases were analyzed. After clipping of 5′-linker and poly(A)-tail all reads shorter than 16 nt were removed. Resulting reads were further processed via a hitherto unpublished software *sncRAS* (A. Billion, manuscript in preparation) which in general requires no clipping, which makes no significant difference during the mapping step, except for short reads combined with a poly(A)-tail which may lead to a deletion of this sequence. The remaining reads were mapped with NCBI BLASTN 2.2.17 (38) against *L. monocytogenes* EGD-e genome (GenBank Accession: NC_003210) with an e-value of 0.001, default word size and rewards for a nucleotide match had been set to 2. Additionally, nucleotide identity was required to be >60% combined with coverage of 80% between query and subject sequence. Sequencing reads which did not fulfil these requirements were not taken into account. Reads ranged in size from 21 to 521 nt and averaged in length in both conditions around 74 nt. 28 49% of used reads matched perfectly to the genome. An additional 4 9% of remaining reads contain one mismatch.

sRNA detection

sRNA detection was carried out by the software *sncRAS* as well. The first step removes transcripts of rRNA and tRNA, plus all reads from intergenic regions (IGR) of rRNA and tRNA genes. The second step identifies locations for potential sRNAs. Regulatory RNAs were grouped in three general classes: sRNAs, antisense small non-coding RNAs (asRNAs) and riboswitches including *cis*-regulatory RNAs (Supplementary Table S3). Valid reads, which mapped within an IGR and fulfilled the cut-off criteria, were used to identify a pile-up of sequences which combined have a minimal length of 50 nt and do not overlap with adjacent genes. Furthermore putative sRNA candidates were required to have ≥10 reads in one condition. For asRNA detection similar criteria were applied, however a 5′- or 3′-overlap of its possible target gene was allowed. The resulting candidates from automatic classification were re-evaluated by visual inspection with the Integrative Genomics Viewer (IGV) version 1.4.2 (http://www.broadinstitute.org/igv). Detection of *cis*-regulatory RNAs was performed with manually evaluated candidates by searching against the Rfam 9.1 database (39). All hits with a covariance model score of more than 50 were classified as *cis*-regulatory RNAs including riboswitches.

Transcription factor binding site analysis

In order to identify putative regulatory motifs within regulatory RNAs candidate promoter regions, a list of known consensus binding boxes was compiled from the literature (Supplementary Table S4). For each putative regulatory sRNA, a region of 50 nt upstream was scanned for transcription factor binding sites. Most binding sites were detected using regular expression search (FUZZNUC, EMBOSS package) allowing one mismatch. Sigma70 promoters were identified using HMMER3 (40) on six different HMMs from the prokaryotic promoter prediction site (http://bioinformatics.biol.rug.nl/websoftware/ppp/).

Terminator identification

Terminators were identified using a pre-calculated prediction of TransTermHP2.0 (41). All terminator-like hairpins with confidence score ≥30 were taken into account. Sliding windows of ±10 nt around the 3′-end of each putative regulatory RNA were used to annotate putative terminators which additionally required at least one base overlap with this defined area. If several candidates within the search window were found, the one located closest to the 3′-end was used.

Downloaded from nar.oxfordjournals.org at Universitaetsbibliothek on February 9, 2011

Small open reading frame detection

Small open reading frame (ORF) detection for *L. monocytogenes* (36) was based on predictions from GenDB (42) as described previously for *L. welshimeri* (43) and *L. seeligeri* (44) applying a lowered ORF detection cut off (>10 amino acids).

Data visualization

For visualization BLAST results were pre-processed by SAMtools package version 0.1.7 (45) and uploaded into the IGV version 1.4.2 (http://www.broadinstitute.org/igv) for further manual inspection. Circular chromosomes were precalculated and visualized by GenomeViz 1.3 (46). Venn diagrams were drawn by VENNY (http://bioinfogp.cnb.csic.es/tools/venny/index.html).

Northern blot analysis

For northern blotting, 15 µg of total RNA was separated on 10% polyacrylamide gels containing 7 M urea. RNA was transfered to Biodyne nylon transfer membrane (VWR) via electroblotting and UV cross linked. Oligodeoxynucleotides (Supplementary Table S1B) were end-labeled with γ^{32}-P-ATP (Hartmann Analytik) using T4 polynucleotide kinase as recommended by the manufactured protocol (Fermentas). Pre-hybridization was performed for 3 h in DEPC-treated water containing 0.1 mg/ml salmon sperm DNA, 6 × SSC, 0.5% SDS and 2.5 × Denhardt's solution. Hybridization was performed overnight after adding γ^{32}-P-ATP end-labeled oligodeoxynucleotide probes to northern membranes under the same conditions used for pre-hybridization. The hybridization temperature for oligodeoxynucleotides ranged between 50 and 65°C. Membranes were exposed on phosphor imaging screens (GE Health Care) and analyzed with Typhoon 9200 (GE Health Care). To evaluate the sizes of sRNAs we constructed a ladder containing different PCR fragments between 150 and 550 nt (Supplementary Table S1B).

sRNA mutant construction

To investigate the role of sRNA in intracellular survival *in vitro* and *in vivo*, chromosomal deletion mutants were generated. Primer sequences used to generate the isogenic mutants are presented in Supplementary Table S1B. Chromosomal deletion mutants were constructed by generating the 5′ (with primers '1-for' and '2-rev') and the 3′ (with primers '3-for' and '4-rev') flanking regions of the sRNA concerned. Construction of the fusion construct was achieved by SOEing as described earlier (47). Generation of the deletion mutants was performed as described previously (48). The deletions in the targeted genes were confirmed by PCR with according primer '5-for' and '6-rev' after verifying the sequence by automated DNA sequencing.

The complementation of the Δ*rli31*, Δ*rli33-1* and Δ*rli50** deletion mutants was carried out using the *L. monocytogenes* site specific phage integration vector pPL2 (49). Primer sequences used for the complementation of the deleted genes are listed in Supplementary Table S1B.

In vitro infection experiment

Infection experiments with *L. monocytogenes* EGD-e and its sRNA deletion mutants were performed using P388D1 murine macrophage as described previously (37).

Galleria mellonella infection model

Bacterial inoculums of wild-type *L. monocytogenes* EGD-e and its sRNA deletion mutants were injected dorsolaterally into the hemocoel of last instar larvae using 1 ml disposable syringes and 0.4 × 20 mm needles mounted on a microapplicator as described previously (50). After injection, larvae were incubated at 37°C. Larvae were considered dead when they showed no movement in response to touch. No mortality of *Galleria* larvae were recorded when injected with 0.9% NaCl. For each experiment 20 animals per sRNA deletion mutant were used and three biological replicates were performed independently.

Murine infection model

Six- to 8-week-old female BALB/c mice, purchased from Harlan Winkelmann (Borchen, Germany), were used in all experiments. *In vivo* growth kinetic and survival of wild-type *L. monocytogenes* EGD-e and its sRNA mutants were tested in a mouse infection model. Infection was performed by intravenous injection of approximately 2000 viable bacteria in a volume of 0.2 ml of PBS. After 3 days, bacterial growth in spleens and livers was determined by plating 10-fold serial dilutions of organ homogenates on BHI. The detection limit of this procedure was 10^2 CFU per organ. Colonies were counted after 24 h of incubation at 37°C.

Ethics statement

This study was carried out in strict accordance with the regulation of the National Protection Animal Act (§7-9a Tierschutzgesetz). The protocol was approved by the local Committee on the Ethics of Animal Experiments (Regierungsbezirk Mittelhessen) and permission was given by the local authority (Regierungspraesidium Giessen, Permit Number: GI 15/5-Nr.63/2007).

Statistical data analysis of infection experiments

All infection experiments were performed a minimum of three times. Significant differences between two values were compared with a paired Student's *t*-test. Values were considered significantly different when $P < 0.05$.

Comparative genomics

In order to assess the dissemination of regulatory RNAs across the genus *Listeria* the GenBank formatted files of four pathogenic (*L. monocytogenes* 1/2a EGD-e, *L. monocytogenes* 4b F2365, *L. monocytogenes* 1/2a 08-5578, *L. monocytogenes* 1/2a 08-5923, *L. monocytogenes* 4a HCC23) and three apathogenic strains (*L. innocua* 6a Clip11262, *L. welshimeri* 6b

Downloaded from nar.oxfordjournals.org at Universitaetsbibliothek on February 9, 2011

SLCC5334, *L. seeligeri* 1/2b SLCC3954) were retrieved from the GenBank repository (http://www.ncbi.nlm.nih.gov/genbank/index.html). These genome data as well as all candidate regulatory RNAs were then introduced to the sRNAdb database (J. Pischimarov, manuscript in preparation) which employs BLASTN to find sequence similarity. Using a cutoff of 60% nucleotide identity and 80% coverage homologs of all candidate regulatory sRNAs were identified in the aforementioned genome sequences.

RESULTS

RNA sequencing

To investigate the intracellular sRNA transcriptome profile of *L. monocytogenes* EGD-e, total RNA was isolated from bacteria grown extracellularly in BHI or from the cytosol 4 h post-infection in P388D1 murine macrophages. Following size-fractionation, cDNA with a size <500 nt was used for the cDNA sequencing studies. Samples were subjected to 454-based pyro-sequencing and base-called reads from two cDNA libraries generated from total RNA of either extracellularly (EC) or intracellularly (IC) grown *L. monocytogenes* were analysed in this study. These libraries generated a total 189 381 reads of which 116 158 (61%) were derived from RNA isolated from intracellular bacteria and 73 223 (39%) from bacteria grown in broth cultures. An overview of the filtered and mapped sequencing reads is given in Supplementary Table S2.

Sequencing reads were mapped to the genome of *L. monocytogenes* using BLASTN with an e-value of 0.001 and default word size with rewards for a nucleotide match that had been set to two. Additionally nucleotide identity was required to be >60% combined with coverage of 80% between query and subject sequence. Reads that did not fulfil these requirements were removed from the dataset. After clipping, additional linker removal, quality control and mapping against the genome of *L. monocytogenes* EGD-e, 114 459 unique reads with at least 80% sequence identity with 80% coverage and a minimum length of 21 nt remained for detailed analysis. We observed that ~49% of the IC reads perfectly match the genome with 100% sequence identity compared with only 28% of the reads from the EC cDNA library. The 'intergenome' of *L. monocytogenes*, i.e. the non-coding sequence between annotated ORFs, comprises ~10% (~300 000 bp) of the entire genome. In the intergenome fraction we observed expression of nearly one-third under intracellular condition which additionally shows the importance of the intergenome (Figure 1A). Approximately 60% of all sequence reads mapped to annotated rRNA and tRNA genes (Figure 1B and C).

Identification of sRNAs, asRNAs and *cis*-regulatory RNAs including riboswitches

Bioinformatics analysis using previously described methods ('Materials and Methods' section) for analysing RNA-Seq data for sRNAs in *L. monocytogenes* EGD-e identified 150 putative regulatory RNAs expressed under either or both growth conditions (Figure 1A and Supplementary Table S3). The putative candidates were placed into three different classes: class I comprises sRNAs, which are located within IGRs without overlapping adjacent genes; class II includes antisense RNAs, referred as asRNAs, located antisense to an annotated ORFs and class III comprises *cis*-regulatory RNAs including riboswitches.

A total of 121 regulatory RNA elements were mapped within IGRs of the *L. monocytogenes* chromosome and included 88 sRNA and 33 *cis*-regulatory RNAs including riboswitches. In addition, we detected 29 putative asRNAs (Figure 1A and Supplementary Table S3). Of the 150 candidates determined, 142 were intracellularly expressed and 121 were transcribed extracellularly (Figure 1A and Supplementary Table S3). A common set of 113 regulatory RNAs are detected in both conditions. Of the 142 candidates expressed intracellularly ~20% (29 candidates) were specific for intracellular growth and 35% (50 candidates) showed enhanced expression which was evident through an increased cDNA read number (IC/EC ratio >2) under intracellular growth conditions. In contrast ~7% (8) of the 121 extracellularly expressed transcripts were specific for extracellular growth, while 49% (60) showed enhanced expression which was evident through an increased cDNA read number (EC/IC ratio >2) under extracellular growth conditions. The sizes of the putative regulatory RNAs ranged from 50 to 517 nt with a minimum of 10 reads and a maximum read count of 1603 in at least one growth condition.

Of the 88 class I sRNAs, 85 showed intracellular, and 77 extracellular, expression. Eleven sRNAs are specific for intracellular growth whereas only three sRNAs are specific to the extracellular lifecycle. The overall numbers of reads covering the 11 intracellular specific sRNAs was in general at least 25% higher than that of the 113 common regulatory RNAs which were found in both conditions.

The class II asRNA family revealed an even higher variation in expression than the sRNAs. Of a total of 29 asRNAs detected, approximately twice that number i.e. 25 asRNAs were expressed intracellularly as compared with only 11 extracellularly. Of these 18 were specifically expressed intracellularly as compared with only 4 for the extracellular growth.

The third class of sRNAs showed no major differences in occurrence between the growth conditions. We identified 33 *cis*-regulatory RNAs in total (32 IC versus 33 EC) with only one candidate being specifically expressed extracellularly. The latter one was identified as a lysine riboswitch and is located upstream of a gene encoding an amino acid permease. However it was remarkable that the geometric mean of extracellularly expressed *cis*-regulatory RNA is at least twice as high as the one of those expressed under intracellular growth conditions.

Identification of eight major transcription factor binding sites (Supplementary Table S4) in the candidate pool suggests regulation of 24% of all candidates. A putative housekeeping promoter was found for 36 of 142 intracellular and 39 of 121 extracellular transcripts.

Downloaded from nar.oxfordjournals.org at Universitaetsbibliothek on February 9, 2011

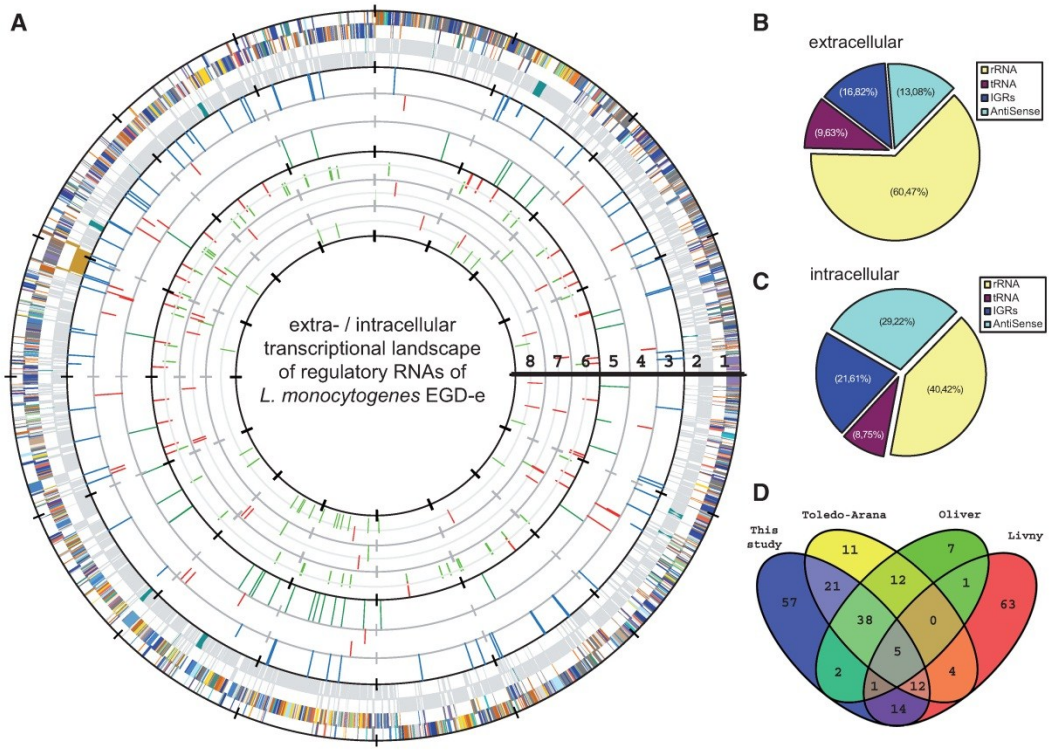


Figure 1. Discovery of the intracellular sRNome of *L. monocytogenes* using RNA-Seq. (**A**) Extracellular and intracellular transcriptional landscape of *L. monocytogenes* is represented using GenomeViz (46). Circles display following information from outside to inside: (1) COG categories; (2) rRNAs and tRNAs (blue), a prophage-like locus (light brown) and the virulence gene cluster (red); (3) intracellular regulatory RNAs (outer circle) and extracellular regulatory RNAs (inner circle); (4) intracellular asRNAs (outer circle) and extracellular asRNAs (inner circle); (5) intracellular *cis*-regulatory RNAs including riboswitches (outer circle) and extracellular *cis*-regulatory RNAs including riboswitches (inner circle); (6) regulation of intracellular sRNAs; (7) regulation of intracellular asRNAs and (8) intracellular *cis*-regulatory RNAs including riboswitches; (**B** and **C**) Distribution of mapped sequence reads used for extracellular and intracellular transcriptome analysis; (**D**) Comparative analysis of sRNA transcriptome data using 'cumulative' values which can be summarized since sRNA candidates would not be counted multiple times (see Supplementary Figure S1 for a 'non-cumulative' version). Comparison of our RNA-seq results, whole genome tiling array from Toledo-Arana and coworkers (31), RNA-seq data of *L. monocytogenes* 10403S (32) and *in silico* regulatory RNA predictions (1).

The vast majority of the putative promoters were accounted for by the sigma70 promoter which was identified preceding 28 intracellular and 34 extracellular regulatory RNAs, respectively. For six intracellular and four extracellular candidates a SigB box was detected in the upstream region suggesting an involvement in the general stress response. The PrfA regulator which is involved in virulence (51,52), putatively binds to the upstream region of two intracellular and one extracellular candidates. We used a pre-computed rho-independent transcription terminator prediction from TransTermHP (41) to identify transcriptional units. For 55% of the sRNAs we were able to define a putative terminator (Supplementary Table S3).

Chromosomal distribution of regulatory RNA

We used GenomeViz (46) to visualize the location and distribution of the regulatory sRNAs on the *L. monocytogenes* EGD-e genome. We found that they are neither located within specific chromosomal clusters nor do they exhibit a specific strand prevalence (Figure 1A).

Comparative analysis with whole genome tiling array data, RNA-seq and *in silico* predictions

We compared 103 putative sRNAs previously published by Toledo-Arana and colleagues to our 150 identified regulatory RNA candidates (53). The comparison revealed 75% of previously published sRNA candidates by Toledo-Arana (31) to be overlapping with the candidates of the current study (Figure 1D and Supplementary Table S3). Of 26 candidates identified by Toledo-Arana but not identified in our study no expression was observed for 12 candidates, 12 regulatory RNAs failed to reach predefined cutoffs and two were either expressed on the opposite strand or were found in the same IGR but without any overlap.

Downloaded from nar.oxfordjournals.org at Universitaetsbibliothek on February 9, 2011

Comparison of our data to that of Toledo-Arana (31) from bacteria grown in BHI, revealed an overlap of 68% of all sRNAs. A comparison of sRNAs expressed when grown in blood by Toledo-Arana to the intracellular sRNA repertoire detected in our study showed considerable overlap of detected regulatory RNAs (72%). In some cases such as for sRNA candidate *rli33*, which is 534 nt in size and represents the longest sRNA in the published work (31), our analysis yielded two different sRNA candidates. In order to avoid confusion we designate them *rli33-1* and *rli33-2*.

In addition we compared Illumina deep sequencing results previously reported by Oliver *et al.* (32) and detected an overlap of 69% with our work (Figure 1D and Supplementary Table S3). Computational analysis using the SIPHT workflow has previously predicted 100 regulatory RNAs for *L. monocytogenes* EGD-e (1). Of these, only 33% could be recovered by our experiment (Figure 1D and Supplementary Table S3) whereas the remaining 77 predicted candidates show no expression or failed to attain the desired cutoffs.

For 14 regulatory RNA candidates we predicted small ORFs (Supplementary Table S3) applying a gene calling cutoff more than 10 amino acids, which may not be represented in the current version of the *L. monocytogenes* genome annotation (36).

We reanalyzed the *rliB* locus which is involved in virulence (31) using the RNA-seq data described herein. This locus harbors five small copies of expressed CRISPR repeats (29 nt long) and spacers (36 nt long) sequences (crRNA) predicted by the CRISPR prediction software PILER-CR (54) and could be confirmed by data present in this study and which was originally noticed by Mandin *et al.* (29). However, no homologous of CRISPR-associated (CAS) genes were detected in the flanking regions. The predicted CRISPR spacer and repeats have a range of between 24–47 bp and 26–72 bp, respectively, and are currently the smallest predicted sRNAs for *L. monocytogenes*.

Verification of selected sRNA candidates

To verify the transcription of sRNA candidates at the locations indicated by deep sequencing, we used northern blot analysis. Four sRNA candidates namely *rli31*, *rli33-1*, *rli50* and *rli112* were selected for verification according to their high read numbers revealed by sequencing results.

The results of the northern blotting, shown in Figure 2, demonstrate signals corresponding to small transcripts from each of the candidates. According to the signals in the northern blot analysis, some of the sRNA candidates were differentially expressed under the tested conditions. *rli31* and *rli33-1* showed variation in transcript concentration during extra- and intra-cellular growth. Both sRNA candidates exhibited significantly higher transcript numbers during intracellular growth. The northern blot results correlate with read-numbers from transcriptome deep sequencing (Supplementary Table S3). The determined size of *rli33-1* by northern blot analysis of ~140 nt confirms the read length (144 nt) assessed by

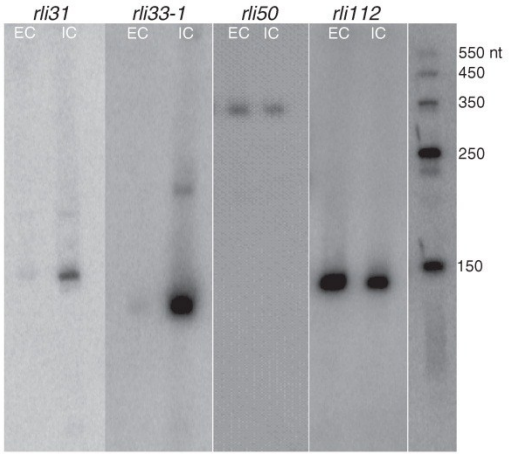


Figure 2. Northern blots of sRNA candidates. Validation of RNA-seq data with northern blot analysis of extracellularly expressed sRNAs from bacteria at mid exponential growth phase in BHI (EC) compared with intracellularly expressed sRNAs at 4 h post-infection in murine macrophages (IC) of *rli31*, *rli33-1*, *rli50* and *rli112*.

RNA-seq. *rli33-1* showed two bands in the northern blot at ~200 and 130 nt in length, that probably originate from processing of the larger co-transcript. Both fragments were also detected by transcriptome sequencing excluding the possibility of cross-hybridization (Figure 2). *rli50* appears to be longer than predicted by sequencing (306 nt) and has a size of ~350 nt as determined by northern blot. We observed with an *rli112* specific probe strong northern blot signals (~140 nt) under extra- and intra-cellular conditions indicating high RNA abundance of this new unpublished putative sRNA.

Noteworthy is the observation that a homolog of *rli112, rli78,* is highly expressed under extracellular growth conditions compared with intracellular growth. To exclude effects of cross hybridization in the northern blot analysis, qRT–PCR was conducted with isogenic mutants lacking *rli78* and *rli112*, respectively, using a primer pair that binds to the homologous region in the same location. The results confirmed higher expression levels of *rli78* under extracellular growth conditions (data not shown). This result correlates with the transcript numbers determined by RNA sequencing and reflect higher cDNA reads extracellularly as indicated by sequencing data (Supplementary Table S3). Additionally, six sRNA candidates including three so far new unpublished sRNAs (*rli80*, *rli91* and *rli105*) were confirmed (Supplementary Figure S2). Generally, northern blotting confirmed the existence of these sRNA candidates and showed that sizes, except *rli91*, correlate with those predicted by RNA-seq. Quantitation based on sequencing frequently differs significantly from that observed on northern blots which is exemplified by *rli80* and *rli105* (Supplementary Figure S2). Further investigations are required to investigate the basis of these differences.

Downloaded from nar.oxfordjournals.org at Universitaetsbibliothek on February 9, 2011

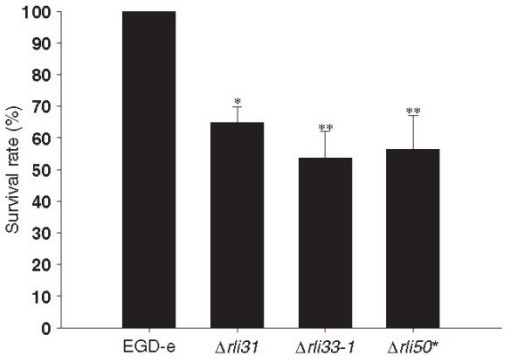


Figure 3. Survival of *L. monocytogenes* sRNA mutants in P388D1 murine macrophage cells. The macrophages were infected with the wild-type *L. monocytogenes* EGD-e and its isogenic deletion mutants, *rli31*, *rli33-1* and *rli50**, with an MOI of 10 in 24-well plates and bacterial CFU counts were measured on agar plated following lysis of the P388D1 cells after 4 h post-infection. $n = 5$; error bars indicate standard deviations ($*P \leq 0.005$, $**P \leq 0.05$).

Moreover, our study provides the first experimental evidence for the presence of two antisense RNAs, *anti2394* and *anti2095*, in *L. monocytogenes* EGD-e (Supplementary Figure S3).

Mutants lacking *rli31*, *rli33-1* and *rli50** are attenuated for infection

Due to an overlap between *rli50* (minus strand) and *rli112* (plus strand) of ~43% in the IGR between *lmo2709* and *lmo2710* we decided to delete the distal end of *rli112* to create the mutant Δ*rli50** (Δ*rli50** refers to the deletion mutant and *rli50* to the sRNA). We generated isogenic deletion mutants of the sRNA candidates *rli31*, *rli33-1* and *rli50** with the highest intracellular transcription levels (Supplementary Figures S4 and S5). Additionally, these sRNA candidates had higher intracellular cDNA read numbers in comparison to extracellular growth conditions. A schematic overview of Δ*rli50** and Δ*rli33-1* and is presented in Supplementary Figure S4.

We assessed the mutant strains' capability to grow in P388D1 murine macrophages together with the wild-type strain. Whereas the mutants Δ*rli31*, Δ*rli33-1* and Δ*rli50** were significantly impaired in their abilities to proliferate intracellularly in macrophages as compared with the wild-type strain (Figure 3), no differences were observable in their *in vitro* growth in BHI (Supplementary Figure S6).

We investigated the survival of larvae from *Galleria* following injection with different sRNA mutants of *L. monocytogenes*. Bacterial cultures grown to exponential phase were injected dorsolaterally into the hemocoel at 10^6 CFU/larva. We observed significant attenuation in the mortality rates of larvae when injected with Δ*rli31*, Δ*rli33-1* and Δ*rli50** in comparison to wild-type EGD-e (Figure 4A C).

We used the mouse infection model to assess the virulence properties of the sRNA mutants Δ*rli31*, Δ*rli33-1*

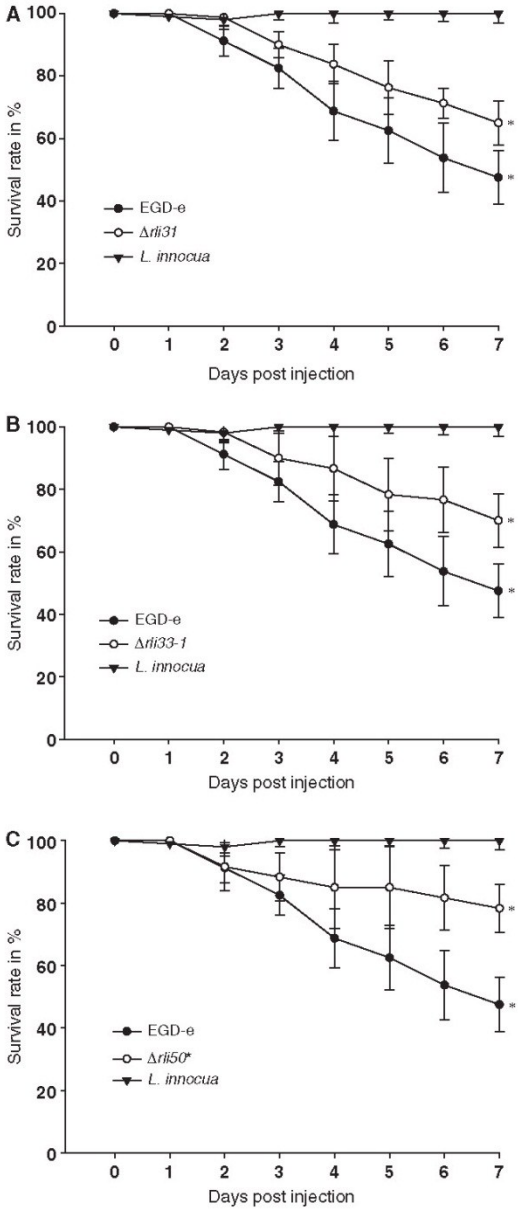


Figure 4. Survival of *Galleria mellonella* larvae after inoculation with different *L. monocytogenes* sRNA mutants and *L. innocua*. Time course of survival of the larvae varies with the type of sRNA mutants employed for inoculation. Inoculation with 10^6 CFU/larvae EGD-e resulted in significantly higher killing rate of larvae in comparison to (**A**) *rli31*, (**B**) *rli33-1* and (**C**) *rli50**. The non-pathogenic *L. innocua* showed no mortality. Values represent means of at least three independent experiments ± standard deviations for 20 larvae per treatment ($*P \leq 0.005$).

Downloaded from nar.oxfordjournals.org at Universitaetsbibliothek on February 9, 2011

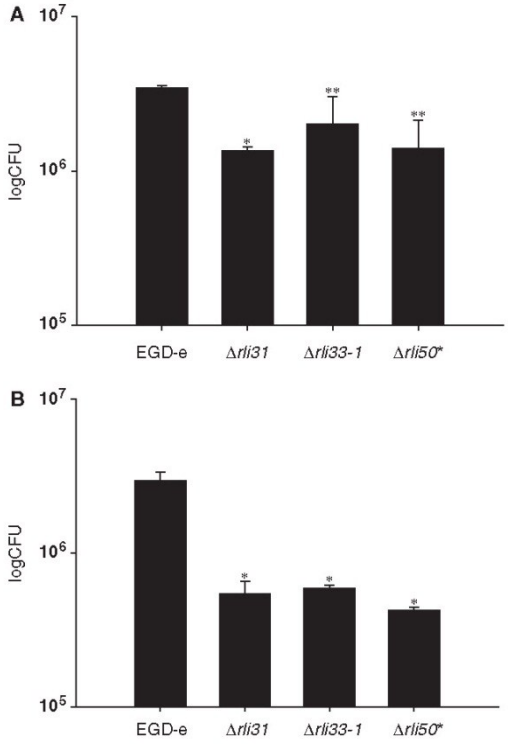


Figure 5. Mice infection studies with sRNA deletion mutants and of *L. monocytogenes*. Bacterial load in mice organs were also determined following *in vitro* infection with 2000 CFU of *L. monocytogenes* EGD-e wild-type strain as well as its isogenic sRNA mutants *rli31*, *rli33-1* and *rli50**. On day 3 after infection, the numbers of viable bacteria in spleens (**A**) and livers (**B**) of three animals per group were determined of wild-type EGD-e versus *rli31*, *rli33-1* and *rli50** in spleen and liver, respectively ($n = 4$). Error bars indicate standard deviations ($^*P \leq 0.005$, $^{**}P < 0.05$).

and Δ*rli50** to that of wild-type *L. monocytogenes* EGD-e. For all three mutants, survival and/or growth in the spleen, and in particular in the liver, were significantly reduced at day 3 post-infection when compared with wild-type (Figure 5A and B).

To exclude that the phenotypes of the deletion mutants were due to polar effects on flanking genes, we generated complemented strains of the Δ*rli31*, Δ*rli33-1* and Δ*rli50** sRNA deletion mutants and examined their abilities to proliferate intracellularly in P388D1 murine macrophage cells. The complementation largely restored the intracellular growth impairment in macrophages (Supplementary Figure S7, see also Figure 3). Additionally, we checked the expression of the flanking genes by quantitative real time PCR under both conditions used in the experiment (extracellular versus intracellular). The results presented in Supplementary Figure S8 show clearly that the deletion of the putative sRNA candidates has no consequences on the expression of the flanking genes.

Comparative analysis of putative regulatory RNAs among members of the genus *Listeria*

Finally, we compared our RNA sequencing identified regulatory RNA to the genomes of four human pathogenic strains of *L. monocytogenes* (36,53,54), an attenuated *L. monocytogenes* 4a strain (HCC23) (56) and the three apathogenic species of *L. innocua*, *L. welshimeri* and *L. seeligeri* (36,43,44) to investigate the regulation of virulence. Here we detected that the number of regulatory RNAs among the *L. monocytogenes* serotypes 4b and 1/2a belonging to Lineage I and II were highly similar compared with *L. monocytogenes* 1/2a EGD-e as reference (II > I > III), but decreased rapidly, when listerial species were more distantly related (Figure 6).

DISCUSSION

In this study, we identified that a total of 150 putative regulatory RNAs were expressed during growth of *L. monocytogenes* in extracellular and intracellular environments. The putative regulatory RNAs can be divided into three groups including 88 putative sRNAs, 29 asRNAs and 33 *cis*-regulatory elements including riboswitches which are expressed during growth of *L. monocytogenes* in extracellular and intracellular environments.

Previously, Toledo-Arana used genome-wide tiling arrays and reported a total 103 sRNAs expressed by *L. monocytogenes* growing under a wide range of conditions from broth culture to blood and the intestinal lumen (31). In this study we detected 79 of the regulatory RNAs previously described suggesting good correlation of the data, given the inherent differences in comparing sequencing-based technologies with that of hybridization analyses. Nevertheless, even though we have only compared two growth conditions using deep sequencing of cDNA derived from size-fractionated RNA (<500 nt), our data revealed 71 new candidates, none of which have been previously described. Our findings suggest that diversity of regulation at the post-transcriptional level is an important component of adaptation to niche specific growth.

For functional analysis we focused on the intracellularly up-regulated sRNA candidates Δ*rli31*, Δ*rli33-1* and Δ*rli50**. Isogenic deletion mutants of these sRNA loci resulted in reproducibly reduced growth properties following infection of P388D1 murine macrophages and in virulence attenuation in both insect and mice models of infection. Because *G. mellonella* is a model for innate immunity responses, it is likely that *rli31*, *rli33-1* and *rli50** represent sRNA loci required for adaptation to intracellular growth in vertebrate and invertebrate hosts.

The gene *lmo0559* located downstream of *rli31* encodes a putative transporter of the CorA superfamily involved in magnesium and cobalt uptake. Intracellular transcriptomic analysis of *L. monocytogenes* has previously indicated that the host cytosol is a rich source of ions as suggested by the down-regulation of several ion transport systems (37). An involvement of the 5′-UTR in the regulation of *mgtA*, a magnesium transporter, has

Downloaded from nar.oxfordjournals.org at Universitaetsbibliothek on February 9, 2011

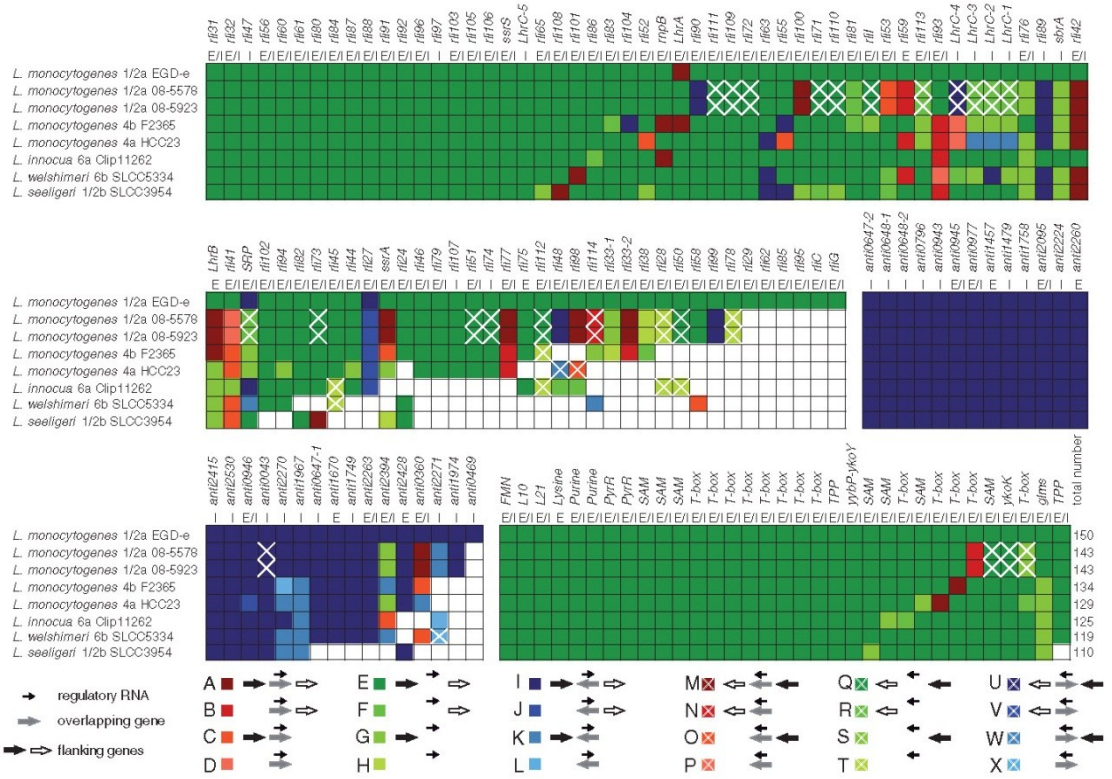


Figure 6. Comparative overview of known and putative regulatory RNAs of *L. monocytogenes* EGD-e. EGD-e was compared with 3 *L. monocytogenes* serotypes (3 × 1/2a, 1 × 4b and 1 × 4a) and three non-pathogenic *Listeria* species (*L. innocua*, *L. welshimeri* and *L. seeligeri*). To determine the distribution of regulatory RNA inside the genus a BLAST analysis was conducted using sRNAdb (unpublished software). Candidates were considered present inside a strain in the case of a sequence identity of 60% and a coverage of 80%. Since the surrounding locus is often important for the function of the regulatory RNA, information about the conservation of adjacent genes was included using the same cutoff. Possible cases for direction, presence and absence of each regulatory RNA and its flanking genes was color-coded below. A white square indicates the absence of the regulatory RNA. As a reference for this analysis the relevant loci of *L. monocytogenes* EGD-e were chosen. The small black arrow depicted in the legend indicates the regulatory RNA, while larger arrows in black and white symbolize the left and right flanking gene, respectively. A large gray arrow denotes a gene overlapping the regulatory RNA in sense or antisense direction. The arrow direction is not representative for the strand but for the relation to the locus in the reference genome of *L. monocytogenes* EGD-e. (E) indicates extracellular and (I) intracellular expression of the regulatory RNA.

been reported previously in *Salmonella enterica* (57), where high Mg^{2+} concentration induced the formation of a stem loop structure leading to the transcriptional termination of *mgtA*.

Interestingly, a putative sRNA located in the intergenic region between *lmo0671* and *lmo0672* was previously reported as *rli33* encoding a transcript of 534 nt in length (31). However, our RNA-seq revealed two smaller fragments of 186 nt (*rli33-1*) and 274 nt (*rli33-2*) in this region. In support of this observation, northern blot analyses confirmed the presence of *rli33-1* and *rli33-2* suggesting either the presence of an internal start site within the larger transcript or an unknown RNA processing mechanism of the transcript under the experimental conditions used. *rli33* is highly induced in the stationary growth phase and in blood indicating a potential role in virulence (31). This has been shown by *in vitro* and *in vivo* experiments with the isogenic deletion mutant *rli33-1* which completely removes the 5′-end of *rli33*.

The *rli50** deletion mutant showed the strongest effect in the insect model and in the liver of mice. Interestingly, our RNA-seq data revealed two sRNA candidates in the chromosomal region between *lmo2709* and *lmo2710*, a new one on the sense (*rli112*) and the other on the antisense strand (*rli50*), which was detected by Toledo-Arana and coworkers (31). As mentioned above the isogenic mutant for *rli50**, removes the distal end of the *rli112* sRNA candidate resulting in a double mutant that may be responsible for the impaired virulence effects obtained. A homolog of *rli112* with 94% similarity is located within the intergenic region between the genes *lmo0470* and *lmo0471*, designated as sRNA candidate *rli78*, which has higher extracellular expression than intracellularly. However, this sRNA is clearly not able to compensate

Downloaded from nar.oxfordjournals.org at Universitaetsbibliothek on February 9, 2011

for the loss of sRNA *rli50* in the isogenic mutant of *rli50**. Northern blot analysis of *rli29* and *rli78* confirmed the presence of the sense and antisense strand transcripts (Supplementary Figure S2). This indicates different regulatory roles of these homologous sRNAs in their environmental niche.

Similar examples for the existence of several copies of sRNA were reported for *V. cholerae*, which harbours four redundant copies of the regulatory RNA Qrr (Qrr1-4) interacting with two feedback loops and which promotes gene dosage compensation among Qrr1-4 for proper control of quorum-sensing processes (58,59). Also, multiple copies of other listerial sRNAs are currently known for the Hfq-dependent LhrC (28) with five copies in total and present at two different chromosomal locations. We find that LhrC is up-regulated in our intracellular data, but the role of these intracellularly induced sRNAs has still to be elucidated. Surprisingly, the chromosomal locus of *lmo0459 lmo0479*, including *rli28*, *rli29* and *rli78*, revealed a lower GC-content suggesting horizontal gene transfer events. An inspection of the surrounding loci indicated a gene encoding a transposase of the IS3 family (*lmo0464*) in this region. Thus, we speculate that horizontal gene transfer (HGT) caused by transposition might be involved in chromosomal spreading of listerial regulatory RNAs.

Our data also indicate differential regulation of sRNA candidates within the A118 prophage-like regions (*rliG*, *rli48*, *rli62*, *rli98* and *rli99*). Bacteriophages play an important role for horizontal gene transfer to shape their microbial host genomes with new genetic functions including sRNA, such as has been described for *ipeX* which is responsible for OmpC porin regulation in *E. coli* (60). Transcriptional activation of phage-related genes has been previously reported from different groups by *in vivo* transcriptomic studies (37,61,62). Further investigation is required to understand the contribution of these phage-related sRNA transcripts to the infection process of *L. monocytogenes*.

In addition, CRISPR systems, responsible for microbial phage defense, are elements that have been described in a large number of prokaryotic genomes (63). We have confirmed as originally noticed by Mandin *et al.* (29) that *rliB* is related to CRISPR elements, but CRISPR-associated CAS genes were absent suggesting that these transcripts are produced by a novel endonucleolytic mechanism.

An emerging class of regulatory RNAs are antisense RNAs which have been observed in different bacteria and archaea (17,64 67). This class of molecules was previously known only for transposons, plasmids and phages (68 70). Transcriptome analysis of *E. coli* and *H. pylori* indicated antisense transcription across the entire genome (17,71). In addition, several long and short antisense RNAs were previously reported for *Listeria* (31). We have identified a number of novel short antisense RNAs in our experiments (Supplementary Table S3). The vast majority of these were intracellularly up-regulated, although there was no strict correlation to down-regulation of their potential targeted genes when inspecting microarray-based intracellular transcriptome studies (37,61). Studies of individual plasmid-encoded and chromosomally encoded asRNAs in a variety of bacterial species have demonstrated that asRNAs can regulate gene expression at the level of translation, mRNA stability or transcription (68).

We were able to demonstrate, for the first time, the existence of two asRNAs (*anti2095* and *anti2394*) with increased intracellular expression using northern blots. Both asRNAs were previously reported to have short and long variants (31). Due to RNA-seq using RNA fraction <500 nt, only the short asRNAs were confirmed in our study (Supplementary Figure S3). The genes transcribed from the opposite strand encoding a hypothetical protein (*lmo2394*) and a phosphofructokinase (*lmo2095*) did not show significant expression changes after shifting from extra- to intra-cellular conditions (37,61). Also a genome-wide COG analysis of asRNA mRNA targets revealed no obvious correlation to specific classes of genes suggesting that asRNAs might be used to control general processes within the bacterial cell.

Riboswitches are *cis*-acting RNA structures responsible for downstream regulation of gene expression in bacteria (15). In *B. subtillis* 2% of the genes are regulated by these RNA elements (72). Out of 42 known listerial riboswitches in the RFAM database, 33 were detected in our study and the majority of them were down-regulated intracellularly. Interestingly, the lysine riboswitch (31) was not expressed during intracellular survival of bacteria indicating a lysine restricted intracellular environment for the bacterial pathogen.

Loh and colleagues have reported on the ability of two S-adenosylmethionine (SAM) riboswitches, SreA and SreB, to act *in trans* to modulate the expression of the critical listerial virulence regulator PrfA (73). We have observed that the SreA and SreB are down-regulated within the host cytosol indicating that the decreased copy number of SreA and SreB allow an intracellular induction of the PrfA regulator. The T-box class of riboswitches represents the largest class of riboswitches in *L. monocytogenes* sensing the level of uncharged tRNAs in the bacterial cell, which is induced in our data. The expression of the corresponding tRNA synthetase genes have previously been found to be decreased in the host cytosolic phase (37), which suggests the involvement of T-box regulation due to infection.

Finally, we found that listerial species have highly conserved riboswitches for adaption to physiological processes, but have clear differences in their sRNA and asRNA repertoire suggesting adaptation to their potential ecological habitats (Figure 6). Indeed comparative analysis indicated several strain and serotype-specific putative regulatory RNAs such as *rli112* which enables this sRNA to become a potential diagnostic marker for lineage II.

CONCLUSION

Here we show that extensive expression of sRNA candidates in *L. monocytogenes* occurs during intracellular growth. Our studies uncovered 71 previously undescribed

Downloaded from nar.oxfordjournals.org at Universitaetsbibliothek on February 9, 2011

putative regulatory RNAs and revealed 29 candidates that are specifically expressed during intracellular growth. Although we have shown that several of these sRNAs are indeed required for virulence of the bacterium their precise role in promoting bacterial survival under these conditions remains to be studied. As a next step, the identification of targets of these putative regulatory RNAs will be required and will enable us to understand the regulatory response triggered by the bacterium when shifting from extracellular to intracellular growth conditions.

ACCESSION NUMBERS

RNA sequencing data have been deposited to ArrayExpress (http://www.ebi.ac.uk/ arrayexpress), accession number E-MTAB-329.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Alexandra Amend, Juri Schklarenko, Nelli Schklarenko and Martina Hudel for technical assistance.

FUNDING

Grants from the German Federal Ministry of Education and Research (BMBF ERA-NET Pathogenomics Network to the sncRNAomics project 62080061 to T.H.). Funding for open access charge: Justus-Liebig University Giessen Institute of Medical Microbiology Frankfurter Strasse 107.

Conflict of interest statement. None declared.

REFERENCES

1. Livny,J., Teonadi,H., Livny,M. and Waldor,M.K. (2008) High-throughput, kingdom-wide prediction and annotation of bacterial non-coding RNAs. *PLoS ONE*, **3**, e3197.
2. Storz,G., Altuvia,S. and Wassarman,K.M. (2005) An abundance of RNA regulators. *Annu. Rev. Biochem.*, **74**, 199–217.
3. Vogel,J. and Sharma,C.M. (2005) How to find small non-coding RNAs in bacteria. *Biol. Chem.*, **386**, 1219–1238.
4. Gottesman,S. (2005) Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends Genet.*, **21**, 399–404.
5. Sorek,R. and Cossart,P. (2010) Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat. Rev. Genet.*, **11**, 9–16.
6. Livny,J. and Waldor,M.K. (2010) Mining regulatory 5′UTRs from cDNA deep sequencing datasets. *Nucleic Acids Res.*, **38**, 1504–1514.
7. Vogel,J., Bartels,V., Tang,T.H., Churakov,G., Slagter-Jager,J.G., Huttenhofer,A. and Wagner,E.G. (2003) RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.*, **31**, 6435–6443.
8. Washietl,S., Hofacker,I.L., Lukasser,M., Huttenhofer,A. and Stadler,P.F. (2005) Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat. Biotechnol.*, **23**, 1383–1390.
9. Washietl,S., Hofacker,I.L. and Stadler,P.F. (2005) Fast and reliable prediction of noncoding RNAs. *Proc. Natl Acad. Sci. USA*, **102**, 2454–2459.
10. Zhang,A., Wassarman,K.M., Rosenow,C., Tjaden,B.C., Storz,G. and Gottesman,S. (2003) Global analysis of small RNA and mRNA targets of Hfq. *Mol. Microbiol.*, **50**, 1111–1124.
11. Mraheil,M.A., Billion,A., Kuenne,C., Pischimarov,J., Kreikemeyer,B., Engelmann,S., Hartke,A., Giard,J.C., Rupnik,M., Vorwerk,S. *et al.* (2010) Comparative genome-wide analysis of small RNAs of major Gram-positive pathogens: from identification to application. *Microb. Biotechnol.*, **3**, 658–676.
12. Marker,C., Zemann,A., Terhorst,T., Kiefmann,M., Kastenmayer,J.P., Green,P., Bachellerie,J.P., Brosius,J. and Huttenhofer,A. (2002) Experimental RNomics: identification of 140 candidates for small non-messenger RNAs in the plant *Arabidopsis thaliana*. *Curr. Biol.*, **12**, 2002–2013.
13. Gottesman,S. (2004) The small RNA regulators of *Escherichia coli*: roles and mechanisms*. *Annu. Rev. Microbiol.*, **58**, 303–328.
14. Repoila,F. and Darfeuille,F. (2009) Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects. *Biol. Cell*, **101**, 117–131.
15. Waters,L.S. and Storz,G. (2009) Regulatory RNAs in bacteria. *Cell*, **136**, 615–628.
16. Romby,P., Vandenesch,F. and Wagner,E.G. (2006) The role of RNAs in the regulation of virulence-gene expression. *Curr. Opin. Microbiol.*, **9**, 229–236.
17. Sharma,C.M., Hoffmann,S., Darfeuille,F., Reignier,J., Findeiss,S., Sittka,A., Chabas,S., Reiche,K., Hackermuller,J., Reinhardt,R. *et al.* (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature*, **464**, 250–255.
18. Toledo-Arana,A., Repoila,F. and Cossart,P. (2007) Small noncoding RNAs controlling pathogenesis. *Curr. Opin. Microbiol.*, **10**, 182–188.
19. Franze de Fernandez,M.T., Eoyang,L. and August,J.T. (1968) Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. *Nature*, **219**, 588–590.
20. Franze de Fernandez,M.T., Hayward,W.S. and August,J.T. (1972) Bacterial proteins required for replication of phage Q ribonucleic acid. Pruification and properties of host factor I, a ribonucleic acid-binding protein. *J. Biol. Chem.*, **247**, 824–831.
21. Valentin-Hansen,P., Eriksen,M. and Udesen,C. (2004) The bacterial Sm-like protein Hfq: a key player in RNA transactions. *Mol. Microbiol.*, **51**, 1525–1533.
22. Aiba,H. (2007) Mechanism of RNA silencing by Hfq-binding small RNAs. *Curr. Opin. Microbiol.*, **10**, 134–139.
23. Ding,Y., Davis,B.M. and Waldor,M.K. (2004) Hfq is essential for *Vibrio cholerae* virulence and downregulates sigma expression. *Mol. Microbiol.*, **53**, 345–354.
24. McNealy,T.L., Forsbach-Birk,V., Shi,C. and Marre,R. (2005) The Hfq homolog in *Legionella pneumophila* demonstrates regulation by LetA and RpoS and interacts with the global regulator CsrA. *J. Bacteriol.*, **187**, 1527–1532.
25. Roop,R.M., Robertson,G.T., Ferguson,G.P., Milford,L.E., Winkler,M.E. and Walker,G.C. (2002) Seeking a niche: putative contributions of the *hfq* and *bacA* gene products to the successful adaptation of the brucellae to their intracellular home. *Vet. Microbiol.*, **90**, 349–363.
26. Sonnleitner,E., Hagens,S., Rosenau,F., Wilhelm,S., Habel,A., Jager,K.E. and Blasi,U. (2003) Reduced virulence of a *hfq* mutant of *Pseudomonas aeruginosa* O1. *Microb. Pathog.*, **35**, 217–228.
27. Christiansen,J.K., Larsen,M.H., Ingmer,H., Sogaard-Andersen,L. and Kallipolitis,B.H. (2004) The RNA-binding protein Hfq of *Listeria monocytogenes*: role in stress tolerance and virulence. *J. Bacteriol.*, **186**, 3355–3362.
28. Christiansen,J.K., Nielsen,J.S., Ebersbach,T., Valentin-Hansen,P., Sogaard-Andersen,L. and Kallipolitis,B.H. (2006) Identification of small Hfq-binding RNAs in *Listeria monocytogenes*. *RNA*, **12**, 1383–1396.
29. Mandin,P., Repoila,F., Vergassola,M., Geissmann,T. and Cossart,P. (2007) Identification of new noncoding RNAs in *Listeria monocytogenes* and prediction of mRNA targets. *Nucleic Acids Res.*, **35**, 962–974.
30. Nielsen,J.S., Olsen,A.S., Bonde,M., Valentin-Hansen,P. and Kallipolitis,B.H. (2008) Identification of a sigma B-dependent

Downloaded from nar.oxfordjournals.org at Universitaetsbibliothek on February 9, 2011

Appendix

small noncoding RNA in *Listeria monocytogenes*. *J. Bacteriol.*, **190**, 6264–6270.
31. Toledo-Arana,A., Dussurget,O., Nikitas,G., Sesto,N., Guet-Revillet,H., Balestrino,D., Loh,E., Gripenland,J., Tiensuu,T., Vaitkevicius,K. *et al.* (2009) The *Listeria* transcriptional landscape from saprophytism to virulence. *Nature*, **459**, 950–956.
32. Oliver,H.F., Orsi,R.H., Ponnala,L., Keich,U., Wang,W., Sun,Q., Cartinhour,S.W., Filiatrault,M.J., Wiedmann,M. and Boor,K.J. (2009) Deep RNA sequencing of *L. monocytogenes* reveals overlapping and extensive stationary phase and sigma B-dependent transcriptomes, including multiple highly transcribed noncoding RNAs. *BMC Genomics*, **10**, 641.
33. Duffy,L.L., Vanderlinde,P.B. and Grau,F.H. (1994) Growth of *Listeria monocytogenes* on vacuum-packed cooked meats: effects of pH, aw, nitrite and ascorbate. *Int. J. Food Microbiol.*, **23**, 377–390.
34. Cossart,P. and Toledo-Arana,A. (2008) *Listeria monocytogenes*, a unique model in infection biology: an overview. *Microbes. Infect.*, **10**, 1041–1050.
35. Hain,T., Chatterjee,S.S., Ghai,R., Kuenne,C.T., Billion,A., Steinweg,C., Domann,E., Karst,U., Jansch,L., Wehland,J. *et al.* (2007) Pathogenomics of *Listeria* spp. *Int. J. Med. Microbiol.*, **297**, 541–557.
36. Glaser,P., Frangeul,L., Buchrieser,C., Rusniok,C., Amend,A., Baquero,F., Berche,P., Bloecker,H., Brandt,P., Chakraborty,T. *et al.* (2001) Comparative genomics of *Listeria* species. *Science*, **294**, 849–852.
37. Chatterjee,S.S., Hossain,H., Otten,S., Kuenne,C., Kuchmina,K., Machata,S., Domann,E., Chakraborty,T. and Hain,T. (2006) Intracellular gene expression profile of *Listeria monocytogenes*. *Infect. Immun.*, **74**, 1323–1338.
38. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
39. Griffiths-Jones,S., Moxon,S., Marshall,M., Khanna,A., Eddy,S.R. and Bateman,A. (2005) Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.*, **33**, D121–D124.
40. Eddy,S.R. (2009) A new generation of homology search tools based on probabilistic inference. *Genome Inform.*, **23**, 205–211.
41. Kingsford,C.L., Ayanbule,K. and Salzberg,S.L. (2007) Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol.*, **8**, R22.
42. Meyer,F., Goesmann,A., McHardy,A.C., Bartels,D., Bekel,T., Clausen,J., Kalinowski,J., Linke,B., Rupp,O., Giegerich,R. *et al.* (2003) GenDB–an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res.*, **31**, 2187–2195.
43. Hain,T., Steinweg,C., Kuenne,C.T., Billion,A., Ghai,R., Chatterjee,S.S., Domann,E., Karst,U., Goesmann,A., Bekel,T. *et al.* (2006) Whole-genome sequence of *Listeria welshimeri* reveals common steps in genome reduction with *Listeria innocua* as compared to *Listeria monocytogenes*. *J. Bacteriol.*, **188**, 7405–7415.
44. Steinweg,C., Kuenne,C.T., Billion,A., Mraheil,M.A., Domann,E., Ghai,R., Barbuddhe,S.B., Karst,U., Goesmann,A., Puhler,A. *et al.* (2010) Complete genome sequence of Listeria seeligeri, a nonpathogenic member of the genus *Listeria*. *J. Bacteriol.*, **192**, 1473–1474.
45. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics.*, **25**, 2078–2079.
46. Ghai,R., Hain,T. and Chakraborty,T. (2004) GenomeViz: visualizing microbial genomes. *BMC Bioinformatics.*, **5**, 198.
47. Thedieck,K., Hain,T., Mohamed,W., Tindall,B.J., Nimtz,M., Chakraborty,T., Wehland,J. and Jansch,L. (2006) The MprF protein is required for lysinylation of phospholipids in listerial membranes and confers resistance to cationic antimicrobial peptides (CAMPs) on *Listeria monocytogenes*. *Mol. Microbiol.*, **62**, 1325–1339.
48. Schaferkordt,S. and Chakraborty,T. (1995) Vector plasmid for insertional mutagenesis and directional cloning in *Listeria* spp. *Biotechniques*, **19**, 720–725.
49. Lauer,P., Chow,M.Y., Loessner,M.J., Portnoy,D.A. and Calendar,R. (2002) Construction, characterization, and use of two *Listeria monocytogenes* site-specific phage integration vectors. *J. Bacteriol.*, **184**, 4177–4186.
50. Mukherjee,K., Altincicek,B., Hain,T., Domann,E., Vilcinskas,A. and Chakraborty,T. (2010) *Galleria mellonella* as a model system for studying *Listeria* pathogenesis. *Appl. Environ. Microbiol.*, **76**, 310–317.
51. Chakraborty,T., Leimeister-Wachter,M., Domann,E., Hartl,M., Goebel,W., Nichterlein,T. and Notermans,S. (1992) Coordinate regulation of virulence genes in *Listeria monocytogenes* requires the product of the *prfA* gene. *J. Bacteriol.*, **174**, 568–574.
52. Hamon,M., Bierne,H. and Cossart,P. (2006) *Listeria monocytogenes*: a multifaceted model. *Nat. Rev. Microbiol.*, **4**, 423–434.
53. Nelson,K.E., Fouts,D.E., Mongodin,E.F., Ravel,J., DeBoy,R.T., Kolonay,J.F., Rasko,D.A., Angiuoli,S.V., Gill,S.R., Paulsen,I.T. *et al.* (2004) Whole genome comparisons of serotype 4b and 1/2a strains of the food-borne pathogen *Listeria monocytogenes* reveal new insights into the core genome components of this species. *Nucleic Acids Res.*, **32**, 2386–2395.
54. Edgar,R.C. (2007) PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics*, **8**, 18.
55. Gilmour,M.W., Graham,M., Van Domselaar,G., Tyler,S., Kent,H., Trout-Yakel,K.M., Larios,O., Allen,V., Lee,B. and Nadon,C. (2010) High-throughput genome sequencing of two *Listeria monocytogenes* clinical isolates during a large foodborne outbreak. *BMC Genomics*, **11**, 120.
56. Liu,D., Lawrence,M.L., Pinchuk,L.M., Ainsworth,A.J. and Austin,F.W. (2007) Characteristics of cell-mediated, anti-listerial immunity induced by a naturally avirulent *Listeria monocytogenes* serotype 4a strain HCC23. *Arch. Microbiol.*, **188**, 251–256.
57. Cromie,M.J., Shi,Y., Latifi,T. and Groisman,E.A. (2006) An RNA sensor for intracellular Mg(2+). *Cell*, **125**, 71–84.
58. Lenz,D.H., Mok,K.C., Lilley,B.N., Kulkarni,R.V., Wingreen,N.S. and Bassler,B.L. (2004) The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. *Cell*, **118**, 69–82.
59. Svenningsen,S.L., Tu,K.C. and Bassler,B.L. (2009) Gene dosage compensation calibrates four regulatory RNAs to control *Vibrio cholerae* quorum sensing. *EMBO J.*, **28**, 429–439.
60. Castillo-Keller,M., Vuong,P. and Misra,R. (2006) Novel mechanism of *Escherichia coli* porin regulation. *J. Bacteriol.*, **188**, 576–586.
61. Joseph,B., Przybilla,K., Stuhler,C., Schauer,K., Slaghuis,J., Fuchs,T.M. and Goebel,W. (2006) Identification of *Listeria monocytogenes* genes contributing to intracellular replication by expression profiling and mutant screening. *J. Bacteriol.*, **188**, 556–568.
62. Camejo,A., Buchrieser,C., Couve,E., Carvalho,F., Reis,O., Ferreira,P., Sousa,S., Cossart,P. and Cabanes,D. (2009) *In vivo* transcriptional profiling of *Listeria monocytogenes* and mutagenesis identify new virulence factors involved in infection. *PLoS Pathog.*, **5**, e1000449.
63. Horvath,P. and Barrangou,R. (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science*, **327**, 167–170.
64. Georg,J., Voss,B., Scholz,I., Mitschke,J., Wilde,A. and Hess,W.R. (2009) Evidence for a major role of antisense RNAs in cyanobacterial gene regulation. *Mol. Syst. Biol.*, **5**, 305.
65. Liu,J.M., Livny,J., Lawrence,M.S., Kimball,M.D., Waldor,M.K. and Camilli,A. (2009) Experimental discovery of sRNAs in *Vibrio cholerae* by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res.*, **37**, e46.
66. Romby,P. and Charpentier,E. (2010) An overview of RNAs with regulatory functions in gram-positive bacteria. *Cell Mol. Life Sci.*, **67**, 217–237.
67. Wurtzel,O., Sapra,R., Chen,F., Zhu,Y., Simmons,B.A. and Sorek,R. (2010) A single-base resolution map of an archaeal transcriptome. *Genome Res.*, **20**, 133–141.
68. Brantl,S. (2007) Regulatory mechanisms employed by *cis*-encoded antisense RNAs. *Curr. Opin. Microbiol.*, **10**, 102–109.

Downloaded from nar.oxfordjournals.org at Universitaetsbibliothek on February 9, 2011

69. Wagner,E.G., Altuvia,S. and Romby,P. (2002) Antisense RNAs in bacteria and their genetic elements. *Adv. Genet.*, **46**, 361–398.
70. Weaver,K.E. (2007) Emerging plasmid-encoded antisense RNA regulated systems. *Curr. Opin. Microbiol.*, **10**, 110–116.
71. Selinger,D.W., Cheung,K.J., Mei,R., Johansson,E.M., Richmond,C.S., Blattner,F.R., Lockhart,D.J. and Church,G.M. (2000) RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. *Nat. Biotechnol.*, **18**, 1262–1268.
72. Phan,T.T. and Schumann,W. (2009) Transcriptional analysis of the lysine-responsive and riboswitch-regulated *lysC* gene of *Bacillus subtilis*. *Curr. Microbiol.*, **59**, 463–468.
73. Loh,E., Dussurget,O., Gripenland,J., Vaitkevicius,K., Tiensuu,T., Mandin,P., Repoila,F., Buchrieser,C., Cossart,P. and Johansson,J. (2009) A *trans*-acting riboswitch controls expression of the virulence regulator PrfA in *Listeria monocytogenes*. *Cell*, **139**, 770–779.

Downloaded from nar.oxfordjournals.org at Universitaetsbibliothek on February 9, 2011

13.5 Appendix publication 4

microbial biotechnology

Microbial Biotechnology (2010) doi:10.1111/j.1751-7915.2010.00171.x

Minireview

Comparative genome-wide analysis of small RNAs of major Gram-positive pathogens: from identification to application

Mobarak A. Mraheil,[1†] André Billion,[1†] Carsten Kuenne,[1] Jordan Pischimarov,[1] Bernd Kreikemeyer,[2] Susanne Engelmann,[3] Axel Hartke,[4] Jean-Christophe Giard,[4] Maja Rupnik,[5] Sonja Vorwerk,[6] Markus Beier,[6] Julia Retey,[7] Thomas Hartsch,[7] Anette Jacob,[8] Franz Cemič,[9] Jürgen Hemberger,[9] Trinad Chakraborty[1] and Torsten Hain[1*]

[1]*Institute of Medical Microbiology, Justus-Liebig-University, Frankfurter Strasse 107, 35392 Giessen, Germany.*
[2]*Institute for Medical Microbiology, University of Rostock, Schillingallee 70, 18057 Rostock, Germany.*
[3]*Institute for Microbiology, University of Greifswald, Jahnstrasse 15, 17487 Greifswald, Germany.*
[4]*Laboratoire de Microbiologie de l'Université de Caen, EA956 USC INRA2017, 14032 CAEN cedex, France.*
[5]*University of Maribor, Faculty of Medicine, Slomskov trg 15, 2000 Maribor, Slovenia.*
[6]*Febit biomed Gmbh, Im Neuenheimer Feld 519, 69120 Heidelberg, Germany.*
[7]*Genedata Bioinformatik GmbH, Lena-Christ-Str. 50, D-82152 Planegg-Martinsried, Germany.*
[8]*Functional Genome Analysis, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 580, 69120 Heidelberg, Germany.*
[9]*Institute for Biochemical Engineering and Analytics, University of Applied Sciences Giessen-Friedberg, Wiesenstasse 14, 35390 Giessen, Germany.*

Summary

In the recent years, the number of drug- and multi-drug-resistant microbial strains has increased rapidly. Therefore, the need to identify innovative approaches for development of novel anti-infectives and new therapeutic targets is of high priority in global health care. The detection of small RNAs (sRNAs) in bacteria has attracted considerable attention as an emerging class of new gene expression regulators. Several experimental technologies to predict sRNA have been established for the Gram-negative model organism *Escherichia coli*. In many respects, sRNA screens in this model system have set a blueprint for the global and functional identification of sRNAs for Gram-positive microbes, but the functional role of sRNAs in colonization and pathogenicity for *Listeria monocytogenes*, *Staphylococcus aureus*, *Streptococcus pyogenes*, *Enterococcus faecalis* and *Clostridium difficile* is almost completely unknown. Here, we report the current knowledge about the sRNAs of these socioeconomically relevant Gram-positive pathogens, overview the state-of-the-art high-throughput sRNA screening methods and summarize bioinformatics approaches for genome-wide sRNA identification and target prediction. Finally, we discuss the use of modified peptide nucleic acids (PNAs) as a novel tool to inactivate potential sRNA and their applications in rapid and specific detection of pathogenic bacteria.

Received 20 October, 2009; accepted 17 February, 2010. *For correspondence. E-mail Torsten.Hain@mikrobio.med.uni-giessen.de; Tel. (+49) 641 99 46400; Fax (+49) 641 99 46409. †Both coauthors contributed equally to the work.

© 2010 The Authors
Journal compilation © 2010 Society for Applied Microbiology and Blackwell Publishing Ltd

Introduction

Small non-coding RNAs and especially microRNAs (miRNAs) have been recently identified as key regulators of several cellular processes in multicellular eukaryotes (Garzon *et al.*, 2009). Massive resources are now allocated to understand this extra layer of gene regulation. Involvement of miRNA expression in different types of tumours (e.g. breast, colon or brain cancer) has been reported (Garzon *et al.*, 2009; Negrini *et al.*, 2009). Consequently, blocking of miRNAs has become of therapeutic interest for tumour treatment (Dalmay, 2008; Negrini *et al.*, 2009). In addition, miRNA profiles are now considered as a new tool for diagnostics. Although miRNA profiles contain much less information than an mRNA profile, the potential value may be higher because of the regulatory role of miRNAs (Dalmay, 2008).

In bacteria, small RNAs (sRNAs) have attracted considerable attention as an emerging class of new gene expression regulators. Apart from open reading frames, the genome codes for a number of RNAs with non-coding functions such as rRNAs, tRNAs, and some small non-coding RNAs, which may act as regulators of transcription or translation and influence mRNA stability (Waters and Storz, 2009). Although the majority of sRNAs might indeed be non-coding, there are sRNAs, which have both a coding and a non-coding function (Ji *et al.*, 1995).

Small RNAs interact by pairing with other RNAs, forming parts of RNA–protein complexes, or adopting structures of other nucleic acids (Storz *et al.*, 2004). A toolbox of experimental technologies such as microarray detection, shotgun cloning (RNomics), co-purification with proteins and algorithms to predict sRNAs have been established for Gram-negative bacteria such as *Escherichia coli* (Sharma and Vogel, 2009). Investigation of the role of sRNAs for other Gram-negative pathogens such as *Salmonella typhimurium* and *Pseudomonas aeruginosa* has recently been started (Sharma and Vogel, 2009). sRNAs were detected within genetic islands of *S. typhimurium*, which showed host-induced expression in macrophages and thus contributed to virulence (Padalon-Brauch *et al.*, 2008).

Recently, an extended number of 103 sRNAs was described for *Listeria monocytogenes* (Christiansen *et al.*, 2006; Mandin *et al.*, 2007; Nielsen *et al.*, 2008; Toledo-Arana *et al.*, 2009). Only two sRNAs have been identified and studied in some detail in *Streptococcus pyogenes* (Kreikemeyer *et al.*, 2001; Mangold *et al.*, 2004) and five sRNAs for *Streptococcus pneumoniae* (Halfmann *et al.*, 2007). A first approach identified 12 sRNAs in *Staphylococcus aureus*, seven of which are localized on pathogenicity islands. Some of the sRNAs show remarkable variations of expression levels among pathogenic *S. aureus* strains, which suggest their involvement in the regulation of virulence factors (Pichon and Felden, 2005). Presently, there is no information available for sRNAs from *Clostridium difficile* and *Enterococcus faecalis.*

To date, sRNAs have been found to be implicated in stress response, iron homeostasis, outer membrane protein biogenesis, sugar metabolism and quorum sensing, suggesting that they might also play an essential and central role in the pathogenicity of many bacteria. A global approach to identify sRNAs in *L. monocytogenes* has been recently published (Toledo-Arana *et al.*, 2009), but genome-wide approaches to elucidate the functional role of sRNAs for other high-risk Gram-positive pathogens in pathogenicity are limited.

To address this deficiency, we recently established a European consortium (http://www.pathogenomics-era.net/2ndJointCall/) to perform comparative global analysis of sRNAs for *L. monocytogenes*, *S. aureus*, *S. pyogenes*, *E. faecalis* and *C. difficile*. In this project, we will utilize bioinformatics, novel high-throughput sRNA screening methods, whole-genome transcriptomics and proteomics, coupled with existing robust molecular characterization methods to provide new information regarding production, regulation and pathogenic implications of sRNAs in these five major high-risk Gram-positive pathogens on a global scale.

These analyses will help us to design novel potential therapeutics based on sRNA-complementary peptide nucleic acids (PNAs). Furthermore, the knowledge about sRNA expression will be used to develop a novel ultra-sensitive diagnostic system, which has been conceived for detection of small target samples at extremely low concentrations in a very short time.

Listeria monocytogenes

Listeria monocytogenes is an opportunistic facultative intracellular bacterium which is ubiquitously distributed in nature. This food-borne pathogen belongs to the group of bacteria with low G + C DNA content which also includes other species of genera such as *Streptococcus*, *Staphylococcus*, *Enterococcus* and *Clostridium*. The genus *Listeria* consists of seven different species, namely *Listeria monocytogenes*, *Listeria ivanovii*, *Listeria innocua*, *Listeria welshimeri*, *Listeria seeligeri*, *Listeria grayi* and *Listeria marthii* (Hain *et al.*, 2007; Graves *et al.*, 2009), of which *L. ivanovii* is predominantly a serious animal pathogen while *L. monocytogenes* can cause fatal infections in humans and animals. With regard to its transmission, a majority of listerial cases have been documented through contaminated food products. The major clinical symptoms exhibited by this pathogen in humans include meningitis, septicaemia, abortion, prenatal infection and gastroenteritis. In spite of an appropriate antibiotic therapy, approximately 20–30% of deaths have been reported in patients suffering from listeriosis (Hof *et al.*, 2007).

A hallmark of this human pathogen is its ability to invade and survive inside vertebrate and invertebrate host cells, wherein the bacterium can freely multiply within the cytosol and can induce actin-based movement and cell-to-cell spreading. Prior to infection, Internalin A and B induce the first step of the infection process by interacting with the eukaryotic host cell and promote intracellular uptake of the pathogen after binding with E-Cadherin and c-Met as interacting receptors in mammals. However, the main virulence genes, responsible for the intracellular life cycle of *L. monocytogenes*, are clustered in a ~9 kb chromosomal region (Hain *et al.*, 2007; Cossart and Toledo-Arana, 2008).

Over the last decade, regulatory RNA elements have gained increasing importance in the physiology and pathogenesis of prokaryotes. They are divided into two

© 2010 The Authors
Journal compilation © 2010 Society for Applied Microbiology and Blackwell Publishing Ltd, *Microbial Biotechnology*

major groups based on their mode of action: *cis* (untranslated region, UTR) and *trans* acting regulatory RNAs (small non-coding RNAs). Riboswitches and RNA thermometers (Narberhaus *et al.*, 2006) belong to the class of *cis* acting regulatory RNAs located at the 5′UTR of their genes. In general, riboswitches modulate their regulatory structure in response to metabolite binding, which are available in their own environment whereas, RNA thermometers are involved in sensing global signals, e.g. the intracellular temperature. Both *cis* regulatory RNA structures are essential for fine regulatory tuning so as to ensure an immediate physiological response of the bacterial cell to a varying habitat.

The first *cis* regulatory RNA was described for the master virulence regulator PrfA (Johansson *et al.*, 2002). This 5′UTR acts as RNA thermometer which is turned off under low-temperature conditions. In addition, the importance of 5′UTRs has also been reported for other virulence genes (Wong *et al.*, 2004; Shen and Higgins, 2005; Stritzker *et al.*, 2005). In general, the effects of UTRs on the expression of ActA, Hly and InlA were analysed by deleting the 5′UTR, while their ribosome binding sites were retained. Further computational analysis has revealed the presence of putative UTRs in surface proteins encoding operons associated with virulence in *L. monocytogenes* and suggests a potential post transcriptional function for these long RNA sequences in pathogenesis (Loh *et al.*, 2006).

The role of *trans* regulatory RNAs such as small non-coding RNAs has been addressed recently. The housekeeping sRNA 4.5S which binds to the signal recognition particle (SRP) was the first sRNA identified in *L. monocytogenes* (Barry *et al.*, 1999). SRPs are highly conserved in most bacteria and are involved in translation and targeting of proteins for cellular secretion.

In *L. monocytogenes*, the RNA chaperone Hfq is involved in stress tolerance and virulence (Christiansen *et al.*, 2004). In order to investigate sRNA–Hfq interaction, a co-immunoprecipitation approach identified three Hfq-binding sRNAs (*lhrA*, *lhrB* and *lhrC*; *lhrC* reveals five copies in the chromosome), which are growth phase-dependently expressed and induced during intracellular multiplication within HepG2 cells. In a later study, new algorithms were used to predict and functionally map novel sRNA for *L. monocytogenes*. This allowed the identification of 12 additional sRNA (*rliA–I*, *ssrS*, *ssrA* and *rnpB*), and for three of them (*rliB*, *rilE* and *rliI*), potential mRNA targets were predicted *in silico* and experimentally verified (Mandin *et al.*, 2007).

Listeria monocytogenes is well known for its robust physiology because of its capability to grow under refrigeration temperature, low pH and also at high osmolarity. However, relatively little is known about the role of sRNA for the adaptive physiological response in promoting survival and growth of the bacterium in such hostile environments. Besides, the involvement of the alternative sigma factor σ^B in the regulation of sRNA was elucidated not until recently when a 70-nt-long sRNA (*sbrA*) was identified which was reported to be under control of this stress tolerance factor (Nielsen *et al.*, 2008).

In order to study global transcriptional profiling in response to the changing environmental niche of the pathogen, several different stress and *in vitro/in vivo* infection conditions were exploited using genome-wide microarray approaches (Joseph *et al.*, 2006; van der Veen *et al.*, 2007; Hain *et al.*, 2008; Raengpradub *et al.*, 2008; Camejo *et al.*, 2009). We have used transcriptome profiling to examine the expression profile of *L. monocytogenes* inside the vacuolar and the cytosolic environments of the host cell using whole-genome microarray and mutant analysis. We found that ~17% of the total genome was mobilized to enable adaptation to intracellular growth (Chatterjee *et al.*, 2006), but the role of sRNA during host cell infection was not addressed in this study.

A first genome-wide approach has been recently reported by Toledo-Arana and colleagues (2009) using tiling arrays to detect novel sRNA. Thus, the transcriptional profile of *L. monocytogenes* was analysed under several different conditions including bacteria growing in BHI exponentially as well as to the stationary phase, under low oxygen and temperature (30°C), in the murine intestine and human blood and compared with the transcriptome of Δ*prfA*, Δ*sigB* and Δ*hfq* isogenic mutants. Under these conditions a complete operon map with 5′ and 3′ end boundaries was determined as well as 103 small regulatory RNA were identified (Table 1). Among the regulatory RNAs, 29 novel sRNAs, 13 *cis* regulatory RNAs (5′- and 3′UTR, putative riboswitches) and 40 *cis* regulatory RNAs including known riboswitches were identified. Isogenic mutant analyses of *rliB* and *rli38* indicated their contribution in virulence in mice. This initial study describes the global transcriptional landscape in *L. monocytogenes* under various growth conditions and provides insights into strategies of extracellular survival of the human pathogen. Nevertheless, details on the precise role of sRNAs in the pathogenesis of *L. monocytogenes* remain limited and additional studies on physiological function are needed in order to understand the sRNA regulatory function.

Staphylococcus aureus

Staphylococcus aureus is one main cause of nosocomial infections worldwide. Because of the high occurrence of multiple resistant strains in hospital settings, infections caused by this pathogen are often difficult to treat and new therapeutic strategies are urgently needed. The post genomic era of *S. aureus* started in 2001 with the publication of the genome sequences of two strains, N315 and

© 2010 The Authors
Journal compilation © 2010 Society for Applied Microbiology and Blackwell Publishing Ltd, *Microbial Biotechnology*

Appendix

Table 1. Current overview of published Gram-positive sRNAs of the genera *Staphylococcus*, *Streptococcus*, *Enterococcus*, *Clostridium* and *Listeria*.

		SIPHT[a]		Experimentally verified	
Genus	Species	Chromosome	Plasmid	Chromosome	Plasmid
Staphylococcus	*aureus*	32–79 (12)	0–5 (9)	23[b]	1[c]
	epidermidis	116–127 (2)	0–4 (7)		
	haemolyticus	74 (1)	–		
	saprophyticus	38 (1)	1 (2)		
Streptococcus	*agalactiae*	29–34 (3)	–		
	mutans	18 (1)	–		
	pneumoniae	28–66 (3)	–	5[d]	
	pyogenes	18–29 (12)	–	3[e]	
	thermophilus	31–36 (3)	0 (2)		
	sanguinis	34 (1)	–		
	suis	23–24 (2)	–		
Enterococcus Clostridium	*faecalis*	14 (1)	0–2 (3)		2[f]
	acetobutylicum	18 (1)	2 (1)	1[g]	
	beijerinckii	31 (1)	–		
	botulinum	54–68 (4)	0 (2)		
	difficile	2 (1)	0 (1)		
	kluyveri	46 (1)	0 (1)		
	novyi	26 (1)	–		
	perfringens	14–18 (3)	0–2 (3)	1[h]	
	tetani	45 (1)	0 (1)		
	thermocellum	6 (1)	–		
Listeria	*innocua*	115 (1)	2 (1)		
	monocytogenes	94–124 (2)	–	27[i]	
	welshimeri	100 (1)	–		

a. Livny *et al.* (2008).
b. Novick *et al.* (1993); Morfeldt *et al.* (1995); Pichon and Felden (2005); Geissmann *et al.* (2009).
c. Kwong *et al.* (2004; 2006).
d. Halfmann *et al.* (2007).
e. Kreikemeyer *et al.* (2001); Mangold *et al.* (2004); Roberts and Scott (2007).
f. Weaver (2007).
g. Fierro-Monti *et al.* (1992).
h. Shimizu *et al.* (2002).
i. Barry *et al.* (1999); Christiansen *et al.* (2006); Mandin *et al.* (2007); Nielsen *et al.* (2008); Toledo-Arana *et al.* (2009).
The SIPHT columns show the minimum and maximum number of annotated sRNAs of each species.
The quantity of analysed strains is depicted in brackets.

Mu50 (Kuroda *et al.*, 2001). In the *S. aureus* N315 genome, 110 non-coding RNAs are annotated. Among them are 78 tRNAs and rRNAs and 32 putative sRNAs and riboswitches (Geissmann *et al.*, 2009).

A large number of virulence factors are known to be involved in the pathogenesis of *S. aureus* whose expression is subject of temporal control and mainly affected by RNAIII. RNAIII was the first regulatory RNA shown to be involved in bacterial virulence and is part of the *agr* locus in *S. aureus* that acts as a quorum sensing system (Recsei *et al.*, 1986; Novick *et al.*, 1993; Ji *et al.*, 1995) RNAIII is a 514 nt regulatory RNA and controls the switch between production of cell wall adhesins and that of extracellular proteins (Dunman *et al.*, 2001; Ziebandt *et al.*, 2004). The mechanism by which RNAIII regulates such a large number of genes has been unknown for a long time. In the meantime, a direct regulatory effect of RNAIII has been shown for *hla* expression (Morfeldt *et al.*, 1995) and for *spa* expression (Huntzinger *et al.*, 2005). However, RNAIII possibly affects the expression of most of the virulence genes in *S. aureus* indirectly via the pleiotropic regulator Rot (for repressor of toxins). RNAIII antagonizes Rot activity by blocking its synthesis, with RNAIII acting as an antisense RNA, pairing with complementary regions in the 5′ end of *rot* mRNA. Binding of RNAIII induces cleavage of the *rot* transcript (Geisinger *et al.*, 2006; Boisset *et al.*, 2007).

Pichon and Felden (2005) described the first global approach to identify sRNAs in *S. aureus*. They used a comparative genomic approach to characterize the RNome of *S. aureus* on the basis of the genome sequence of strain N315 and detected at least 12 sRNAs: five are encoded in the core genome and seven are localized on pathogenicity islands. Some of the sRNAs show remarkable variations of expression among pathogenic strains and it is postulated that they might be involved in the regulation of virulence factors as it is described for RNAIII.

Most recently, a more comprehensive search for sRNAs in *S. aureus* was performed by Geissmann and colleagues (2009). They identified 11 new sRNAs (RsaA to

© 2010 The Authors
Journal compilation © 2010 Society for Applied Microbiology and Blackwell Publishing Ltd, *Microbial Biotechnology*

K). Transcriptional analyses using three reference strains of *S. aureus* (RN6390, Newman and COL) also showed different transcription profiles of these sRNAs between the strains. Most of the sRNAs were growth phase-dependently expressed and transcribed in response to environmental changes such as oxidative stress, heat stress, osmotic stress and acidic pH. Transcription of three of them was mediated by the alternative sigma factor σ^B and one was regulated by the quorum sensing system *agr*.

Details on the regulatory function of sRNAs in physiology and pathogenesis of *S. aureus* are still limited. By using transcriptomic and proteomic analyses, Geissmann and colleagues (2009) have shown that RsaE which was expressed under various stress conditions is involved in the regulation of several genes involved in amino acid and peptide transport, cofactor synthesis, lipid and carbohydrate metabolism and the TCA cycle. RsaE binds to its target mRNAs via a specific sequence motif and inhibits the formation of the translational initiation complex. Direct interactions of RsaE with *oppB*, *sucD* and *SA0873* mRNAs have been shown. The specific sequence motif seems to be highly conserved among sRNAs in *S. aureus* indicating a very similar mode of action (Geissmann *et al.*, 2009).

These studies (Pichon and Felden, 2005; Geissmann *et al.*, 2009), however, can only represent the blueprint for the prediction, detection and characterization of sRNAs in *S. aureus.* It can be postulated that the *S. aureus* genome likely codes for more regulatory sRNAs that affect gene expression than have been identified to date. By using tiling DNA arrays and new sequencing techniques, we will certainly get a more comprehensive picture of regulatory sRNAs in *S. aureus* expressed under certain conditions. Characterization of their regulatory role in global gene expression could shed more light on processes of *S. aureus* involved in adaptation to the host environment and pathogenicity and may provide new targets for therapeutic strategies to treat infections caused by this pathogen.

Streptococcus pyogenes

Streptococcus pyogenes (group A streptococci, GAS) is another important exclusively human bacterial pathogen. The annual global burden of GAS diseases, including 111 million cases of pyoderma, over 616 million cases of pharyngitis, and at least over 500 000 deaths due to severe invasive diseases, places this pathogen among the most important Gram-positive bacterial species with major impact in global mortality and morbidity (Carapetis *et al.*, 2005). GAS encodes a large number of virulence factors that are involved in adhesion, colonization, immune evasion and long-term survival within the host (Courtney *et al.*, 2002; Kreikemeyer *et al.*, 2003; 2004). For successful adaptation and survival in various targeted host compartments, about 30 orphan transcriptional regulators (RR), together with 13 two-component signal transduction systems (TCSs) (of which 11 are conserved across all GAS serotypes), sense and integrate the environmental information in GAS (Kreikemeyer *et al.*, 2003; Beyer-Sehlmeyer *et al.*, 2005; Musser and DeLeo, 2005).

As in many other bacterial species, a novel level of potential virulence regulation has emerged from the discovery and activity of sRNAs. Two bioinformatics-based approaches have identified candidate sRNAs in the *S. pyogenes* genome. Livny and colleagues developed SIPTH (sRNA identification protocol using high-throughput technologies) to identify sRNAs in intergenic loci based on colocalization of intergenic conservation and the presence of Rho-independent terminators (Livny *et al.*, 2008). In order to integrate all available sRNA identification methods we have recently developed MOSES (modular sequence suit), a Java-based framework, to integrate detection approaches for sRNAs. Using MOSES on the genome sequence of the *S. pyogenes* M49 NZ131 strain we identified four highly probable sRNA candidate genes which were verified by RT-PCR (P. Raasch, U. Schmitz, N. Patenge, B. Kreikemeyer and O. Wolkenhauer, submitted).

Experimental approaches prior to the availability of tiling array technology allowed identification and partial characterization of three sRNAs (SLS, FasX and RivX) in *S. pyogenes* (Kreikemeyer *et al.*, 2001; Mangold *et al.*, 2004; Roberts and Scott, 2007).

A first GAS sRNA *sagA* was discovered during the identification and functional characterization of the major haemolysin of GAS, streptolysin S (SLS/SagA) (Li *et al.*, 1999; Carr *et al.*, 2001). Mangold and colleagues later reported that the mRNA transcribed from the *sls* gene most likely acts as sRNA and effectively regulates expression of *emm* (M protein), *sic* (streptococcal inhibitor of complement) and *nga* (NAD-glycohydrolase) at the transcriptional level, and expression of SpeB (cystein protease) at the post-transcriptional level (Mangold *et al.*, 2004).

The second putative sRNA experimentally identified was *fasX*, a 300-nucleotide-long transcript, which is associated with the FasBCAX (fibronectin/fibrinogen binding/haemolytic activity/streptokinase regulator) system (Kreikemeyer *et al.*, 2001).

Measurement of a luciferase–promotor fusion revealed a growth phase-associated transcription of *fasX* with peak activities during late exponential phase. Mutation of *fasX* uncovered a reduced expression of the secreted virulence factors *sls* and streptokinase (*ska*), and simultaneously a prolonged expression of fibronectin- (*fbp54*) and fibrinogen-binding (*mrp*) adhesins. Analysis of the role of *fasX* in *S. pyogenes* adherence to and internalization into

© 2010 The Authors
Journal compilation © 2010 Society for Applied Microbiology and Blackwell Publishing Ltd, *Microbial Biotechnology*

HEp-2 cells (a human laryngeal carcinoma cell line) revealed that this sRNA apparently promotes high adherence and internalization rates, leading to massive cytokine gene transcription and cytokine release, host cell apoptosis via a novel caspase 2 activation pathway and cytotoxicity (Klenk *et al.*, 2005). It is evident that the Fas two-component signal transduction system, with the *fasX* sRNA as its integral and main effector part, mediates virulence gene regulation and could be involved in GAS aggressiveness and local tissue destruction (Klenk *et al.*, 2005).

Recently, Roberts and Scott showed that *rivX* encodes another sRNA and microarray analysis revealed that products of the *rivRX* locus exert positive control over transcription of members of the Mga regulon (Roberts and Scott, 2007).

The first genome-wide and comprehensive approach to identify novel sRNAs in the M1T1 MGAS2221 *S. pyogenes* strain, using a combination of bioinformatics and intergenic region tiling arrays, was most recently performed by Sumby and colleagues (Perez *et al.*, 2009). The tiling array approach identified 40 candidate sRNAs, of which only seven were also identified by the bioinformatics approach of Livny and co-workers. This small number of cumulatively identified sRNAs stresses the necessity for multifaceted and integrated approaches. Most of the investigated sRNAs varied in their stability and inter- and/or intra-serotype-specific levels of abundance. Most interestingly, Sumby and colleagues could not confirm earlier results from Mangold and colleagues who postulated a regulatory function for sagA/Pel (Mangold *et al.*, 2004).

In summary, three candidate sRNAs were identified experimentally, 29 candidate sRNAs were identified by Livny using a bioinformatic approach (Table 1), a further four appeared from the novel MOSES sRNA software and 40 (of which only seven matched candidates from the Livny paper) were found by the first tiling array approach performed by Sumby and co-workers (Perez *et al.*, 2009).

Enterococcus faecalis

Enterococcus faecalis is a human commensal and member of the lactic acid bacteria. It can be used as a starter in food industry and for some strains, probiotic effects have been claimed (Domann *et al.*, 2007). However, *E. faecalis* is also used as an indicator of faecal contamination and represent one of the principal causes of nosocomial infections (Ogier and Serror, 2008). Despite the increasing number of infections due to *E. faecalis*, mechanisms of virulence remain poorly understood. These infections affect mainly young and immunodepressed subjects causing endocarditis, meningitis, pneumonias, peritonitis, visceral abscesses, urinary infections and septicaemias (Gilmore *et al.*, 2002). This makes *E. faecalis* an ambiguous microorganism and, consequently, the use of enterococci in the food industry has been viewed more critically (Eaton and Gasson, 2001). Thus, efforts are urgently needed in order to get a better understanding of the molecular reasons why this normally harmless bacterium can transform into a dangerous pathogen. In this context, analysis of regulation of gene expression is of prime importance.

From the 3337 predicted protein-encoding open reading frames in *E. faecalis* V583, 214 have or may have regulatory functions (Paulsen *et al.*, 2003) but only few transcriptional regulators have been studied so far. Some of them have been shown to be correlated with stress response and/or virulence such as Fsr, EtaRS, CylR, HypR, PerR or Ers (Qin *et al.*, 2001; Gilmore *et al.*, 2002; Teng *et al.*, 2002; Verneuil *et al.*, 2004; Verneuil *et al.*, 2005; Riboulet-Bisson *et al.*, 2008). However, in recent years it has been discovered that not only proteins but also small non-coding RNAs are key players in the control of bacterial gene expression. At present, only one mechanism based on two sRNAs (RNA I and RNA II), required for the stable inheritance of the plasmid pAD1, has been reported for *E. faecalis* (Weaver, 2007). The two sRNAs are convergently transcribed towards a bidirectional intrinsic terminator. RNA I encodes the Fst toxin the translation of which is inhibited by the interaction with RNA II. RNA II acts then as the antitoxin and is less stable than RNA I. This suggests that in absence of continued transcription from the resident plasmid, RNA I is removed from the complex allowing the translation of *fst* that kills the cell. One recent study shows that the stability of the toxin-encoding RNA (RNA I) is due to an intramolecular helix sequestering the 5′ end of the RNA (Shokeen *et al.*, 2009).

Recently, an *in silico* study led to the prediction and annotation of 17 further putative sRNA-encoding genes in *E. faecalis* V583 (Livny *et al.*, 2008), 14 on the chromosome and three on plasmids (Table 1). However, in comparison with the number of sRNA identified in the same study in model bacteria such as *E. coli* and *Bacillus subtilis*, the number of potential candidates in *E. faecalis* is roughly 10-fold lower.

Clostridium difficile

Clostridium difficile is currently one of the increasingly important nosocomial pathogens, but also rising are the numbers of human community-associated infections and animal infections (Rupnik *et al.*, 2009). The disease spectrum ranges from mild diarrhoea to severe forms of colitis, pseudomembranous colitis and bowel perforation. Most of the symptoms can be explained by effects of two large toxins produced by the bacterium, toxin A (TcdA) and toxin

© 2010 The Authors
Journal compilation © 2010 Society for Applied Microbiology and Blackwell Publishing Ltd, *Microbial Biotechnology*

B (TcdB). While both toxins, TcdA and TcdB, are well characterized from molecular and biochemical perspective, our understanding on the regulation mechanisms involved in their expression is only partially elucidated (Dupuy *et al.*, 2008). Even less is known about regulation of additional factors influencing the virulence such as binary toxin CDT, adhesion molecules, sporulation properties and antibiotic resistance.

The regulatory role of small non-coding RNAs was so far not studied or reported in *C. difficile*. Only two candidate loci (Table 1) were found in a large computational study by Livny and colleagues (2008). In the same study candidate sRNAs were detected in some other pathogenic and non-pathogenic *Clostridia*, e.g. *Clostridium tetani*, *Clostridium botulinum*, *Clostridium perfringens*, *Clostridium novyi* (Table 1). None of them is a closer relative to *C. difficile*; however, *C. novyi* has a toxin TcnA that is homologous to *C. difficile* large toxins.

So far, *C. perfringens* is the only pathogenic clostridial species where the role of sRNA was studied in detail (Shimizu *et al.*, 2002). VR-RNA is a region expressed under the control of the VirR/VirS system. It encodes an RNA molecule involved in regulation of some *C. perfringens* toxins and other virulence factors. This strongly suggests that regulatory mechanism involving sRNA could also be present in other toxigenic *Clostridia* and justifies an extensive search for and characterization of novel sRNAs in *C. difficile*.

Facultative requirement of Hfq for sRNA–mRNA interaction in Gram-positive bacteria

The bacterial RNA chaperone Hfq facilitates pairing interactions between sRNAs and their mRNA targets in several bacteria. However, approximately only half of the sequenced bacterial genomes encode a Hfq homologue (Sun *et al.*, 2002). Hfq modulates the stability, translation and polyadenylation of many mRNAs in *E. coli* (Folichon *et al.*, 2003). In *L. monocytogenes*, Hfq is required for several sRNAs (LhrA–C) (Christiansen *et al.*, 2006), whereas other identified sRNAs in the same bacterium are able to interact with their mRNA without the contribution of Hfq (Mandin *et al.*, 2007; Toledo-Arana *et al.*, 2009). Similar examples from *Vibrio cholerae* (Qrr1–4) illustrate how heterogeneous the role of Hfq can be in sRNA–mRNA-mediated regulations within the same bacterium (Lenz *et al.*, 2004).

Although present in *S. aureus*, Hfq seems not to be required for sRNA–mRNA interactions (Bohn *et al.*, 2007). For the sporogenic bacterium *C. difficile* less is known about the involvement of the clostridial Hfq homologue in the regulation of sRNAs.

Surprisingly, a gene encoding a homologue of the protein Hfq is not present in the in *S. pyogenes* and *E. faecalis* genome sequence (Sun *et al.*, 2002). From this finding arose the question of how sRNAs are stabilized and annealed to their targets in this species? It is unlikely that non-coding RNAs are 'naked' in a cellular environment and are more likely to be associated with RNA-binding proteins throughout its lifetime (Vogel, 2009). Thus, an important step for the molecular characterization of sRNAs and identification of their targets will also need the discovery of these RNA-binding proteins in *S. pyogenes* and *E. faecalis*.

Whole-genome sRNA identification methods

Small RNAs have been hard to detect both computationally and experimentally. The first sRNAs were discovered four decades ago serendipitously by radiolabelling of total RNA and subsequent isolation via gel separation (Hindley, 1967; Griffin, 1971; Ikemura and Dahlberg, 1973). The sRNA discovery in bacteria over the last decade has shifted from the direct labelling and sequencing to genome-wide technologies including, for example, the genomic SELEX approach (Lorenz *et al.*, 2006), collecting of sRNA genes by shotgun cloning (RNomics) (Vogel *et al.*, 2003), detection of sRNAs by DNA microarrays (Selinger *et al.*, 2000) and next-generation sequencing (NGS) technologies to provide new means to sRNA discovery (Meyers *et al.*, 2006).

DNA microarrays are a powerful tool for screening sRNA genes on a genome-wide scale. Employing the so-called tiling array approach for sRNA gene discovery, probes specific to the region of interest have to be designed. These regions can span the whole genome or just the intergenic regions, where most of the currently known bacterial sRNAs are encoded. Critical design parameters are probe length and their overlap. As the probe sequence is defined by the sequence of the genome, options for probe optimization are limited. By varying the probe lengths, minor adjustments of melting temperatures can be obtained. Also incorporation of modified nucleotides, such as locked nucleic acids (LNAs), might result in more uniform melting temperatures. In any case, sensitive labelling and optimized hybridization conditions are essential. Bacterial sRNAs are highly heterogeneous in length (ranging from approximately 35 to several hundred nucleotides) and do not contain a common sequence such as the polyA tail of eukaryotic mRNAs. Therefore, careful size selection ensures that the complexity of the input RNA sample is reduced and excludes RNA species (e.g. mRNA, rRNA) for downstream microarray analysis. As amplification methods tend to be biased, direct labelling is preferred, even though lower signal intensities are obtained. Due to the fact that only a small number of sRNA genes (~20–150) are expected to be discovered from bacteria, this results

© 2010 The Authors
Journal compilation © 2010 Society for Applied Microbiology and Blackwell Publishing Ltd, *Microbial Biotechnology*

in a relatively low number of signals on the tiling microarrays. Therefore, it is vitally important to provide a range of controls upon design of the microarrays to discriminate between signals arising from other RNA species (e.g. mRNA, rRNA, tRNA) or truncated RNA fragments and existing sRNA genes.

Such tiling experiments can also be performed in more than one round, e.g. by running a first 'rough' mapping approach using long probes (around 50-mer) and then taking promising regions for a second 'fine-mapping' with shorter probes (around 25-mer) to narrow down the 5′ and 3′ ends of the sRNA genes. Although such a second round of screening increases confidence, results always have to be verified by independent techniques, such as real-time PCR, NGS, RACE-PCR, parallelized gene knock-outs or microtitre plate-based phenotype screens (Bochner, 2003).

Microarrays for such tiling experiments are usually not available off the shelf, but need to be individually designed. Systems such as the flexible Febit® microarray technology (Baum *et al.*, 2003) however provide cost-efficient and easy-to-use solutions for sRNA genes discovery employing the aforementioned tiling approach.

Next-generation sequencing technology provides new means in sRNA discovery and gives access to a screenshot of the global transcriptional landscape without complicated cloning (Sittka *et al.*, 2008; Simon *et al.*, 2009). Methods like sRNA sequencing allow for interrogation of the entire small, non-coding RNA (sRNA) repertoire. It includes a novel treatment that depletes total RNA fractions of highly abundant tRNAs and small subunit rRNA, thereby enriching the starting pool for sRNA transcripts with novel functionality (Liu *et al.*, 2009). An additional advantage of the NGS technology is the immediate information about 5′ and 3′ ends of the sRNAs. However, data analysis is still in its infancy and might make data interpretation difficult and the library preparation necessary for NGS might introduce bias. Thus, the combination of both array-based and NGS approaches appears to be the most promising experimental procedure for genome-wide identification of novel sRNAs in bacteria.

Computational approaches on genome-wide sRNA prediction and target identification

Despite the new era of NGS technologies, the task to search for and predict candidate sRNA genes on genome level remains a challenge and an open problem. Over the last years tremendous progress has been made in this field and sRNA gene prediction has been implemented based on genome scanners or on classifiers. Sequence-based classifiers use input information such as sequence, motif or alignment to assign a homology-based label to the output. Genome scanning uses combinations of assumptions and a sliding window to analyse entire genome sequences.

These genome scanning and *ab initio* methods to predict new sRNA candidates are based on the assumptions that structured RNAs have a lower free energy than random sequences (Le *et al.*, 1988). Rivas and Eddy (2000) demonstrated that minimum free energy (MFE) difference is not significant enough to be used as a general sRNA discriminator. Taken the fact that MFE calculation is based on base stacking, not only the nucleotide frequency but the di-nucleotide frequency is of interest (Bonnet *et al.*, 2004; Clote, 2005). A thermodynamic calculation to predict the MFE is highly affected by an additional or missing nucleotide related to the window size applied for scanning the genome. Combining all the information obtained from all individual windows decreases the problem (Bernhart *et al.*, 2006).

A common way to analyse nucleotide sequences is in terms of its nucleotide composition. This ranges from mononucleotide, di-nucleotide to *n*-nucleotide analyses and leads to their frequencies. Applying compositional statistics to predict new candidates was driven by reports that sRNAs have an average of 50% GC content (Rivas and Eddy, 2000). This has been successfully tested when searching GC-rich islands in some AT-rich organisms (Klein *et al.*, 2002) and applying di-nucleotide statistics in other organisms (Schattner, 2002). Also machine learning techniques have been performed to classify a sequence either as a coding or non-coding RNA sequence (Liu *et al.*, 2006).

Comparative methods use genome sequences of two or more related species performing alignments in order to find maximum regions of similarity between them (Rivas and Eddy, 2001; di Bernardo *et al.*, 2003; Coventry *et al.*, 2004; Rivas, 2005; Washietl *et al.*, 2005). Once the regions of similarity are identified, either thermodynamic information or covariate analysis can provide more evidence on sRNA assignment (Washietl *et al.*, 2005; Uzilov *et al.*, 2006).

Since the first experimentally identified sRNA in 1981 (Stougaard *et al.*, 1981; Tomizawa *et al.*, 1981), sRNA gene-finders based on a set of known sequences have been applied to identify genes belonging to the same gene family. Rfam is a collection of multiple sequence alignments, consensus secondary structures and covariance models representing predicted sRNA families. Starting with 25 sRNA gene families in 2003, the Rfam 9.1 release covers 1372 families in January 2009 (Griffiths-Jones *et al.*, 2003; Gardner *et al.*, 2009).

SIPHT is a high-throughput computational tool focusing on genome-wide bacterial sRNA prediction and annotation (Livny *et al.*, 2008). Candidate sRNA-encoding loci are identified based on the presence of putative Rho-independent terminators downstream of conserved intergenic sequences and each locus is annotated for several

© 2010 The Authors
Journal compilation © 2010 Society for Applied Microbiology and Blackwell Publishing Ltd, *Microbial Biotechnology*

features, including conservation in other species, association with one of several transcription factor binding sites and homology to any of over 300 previously identified sRNAs and *cis* regulatory RNA elements. The procedure is sufficiently automated to be applied to all available bacterial genomes; however, the requirement for a terminator motif introduces a bias against the significant fraction of RNA elements that are not followed by a detectable terminator.

Phylogenetic analysis is one of the most informative computational methods for the annotation of genes and identification of evolutionary modules of functionally related genes and sequences (Pellegrini *et al.*, 1999). Nucleic acid phylogenetic profiling is successfully applied to protein annotation and has recently been applied for high-throughput sRNA identification and their functional characterization to *S. aureus* (Marchais *et al.*, 2009).

Given the scores of sRNAs of unknown function, the identification of their cellular targets has become essential and requires both experimental technologies and computational approaches (Vogel and Wagner, 2007). Mandin and colleagues described such a tool that has been applied successfully in the genome of *L. monocytogenes*, but it is not available as a webserver (Mandin *et al.*, 2007). TargetRNA is a web tool which uses a dynamic programming algorithm to search each annotated transcript to generate base-pair-binding potential of the input sRNA sequence and rank list of candidate genes. Many existing target prediction programs except IntaRNA neglect the accessibility of target sites and intra-molecular base pairs using free energy of the hybridized duplex to predict the potential target site (Busch *et al.*, 2008).

There is no general toolset that provides a comprehensive solution to the prediction of sRNA. A good practice when starting with computational tools is to carefully choose a set of applicable methods with different approaches and compare the results. An example for such an integrative prediction framework and modular sequence suite is MOSES (P. Raasch, U. Schmitz, N. Patenge, B. Kreikemeyer and O. Wolkenhauer, submitted). Finally, we reiterate that while all of the outlined tools predict sRNAs, the planning of experiments and integration of predictions and experimental data is needed for a reliable validation of *in silico* sRNA and target prediction.

In order to streamline and support the iterative processes, the consortium uses a central data management solution for data integration provided by Genedata. The infrastructure platform, based on Genedata Phylosopher® for sharing and interpreting the experimental data, covers the required comparative genomics and cross-technology approach. Furthermore, the consortium reconstructs genome-wide regulatory networks and pathogenesis to identify and validate targets for diagnostics and novel anti-infective therapies within the Genedata platform.

Antisense reagents as an alternative treatment of microbial infections

The growing knowledge on sRNAs in bacteria, their regulatory role and function and their possible involvement in pathogen infection makes this class of molecules extremely interesting as new targets for antisense therapy and drug development.

Antisense therapy is a form of treatment where genes which are known to be causative for a particular disease are inactivated ('turned off') by a small synthetic oligonucleotide analogue (RNA or DNA, 14–25 nt in length). This analogue is designed to bind either to the DNA of the gene and inhibit transcription or to the respective mRNA produced by this gene, thereby inhibiting translation. For eukaryotes, the strand might also be targeted to bind a splicing site on pre-mRNA (Morcos, 2007). Recently, miRNAs with their regulatory role have gained interest as targets, too (Fabani and Gait, 2008). Here, in most cases the oligonucleotide is designed to inhibit micro-RNA function essentially by a steric block, RNase H- and RISC-independent antisense mechanism through complementary binding to the micro-RNA sequence.

To date, antisense drugs are studied to treat diseases such as diabetes, different types of cancer, amyothrophic lateral sclerosis, Duchenne muscular dystrophy as well as asthma and arthritis.

Apart from finding suitable and promising targets, the design of such drugs themselves represents another challenge. Issues like *in vivo* stability and efficient cell delivery have high impact on the dose–response relationship and need to be addressed. In order to avoid enzymatic *in vivo* degradation, modified oligonucleotides or completely synthetic analogues are used. For the antiviral Fomivirsen (marketed as Vitravene) (Roush, 1997; Mulamba *et al.*, 1998), for example, which is the first antisense reagent approved by the FDA (US Food and Drug Administration), a 21-mer oligonucleotide with phosphorothioate linkages is used. Phosphorothioates are also employed for a few other promising antisense reagents, some of which already entered preclinical trials, such as AP12009 (Hau *et al.*, 2007). Nevertheless, this class of compounds exhibits several disadvantages: they have a comparable low binding affinity to complementary nucleic acids and show non-specific binding to proteins (Brown *et al.*, 1994; Guvakova *et al.*, 1995), causing toxic side-effects that limit many applications. The toxicity is reduced but not absent in second generation antisense agents with mixed backbone oligonucleotides.

Meeting the challenge to find modifications that provide efficient and specific antisense activity *in vivo* without being toxic, a third generation of antisense agents have emerged (Fig. 1). Analogues such as LNAs, phosphorodiamidate morpholino oligomers (PMOs) and PNAs are

© 2010 The Authors
Journal compilation © 2010 Society for Applied Microbiology and Blackwell Publishing Ltd, *Microbial Biotechnology*

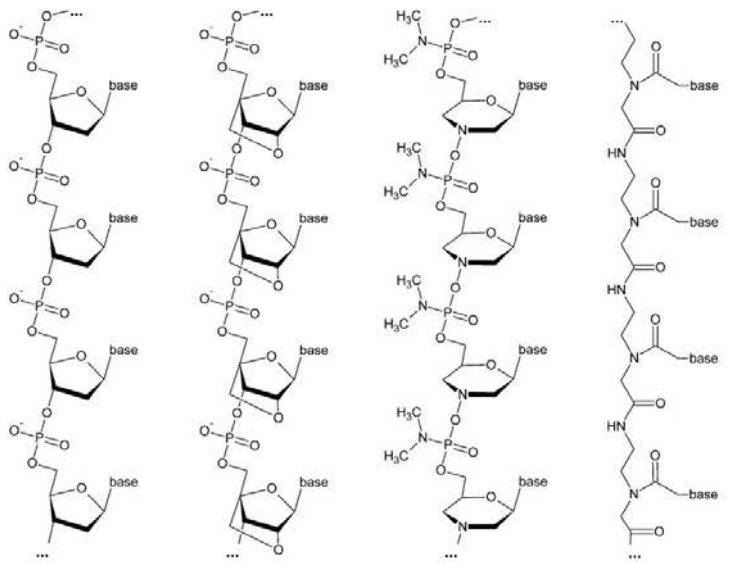


Fig. 1. Overview of chemical structures of DNA, LNA (locked nucleic acid), PMO (phosphorodiamidate morpholino oligomer) and PNA (peptide nucleic acid).

altogether no substrates for enzymatic degradation and hybridize with exceptional affinity and target specificity, thereby forming duplexes which are even more stable than the respective RNA:RNA duplexes themselves. A short general overview on PNAs, LNAs and morpholinos was published recently (Karkare and Bhatnagar, 2006).

As shown in Fig. 1, LNAs are RNA nucleotide analogues in which the ribose ring is constrained by a methylene linkage between the 2′-oxygen and the 4′-carbon, resulting in the locked C3′-*endo* conformation. The introduction of LNA residues in oligonucleotides stabilizes the LNA:RNA duplex by either pre-organization or increased base stacking (increasing T_m by 1–8°C against DNA and 2–10°C against RNA by one LNA substitution in an oligonucleotide). This results in increased RNA target accessibility and higher selectivity. In general, for efficient gene silencing *in vitro* and *in vivo*, fully modified or chimeric LNA oligonucleotides (LNA mixmers) have been applied.

Locked nucleic acids and LNA mixmers can be transfected using standard transfection reagents based on ionic interactions of the negatively charged backbone. However, due to the very strong base pairing in LNA:LNA duplexes it is important to design LNAs without extensive self-complementary sequences or to apply chimeric LNAs (LNA mixmers), thereby diminishing biological stability to some degree. Furthermore, unless very short probes are required, stretches of more than four LNAs should be avoided.

Contrary to LNAs, PMOs and PNAs have a neutral backbone and thereby different physiochemical properties. Here, high specificity and increased duplex stability with RNAs and DNAs are due to the missing interstrand repulsion between the neutral PMOs (or PNAs) and the negatively charged RNA or DNA target. Thus, hybridization is nearly independent of salt concentration because no counter-ions are needed to stabilize the duplex.

Phosphorodiamidate morpholino oligomers represent the closest relatives to the previously mentioned first-generation phosphorothioates. But instead of the negatively charged sulfur, they comprise a neutral dimethylamino-group bound to the phosphor. In addition to their excellent base stacking, standard PMOs are highly soluble in water, but their thermodynamic properties are not well defined, yet.

For PNAs it is the other way round, thermodynamics are well defined, but water solubility is limited, at least for longer purine-rich oligomers which tend to aggregate. However, water solubility can be adjusted by introduction of one or more lysines which can be easily employed, as PNA synthesis strictly follows Fmoc-peptide synthesis. PNA is a fully synthetic DNA analogue, which consists of the four nucleobases attached to a neutral peptide-like amide backbone (Fig. 1). Despite the totally different backbone, PNA hybridizes with very high specificity to RNA and DNA via Watson-Crick as well as Hoogsteen base pairing. *In vitro* studies indicate that PNAs could

© 2010 The Authors
Journal compilation © 2010 Society for Applied Microbiology and Blackwell Publishing Ltd, *Microbial Biotechnology*

inhibit both transcription and translation of genes to which it has been targeted (Good and Nielsen, 1998; Nekhotiaeva *et al.*, 2004). Due to higher duplex stability, probes can be much shorter in length than DNA probes. Very high chemical as well as biological stability make them excellent probes for antisense drug development and diagnostics.

In summary, LNAs, PMOs and PNAs have great potential as antisense drugs, each of them having its own advantages and draw-backs. Even though there are some recent reports comparing these analogues in experiments (Gruegelsiepe *et al.*, 2006; Fabani and Gait, 2008), at the time being, the choice of analogue still depends on the experimental conditions and detailed applications.

For all analogues, cell-wall permeability is very limited, but cellular uptake can be improved by shortening the length of the antisense oligomer and/or attaching cell-penetrating peptides (CPPs) (Tan *et al.*, 2005; Gruegelsiepe *et al.*, 2006; Kurupati *et al.*, 2007) or other modifications which increase uptake (Koppelhus and Nielsen, 2003; Koppelhus *et al.*, 2008). The CPP approach is currently a major avenue in engineering delivery systems that are hoped to mediate the non-invasive import of cargos into cells. The large number of cargo molecules that have been efficiently delivered by CPPs ranges from small molecules to proteins, oligonucleotide analogues and even liposomes and particles.

To date, the most common reported CPP for bacteria is $(KFF)_3K$. It is mainly used in combination with PNAs or PMOs. These compounds show efficient inhibition of gene expression in a sequence specific and dose-dependent manner at low micromolar concentrations in different bacteria (Nekhotiaeva *et al.*, 2004; Tan *et al.*, 2005; Gruegelsiepe *et al.*, 2006; Kurupati *et al.*, 2007; Goh *et al.*, 2009).

The search of efficient CPPs, which may improve cell delivery even further, is still an emerging field of research and has high potential for optimization. As demonstrated recently by Mellbye and colleagues variations in amino acid compositions of CPPs and linker molecules show clear effects on the efficiency of cell delivery (Mellbye *et al.*, 2009). Broader systematic studies like this are still missing, but may open the road to a toolbox of different CPPs with optimized and well-characterized efficiencies to transfer drugs specifically into different cell types.

Due to similar synthesis methods, PNA–peptide conjugates can easily be produced in microwell plates in a parallel manner (Brandt *et al.*, 2003), thereby allowing for a fast optimization of the antisense as well as the cell-delivery sequences for antisense drug design. A proof of principle that these compounds can be used as therapeutic drugs against bacterial infection was demonstrated by Tan and colleagues (2005) in a mouse model. In combination with promising targets such as sRNAs, which are suggested to play a major regulatory role in bacterial pathogenesis, the development of highly efficient antisense drugs becomes feasible as an excellent alternative to classical antibiotic treatment.

sRNAs for microbial diagnosis

Newly discovered sRNA sequences that were verified to be involved in virulence mechanisms of the different Gram-positive bacteria may not only be used as potential antisense drugs but also exploited as targets for diagnostic systems.

In this respect sRNAs represent a useful alternative to mRNA or DNA oligonucleotides, because of his putative regulatory function. The potential of this class of compounds as diagnostic markers is currently a very active field of research (Hartmann & Consorten).

Several different formats for rapid diagnostic tests based on nucleic acid interactions have been described in the literature. Many of them are PCR-based with the Genexpert system produced by Cepheid (Sunnyvale, CA) being the most advanced as it integrates sample preparation, amplification and detection (Chandler *et al.*, 2001). These methods require experienced personnel and need 30–60 min for an assay to be completed. Therefore they are not suitable for point-of-care (POC) testing, the ultimate goal in the development of rapid diagnostic screening tests. An alternative and more cost-effective approach to nucleic acid testing is the use of a lateral-flow platform. Lateral-flow assays (LFAs) were originally developed as immunoassays and combine chromatographic purification with immunodetection at high speed (Fig. 2A). They are very simple in handling and represent the only true POC test up to now (Seal *et al.*, 2006).

Nucleic acid lateral flow uses nucleic acid hybridization to capture and detect the nucleic acid analyte in a manner akin to lateral-flow immunoassays. The most favourable approach is to immobilize oligonucleotide probes directly onto the nitrocellulose membrane used for the chromatographic step of the LFA. This may be done by covalent chemical bonding or by passive adsorption of an albumin–oligonucleotide conjugate to the membrane.

In principle any detection chemistry used in traditional LFA systems may also be applied for nucleic acid detection. Reverse hybridization enzymatic strip assays are the most advanced methods for the detection of nucleic acid interactions. In these tests, enzyme-labelled probes are hybridized to complementary nucleic acid target species on the surface of the nitrocellulose membrane. The result is a hapten–antibody–enzyme complex such as for example biotin–streptavidine–alkaline phosphatase. But complex wash and substrate incubation steps are necessary to develop a readable colorimetric signal. This vastly

© 2010 The Authors
Journal compilation © 2010 Society for Applied Microbiology and Blackwell Publishing Ltd, *Microbial Biotechnology*

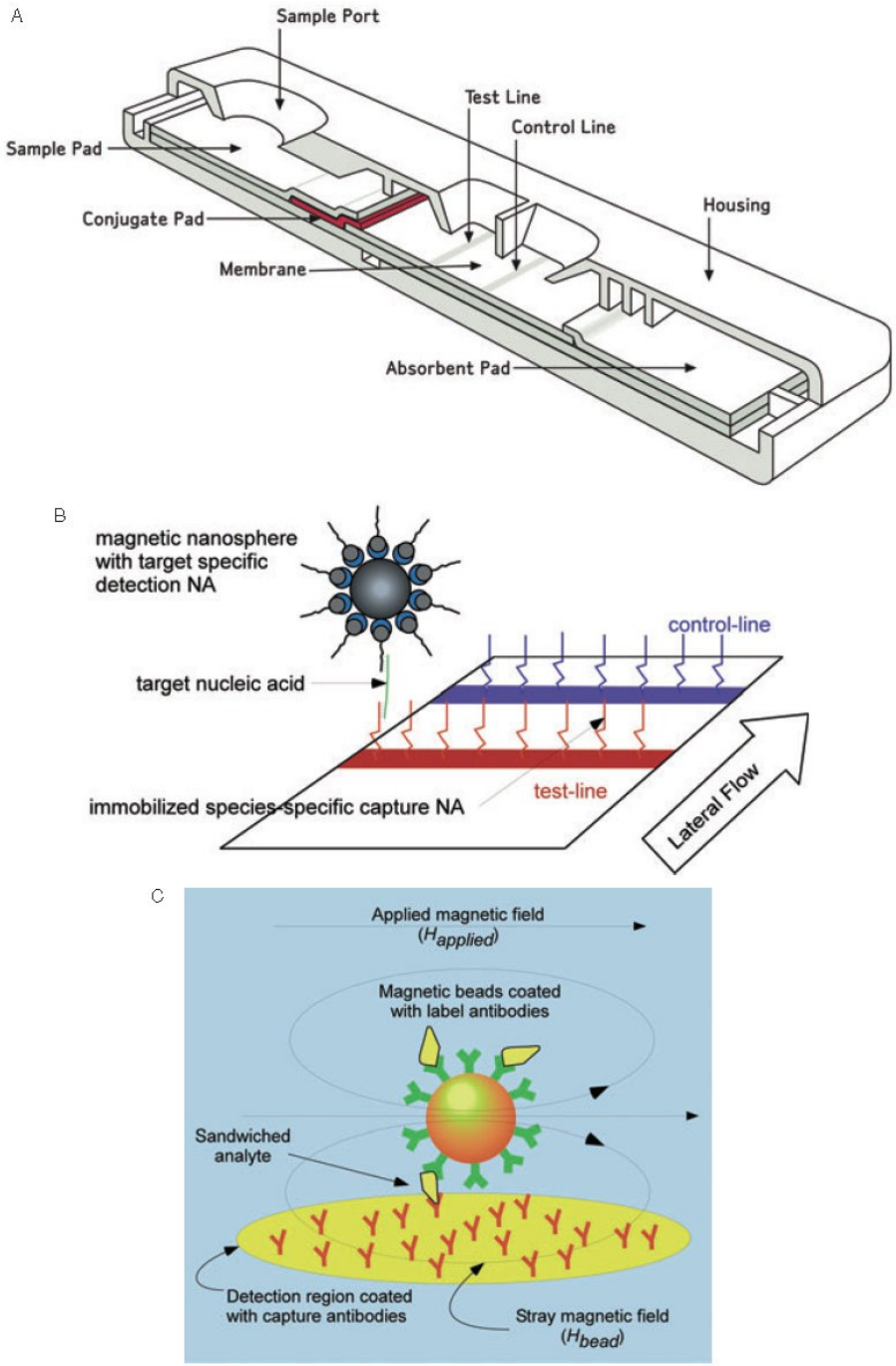


Fig. 2. A. Schematic outline of a standard lateral-flow assay device (Carney *et al.*, 2006).
B. Lateral-flow assay of sRNA captured by PNAs using superparamagnetic magnet nanoparticles.
C. Magnetoresistive detection of captured magnet nanoparticles (Tondra, 2007).

© 2010 The Authors
Journal compilation © 2010 Society for Applied Microbiology and Blackwell Publishing Ltd, *Microbial Biotechnology*

limits the speed and simplicity of the assay and makes the POC use of such a system very difficult to achieve.

As an alternative, bead technologies mainly in the form of nanoparticles were successfully applied to nucleic acid LFAs. In this context, gold nanoparticles are the beads of choice, mainly because of their small size, good visibility and robustness in manufacturing (Carney *et al.*, 2006). They can be conjugated with oligonucleotides and labelled with small binding moieties such as biotin. Gold NA-LFA systems generally use 30–80 nm particles conjugated to an anti-biotin antibody. This conjugate is complexed with a biotin-oligonucleotide detection probe and forms a visible red line on the membrane strip when captured by a complementary NA sequence. Typical detection limits of such assay formats are about 2.5 mg DNA ml^{-1} (Mao *et al.*, 2009).

A number of methods for improving the sensitivity of nanoparticle NA-LFA have been investigated. Detection probes may be labelled with multiple hapten moieties to form large signal-enhancing complexes or make use of DNA dendrimeres, which may result in a detection limit up to 200-fold lower compared with standard assay conditions (Dineva *et al.*, 2005). Other approaches use fluorescence labels to achieve a better detection. For example, a dual fluorescein-biotin-labelled oligonucleotide probe has been used to detect single-stranded nucleic acids. However the utility of common fluorophores is limited by high background and the need for a complex reader. These problems may be overcome by the application of quantum dots (Lambert and Fischer, 2005). They consist of nanometre-sized semiconductor nanocrystals and have superior fluorescence properties that enable them to emit about 1000-fold brighter light than conventional dyes. They have narrow, symmetrical emission profiles, but have to be excited with UV light. Especially the development of water-soluble quantum dots has enabled their use in LFAs. However, the price of these fluorescent nanocrystals is quite high, so that no commercial application has been developed so far.

In summary, specificity, sensitivity and cost-effectiveness are key issues in the development of new nucleic acid-based diagnostic systems.

To overcome these problems, we use PNA as capture molecules to increase the specificity of the interaction. The peptide backbone of the nucleic acid derivatives makes them much more stable compared with RNA or DNA (Fig. 1). Therefore they are much easier to handle, a great advantage for the development of an appropriate assay as well as for the shelf half life of the reagents in such a diagnostic kit. Equally important, the loss of the negative charge in the phosphate diester of the standard nucleic acid in PNAs led to a higher binding energy, mainly because electrostatic repulsion between the two strands does not exist in PNA/NA duplexes. This effect largely increases the specificity of the interaction. In fact, the melting temperature of a PNA/DNA duplex was measured to be about 15°C higher compared with an isosequential DNA/DNA duplex (Sen and Nielsen, 2007).

To increase the sensitivity in the detection of the sRNAs captured by the complementary PNA, we use superparamagnetic nanoparticles combined with a suitable magnetoelectronic device (Fig. 2B). This set-up offers a number of advantages for LFA-based POC diagnostic measurements. A comprehensive review on magnetic labelling, detection and system integration has appeared recently (Tamanaha *et al.*, 2008). A huge benefit of magnetic labelling and detection over fluorescence labelling and optical detection is that there is usually very little magnetic background present in most samples, thereby making single magnetic particle detection both possible and technically feasible (Perez *et al.*, 2002). Nanoscale magnetic beads have been extensively developed over the last decade and are commercially available from a number of sources and in different varieties of sizes and surface chemistry. A prime requirement for use as diagnostic labels is that beads must be paramagnetic or non-remanent to avoid clustering caused by residual magnetic moment in the absence of magnetic fields. Magnetic nanoparticles used in LFA-based diagnostic devices are preferred to be single domain and small enough to be superparamagnetic, i.e. the thermal fluctuations at room temperature overcome magnetic anisotropy forces spontaneously randomizing magnetic directions. The beads consist of iron oxide cores coated by organic polymers.

Magnetic detection can be established with a number of different sensor types. The simplest devices are based on Maxwell-Wien bridges or frequency dependent magnetometers made of LC circuits. With these devices, to our best knowledge, single magnetic particle detection has not yet been accomplished. This goal has however been reached with devices based on superconducting quantum interference devices (SQUID), magnetoresistance or Hall sensors. For an excellent review on magnetic sensing see for example Tamanaha and colleagues (2008). SQUID-based sensors, although being the most sensitive ones, are not apt for POC applications because they need extensive cooling down to at least N_2 temperatures (77 K).

A schematic illustration of the concept of magnetic detection in traditional lateral-flow immunoassays is shown in Fig. 2C. Magnetic beads coated with antibodies bind the analyte and were captured by an appropriate antibody immobilized on a magnetoresistive detector. This way magnetic beads loaded with analyte accumulate on the sensor area. Stray magnetic fields from the beads change the electric resistance of the sensor. This change is used to determine the amount of analyte present in the sample. This detection principle is adapted to PNA as capture molecule and sRNA analyte without great difficulties.

© 2010 The Authors
Journal compilation © 2010 Society for Applied Microbiology and Blackwell Publishing Ltd, *Microbial Biotechnology*

Table 2. Comparative genome analysis of 21 of 100 published listerial sRNAs having orthologues (Toledo-Arana *et al.*, 2009) in *Staphylococcus aureus*, *Streptococcus pyogenes*, *Enterococcus faecalis* and *Clostridium difficile*.

Regulatory RNA	Length (bp)	Position in *Listeria monocytogenes* EGD-e (bp)	*Clostridium difficile* 630	*Enterococcus faecalis* V583	*Staphylococcus aureus* COL	*Streptococcus pyogenes* M1 GAS
FMN	123	2020609–2020487	+ (2)	+	+ (2)	+
Glycine	92	1372840–1372931	+		+	+
L13	53	2685233–2685181		+	+	
L19	39	1862308–1862270	+ (4)	+	+ (9)	+
L21	79	1577146–1577068		+		
M-box	165	2766104–2765940		+ (5)		
PreQ1	48	907926–907973		+		
SAM2	120	309385–309266	+ (5)			
SAM3	101	637926–637826	+ (4)			
SAM5	109	1716651–1716543	+		+	
SAM6	107	1739597–1739491	+ (5)			
SAM7	119	2491176–2491058	+			
SRP	333	2784604–2784272	+		+	
glmS	195	756458–756652			+	
rli31	115	597812–597926	+			
rli36	84	859527–859444	+			
rli45	78	2154775–2154852		+	+	
rli50	177	2783274–2783098			+	
rliD	328	1359529–1359202		+	+	
rnpB	385	1965188–1961804		+	+	+
ssrA	367	2510220–2509854		+		

Comparative analysis of five genomes: *L. monocytogenes* EGD-e, *C. difficile* 630, *E. faecalis* V583, *S. aureus* COL and *S. pyogenes* M1 GAS.
Genome sequences were downloaded from NCBI http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi.
One hundred sRNAs from *L. monocytogenes* EGD-e (Toledo-Arana *et al.*, 2009) were used as reference and compared against the other four genomes using BLAST.
Default settings for BLAST were used except reward match penalty (flag-r), which was set to 2 to accommodate the distant relation among the genomes.
The BLAST results were filtered by identity = 30% and query coverage = 50–150%.
Query coverage is defined as the percentage of the sRNA that is covered by the alignment [(alignment length * 100)/query length].
The number of hits is depicted in brackets.

Conclusion

The current knowledge about sRNAs of *S. aureus*, *S. pyogenes*, *E. faecalis*, *C. difficile* and *L. monocytogenes* represents only the blueprint for the prediction, detection and characterization of this group of RNAs and will be extended by the inclusion of additional strains of these five Gram-positive genomes. Currently (October 2009) genome sequences of 39 *S. aureus*, 14 *S. pyogenes*, 23 *E. faecalis*, 12 *C. difficile* and 24 *L. monocytogenes* strains have become available in the NCBI databases, and more serotype- and strain-specific genome sequences will be analysed in future studies.

In silico prediction of sRNA showed that the genus of *Listeria* bears a high number of putative sRNA (Livny *et al.*, 2008) compared with other genera such as *Streptococcus*, *Staphylococcus*, *Enterococcus* and *Clostridium* (Table 1), which might reflect their potential ubiquitous adaption ability in nature and mammals. Comparative analyses of the recently reported 103 regulatory RNAs of *L. monocytogenes* among the genomes of *S. aureus*, *S. pyogenes*, *E. faecalis* and *C. difficile* revealed that riboswitches seem to be more conserved among these Gram-positive pathogens than sRNA (Table 2). Thus it could be speculated that a common ancient mechanism of *cis* acting RNA regulation might exist in Gram-positive bacteria, while *trans* acting RNAs might be involved in speciation of bacteria in response to their host adaptation processes.

Combating of emerging antibiotic-resistant microbial strains remains an urgent and continuous task worldwide. Thus the search for novel sRNAs on a global scale is another facet for a better and systematic understanding of the pathogenesis of Gram-positive bacteria and for the identification of potential novel drug targets and intervention strategies. Newly identified sRNAs are also potential markers and will be used to establish diagnostic tests that are fast, sensitive and suitable for POC applications at low cost.

The most urgent questions for the future will be: (i) what is the mode of action of many of the candidate sRNAs, (ii) how many of them are relevant for pathogenesis and regulatory networks, and most importantly (iii) are some of them targets for novel anti-infective therapy by PNA technology and (iv) can they be used for the development of novel diagnostic systems.

Acknowledgements

The authors want to thank Alexandra Amend for excellent technical assistance and Deepak Rawool for critical reading

© 2010 The Authors
Journal compilation © 2010 Society for Applied Microbiology and Blackwell Publishing Ltd, *Microbial Biotechnology*

of the manuscript. The work was supported by grants from the Bundesministerium für Bildung und Forschung (BMBF), the Agence National de Recherche (ANR) and the ERA-NET Pathogenomics Network to the sRNAomics project.

References

Barry, T., Kelly, M., Glynn, B., and Peden, J. (1999) Molecular cloning and phylogenetic analysis of the small cytoplasmic RNA from *Listeria monocytogenes*. *FEMS Microbiol Lett* **173:** 47–53.

Baum, M., Bielau, S., Rittner, N., Schmid, K., Eggelbusch, K., Dahms, M., *et al.* (2003) Validation of a novel, fully integrated and flexible microarray benchtop facility for gene expression profiling. *Nucleic Acids Res* **31:** e151.

Bernhart, S.H., Hofacker, I.L., and Stadler, P.F. (2006) Local RNA base pairing probabilities in large sequences. *Bioinformatics* **22:** 614–615.

Beyer-Sehlmeyer, G., Kreikemeyer, B., Horster, A., and Podbielski, A. (2005) Analysis of the growth phase-associated transcriptome of *Streptococcus pyogenes*. *Int J Med Microbiol* **295:** 161–177.

Bochner, B.R. (2003) New technologies to assess genotype–phenotype relationships. *Nat Rev Genet* **4:** 309–314.

Bohn, C., Rigoulay, C., and Bouloc, P. (2007) No detectable effect of RNA-binding protein Hfq absence in *Staphylococcus aureus*. *BMC Microbiol* **7:** 10.

Boisset, S., Geissmann, T., Huntzinger, E., Fechter, P., Bendridi, N., Possedko, M., *et al.* (2007) *Staphylococcus aureus* RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. *Genes Dev* **21:** 1353–1366.

Bonnet, E., Wuyts, J., Rouze, P., and Van de, P.Y. (2004) Evidence that microRNA precursors, unlike other noncoding RNAs, have lower folding free energies than random sequences. *Bioinformatics* **20:** 2911–2917.

Brandt, O., Feldner, J., Stephan, A., Schroder, M., Schnolzer, M., Arlinghaus, H.F., *et al.* (2003) PNA microarrays for hybridisation of unlabelled DNA samples. *Nucleic Acids Res* **31:** e119.

Brown, D.A., Kang, S.H., Gryaznov, S.M., DeDionisio, L., Heidenreich, O., Sullivan, S., *et al.* (1994) Effect of phosphorothioate modification of oligodeoxynucleotides on specific protein binding. *J Biol Chem* **269:** 26801–26805.

Busch, A., Richter, A.S., and Backofen, R. (2008) IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics* **24:** 2849–2856.

Camejo, A., Buchrieser, C., Couve, E., Carvalho, F., Reis, O., Ferreira, P., *et al.* (2009) *In vivo* transcriptional profiling of *Listeria monocytogenes* and mutagenesis identify new virulence factors involved in infection. *PLoS Pathog* **5:** e1000449.

Carapetis, J.R., Steer, A.C., Mulholland, E.K., and Weber, M. (2005) The global burden of group A streptococcal diseases. *Lancet Infect Dis* **5:** 685–694.

Carney, J., Braven, H., Seal, J., and Whitworth, E. (2006) Present and future applications of gold in rapid assays. *IVD Technol* **12:** 41–50.

Carr, A., Sledjeski, D.D., Podbielski, A., Boyle, M.D., and Kreikemeyer, B. (2001) Similarities between complement-mediated and streptolysin S-mediated hemolysis. *J Biol Chem* **276:** 41790–41796.

Chandler, D.P., Brown, J., Bruckner-Lea, C.J., Olson, L., Posakony, G.J., Stults, J.R., *et al.* (2001) Continuous spore disruption using radially focused, high-frequency ultrasound. *Anal Chem* **73:** 3784–3789.

Chatterjee, S.S., Hossain, H., Otten, S., Kuenne, C., Kuchmina, K., Machata, S., *et al.* (2006) Intracellular gene expression profile of *Listeria monocytogenes*. *Infect Immun* **74:** 1323–1338.

Christiansen, J.K., Larsen, M.H., Ingmer, H., Sogaard-Andersen, L., and Kallipolitis, B.H. (2004) The RNA-binding protein Hfq of *Listeria monocytogenes*: role in stress tolerance and virulence. *J Bacteriol* **186:** 3355–3362.

Christiansen, J.K., Nielsen, J.S., Ebersbach, T., Valentin-Hansen, P., Sogaard-Andersen, L., and Kallipolitis, B.H. (2006) Identification of small Hfq-binding RNAs in *Listeria monocytogenes*. *RNA* **12:** 1383–1396.

Clote, P. (2005) An efficient algorithm to compute the landscape of locally optimal RNA secondary structures with respect to the Nussinov-Jacobson energy model. *J Comput Biol* **12:** 83–101.

Cossart, P., and Toledo-Arana, A. (2008) *Listeria monocytogenes*, a unique model in infection biology: an overview. *Microbes Infect* **10:** 1041–1050.

Courtney, H.S., Hasty, D.L., and Dale, J.B. (2002) Molecular mechanisms of adhesion, colonization, and invasion of group A streptococci. *Ann Med* **34:** 77–87.

Coventry, A., Kleitman, D.J., and Berger, B. (2004) MSARI: multiple sequence alignments for statistical detection of RNA secondary structure. *Proc Natl Acad Sci USA* **101:** 12102–12107.

Dalmay, T. (2008) MicroRNAs and cancer. *J Intern Med* **263:** 366–375.

di Bernardo, D., Down, T., and Hubbard, T. (2003) ddbRNA: detection of conserved secondary structures in multiple alignments. *Bioinformatics* **19:** 1606–1611.

Dineva, M.A., Candotti, D., Fletcher-Brown, F., Allain, J.P., and Lee, H. (2005) Simultaneous visual detection of multiple viral amplicons by dipstick assay. *J Clin Microbiol* **43:** 4015–4021.

Domann, E., Hain, T., Ghai, R., Billion, A., Kuenne, C., Zimmermann, K., and Chakraborty, T. (2007) Comparative genomic analysis for the presence of potential enterococcal virulence factors in the probiotic *Enterococcus faecalis* strain Symbioflor 1. *Int J Med Microbiol* **297:** 533–539.

Dunman, P.M., Murphy, E., Haney, S., Palacios, D., Tucker-Kellogg, G., Wu, S., *et al.* (2001) Transcription profiling-based identification of *Staphylococcus aureus* genes regulated by the *agr* and/or *sarA* loci. *J Bacteriol* **183:** 7341–7353.

Dupuy, B., Govind, R., Antunes, A., and Matamouros, S. (2008) *Clostridium difficile* toxin synthesis is negatively regulated by TcdC. *J Med Microbiol* **57:** 685–689.

Eaton, T.J., and Gasson, M.J. (2001) Molecular screening of *Enterococcus* virulence determinants and potential for genetic exchange between food and medical isolates. *Appl Environ Microbiol* **67:** 1628–1635.

© 2010 The Authors
Journal compilation © 2010 Society for Applied Microbiology and Blackwell Publishing Ltd, *Microbial Biotechnology*

Appendix

Fabani, M.M., and Gait, M.J. (2008) miR-122 targeting with LNA/2′-O-methyl oligonucleotide mixmers, peptide nucleic acids (PNA), and PNA–peptide conjugates. *RNA* **14:** 336–346.

Fierro-Monti, I.P., Reid, S.J., and Woods, D.R. (1992) Differential expression of a Clostridium acetobutylicum antisense RNA: implications for regulation of glutamine synthetase. *J Bacteriol* **174:** 7642–7647.

Folichon, M., Arluison, V., Pellegrini, O., Huntzinger, E., Regnier, P., and Hajnsdorf, E. (2003) The poly(A) binding protein Hfq protects RNA from RNase E and exoribonucleolytic degradation. *Nucleic Acids Res* **31:** 7302–7310.

Gardner, P.P., Daub, J., Tate, J.G., Nawrocki, E.P., Kolbe, D.L., Lindgreen, S., *et al.* (2009) Rfam: updates to the RNA families database. *Nucleic Acids Res* **37:** D136–D140.

Garzon, R., Calin, G.A., and Croce, C.M. (2009) MicroRNAs in cancer. *Annu Rev Med* **60:** 167–179.

Geisinger, E., Adhikari, R.P., Jin, R., Ross, H.F., and Novick, R.P. (2006) Inhibition of rot translation by RNAIII, a key feature of agr function. *Mol Microbiol* **61:** 1038–1048.

Geissmann, T., Chevalier, C., Cros, M.J., Boisset, S., Fechter, P., Noirot, C., *et al.* (2009) A search for small noncoding RNAs in *Staphylococcus aureus* reveals a conserved sequence motif for regulation. *Nucleic Acids Res* **37:** 7239–7257.

Gilmore, M.S., Coburn, P.S., Nallapareddy, R.S., and Murray, B.E. (2002) Enterococcal virulence. In *The Enterococci, Pathogenesis, Molecular Biology, and Antibiotic Resistance.* Clewell, D.B., Courvalin, P., Dunny, G.M., Murray, B.E., and Rice, L.B. (eds). Washington, DC, USA: American Society for Microbiology, pp. 301–354.

Goh, S., Boberek, J.M., Nakashima, N., Stach, J., and Good, L. (2009) Concurrent growth rate and transcript analyses reveal essential gene stringency in *Escherichia coli. PLoS ONE* **4:** e6061.

Good, L., and Nielsen, P.E. (1998) Inhibition of translation and bacterial growth by peptide nucleic acid targeted to ribosomal RNA. *Proc Natl Acad Sci USA* **95:** 2073–2076.

Graves, L.M., Helsel, L.O., Steigerwalt, A.G., Morey, R.E., Daneshvar, M.I., Roof, S.E., *et al.* (2009) *Listeria marthii* sp. nov., isolated from the natural environment, Finger Lakes National Forest. *Int J Syst Evol Microbiol* (in press): doi: ijs.0.014118-0

Griffin, B.E. (1971) Separation of 32P-labelled ribonucleic acid components. The use of polyethylenimine-cellulose (TLC) as a second dimension in separating oligoribonucleotides of '4.5 S' and 5 S from *E. coli. FEBS Lett* **15:** 165–168.

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S.R. (2003) Rfam: an RNA family database. *Nucleic Acids Res* **31:** 439–441.

Gruegelsiepe, H., Brandt, O., and Hartmann, R.K. (2006) Antisense inhibition of RNase P: mechanistic aspects and application to live bacteria. *J Biol Chem* **281:** 30613–30620.

Guvakova, M.A., Yakubov, L.A., Vlodavsky, I., Tonkinson, J.L., and Stein, C.A. (1995) Phosphorothioate oligodeoxynucleotides bind to basic fibroblast growth factor, inhibit its binding to cell surface receptors, and remove it from low affinity binding sites on extracellular matrix. *J Biol Chem* **270:** 2620–2627.

Hain, T., Chatterjee, S.S., Ghai, R., Kuenne, C.T., Billion, A., Steinweg, C., *et al.* (2007) Pathogenomics of *Listeria* spp. *Int J Med Microbiol* **297:** 541–557.

Hain, T., Hossain, H., Chatterjee, S.S., Machata, S., Volk, U., Wagner, S., *et al.* (2008) Temporal transcriptomic analysis of the *Listeria monocytogenes* EGD-e sigmaB regulon. *BMC Microbiol* **8:** 20.

Halfmann, A., Kovacs, M., Hakenbeck, R., and Bruckner, R. (2007) Identification of the genes directly controlled by the response regulator CiaR in *Streptococcus pneumoniae*: five out of 15 promoters drive expression of small noncoding RNAs. *Mol Microbiol* **66:** 110–126.

Hau, P., Jachimczak, P., Schlingensiepen, R., Schulmeyer, F., Jauch, T., Steinbrecher, A., *et al.* (2007) Inhibition of TGF-beta2 with AP 12009 in recurrent malignant gliomas: from preclinical to phase I/II studies. *Oligonucleotides* **17:** 201–212.

Hindley, J. (1967) Fractionation of 32P-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting. *J Mol Biol* **30:** 125–136.

Hof, H., Szabo, K., and Becker, B. (2007) [Epidemiology of listeriosis in Germany: a changing but ignored pattern]. *Dtsch Med Wochenschr* **132:** 1343–1348.

Huntzinger, E., Boisset, S., Saveanu, C., Benito, Y., Geissmann, T., Namane, A., *et al.* (2005) *Staphylococcus aureus* RNAIII and the endoribonuclease III coordinately regulate spa gene expression. *EMBO J* **24:** 824–835.

Ikemura, T., and Dahlberg, J.E. (1973) Small ribonucleic acids of *Escherichia coli.* I. Characterization by polyacrylamide gel electrophoresis and fingerprint analysis. *J Biol Chem* **248:** 5024–5032.

Ji, G., Beavis, R.C., and Novick, R.P. (1995) Cell density control of staphylococcal virulence mediated by an octapeptide pheromone. *Proc Natl Acad Sci USA* **92:** 12055–12059.

Johansson, J., Mandin, P., Renzoni, A., Chiaruttini, C., Springer, M., and Cossart, P. (2002) An RNA thermosensor controls expression of virulence genes in *Listeria monocytogenes. Cell* **110:** 551–561.

Joseph, B., Przybilla, K., Stuhler, C., Schauer, K., Slaghuis, J., Fuchs, T.M., and Goebel, W. (2006) Identification of *Listeria monocytogenes* genes contributing to intracellular replication by expression profiling and mutant screening. *J Bacteriol* **188:** 556–568.

Karkare, S., and Bhatnagar, D. (2006) Promising nucleic acid analogs and mimics: characteristic features and applications of PNA, LNA, and morpholino. *Appl Microbiol Biotechnol* **71:** 575–586.

Klein, R.J., Misulovin, Z., and Eddy, S.R. (2002) Noncoding RNA genes identified in AT-rich hyperthermophiles. *Proc Natl Acad Sci USA* **99:** 7542–7547.

Klenk, M., Koczan, D., Guthke, R., Nakata, M., Thiesen, H.J., Podbielski, A., and Kreikemeyer, B. (2005) Global epithelial cell transcriptional responses reveal *Streptococcus pyogenes* Fas regulator activity association with bacterial aggressiveness. *Cell Microbiol* **7:** 1237–1250.

Koppelhus, U., and Nielsen, P.E. (2003) Cellular delivery of peptide nucleic acid (PNA). *Adv Drug Deliv Rev* **55:** 267–280.

Koppelhus, U., Shiraishi, T., Zachar, V., Pankratova, S., and Nielsen, P.E. (2008) Improved cellular activity of antisense

© 2010 The Authors
Journal compilation © 2010 Society for Applied Microbiology and Blackwell Publishing Ltd, *Microbial Biotechnology*

Appendix

peptide nucleic acids by conjugation to a cationic peptide-lipid (CatLip) domain. *Bioconjug Chem* **19:** 1526–1534.
Kreikemeyer, B., Boyle, M.D., Buttaro, B.A., Heinemann, M., and Podbielski, A. (2001) Group A streptococcal growth phase-associated virulence factor regulation by a novel operon (Fas) with homologies to two-component-type regulators requires a small RNA molecule. *Mol Microbiol* **39:** 392–406.
Kreikemeyer, B., McIver, K.S., and Podbielski, A. (2003) Virulence factor regulation and regulatory networks in *Streptococcus pyogenes* and their impact on pathogen-host interactions. *Trends Microbiol* **11:** 224–232.
Kreikemeyer, B., Klenk, M., and Podbielski, A. (2004) The intracellular status of *Streptococcus pyogenes*: role of extracellular matrix-binding proteins and their regulation. *Int J Med Microbiol* **294:** 177–188.
Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., *et al.* (2001) Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. *Lancet* **357:** 1225–1240.
Kurupati, P., Tan, K.S., Kumarasinghe, G., and Poh, C.L. (2007) Inhibition of gene expression and growth by antisense peptide nucleic acids in a multiresistant beta-lactamase-producing *Klebsiella pneumoniae* strain. *Antimicrob Agents Chemother* **51:** 805–811.
Kwong, S.M., Skurray, R.A., and Firth, N. (2004) *Staphylococcus aureus* multiresistance plasmid pSK41: analysis of the replication region, initiator protein binding and antisense RNA regulation. *Mol Microbiol* **51:** 497–509.
Kwong, S.M., Skurray, R.A., and Firth, N. (2006) Replication control of staphylococcal multiresistance plasmid pSK41: an antisense RNA mediates dual-level regulation of Rep expression. *J Bacteriol* **188:** 4404–4412.
Lambert, J.L., and Fischer, A.L. (2005) Diagnostic assays including multiplexed lateral flow immunoassays with Quantum dots. 20050250141. 10-11-2005. United States.
Le, S.V., Chen, J.H., Currey, K.M., and Maizel, J.V., Jr (1988) A program for predicting significant RNA secondary structures. *Comput Appl Biosci* **4:** 153–159.
Lenz, D.H., Mok, K.C., Lilley, B.N., Kulkarni, R.V., Wingreen, N.S., and Bassler, B.L. (2004) The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. *Cell* **118:** 69–82.
Li, Z., Sledjeski, D.D., Kreikemeyer, B., Podbielski, A., and Boyle, M.D. (1999) Identification of *pel*, a *Streptococcus pyogenes* locus that affects both surface and secreted proteins. *J Bacteriol* **181:** 6019–6027.
Liu, J., Gough, J., and Rost, B. (2006) Distinguishing protein-coding from non-coding RNAs through support vector machines. *PLoS Genet* **2:** e29.
Liu, J.M., Livny, J., Lawrence, M.S., Kimball, M.D., Waldor, M.K., and Camilli, A. (2009) Experimental discovery of sRNAs in *Vibrio cholerae* by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res* **37:** e46.
Livny, J., Teonadi, H., Livny, M., and Waldor, M.K. (2008) High-throughput, kingdom-wide prediction and annotation of bacterial non-coding RNAs. *PLoS ONE* **3:** e3197.
Loh, E., Gripenland, J., and Johansson, J. (2006) Control of *Listeria monocytogenes* virulence by 5′-untranslated RNA. *Trends Microbiol* **14:** 294–298.
Lorenz, C., von Pelchrzim, F., and Schroeder, R. (2006) Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels. *Nat Protoc* **1:** 2204–2212.
Mandin, P., Repoila, F., Vergassola, M., Geissmann, T., and Cossart, P. (2007) Identification of new noncoding RNAs in *Listeria monocytogenes* and prediction of mRNA targets. *Nucleic Acids Res* **35:** 962–974.
Mangold, M., Siller, M., Roppenser, B., Vlaminckx, B.J., Penfound, T.A., Klein, R., *et al.* (2004) Synthesis of group A streptococcal virulence factors is controlled by a regulatory RNA molecule. *Mol Microbiol* **53:** 1515–1527.
Mao, X., Ma, Y., Zhang, A., Zhang, L., Zeng, L., and Liu, G. (2009) Disposable nucleic acid biosensors based on gold nanoparticle probes and lateral flow strip. *Anal Chem* **81:** 1660–1668.
Marchais, A., Naville, M., Bohn, C., Bouloc, P., and Gautheret, D. (2009) Single-pass classification of all noncoding sequences in a bacterial genome using phylogenetic profiles. *Genome Res* **19:** 1084–1092.
Mellbye, B.L., Puckett, S.E., Tilley, L.D., Iversen, P.L., and Geller, B.L. (2009) Variations in amino acid composition of antisense peptide-phosphorodiamidate morpholino oligomer affect potency against *Escherichia coli in vitro* and *in vivo*. *Antimicrob Agents Chemother* **53:** 525–530.
Meyers, B.C., Souret, F.F., Lu, C., and Green, P.J. (2006) Sweating the small stuff: microRNA discovery in plants. *Curr Opin Biotechnol* **17:** 139–146.
Morcos, P.A. (2007) Achieving targeted and quantifiable alteration of mRNA splicing with Morpholino oligos. *Biochem Biophys Res Commun* **358:** 521–527.
Morfeldt, E., Taylor, D., von Gabain, A., and Arvidson, S. (1995) Activation of alpha-toxin translation in *Staphylococcus aureus* by the *trans*-encoded antisense RNA, RNAIII. *EMBO J* **14:** 4569–4577.
Mulamba, G.B., Hu, A., Azad, R.F., Anderson, K.P., and Coen, D.M. (1998) Human cytomegalovirus mutant with sequence-dependent resistance to the phosphorothioate oligonucleotide fomivirsen (ISIS 2922). *Antimicrob Agents Chemother* **42:** 971–973.
Musser, J.M., and DeLeo, F.R. (2005) Toward a genome-wide systems biology analysis of host-pathogen interactions in group A *Streptococcus*. *Am J Pathol* **167:** 1461–1472.
Narberhaus, F., Waldminghaus, T., and Chowdhury, S. (2006) RNA thermometers. *FEMS Microbiol Rev* **30:** 3–16.
Negrini, M., Nicoloso, M.S., and Calin, G.A. (2009) MicroRNAs and cancer – new paradigms in molecular oncology. *Curr Opin Cell Biol* **21:** 470–479.
Nekhotiaeva, N., Awasthi, S.K., Nielsen, P.E., and Good, L. (2004) Inhibition of *Staphylococcus aureus* gene expression and growth using antisense peptide nucleic acids. *Mol Ther* **10:** 652–659.
Nielsen, J.S., Olsen, A.S., Bonde, M., Valentin-Hansen, P., and Kallipolitis, B.H. (2008) Identification of a sigma B-dependent small noncoding RNA in *Listeria monocytogenes*. *J Bacteriol* **190:** 6264–6270.
Novick, R.P., Ross, H.F., Projan, S.J., Kornblum, J., Kreiswirth, B., and Moghazeh, S. (1993) Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. *EMBO J* **12:** 3967–3975.

© 2010 The Authors
Journal compilation © 2010 Society for Applied Microbiology and Blackwell Publishing Ltd, *Microbial Biotechnology*

Ogier, J.C., and Serror, P. (2008) Safety assessment of dairy microorganisms: the *Enterococcus* genus. *Int J Food Microbiol* **126:** 291–301.

Padalon-Brauch, G., Hershberg, R., Elgrably-Weiss, M., Baruch, K., Rosenshine, I., Margalit, H., and Altuvia, S. (2008) Small RNAs encoded within genetic islands of *Salmonella typhimurium* show host-induced expression and role in virulence. *Nucleic Acids Res* **36:** 1913–1927.

Paulsen, I.T., Banerjei, L., Myers, G.S., Nelson, K.E., Seshadri, R., Read, T.D., *et al.* (2003) Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. *Science* **299:** 2071–2074.

Pellegrini, M., Marcotte, E.M., Thompson, M.J., Eisenberg, D., and Yeates, T.O. (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci USA* **96:** 4285–4288.

Perez, J.M., Josephson, L., O'Loughlin, T., Hogemann, D., and Weissleder, R. (2002) Magnetic relaxation switches capable of sensing molecular interactions. *Nat Biotechnol* **20:** 816–820.

Perez, N., Trevino, J., Liu, Z., Ho, S.C., Babitzke, P., and Sumby, P. (2009) A genome-wide analysis of small regulatory RNAs in the human pathogen group A *Streptococcus*. *PLoS ONE* **4:** e7668.

Pichon, C., and Felden, B. (2005) Small RNA genes expressed from *Staphylococcus aureus* genomic and pathogenicity islands with specific expression among pathogenic strains. *Proc Natl Acad Sci USA* **102:** 14249–14254.

Qin, X., Singh, K.V., Weinstock, G.M., and Murray, B.E. (2001) Characterization of fsr, a regulator controlling expression of gelatinase and serine protease in *Enterococcus faecalis* OG1RF. *J Bacteriol* **183:** 3372–3382.

Raengpradub, S., Wiedmann, M., and Boor, K.J. (2008) Comparative analysis of the sigma B-dependent stress responses in *Listeria monocytogenes* and *Listeria innocua* strains exposed to selected stress conditions. *Appl Environ Microbiol* **74:** 158–171.

Recsei, P., Kreiswirth, B., O'Reilly, M., Schlievert, P., Gruss, A., and Novick, R.P. (1986) Regulation of exoprotein gene expression in *Staphylococcus aureus* by *agrA*. *Mol Gen Genet* **202:** 58–61.

Riboulet-Bisson, E., Sanguinetti, M., Budin-Verneuil, A., Auffray, Y., Hartke, A., and Giard, J.C. (2008) Characterization of the Ers regulon of *Enterococcus faecalis*. *Infect Immun* **76:** 3064–3074.

Rivas, E. (2005) Evolutionary models for insertions and deletions in a probabilistic modeling framework. *BMC Bioinformatics* **6:** 63.

Rivas, E., and Eddy, S.R. (2000) Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs. *Bioinformatics* **16:** 583–605.

Rivas, E., and Eddy, S.R. (2001) Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics* **2:** 8.

Roberts, S.A., and Scott, J.R. (2007) RivR and the small RNA RivX: the missing links between the CovR regulatory cascade and the Mga regulon. *Mol Microbiol* **66:** 1506–1522.

Roush, W. (1997) Antisense aims for a renaissance. *Science* **276:** 1192–1193.

Rupnik, M., Wilcox, M.H., and Gerding, D.N. (2009) *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nat Rev Microbiol* **7:** 526–536.

Schattner, P. (2002) Searching for RNA genes using base-composition statistics. *Nucleic Acids Res* **30:** 2076–2082.

Seal, J., Braven, H., and Wallace, P. (2006) Point-of-care nucleic acid lateral-flow tests. *In Vitro Diagn Technol* 41–54.

Selinger, D.W., Cheung, K.J., Mei, R., Johansson, E.M., Richmond, C.S., Blattner, F.R., *et al.* (2000) RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. *Nat Biotechnol* **18:** 1262–1268.

Sen, A., and Nielsen, P.E. (2007) On the stability of peptide nucleic acid duplexes in the presence of organic solvents. *Nucleic Acids Res* **35:** 3367–3374.

Sharma, C.M., and Vogel, J. (2009) Experimental approaches for the discovery and characterization of regulatory small. RNA. *Curr Opin Microbiol* **12:** 536–546.

Shen, A., and Higgins, D.E. (2005) The 5′ untranslated region-mediated enhancement of intracellular listeriolysin O production is required for *Listeria monocytogenes* pathogenicity. *Mol Microbiol* **57:** 1460–1473.

Shimizu, T., Yaguchi, H., Ohtani, K., Banu, S., and Hayashi, H. (2002) Clostridial VirR/VirS regulon involves a regulatory RNA molecule for expression of toxins. *Mol Microbiol* **43:** 257–265.

Shokeen, S., Greenfield, T.J., Ehli, E.A., Rasmussen, J., Perrault, B.E., and Weaver, K.E. (2009) An intramolecular upstream helix ensures the stability of a toxin-encoding RNA in *Enterococcus faecalis*. *J Bacteriol* **191:** 1528–1536.

Simon, S.A., Zhai, J., Nandety, R.S., McCormick, K.P., Zeng, J., Mejia, D., and Meyers, B.C. (2009) Short-read sequencing technologies for transcriptional analyses. *Annu Rev Plant Biol* **60:** 305–333.

Sittka, A., Lucchini, S., Papenfort, K., Sharma, C.M., Rolle, K., Binnewies, T.T., *et al.* (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet* **4:** e1000163.

Storz, G., Opdyke, J.A., and Zhang, A. (2004) Controlling mRNA stability and translation with small, noncoding RNAs. *Curr Opin Microbiol* **7:** 140–144.

Stougaard, P., Molin, S., and Nordstrom, K. (1981) RNAs involved in copy-number control and incompatibility of plasmid R1. *Proc Natl Acad Sci USA* **78:** 6008–6012.

Stritzker, J., Schoen, C., and Goebel, W. (2005) Enhanced synthesis of internalin A in *aro* mutants of *Listeria monocytogenes* indicates posttranscriptional control of the *inlAB* mRNA. *J Bacteriol* **187:** 2836–2845.

Sun, X., Zhulin, I., and Wartell, R.M. (2002) Predicted structure and phyletic distribution of the RNA-binding protein Hfq. *Nucleic Acids Res* **30:** 3662–3671.

Tamanaha, C.R., Mulvaney, S.P., Rife, J.C., and Whitman, L.J. (2008) Magnetic labeling, detection, and system integration. *Biosens Bioelectron* **24:** 1–13.

Tan, X.X., Actor, J.K., and Chen, Y. (2005) Peptide nucleic acid antisense oligomer as a therapeutic strategy against bacterial infection: proof of principle using mouse intraperitoneal infection. *Antimicrob Agents Chemother* **49:** 3203–3207.

© 2010 The Authors
Journal compilation © 2010 Society for Applied Microbiology and Blackwell Publishing Ltd, *Microbial Biotechnology*

Appendix

Teng, F., Wang, L., Singh, K.V., Murray, B.E., and Weinstock, G.M. (2002) Involvement of PhoP–PhoS homologs in *Enterococcus faecalis* virulence. *Infect Immun* **70:** 1991–1996.

Toledo-Arana, A., Dussurget, O., Nikitas, G., Sesto, N., Guet-Revillet, H., Balestrino, D., *et al.* (2009) The *Listeria* transcriptional landscape from saprophytism to virulence. *Nature* **459:** 950–956.

Tomizawa, J., Itoh, T., Selzer, G., and Som, T. (1981) Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA. *Proc Natl Acad Sci USA* **78:** 1421–1425.

Tondra, M. (2007) Using integrated magnetic microchip devices in IVDs. *IVD Technol* **13:** 33–38.

Uzilov, A.V., Keegan, J.M., and Mathews, D.H. (2006) Detection of non-coding RNAs on the basis of predicted secondary structure formation free energy change. *BMC Bioinformatics* **7:** 173.

van der Veen, S., Hain, T., Wouters, J.A., Hossain, H., de Vos, W.M., Abee, T., *et al.* (2007) The heat-shock response of *Listeria monocytogenes* comprises genes involved in heat shock, cell division, cell wall synthesis, and the SOS response. *Microbiology* **153:** 3593–3607.

Verneuil, N., Sanguinetti, M., Le Breton, Y., Posteraro, B., Fadda, G., Auffray, Y., *et al.* (2004) Effects of the *Enterococcus faecalis hypR* gene encoding a new transcriptional regulator on oxidative stress response and intracellular survival within macrophages. *Infect Immun* **72:** 4424–4431.

Verneuil, N., Rince, A., Sanguinetti, M., Posteraro, B., Fadda, G., Auffray, Y., *et al.* (2005) Contribution of a PerR-like regulator to the oxidative-stress response and virulence of *Enterococcus faecalis*. *Microbiology* **151:** 3997–4004.

Vogel, J. (2009) A rough guide to the non-coding RNA world of *Salmonella*. *Mol Microbiol* **71:** 1–11.

Vogel, J., and Wagner, E.G. (2007) Target identification of small noncoding RNAs in bacteria. *Curr Opin Microbiol* **10:** 262–270.

Vogel, J., Bartels, V., Tang, T.H., Churakov, G., Slagter-Jager, J.G., Huttenhofer, A., and Wagner, E.G. (2003) RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res* **31:** 6435–6443.

Washietl, S., Hofacker, I.L., and Stadler, P.F. (2005) Fast and reliable prediction of noncoding RNAs. *Proc Natl Acad Sci USA* **102:** 2454–2459.

Waters, L.S., and Storz, G. (2009) Regulatory RNAs in bacteria. *Cell* **136:** 615–628.

Weaver, K.E. (2007) Emerging plasmid-encoded antisense RNA regulated systems. *Curr Opin Microbiol* **10:** 110–116.

Wong, K.K., Bouwer, H.G., and Freitag, N.E. (2004) Evidence implicating the 5′ untranslated region of *Listeria monocytogenes actA* in the regulation of bacterial actin-based motility. *Cell Microbiol* **6:** 155–166.

Ziebandt, A.K., Becher, D., Ohlsen, K., Hacker, J., Hecker, M., and Engelmann, S. (2004) The influence of *agr* and sigmaB in growth phase dependent regulation of virulence factors in *Staphylococcus aureus*. *Proteomics* **4:** 3034–3047.

© 2010 The Authors
Journal compilation © 2010 Society for Applied Microbiology and Blackwell Publishing Ltd, *Microbial Biotechnology*

13.6 Appendix publication 5

JB Journals.ASM.org

GENOME ANNOUNCEMENT

Complete Sequences of Plasmids from the Hemolytic-Uremic Syndrome-Associated *Escherichia coli* Strain HUSEC41

Carsten Künne,[a] Andre Billion,[a] Stephen E. Mshana,[b] Judith Schmiedel,[a] Eugen Domann,[a] Hamid Hossain,[a] Torsten Hain,[a] Can Imirzalioglu,[a] and Trinad Chakraborty[a]

Institute of Medical Microbiology, Justus-Liebig University, Giessen, Germany,[a] and Department of Microbiology, Weill Bugando Medical, College of Health Sciences, Mwanza, Tanzania[b]

The complete and annotated sequences of four plasmids from a historical enteroaggregative Shiga toxin-producing *Escherichia coli* (HUSEC) serotype O104:H4 strain, HUSEC41/01-09591, isolated in 2001 in Germany are reported.

The *Escherichia coli* serotype O104:H4 sequence type (ST) 678 strain which caused a disease outbreak in Germany in 2011 harbors three plasmids encoding a putative autotransporter serine protease, the aggregative adherence regulator *aggR*, and the extended-spectrum beta-lactamase CTX-M-15, all of which have contributed to the evolution of this virulent strain (1, 3, 6). A subsequent study of plasmids from the historical enteroaggregative Shiga toxin-producing serotype O104:H4 strain HUSEC41, isolated in 2001 from a child presenting with hemolytic-uremic syndrome, reported the detection and partial sequence of two plasmids, 95 and 75 kb (5). Our analysis of HUSEC41, however, indicated the presence of four plasmids, with sizes of 92, 74, 8, and 5 kb. This observation prompted us to determine the complete sequences of all plasmids in strain HUSEC41. These sequences provide a backdrop for the comparative analysis of the genealogy and evolution of plasmids that have contributed to the virulence properties of the HUSEC strain responsible for the recent outbreak of 2011.

Genomic DNA was isolated as described by Pitcher et al. (7). Sequencing was performed by Vertis (Germany) on a 454 GS-FLX system. Reads were assembled *de novo* with the 454 Newbler assembler, and resulting contigs were mapped against reference plasmids to determine a plasmid context. PCR-based techniques were used to close the gaps, followed by sequencing with ABI BigDye 3.0 technology (Applied Biosystems, Germany). A total of 17 contigs were assembled in four circular replicons with an average coverage of 45×. ORF calling and a first-pass automatic annotation were performed using RAST (rast.nmpdr.org) followed by manual comparative curation (4) and sequence similarity searches versus the NCBI (www.ncbi.nlm.nih.gov/BLAST), PFAM, and IS Finder (www-is.biotoul.fr) databases.

Four plasmids were detected: pHUSEC41-1, a large conjugative IncI1-type plasmid of 91,942 bp; pHUSEC41-2, a 73,564-bp nonconjugative IncF-type plasmid; and two small plasmids of 7,930 bp and 5,153 bp, designated pHUSEC41-3 and -4, respectively.

Plasmid pHUSEC41-1 displays 131 ORFs and harbors the resistance genes for streptomycin and sulfonamides. The organization comprising *trbC*, *sul2*, *strA*, bla_{TEM-1}, and *strB* is similar to p3521, an IncB plasmid (GenBank no. GU256641), and the IncQ RSF 1010 plasmid (GenBank no. M28829) (6). The transfer region of pHUSEC41-1 includes *trb*, *tra*, and *pil*, which are most related to plasmid p026vir (GenBank no. FJ386569).

The pHUSEC41-2 IncF plasmid contains 140 coding sequences (CDS) and is related to p55989 from the enteroaggregative *E. coli* strain 55989 (GenBank no. LB226692). Unlike pHUSEC41-1, it was not transferable to *E. coli* C118 by conjugation. The pHUSEC41-2 transfer region was found to exhibit deletions (e.g., *traV*) similar to those previously seen with other IncF plasmids with impaired conjugation properties (2). Plasmid pHUSEC41-3 displays 15 ORFs. Four of these are related to plasmid ColE1 mobilization proteins (MobA to -D) (GenBank no. J01566) (8).

We found 9 CDS on the smallest plasmid pHUSEC41-4, which resembles (~70% identity) plasmid ColE1 (8) minus its mobilization module. Comobilization of pHUSEC41-3 and -1 to *E. coli* CC118 occurred with a frequency of 10^{-5} per donor cell.

Nucleotide sequence accession numbers. The plasmid sequences reported here have been deposited in the EMBL database under accession numbers HE603110 (pHUSEC41-1), HE603111 (pHUSEC41-2), HE603112 (pHUSEC41-3), and HE603113 (pHUSEC41-4).

ACKNOWLEDGEMENTS

This study was supported by grants from the Federal Ministry of Education and Research (BMBF, Germany) within the framework of the RESET research network (contract no. 01KI1013G) and the Innovation Funds of the Ministry of Science and Arts of the State of Hessen to C.I. and T.C., respectively.

We thank Christina Gerstmann and Alexandra Amend-Foerster for excellent technical assistance. Furthermore, we acknowledge the German National Reference Center for Salmonellae and Other Bacterial Enteric Pathogens for providing strain HUSEC41.

REFERENCES

1. **Bielaszewska M, et al.** 2011. Characterisation of the *Escherichia coli* strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. Lancet Infect. Dis. **11**:671–676.
2. **Billard-Pomares T, et al.** 2011. Complete nucleotide sequence of plasmid pTN48, encoding the CTX-M-14 extended-spectrum β-lactamase from

Received 16 October 2011 **Accepted** 21 October 2011

Address correspondence to Can Imirzalioglu, can.imirzalioglu@mikrobio.med.uni-giessen.de, or Trinad Chakraborty, trinad.chakraborty@mikrobio.med.uni-giessen.de.

C. Künne and A. Billion contributed equally.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.06368-11

13.7 Appendix publication 6

JOURNAL OF CLINICAL MICROBIOLOGY, May 2011, p. 1843–1852 Vol. 49, No. 5
0095-1137/11/$12.00 doi:10.1128/JCM.01492-10
Copyright © 2011, American Society for Microbiology. All Rights Reserved.

Highly Specific and Quick Detection of *Mycobacterium avium* subsp. *paratuberculosis* in Feces and Gut Tissue of Cattle and Humans by Multiple Real-Time PCR Assays▽

Can Imirzalioglu,[1] Heinrich Dahmen,[2] Torsten Hain,[1] Andre Billion,[1] Carsten Kuenne,[1] Trinad Chakraborty,[1] and Eugen Domann[1]*

Institute of Medical Microbiology, Frankfurter Strasse 107, D-35392 Giessen, Germany,[1] and Veterinary Practice, Burgring 9, D-54595 Pruem, Germany[2]

Received 23 July 2010/Returned for modification 23 December 2010/Accepted 7 March 2011

Mycobacterium avium* subsp. *paratuberculosis* is the causative agent of Johne's disease (JD) in cattle and may be associated with Crohn's disease (CD) in humans. It is the slowest growing of the cultivable mycobacteria, and culture from clinical, veterinary, food, or environmental specimens can take 4 months or even longer. Currently, the insertion element IS*900* is used to detect *M. avium* subsp. *paratuberculosis* DNA. However, closely related IS*900* elements are also present in other mycobacteria, thus limiting its specificity as a target. Here we describe the use of novel primer sets derived from the sequences of two highly specific single copy genes, MAP2765c and MAP0865, for the quantitative detection of *M. avium* subsp. *paratuberculosis* within 6 h by using real-time PCR. Specificity of the target was established using 40 *M. avium* subsp. *paratuberculosis* isolates, 67 different bacterial species, and two intestinal parasites. Using the probes and methods described, we detected 27 (2.09%) *M. avium* subsp. *paratuberculosis*-positive stool specimens from 1,293 individual stool samples by the use of either IS*900* or probes deriving from the MAP2765c and MAP0865 genes described here. In general, bacterial load due to *M. avium* subsp. *paratuberculosis* was uniformly low in these samples and we estimated 500 to 5,000 *M. avium* subsp. *paratuberculosis* bacteria per gram of stool in assay-positive samples. Thus, the methods described here are useful for rapid and specific detection of *M. avium* subsp. *paratuberculosis* in clinical samples.

Mycobacterium avium subsp. *paratuberculosis* is the causative agent of Johne's disease (JD) in cattle, and it has been suggested that this microorganism may be associated with Crohn's disease (CD) in humans (14, 17). *M. avium* subsp. *paratuberculosis* belongs to the mycobacterial species *M. avium*, which is currently subdivided into three subspecies (33): *M. avium* subsp. *avium* (synonym, *M. avium*), *M. avium* subsp. *paratuberculosis* (synonym, *M. paratuberculosis*), and *M. avium* subsp. *silvaticum* (synonym, *M. silvaticum*). At the subspecies level, *M. avium* subsp. *paratuberculosis* can be differentiated phenotypically from *M. avium* subsp. *avium* and *M. avium* subsp. *silvaticum* by its dependence on mycobactin (51) and genotypically by the presence of multiple copies of the insertion element IS*900* (2, 5, 13, 22, 45).

JD (or paratuberculosis) is a chronic, granulomatous severe form of gastroenteritis with progressive weight loss and emaciation affecting domestic and wild ruminants, e.g., cattle, sheep, goats, red deer, and rabbits worldwide (12, 30, 31). Infected livestock periodically shed *M. avium* subsp. *paratuberculosis* via feces and milk, which results in environmental distribution, where *M. avium* subsp. *paratuberculosis* can survive for extended periods (55). Milk pasteurization trials showed that high-temperature and short-duration standard pasteurization procedures do not effectively kill *M. avium* subsp. *paratuberculosis* in milk, as clinical strains of *M. avium* subsp. *paratuberculosis* have been shown to be more thermally tolerant than either *M. bovis* or *Coxiella burnetii*, the current milk pasteurization standard microorganisms (10, 44, 47, 48). Therefore, human contact can result from the consumption of inadequately pasteurized milk or raw milk or of certain other dairy products, fecally contaminated vegetables, contaminated beef, or even water (12, 14, 34, 41, 47, 48).

CD is a chronic inflammatory disease of the gastrointestinal tract in humans, affecting in particular the terminal ileum, with symptoms of general malaise, weight loss, abdominal pain, and diarrhea (4, 12, 16, 40). A hallmark of CD is the histological proof of a granulomatous inflammation, which is also characteristic of JD and other mycobacterial diseases, leading to suggestions that *M. avium* subsp. *paratuberculosis* may be associated with CD in humans (see references 12 [and references therein], 8, 15, 16, 27, and 28).

Currently, there is controversy as to whether *M. avium* subsp. *paratuberculosis* (i) is an innocent bystander that has merely colonized the intestine of Crohn's patients, (ii) could be a secondary infection but not a cause of the disease, (iii) could be the primary infectious agent and the cause of CD, (iv) acts as a superantigen, or (v) modifies the immune response in CD (6, 12, 15, 25, 32, 37, 42, 43, 50). One of the major obstacles to resolution of the debate on the role of *M. avium* subsp. *paratuberculosis* in CD and the controversial studies published is the requirement of reliable and contemporary detection and identification of *M. avium* subsp. *paratuberculosis* in complex

* Corresponding author. Mailing address: Institute of Medical Microbiology, Frankfurter Strasse 107, D-35392 Giessen, Germany. Phone: 49 641-9941287. Fax: 49 641-9941259. E-mail: eugen.domann@mikrobio.med.uni-giessen.de.
▽ Published ahead of print on 23 March 2011.

Appendix

TABLE 1. Specific primers and TaqMan probes used in this study to detect *M. avium* subsp. *paratuberculosis*

Target (GenBank accession no.)	Oligonucleotide sequence	Reference(s) or source
Amplicon I IS900 (AE016958.1)		
Forward	5′-AAT GAC GGT TAC GGA GGT GGT-3′	21, 22
Reverse	5′-GCA GTA ATG GTC GGC CTT ACC-3′	
Probe	FAM-TCC ACG CCC GCC CAG ACA GG-TAMRA	
Amplicon II 251 (AF445445)		
Forward	5′-GCA AGA CGT TCA TGG GAA CT-3′	2, 38
Reverse	5′-GCG TAA CTC AGC GAA CAA CA-3′	
Probe	FAM-CTG ACT TCA CGA TGC GGT TCT TC-TAMRA	
Amplicon III f57 (X70277)		
Forward	5′-TAC CGA ATG TTG TTG TCA CCG-3′	35, this study
Reverse	5′-TGG CAC AGA CGA CCA TTC AA-3′	
Probe	FAM-CCG GTC CCA GGT GTG TTC GAG TTG-TAMRA	
Amplicon IV MAP0865 (AE016958.1)		
Forward	5′-GCG CGG CCA GTA TGG ATA TA-3′	19, 22, this study
Reverse	5′-GAC TCA ACC CAA CGA GCT CC-3′	
Probe	FAM-AGA TGC CTC TCC GAT GCT CGA TGG-TAMRA	
Amplicon V_1 16S rRNA gene		
Universal forward	5′-TCC TAC GGG AGG CAG CAG T-3′	23
Universal reverse	5′-GGA CTA CCA GGG TAT CTA ATC CTG TT-3′	
Universal probe	FAM-CGT ATT ACC GCG GCT GCT GGC AC-TAMRA	
Amplicon V_2 18S rRNA gene		
Universal forward	5′-GCG GAT CCG CGG CCG CTG GTT GAT CCT GCC AGT-3′	26
Universal reverse	5′-GCG GAT CCG CGG CCG CGG CAG GTT CAC CTA C-3′	

specimens such as blood, biopsy samples, breast milk, and feces. *M. avium* subsp. *paratuberculosis* is the slowest growing of the cultivable mycobacteria, and primary culture from clinical, veterinary, food, or environmental specimens can take 4 months or even longer (12, 24, 44, 55). Moreover, the characteristics of *M. avium* subsp. *paratuberculosis* in JD and in CD seem to be totally different: in CD, *M. avium* subsp. *paratuberculosis* appears as non-acid-fast coccobacilli with the ultrastructure of spheroplasts (cell-wall-deficient forms) that do not transform into characteristic *M. avium* subsp. *paratuberculosis* organisms until after several months of incubation (4).

In this study, we aimed at establishing highly specific multiple quantitative real-time PCR (qrt-PCR) assays based on the published genome sequence of *M. avium* subsp. *paratuberculosis* strain K-10 to enable rapid and unequivocal detection and identification of *M. avium* subsp. *paratuberculosis* directly from clinical, veterinary, food, or environmental specimens as well as from pure cultures. The method developed allows the quick and reliable detection and quantification of *M. avium* subsp. *paratuberculosis* directly from stool samples within 6 h at reasonable costs.

MATERIALS AND METHODS

Human clinical samples and data. Patient fecal and tissue specimens represented clinical routine diagnostic samples for the detection of pathogenic bacteria and parasites. Testing for the presence of *M. avium* subsp. *paratuberculosis* DNA was performed in addition to DNA detection of conventional enteropathogenic bacteria and parasites. Positive results were reported. Clinical data from patients with positive results were obtained after informed consent by record review. The study was approved by the Ethics Board of the Justus-Liebig-University of Giessen, Faculty of Medicine.

Cattle samples. Feces and gut biopsy specimens were obtained from healthy cows and from cows with suspected paratuberculosis. The symptoms of clinical paratuberculosis are chronic diarrhea and progressive weight loss, whereas subclinically infected animals mainly exhibit decreased milk production. The guts were dissected into small pieces and subdivided into inflamed and noninflamed tissue samples based on microscopic examination.

Microorganisms and standard cultivation. The mycobacterial and nonmycobacterial species used in the study are listed here (see Table 2). The strains were obtained from the German Resource Center for Biological Material (DSMZ) and from the strain collection of the Institute of Medical Microbiology, Justus-Liebig University of Giessen.

M. avium subsp. *paratuberculosis* was cultivated on BD BBL Herrold's egg yolk agar with Mycobactin J and ANV (Becton Dickinson, Heidelberg, Germany) and in modified Middlebrook 7H9 medium for up to 4 months as previously described (12, 39). All remaining bacteria were grown under optimal conditions on appropriate media as previously recommended (3).

Acid-fast staining. The acid-fast staining of specimens was done by the Ziehl-Neelsen procedure, and stained specimens examined under conditions of oil immersion. Dissected biopsy samples from gut tissue of cows were stained as previously described for tissue sections of bisons (20).

DNA extraction. DNA was extracted from pure bacterial cultures with a RTP Spin Bacteria DNA kit (Invitek, Berlin, Germany). The total DNA from human and animal fecal specimens was extracted with a PSP Spin Stool DNA kit and from tissue with an Invisorb Spin Tissue minikit (Invitek) as recommended by the vendor.

Parasitic DNA from *Entamoeba histolytica* and from *Giardia lamblia* was obtained from the Bernhard-Nocht Institute for Tropical Medicine (Hamburg, Germany).

Quantitative real-time PCR (qrt-PCR) and conventional PCR. The new primers and probes used were designed with "Primer Express" version 1.0 software (Applied Biosystems, Foster City, CA). The internal probes were labeled with the fluorescent reporter dye 5-carboxyfluoroscein (FAM) on the 5′ end and the quencher dye N′,N′,N′,N′-tetramethyl-6-carboxyrhodamine (TAMRA) on the 3′ end. The qrt-PCR was accomplished as described previously (21, 38). Protozoal universal 18S small-subunit rRNA (SSU-rRNA) detection was performed by conventional PCR using eukaryote-specific primers (Table 1) as described previously (26). The amounts of specific target sequences present in unknown samples were calculated by measuring the threshold cycle (C_T) values and using standard curves generated with a series of known quantities of target sequences. The C_T value represented the cycle at which the copy of the amplified target

Appendix

sequence intersected the threshold or baseline. Briefly, an inoculation loop of the *M. avium* subsp. *paratuberculosis* DSM 44133 type strain grown on Herrold's egg yolk was carefully resuspended in phosphate-buffered saline (PBS) and serially diluted. The total amounts of cells were counted using a Neubauer counting chamber. We determined the total cell numbers, since complex biological material may contain both viable and dead bacteria.

Detection of *Salmonella enterica* serovar *enteritidis*, *Campylobacter jejuni*, *Yersinia enterocolitica*, *Clostridium difficile*, *Entamoeba histolytica*, and *Giardia lamblia*. The total DNA of human fecal specimens was extracted with a PSP Spin Stool kit (Invitek) as described above. *S. enteritidis*, *C. jejuni*, *Y. enterocolitica*, *C. difficile*, *E. histolytica*, and *G. lamblia* were detected by specific real-time PCRs as previously described (18, 29, 52, 53, 54).

Multiple displacement amplification (MDA) of IS*900*-positive fecal specimens. MDA was performed with the extracted DNA of amplicon I (IS*900*)-positive but amplicon II (MAP2765c [251])-, III (MAP0865 [f57])-, and IV (MAP0865)-negative fecal samples. A commercially available GenomiPhi DNA amplification kit (Amersham Biosciences, Uppsala, Sweden), which utilizes MDA to exponentially amplify genomic DNA, was used, following the manufacturer's instructions. Briefly, 1 μl of DNA extract was added to 9 μl of sample buffer containing random hexamer primers and heated to 95°C. The chilled sample was mixed with 9 μl of reaction buffer and 1 μl of enzyme mix. The mixture was incubated for 14 h at 30°C and afterwards subjected to heat inactivation for 10 min at 65°C.

Bioinformatics. DNA sequences were aligned with MegAlign version 5.0 software (DNASTAR Inc., Madison, WI) and compared with sequences deposited in the GenBank, EMBL, DDBJ, and PDB databases using the BLASTn basic local alignment search tool (1). The published genome of *M. avium* subsp. *paratuberculosis* strain K-10 (22; GenBank accession number AE016958.1) was analyzed and depicted with the GenomeViz bioinformatics tool (11). The standard curves to calculate the amounts of *M. avium* subsp. *paratuberculosis* in complex stool and biopsy specimens were generated with SigmaPlot 2000 version 6.00 software (SPSS GmbH Software, Munich, Germany).

RESULTS

Development of MAP0865-specific real-time PCRs. Fragment f57 (GenBank accession no. X70277) has been previously described and used as a highly specific 620-bp-long probe for detection of *M. avium* subsp. *paratuberculosis* and diagnosis of Johne's disease (35, 49). The nucleotide-nucleotide BLAST (BLASTn) analysis we performed revealed that it was located in gene MAP0865, one of the 4,350 predicted open reading frames (ORFs) of *M. avium* subsp. *paratuberculosis* strain K-10 (Fig. 1) (22). Fragment f57 covered 48.7% of this 1,272-bp-long ORF with an identity of 100% (Fig. 1). Another BLASTn analysis using the nucleotide sequence of ORF MAP0865 with at least 3.7 million sequences deposited in databases GenBank, EMBL, DDBJ, and PDB revealed it to be unique. Consensus was detected only with deposited MAP0865 and fragment f57 sequences.

By the use of Primer Express software, two new TaqMan amplicons with primers and probes were designed for this *M. avium* subsp. *paratuberculosis*-specific MAP0865 ORF. The first one, designated amplicon III, covered both fragment f57 and ORF MAP0865, while the second, designated amplicon IV, covered only ORF MAP0865 (Fig. 1, Table 1).

Specificities of the MAP0865-specific real-time PCRs. The specificities of the new real-time PCRs using amplicons III and IV were demonstrated by the analysis of DNA from panels of 40 *M. avium* subsp. *paratuberculosis* isolates, 17 other mycobacterial species, 13 Gram-positive bacterial species, 20 Gram-negative bacterial species (including 8 different species causing gastroenteritis), 7 obligate anaerobic bacterial species, and 2 intestinal parasite species (Table 2). The results showed that the real-time PCRs with amplicons III and IV, as well as those performed with previously developed amplicons I (IS*900*) and II (MAP2765c [251]), specifically amplified only the *M. avium* subsp. *paratuberculosis* DNA but not the DNA of the remaining bacteria (Table 2), indicating that the entire ORF MAP0865 is unique for *M. avium* subsp. *paratuberculosis*.

All DNA extracts gave positive results using 16S rRNA gene universal amplicon V_1 for bacteria or 18S rRNA gene universal amplicon V_2 for parasites (Table 2). The nucleic acid concentrations of the DNA extracts were calculated to be at ~50 ng/μl.

Positions of element IS*900*, ORF MAP2765c (251), and ORF MAP0865 (f57) on the chromosome of *M. avium* subsp. *paratuberculosis* strain K-10. A previously published genome-scale comparison of *M. avium* subsp. *paratuberculosis* with its closely related subspecies *M. avium* subsp. *avium* revealed potential new diagnostic sequences (2). Among these, target 251 with a length of 540 bp has been identified as a valuable new sequence for specific amplification of *M. avium* subsp. *paratuberculosis* DNA (38; GenBank accession no. AF445445). Nucleotide sequence alignments indicated that target 251 is located in gene MAP2765c of *M. avium* subsp. *paratuberculosis* strain K-10 (Fig. 1) (22).

Genome visualization using the GenomeViz bioinformatics tool (11) revealed a random distribution of 17 copies of IS*900* on the *M. avium* subsp. *paratuberculosis* chromosome and that both ORF MAP0865 (f57) and ORF MAP2765c (251) exist as singular sequences at diametrically opposite positions (Fig. 1). Eleven copies of IS*900* and MAP0865 (f57) were located on the positive strand, whereas the remaining six copies of IS*900* and MAP2765c (251) were located on the negative strand (Fig. 1). This underlines the utility of IS*900* as a highly sensitive PCR target for *M. avium* subsp. *paratuberculosis* due to its high copy number on the *M. avium* subsp. *paratuberculosis* chromosome. Nevertheless, ORFs MAP0865 and MAP2765c are highly specific PCR targets that can be used to identify *M. avium* subsp. *paratuberculosis* and to confirm IS*900*-specific PCRs.

Optimization of PCR conditions and annealing temperature of IS*900*-, MAP2765c (251)-, MAP0865 (f57)-, and MAP0865-specific real-time PCRs. In order to determine the optimal PCR buffer conditions and the optimal primer annealing temperature for all four amplicons, we analyzed the efficiencies of the real-time PCRs with $MgCl_2$ concentrations in the range of 1.0 to 5.0 mM and temperatures in the range of 55 to 65°C (data not shown). The DNA used was extracted from ~10^8 cells of *M. avium* subsp. *paratuberculosis*, and the experiments were done in triplicate. For all amplicons (I through IV), the optimal $MgCl_2$ concentration was 3.5 mM and the annealing temperature 57.8°C. The ideal results obtained consisted of a C_T value of 14.4 for amplicon I (IS*900*), a C_T value of 18.2 for amplicon II (MAP2765c [251]), and a C_T value of 16.2 for amplicons III (MAP0865 [f57]) and IV (MAP0865) (Fig. 2). The order with respect to sensitivity and efficiency of the TaqMan-PCR was amplicon I first, amplicons III and IV second, and amplicon II third. In cases of amplicon I-positive but amplicon II- to IV-negative specimens, multiple displacement amplification (MDA) was used to exponentially amplify genomic DNA in the DNA extracts of the tested samples. The results suggest the use of MDA as a routine tool for amplification of specimens when low concentrations of *M. avium* subsp. *paratuberculosis*-specific DNA are suspected.

Appendix

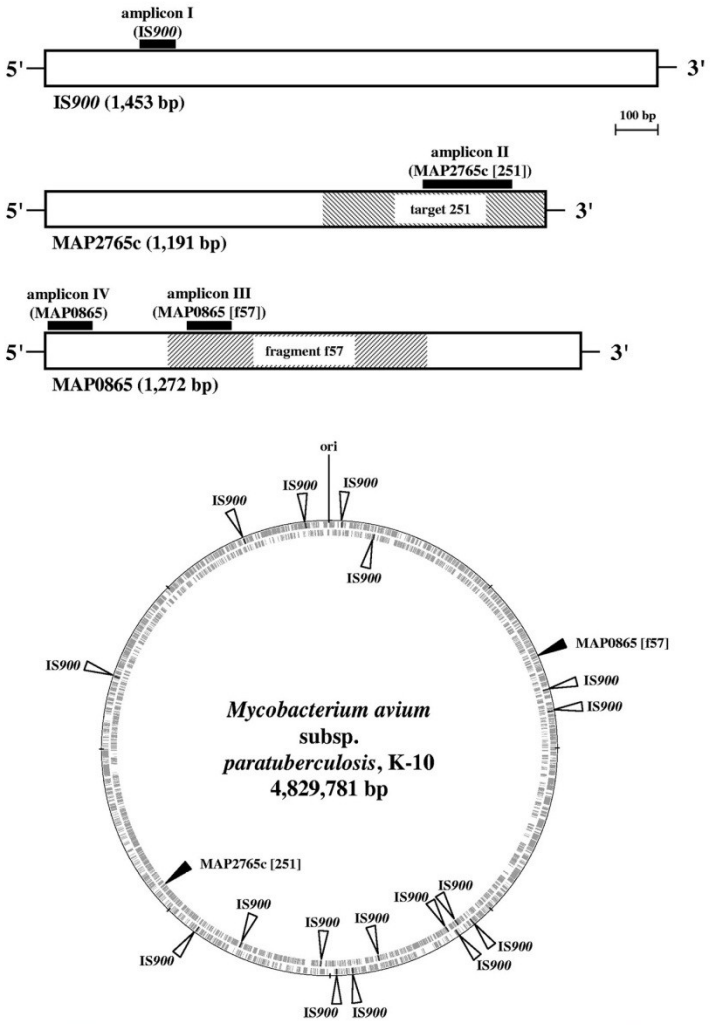


FIG. 1. Graphic presentation of the *M. avium* subsp. *paratuberculosis*-specific chromosomal regions IS*900*, MAP2765c (target 251), and MAP0865 (fragment f57) and their positions in the *M. avium* subsp. *paratuberculosis* K-10 genome (GenBank accession no. AE016958.1). Outer circle, plus strand; inner circle, minus strand. The homology analyses and the graphic presentations were done using the methods BLAST, Clustal W (1), and GenomeViz (11). The black boxes represent specific TaqMan PCR amplicons I through IV, including forward and reverse primers and the internal probe labeled with FAM and TAMRA (Table 1). The amplicons were designed using Primer Express software (Applied Biosystems, Foster City, CA).

The sensitivity of real-time PCR depends on the DNA extraction method used. Since the real-time PCR with amplicon I (IS*900*) showed the highest efficiency, serial dilutions of the genomic DNA of *M. avium* subsp. *paratuberculosis* type strain DSM 44133 were used as templates to assess its sensitivity under the conditions mentioned above. The detection limit of the real-time PCR with amplicon I (IS*900*) was (theoretically) determined to be 1 to 10 CFU/reaction and therefore was identical to the results obtained previously (21). To identify *M. avium* subsp. *paratuberculosis* and to calculate directly the amount of the bacteria in complex biological material such as feces and tissue without cultivation, we assessed the efficiency of the DNA extraction methods with two commercially available kits: a stock solution of 7.3×10^7 *M. avium* subsp. *paratuberculosis* bacteria was serially diluted to 7.3×10^6, 7.3×10^5, 7.3×10^4, 7.3×10^3, 7.3×10^2, and 7.3×10^1 bacteria, and the DNA of the bacteria from each dilution step was extracted using commercially available kits for stool and tissue specimens (Invitek, Berlin, Germany). The stool kit had a detection limit of ≥70 cells/reaction, whereas the detection limit of the tissue kit was ≥10 cells/reaction. Both analyses confirmed very high correlation, as demonstrated by determinations of linear regression (r^2) (Fig. 3).

Appendix

TABLE 2. Bacterial isolates (and their sources) used in this study[a]

Bacterium	Source	PCR result with indicated amplicon					
		I	II	III	IV	V_1	V_2
M. avium subsp. *paratuberculosis* type strain	DSM 44133	+	+	+	+	+	−
M. avium subsp. *paratuberculosis*	DSM 44135	+	+	+	+	+	−
M. avium subsp. *paratuberculosis* (n = 40)	Field isolates	+	+	+	+	+	−
M. avium subsp. *paratuberculosis*	Patient isolate	+	+	+	+	+	−
M. avium subsp. *avium*	DSM 44158	−	−	−	−	+	−
M. avium subsp. *silvaticum*	DSM 44175	−	−	−	−	+	−
Mycobacterium tuberculosis (n = 10)	Patient isolates	−	−	−	−	+	−
Mycobacterium marinum	Patient isolate	−	−	−	−	+	−
Mycobacterium kansasii	Patient isolate	−	−	−	−	+	−
Mycobacterium chelonae	Patient isolate	−	−	−	−	+	−
Mycobacterium intracellulare	Patient isolate	−	−	−	−	+	−
Mycobacterium abscessus	Patient isolate	−	−	−	−	+	−
Staphylococcus aureus	Patient isolate	−	−	−	−	+	−
Streptococcus agalactiae	Patient isolate	−	−	−	−	+	−
Lactobacillus spp. (n = 5)	Patient isolates	−	−	−	−	+	−
Lactococcus lactis	Patient isolate	−	−	−	−	+	−
Enterococcus faecalis (n = 3)	Patient isolates	−	−	−	−	+	−
Enterococcus faecium	Patient isolate	−	−	−	−	+	−
Listeria monocytogenes	Patient isolate	−	−	−	−	+	−
Salmonella enteritidis (n = 2)	Patient isolates	−	−	−	−	+	−
Salmonella typhimurium (n = 2)	Patient isolates	−	−	−	−	+	−
Salmonella infantis	Patient isolate	−	−	−	−	+	−
Salmonella typhi	Patient isolate	−	−	−	−	+	−
Campylobacter jejuni	Patient isolate	−	−	−	−	+	−
Yersinia enterocolitica	Patient isolate	−	−	−	−	+	−
Shigella flexneri	Patient isolate	−	−	−	−	+	−
Enterohemorrhagic *Escherichia coli* (EHEC)	Patient isolate	−	−	−	−	+	−
Proteus mirabilis (n = 3)	Patient isolates	−	−	−	−	+	−
Citrobacter diversus (n = 2)	Patient isolates	−	−	−	−	+	−
Enterobacter cloacae	Patient isolate	−	−	−	−	+	−
Enterobacter sakazakii	Patient isolate	−	−	−	−	+	−
Escherichia coli	Patient isolate	−	−	−	−	+	−
Hafnia alvei	Patient isolate	−	−	−	−	+	−
Helicobacter pylori	Patient isolate	−	−	−	−	+	−
Clostridium difficile	Patient isolate	−	−	−	−	+	−
Fusobacterium nucleatum	Patient isolate	−	−	−	−	+	−
Propionibacterium acnes	Patient isolate	−	−	−	−	+	−
Bacteroides fragilis	Patient isolate	−	−	−	−	+	−
Prevotella intermedia	Patient isolate	−	−	−	−	+	−
Veillonella spp.	DSM 2008	−	−	−	−	+	−
Atopobium spp.	DSM 15829	−	−	−	−	+	−
Entamoeba histolytica	BNI	−	−	−	−	−	+
Giardia lamblia	BNI	−	−	−	−	−	+

[a] Positive results for amplicons I through V are indicated by a plus sign, and negative results are indicated by a minus sign. n, number of isolates; DSM, strain derived from the German Resource Center for Biological Material (DSMZ); BNI, parasite DNA obtained from the Bernhard-Nocht Institute for Tropical Medicine; amplicon I, IS*900*; amplicon II, MAP2765c (251); amplicon III, MAP0865 (f57); amplicon IV, MAP0865; amplicon V_1, universal 16S rRNA gene; amplicon V_2, universal 18S rRNA gene.

Microscopic detection and cultivation of *M. avium* subsp. *paratuberculosis* in feces and gut tissue from cattle with Johne's disease. We examined the feces and dissected gut tissue from 12 butchered cows by Ziehl-Neelsen staining and culture. Three cows were healthy, and the remaining nine were suspected to be afflicted with Johne's disease. Diseased cows 1 to 4, 6 to 9, and 12 showed acid-fast bacteria in both feces and inflamed gut tissue, and pure cultures of *M. avium* subsp. *paratuberculosis* could be obtained after 3 to 6 months of incubation (Fig. 2A, Table 3). The noninflamed gut tissue samples were uniformly negative for *M. avium* subsp. *paratuberculosis*. Healthy cows 5, 10, and 11 were negative for *M. avium* subsp. *paratuberculosis* in all samples examined (Table 3). All *M. avium* subsp. *paratuberculosis* isolates were identified by the specific real-time PCRs with amplicons I (IS*900*), II (MAP2765c [251]), III (MAP0865 [f57]), and IV (MAP0865). The growth of all *M. avium* subsp. *paratuberculosis* isolates was mycobactin dependent. Furthermore, the isolates were assigned to the *M. avium* complex by partial sequencing of the 16S rRNA genes and by performing a Genotype *Mycobacterium CM*|*AS* test (Hain Lifescience GmbH, Nehren, Germany).

Evaluation of the *M. avium* subsp. *paratuberculosis*-specific real-time PCRs using original bovine gut tissue and feces. The total DNA from inflamed and noninflamed dissected gut tissue and from feces of healthy and diseased cows was extracted with commercially available kits for tissue and stool specimens (Invitek, Berlin, Germany). All of the diseased cows with positive *M. avium* subsp. *paratuberculosis* culture results were also positive for specific real-time PCR amplicons I to IV (Table 3).

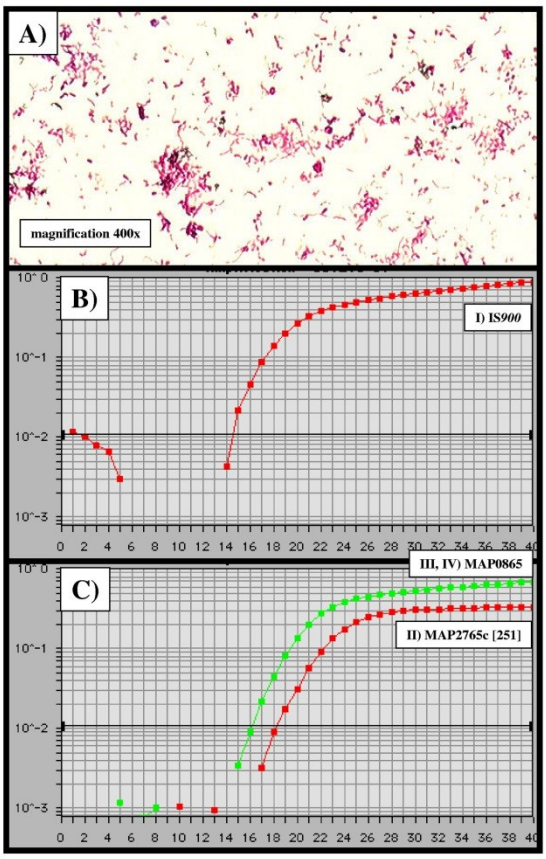


FIG. 2. Temperature optimization and sensitivity of IS*900*-, MAP2765c-, and MAP0865-specific TaqMan PCRs. Using a DNA extract of ~10^8 CFU of the *M. avium* subsp. *paratuberculosis* DSM 44133 type strain, the optimal annealing temperature for all amplicons I through IV was determined to be uniformly 57.8°C. (A) Ziehl-Neelsen staining of the *M. avium* subsp. *paratuberculosis* wild-type isolate obtained from feces of cow 7 (Table 3). (B and C) *x* axis, C_T values; *y* axis, ΔR_n (FAM reporter signal divided by the ROX [carboxy-X-rhodamine] passive reference signal). (B) amplicon I (IS*900*), C_T value = 14.4; (C) amplicon II (MAP2765c [251]), C_T value = 18.2; amplicons III (MAP0865 [f57]) and IV (MAP0865), C_T value = 16.2.

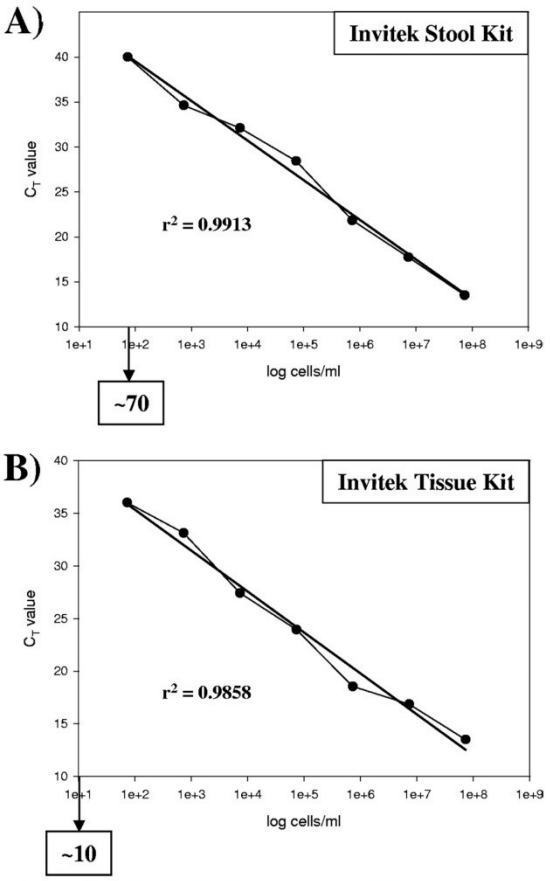


FIG. 3. Calculation of the detection limits for *M. avium* subsp. *paratuberculosis* by using commercially available DNA extraction kits and IS*900* TaqMan PCR (amplicon I). The DNA of 1 ml of serial dilutions of the *M. avium* subsp. *paratuberculosis* DSM 44133 type strain was extracted to generate standard curves and to determine the linear regression (r^2) and the detection limit (indicated by arrows) for each method. (A) Invitek stool kit (r^2 = 0.9913); detection limit, ≥70 cells. (B) Invitek tissue kit (r^2 = 0.9858); detection limit, ≥10 cells. The calculations were done with Sigma Plot software (SPSS Inc. Software, Munich, Germany).

Using the standard curves (Fig. 3), the amount of *M. avium* subsp. *paratuberculosis* was calculated to be in the range of ~2×10^4 to 6×10^7 bacteria per gram of inflamed gut tissue and ~1×10^7 to 2×10^9 bacteria per gram of feces. Thus, the concentration of *M. avium* subsp. *paratuberculosis* in feces was continuously ~30- to 500-fold higher than in inflamed parts of the intestinal tissue. All of the tissue samples from healthy *M. avium* subsp. *paratuberculosis* culture-negative cows as well as the noninflamed tissue samples of diseased cows were also negative in the *M. avium* subsp. *paratuberculosis*-specific real-time PCRs (Table 3).

Detection of *M. avium* subsp. *paratuberculosis* in stool specimens and gut tissue of patients with diarrhea. We examined consecutive stool specimens of 1,293 hospitalized patients with mild or severe symptoms of diarrhea for the presence of *S. enteritidis*, *C. jejuni*, *Y. enterocolitica*, *C. difficile*, *E. histolytica*, *G. lamblia*, and *M. avium* subsp. *paratuberculosis*. The total DNA was extracted with an Invitek stool kit, and the *M. avium* subsp. *paratuberculosis* analysis was done with highly sensitive amplicon I (IS*900*). Twenty-seven patients (2.09%) gave positive results, and the concentrations of *M. avium* subsp. *paratuberculosis* were calculated in the range of ~5×10^2 to 5×10^3 bacteria per gram of stool by the use of the corresponding standard curve (Fig. 3A). The results were confirmed by using amplicons II through IV. For 6 patients, the results for amplicons II to IV remained negative following initial amplification despite a positive amplicon I result, suggesting the presence of non-*M. avium* subsp. *paratuberculosis* mycobacterial DNA.

Appendix

TABLE 3. Detection of *M. avium* subsp. *paratuberculosis* in bovine gut tissue and feces by amplicon I (IS*900*)-, amplicon II (MAP2765c [251])-, amplicon III (MAP0865 [f57])-, and amplicon IV (MAP0865)-specific TaqMan PCR, by acid-fast staining, and by culture (n = 12)[a]

Cattle and specimen category	Acid-fast staining result	Culture result	TaqMan PCR result for indicated amplicon			
			I	II	III	IV
Diseased (n = 9)						
Noninflamed gut tissue	−	−	−	−	−	−
Inflamed gut tissue	+	+	+	+	+	+
Feces	+	+	+	+	+	+
Healthy (n = 3)						
Noninflamed gut tissue	−	−	−	−	−	−
Feces	−	−	−	−	−	−

[a] Diseased cattle (n = 9) were suspected to be infected with *M. avium* subsp. *paratuberculosis* because of typical symptoms of Johne's disease (paratuberculosis).

However, following MDA, the presence of *M. avium* subsp. *paratuberculosis* was then confirmed by positive PCR results for amplicons II to IV for all 6 specimens. All of the 26 *M. avium* subsp. *paratuberculosis*-positive patients gave negative test results for *S. enteritidis*, *C. jejuni*, *Y. enterocolitica*, *E. histolytica*, and *G. lamblia*. Only one of these patients, who suffered from pseudomembranous colitis, subsequently tested positive for *C. difficile*.

Among the 1,293 patients, we identified 11 patients with chronic inflammatory bowel disease. Of these, 6 patients had clinically confirmed CD and an additional 5 were classified with ulcerative colitis (UC). Only two of the 27 *M. avium* subsp. *paratuberculosis*-positive individuals were CD patients. Additionally, a small piece of gut tissue from one CD patient also gave *M. avium* subsp. *paratuberculosis*-positive test results. The identification was confirmed by reamplification with amplicons II through IV, acid-fast staining, and culture. The remaining four CD and five UC patients gave negative results for M. *avium* subsp. *paratuberculosis*.

DISCUSSION

The most prominent target used in several studies to detect DNA of *M. avium* subsp. *paratuberculosis* by PCR is the insertion element IS*900* (19, 21). The multicopy (17 copies) nature of the sequence on the *M. avium* subsp. *paratuberculosis* chromosome makes it ideal as a target sequence for the detection of *M. avium* subsp. *paratuberculosis*, since it exhibits a higher level of sensitivity compared to the use of single-copy genes as targets (22, 45). We analyzed consecutive stool specimens of 1,293 hospitalized patients by the use of target IS*900*. Twenty-seven (2.09%) of the cohort gave positive test results for IS*900*. Only two of these patients suffered from CD. The bacterial load was persistently low and was calculated in the range of 500 to 5,000 *M. avium* subsp. *paratuberculosis* bacteria per gram of stool. In addition, an analyzed section of gut tissue from one CD patient was also positive for IS*900* and the bacterium could be isolated by culture. It was not possible to isolate *M. avium* subsp. *paratuberculosis* from the 26 other patients, since the analyses were performed retrospectively with stored DNA extracts.

Unfortunately, the specificity of target IS*900* is not 100%, since IS*900* insertion elements with close sequence homology are also present on the chromosomes of *M. cookii*, *M. marinum*, *M. paraffinicum*, and *M. scrofulaceum* isolates (7, 9, 21, 37, 40). Furthermore, polymorphisms detected in IS*900* as variants of *M. avium* subsp. *paratuberculosis* have been previously described; such variants should be interpreted as suggestive of the presence of a *Mycobacterium* organism other than *M. avium* subsp. *paratuberculosis* until the detection has been confirmed by independent methods (38). Therefore, to enhance the specificity of *M. avium* subsp. *paratuberculosis* detection, it is indispensable to use multiple *M. avium* subsp. *paratuberculosis*-specific targets. So far, several specific targets have been used by employing different techniques: IS*900* and target 251 by real-time PCR (21, 38, 44, 49), ISMap*02* by a nested PCR method (46), and f57 sequences by hybridization and by PCR (35, 44, 49). Also, the completed genome sequence of *M. avium* subsp. *paratuberculosis* strain K-10 and comparative genome analysis with the closely related species *M. avium* subsp. *avium*, including experimental studies to identify *M. avium* subsp. *paratuberculosis*-specific genomic regions, revealed miscellaneous potential new diagnostic targets (2, 19, 22, 35, 38). In order to improve detection, we used several *M. avium* subsp. *paratuberculosis*-specific targets in one assay for detection and also used quantitative TaqMan real-time PCR, since the technology is highly sensitive and specific.

For a multiple real-time PCR assay, we chose IS*900*, target 251, and the f57 sequence, which are randomly distributed on the *M. avium* subsp. *paratuberculosis* chromosome (Fig. 1, bottom). When BLASTn analysis and bioinformatic GenomeViz software were used, target 251 and sequence f57 were found in genes MAP2765c and MAP0865 of the published *M. avium* subsp. *paratuberculosis* K-10 strain, respectively (Fig. 1, top). Computer-aided analysis of the entire MAP0865 ORF revealed that it was unique for *M. avium* subsp. *paratuberculosis*, since no corresponding sequences were detected in publicly available databases. To establish real-time PCR assays for this sequence, two new TaqMan targets were generated: amplicon III, located in the f57 sequence, and amplicon IV, located at the 5′ end of ORF MAP0865 (Fig. 1, top). Each PCR was run separately and independently in corresponding unique microtiter wells but under the same conditions with respect to PCR buffer and temperature profiles. Even though the conditions were identical, the results showed that the sensitivities of the individual PCRs were different (Fig. 2B and C). The highest sensitivity was obtained with the IS*900* PCR and the lowest with the MAP2765c (251) PCR. Amplicons III (MAP0865 [f57]) and IV (MAP0865) showed identical levels of efficacy and intermediate levels of sensitivity between those of IS*900* PCR and MAP2765c (251) PCR. Because of the identical efficacy results determined under the applied conditions, only one curve, representing both amplicons, is depicted (Fig. 2C). The high sensitivity of the IS*900* PCR is attributed to the multicopy nature of this element, which is present as 17 copies on the *M. avium* subsp. *paratuberculosis* K-10 chromosome (Fig. 1, bottom). The sensitivities determined for amplicons II through IV were lower, since ORFs MAP2765c and MAP0865 are singular targets (Fig. 1, bottom). We also attribute the

Appendix

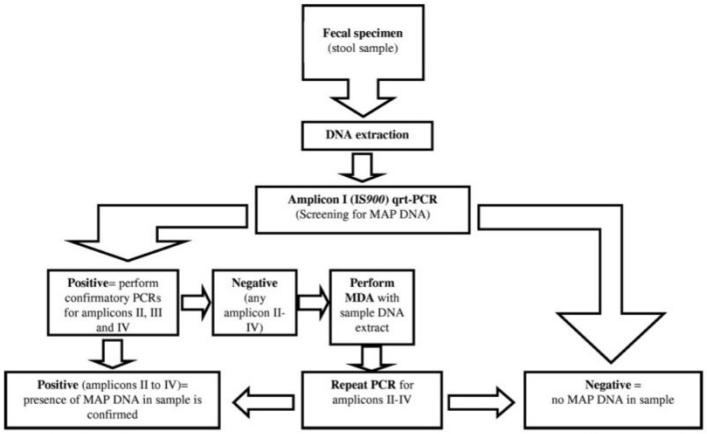


FIG. 4. Suggested workflow for the detection of *M. avium* subsp. *paratuberculosis* DNA in stool samples. MDA, multiple displacement amplification. Amplicon I = IS*900*; amplicon II = MAP2765c (251); amplicon III = MAP0865 (f57); amplicon IV = MAP0865.

lower sensitivity of the MAP2765c PCR to its size (203 bp), which affected the efficacy of the PCR methods used. The recommended size for TaqMan-based amplicons is 70 to 150 bp (36), and both amplicons III and IV of MAP0865 represent an optimal size of 101 bp, which resulted in identical efficacy results for these independent PCRs.

The specificities of newly designed quantitative real-time PCR amplicons III and IV, along with those of the previously described specific amplicons I and II, were demonstrated using 2 *M. avium* subsp. *paratuberculosis* type strains, 40 field isolates of *M. avium* subsp. *paratuberculosis* (human, animal, and environmental origin), and 13 species of Gram-positive bacteria, 20 species of Gram-negative bacteria, 7 species of anaerobic bacteria, and 2 species of intestinal parasites (Table 2). Positive signals were obtained only for the *M. avium* subsp. *paratuberculosis* strains; all of the other microorganisms tested gave uniformly negative results. The results for the pure microorganisms were achieved with DNA extracted by an RTP Spin Bacteria kit. Since stool and tissue specimens are complex and represent difficult samples for DNA extraction and PCR with respect to potential DNA-degrading enzymes and PCR inhibitors and to standardize methodology, we used commercially available stool and tissue kits from Invitek (see Materials and Methods). DNA extraction is generally achieved within 1 h. The detection limits for *M. avium* subsp. *paratuberculosis* determined using the tissue and the stool kits for DNA extraction of tissue and stool specimens were marginally above the theoretical detection limit of 1 to 10 CFU and determined to be ~10 CFU and ~70 CFU, respectively (Fig. 3).

The applicability of the DNA extraction methods and the specific quantitative real-time PCR assays used was assessed by acid-fast staining and by culture using tissue and stool specimens of three healthy cattle and of nine diseased cattle with Johne's disease. The real-time PCR results were 100% identical to the results of microscopic analysis and the culture (Table 3) but could be accomplished in 6 h. The quantification obtained by qrt-PCR showed that the examined cattle shed very large amounts of *M. avium* subsp. *paratuberculosis* calculated to be in the range of approximately 1×10^7 to 2×10^9 bacteria per gram of feces, which was ~20,000- through ~400,000-fold higher than in human feces. Fecal samples from cattle are easily obtained and are therefore superior to other specimens such as biopsy specimens or blood samples for analysis. In particular, the results we obtained from gut tissue of infected cattle strongly depended on the biopsy specimens used. Noninflamed gut tissue from diseased cattle invariably gave negative results, whereas inflamed gut tissue gave positive results (Table 3).

The presence of *M. avium* subsp. *paratuberculosis* in human stool specimens and one biopsy specimen as detected with target IS*900* was confirmed by reamplifying the samples with amplicons II through IV. This means that all samples that gave IS*900*-positive test results truly represented *M. avium* subsp. *paratuberculosis*. We are well aware that our human samples were single snapshots randomly collected and want to emphasize that our results for CD and UC patients are not meant to be representative. However, the results clearly demonstrate that the technology of DNA extraction and qrt-PCR utilized is appropriate for analysis of stool specimens and gut tissue for the presence of *M. avium* subsp. *paratuberculosis* and could therefore substantially contribute to the current debate about the role of *M. avium* subsp. *paratuberculosis* in CD. Further studies are necessary, and we suggest using several consecutive stool specimens, which are noninvasive compared to biopsy specimens and with which we were able to demonstrate high sensitivity and specificity for the detection of *M. avium* subsp. *paratuberculosis*.

In conclusion, we demonstrate the development of qrt-PCR amplicons III and IV, which, in combination with amplicons I and II, enable unequivocal detection, identification, and quantification of *M. avium* subsp. *paratuberculosis* directly from clinical, veterinary, food, and environmental specimens as well as from pure cultures. For IS*900*-positive but amplicon II- to IV-negative specimens, we suggest the use of MDA and the consecutive repetition of the confirmatory PCRs. A suggested workflow is shown in Fig. 4. Fecal specimens are true alterna-

Appendix

tives for the analysis of patients and cattle because of their noninvasiveness and simplicity in collection. Our data presented here provide a basis for further structured studies of the potential role of *M. avium* subsp. *paratuberculosis* in CD and other human diseases.

ACKNOWLEDGMENTS

We thank Martina Klös-Langsdorf and Kirsten-Susann Bommersheim for excellent technical assistance, Isabell Krabs for the preparation of the gut tissue of slaughtered cows, Simone Ries for running qrt-PCRs during a practical course, and Egbert Tannich from the Bernhard-Nocht Institute for Tropical Medicine (Hamburg, Germany) for providing parasite DNA obtained from *Entamoeba histolytica* and from *Giardia lamblia*.

This work was supported by grants from the Bundesministerium fuer Bildung und Forschung, Germany, within the framework of the National Genome Research Network (NGFN) (contract no. 01GS0401).

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Bannantine, J., P. E. Baechler, Q. Zhang, L. Li, and V. Kapur.** 2002. Genome scale comparison of *Mycobacterium avium* subsp. *paratuberculosis* with *Mycobacterium avium* subsp. *avium* reveals potential diagnostic sequences. J. Clin. Microbiol. **40:**1303–1310.
3. **Chapin, K. C., and T.-L. Lauderdale.** 2007. Reagents, stains, and media: bacteriology, p. 334–364. *In* P. R. Murray, E. J. Baron, J. H. Jorgensen, M. L. Landry, and M. A. Pfaller (ed.), Manual of clinical microbiology, 9th ed. ASM Press, Washington, DC.
4. **Chiodini, R. J.** 1989. Crohn's disease and the mycobacterioses: a review and comparison of two disease entities. Clin. Microbiol. Rev. **2:**90–117.
5. **Collins, D. M., and G. W. de Lisle.** 1986. Restriction endonuclease analysis of various strains of *Mycobacterium paratuberculosis* isolated from cattle. Am. J. Vet. Res. **47:**2226–2229.
6. **Committee on Diagnosis and Control of Johne's Disease.** 2003. Johne's disease and Crohn's disease, p. 104–120. *In* Diagnosis and control of Johne's disease: report for the National Research Council of the National Academy of Science. The National Academic Press, Washington, DC.
7. **Cousins, D. V., et al.** 1999. Mycobacteria distinct from *Mycobacterium avium* subsp. *paratuberculosis* isolated from the faeces of ruminants possess IS*900*-like sequences detectable IS*900* polymerase chain reaction: implications for diagnosis. Mol. Cell Probes **13:**431–442.
8. **Dalziel, T. K.** 1913. Chronic interstitial enteritis. Br. Med. J. **2:**1068–1070.
9. **Englund, S., G. Bolske, and K. E. Johansson.** 2002. An IS*900*-like sequence found in a *Mycobacterium* sp. other than *Mycobacterium avium* subsp. *paratuberculosis*. FEMS Microbiol. Lett. **209:**267–271.
10. **Gao, A., L. Mutharia, S. Chen, K. Rahn, and J. Odumeru.** 2002. Effect of pasteurisation on survival of *Mycobacterium paratuberculosis* in milk. J. Dairy Sci. **85:**3198–3205.
11. **Ghai, R., T. Hain, and T. Chakraborty.** 2004. GenomeViz: visualizing microbial genomes. BMC Bioinformatics **5:**198.
12. **Grant, I. R.** 2005. Zoonotic potential of *Mycobacterium avium* subsp. *paratuberculosis*: the current position. J. Appl. Microbiol. **98:**1282–1293.
13. **Green, E. P., et al.** 1989. Sequence and characteristics of IS*900*, an insertion element identified in a human Crohn's disease isolate of *Mycobacterium paratuberculosis*. Nucleic Acids Res. **17:**9063–9073.
14. **Harris, N. B., and R. G. Barletta.** 2001. *Mycobacterium avium* subsp. *paratuberculosis* in veterinary medicine. Clin. Microbiol. Rev. **14:**489–512.
15. **Hermon-Taylor, J.** 2001. Protagonist: *Mycobacterium avium* subspecies *paratuberculosis* is a cause of Crohn's disease. Gut **49:**755–756.
16. **Hermon-Taylor, J., and T. J. Bull.** 2002. Crohn's disease caused by *Mycobacterium avium* subsp. *paratuberculosis*: a public health tragedy whose resolution is long overdue. J. Med. Microbiol. **51:**3–6.
17. **Hermon-Taylor, J., et al.** 2000. Causation of Crohn's disease by *Mycobacterium avium* subspecies *paratuberculosis*. Can. J. Gastroenterol. **14:**521–539.
18. **Hoorfar, J., P. Ahrens, and P. Radström.** 2000. Automated 5′ nuclease PCR assay for identification of *Salmonella enterica*. J. Clin. Microbiol. **38:**3429–3435.
19. **Hruska, K., M. Bartos, P. Kralik, and I. Pavlik.** 2005. *Mycobacterium avium* subsp. *paratuberculosis* in powdered infant milk: paratuberculosis in cattle—the public health problem to be solved. Vet. Med.-Czech **50:**327–335.
20. **Huntley, J. F. J., R. H. Whitlock, J. P. Bannantine, and J. R. Stabel.** 2005. Comparison of diagnostic detection methods for *M. avium* subsp. *paratuberculosis* in North American bison. Vet. Pathol. **42:**42–51.
21. **Kim, S. G., et al.** 2002. Development and application of quantitative polymerase chain reaction assay based on the ABI 7700 system (TaqMan) for detection and quantification of *Mycobacterium avium* subsp. *paratuberculosis*. J. Vet. Diagn. Invest. **14:**126–131.
22. **Li, L., et al.** 2005. The complete genome sequence of *Mycobacterium avium* subspecies *paratuberculosis*. Proc. Natl. Acad. Sci. U. S. A. **102:**12344–12349.
23. **Martin, F. E., M. A. Nadkarni, N. A. Jacques, and N. Hunter.** 2002. Quantitative microbiological study of human carious dentine by culture and real-time PCR: association of anaerobes with histopathological changes in chronic pulpitis. J. Clin. Microbiol. **40:**1698–1704.
24. **McFadden, J. J., P. D. Butcher, R. J. Chiodini, and J. Hermon-Taylor.** 1987. Crohn's disease-isolated mycobacteria are identical to *Mycobacterium paratuberculosis*, as determined by DNA probes that distinguish between mycobacterial species. J. Clin. Microbiol. **25:**796–801.
25. **Mendoza, J. L., R. Lana, and M. Díaz-Rubio.** 2009. *Mycobacterium avium* subspecies *paratuberculosis* and its relationship with Crohn's disease. World J. Gastroenterol. **15:**417–422.
26. **Müller, A., et al.** 2000. Detection of *Isospora belli* by polymerase chain reaction using primers based on small-subunit ribosomal RNA sequences. Eur. J. Clin. Microbiol. Infect. Dis. **19:**631–634.
27. **Naser, S. A., D. Schwartz, and I. Shafran.** 2000. Isolation of *Mycobacterium avium* subsp. *paratuberculosis* from breast milk of Crohn's disease patients. Am. J. Gastroenterol. **95:**1094–1095.
28. **Naser, S. A., G. Ghobrial, C. Romero, and J. F. Valentine.** 2004. Culture of *Mycobacterium avium* subspecies *paratuberculosis* from the blood of patients with Crohn's disease. Lancet **364:**1039–1044.
29. **Nogva, H. K., A. Bergh, A. Holck, and K. Rudi.** 2000. Application of the 5′-nuclease PCR assay in evaluation and development of methods for quantitative detection of *Campylobacter jejuni*. Appl. Environ. Microbiol. **66:**4029–4036.
30. **Olsen, I., G. Sigurgardottir, and B. Djonne.** 2002. Paratuberculosis with special reference to cattle. Vet. Q. **24:**12–28.
31. **Ott, S. L., S. J. Wells, and B. A. Wagner.** 1999. Herd-level economic losses associated with Johne's disease on US dairy operations. Prev. Vet. Med. **40:**179–192.
32. **Parrish, N. M., et al.** 2009. Absence of *Mycobacterium avium* subsp. *paratuberculosis* in Crohn's patients. Inflamm. Bowel Dis. **15:**558–565.
33. **Pfyffer, G. E.** 2007. *Mycobacterium*: general characteristics, laboratory detection, and staining procedures, p. 543–572. *In* P. R. Murray, E. J. Baron, J. H. Jorgensen, M. L. Landry, and M. A. Pfaller (ed.), Manual of clinical microbiology, 9th ed. American Society for Microbiology Press, Washington, DC.
34. **Pierce, E. S.** 2009. Possible transmission of *Mycobacterium avium* subspecies *paratuberculosis* through potable water: lessons from an urban cluster of Crohn's disease. Gut Pathog. **23:**17.
35. **Poupart, P., M. Coene, H. Vanheuverswyn, and C. Cocito.** 1993. Preparation of a specific RNA probe for detection of *Mycobacterium paratuberculosis* and diagnosis of Johne's disease. J. Clin. Microbiol. **31:**1601–1605.
36. **Proudnikov, D., et al.** 2003. Optimizing primer-probe design for fluorescent PCR. J. Neurosci. Methods **123:**31–45.
37. **Quirke, P.** 2001. Antagonist: *Mycobacterium avium* subspecies *paratuberculosis* is a cause of Crohn's disease. Gut **49:**757–760.
38. **Rajeev, S., Y. Zhang, S. Sreevatsan, A. S. Motiwala, and B. Byrum.** 2005. Evaluation of multiple genomic targets for identification and confirmation of *Mycobacterium avium* subsp. *paratuberculosis* isolates using real-time PCR. Vet. Microbiol. **105:**215–221.
39. **Schwartz, D., et al.** 2000. Use of short-term culture for identification of *M. avium* subsp. *paratuberculosis* in tissue from Crohn's disease patients. Clin. Microbiol. Infect. **6:**303–307.
40. **Semret, M., C. Y. Turenne, and M. A. Behr.** 2006. Insertion sequence IS*900* revisited. J. Clin. Microbiol. **44:**1081–1083.
41. **Shankar, H., et al.** 2010. Presence, characterization, and genotype profiles of *Mycobacterium avium* subspecies *paratuberculosis* from unpasteurized individual and pooled milk, commercial pasteurized milk, and milk products in India by culture, PCR, and PCR-REA methods. Int. J. Infect. Dis. **14:**121–126.
42. **Shivananda, S., et al.** 1996. Incidence of inflammatory bowel disease across Europe: is there a difference between north and south? Results of the European Collaborative Study in Inflammatory Bowel Disease (EC-IBD). Gut **39:**690–697.
43. **Sibartie, S., et al.** 2010. *Mycobacterium avium* subsp. *paratuberculosis* (MAP) as a modifying factor in Crohn's disease. Inflamm. Bowel Dis. **16:**296–304.
44. **Slana, I., M. Liapi, M. Moravkova, A. Kralova, and I. Pavlik.** 2009. *Mycobacterium avium* subsp. *paratuberculosis* in cow bulk tank milk in Cyprus detected by culture and quantitative IS*900* and F57 real-time PCR. Prev. Vet. Med. **89:**223–226.
45. **Soumya, M. P., R. M. Pillai, P. X. Antony, H. K. Mukhopadhyay, and V. N. Rao.** 2009. Comparison of faecal culture and IS*900* PCR assay for the detection of *Mycobacterium avium* subsp. *paratuberculosis* in bovine faecal samples. Vet. Res. Commun. **33:**781–791.
46. **Stabel, J. R., and J. P. Bannantine.** 2005. Development of a nested PCR method targeting a unique multicopy element, ISMap*02*, for detection of *Mycobacterium avium* subsp. *paratuberculosis* in fecal samples. J. Clin. Microbiol. **43:**4744–4750.
47. **Sung, N., and M. T. Collins.** 1998. Thermal tolerance of *Mycobacterium paratuberculosis*. Appl. Environ. Microbiol. **64:**999–1005.
48. **Sung, N., and M. T. Collins.** 2000. Effect of three factors in cheese production (pH, salt and heat) on *Mycobacterium avium* subsp. *paratuberculosis* viability. Appl. Environ. Microbiol. **66:**1334–1339.
49. **Tasara, T., and R. Stephan.** 2005. Development of an F57 sequence-based real-time PCR assay for detection of *M. avium* subsp. *paratuberculosis* in milk. Appl. Environ. Microbiol. **71:**5957–5968.
50. **Thomas Dow, C.** 2008. Cows, Crohn's and more: is *Mycobacterium paratuberculosis* a superantigen? Med. Hypotheses **71:**858–861.
51. **Thorel, M. F., M. Krichevsky, and V. V. Levy-Frebault.** 1990. Numerical taxonomy of mycobactin-dependent mycobacteria, emended description of *Mycobacterium avium*, and description of *Mycobacterium avium* subsp. *avium* subsp. nov., *Mycobacterium avium* subsp. *paratuberculosis* subsp. nov., and *Mycobacterium avium* subsp. *silvaticum* subsp. nov. Int. J. Syst. Bacteriol. **40:**254–260.
52. **van den Berg, R. J., et al.** 2005. Prospective multicenter evaluation of a new immunoassay and real-time PCR for rapid diagnosis of *Clostridium difficile*-associated diarrhea in hospitalized patients. J. Clin. Microbiol. **43:**5338–5340.
53. **Verweij, J. J., et al.** 2004. Simultaneous detection of *Entamoeba histolytica*, *Giardia lamblia*, and *Cryptosporidium parvum* in fecal samples by using multiplex real-time PCR. J. Clin. Microbiol. **42:**1220–1223.
54. **Vishnubhatla, A., et al.** 2000. Rapid 5′ nuclease (TaqMan) assay for detection of virulent strains of *Yersinia enterocolitica*. Appl. Environ. Microbiol. **66:**4131–4135.
55. **Whittington, R. J., D. J. Marshall, P. J. Nicholls, I. B. Marsh, and L. A. Reddacliff.** 2004. Survival and dormancy of *Mycobacterium avium* subsp. *paratuberculosis* in the environment. Appl. Environ. Microbiol. **70:**2989–3004.

13.8 Appendix publication 7

OPEN ACCESS Freely available online

PLoS one

Universal Stress Proteins Are Important for Oxidative and Acid Stress Resistance and Growth of *Listeria monocytogenes* EGD-e *In Vitro* and *In Vivo*

Christa Seifart Gomes[1☯], Benjamin Izar[1☯], Farhad Pazan[1☯], Walid Mohamed[1], Mobarak Abu Mraheil[1], Krishnendu Mukherjee[2], André Billion[1], Yair Aharonowitz[3], Trinad Chakraborty[1], Torsten Hain[1]*

1 Institute of Medical Microbiology, Justus-Liebig-University, Giessen, Germany, **2** Institute of Phytopathology and Applied Zoology, Justus-Liebig-University, Giessen, Germany, **3** Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Tel Aviv, Israel

Abstract

Background: Pathogenic bacteria maintain a multifaceted apparatus to resist damage caused by external stimuli. As part of this, the universal stress protein A (UspA) and its homologues, initially discovered in *Escherichia coli* K-12 were shown to possess an important role in stress resistance and growth in several bacterial species.

Methods and Findings: We conducted a study to assess the role of three homologous proteins containing the UspA domain in the facultative intracellular human pathogen *Listeria monocytogenes* under different stress conditions. The growth properties of three UspA deletion mutants (Δ*lmo0515*, Δ*lmo1580* and Δ*lmo2673*) were examined either following challenge with a sublethal concentration of hydrogen peroxide or under acidic conditions. We also examined their ability for intracellular survival within murine macrophages. Virulence and growth of *usp* mutants were further characterized in invertebrate and vertebrate infection models. Tolerance to acidic stress was clearly reduced in Δ*lmo1580* and Δ*lmo0515*, while oxidative stress dramatically diminished growth in all mutants. Survival within macrophages was significantly decreased in Δ*lmo1580* and Δ*lmo2673* as compared to the wild-type strain. Viability of infected *Galleria mellonella larvae* was markedly higher when injected with Δ*lmo1580* or Δ*lmo2673* as compared to wild-type strain inoculation, indicating impaired virulence of bacteria lacking these *usp* genes. Finally, we observed severely restricted growth of all chromosomal deletion mutants in mice livers and spleens as compared to the load of wild-type bacteria following infection.

Conclusion: This work provides distinct evidence that universal stress proteins are strongly involved in listerial stress response and survival under both *in vitro* and *in vivo* growth conditions.

Citation: Seifart Gomes C, Izar B, Pazan F, Mohamed W, Mraheil MA, et al. (2011) Universal Stress Proteins Are Important for Oxidative and Acid Stress Resistance and Growth of *Listeria monocytogenes* EGD-e *In Vitro* and *In Vivo*. PLoS ONE 6(9): e24965. doi:10.1371/journal.pone.0024965

Editor: Pere-Joan Cardona, Fundació Institut Germans Trias i Pujol; Universitat Autònoma de Barcelona CibeRES, Spain

Received May 9, 2011; **Accepted** August 25, 2011; **Published** September 30, 2011

Copyright: © 2011 Seifart Gomes et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the German Federal Ministry of Education and Research (BMBF, ERA-NET PathoGenomics Network projects sncRNAomics 0315437A to TH and LISTRESS 0315907A to TH and TC) and through the LOEWE program of the state Hessia by the collaborative research project Insect Biotechnology to TH and TC. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: torsten.hain@mikrobio.med.uni-giessen.de

☯ These authors contributed equally to this work.

Introduction

Universal stress proteins (Usps) comprise a group of proteins induced by different stress conditions and are found in numerous prokaryotic as well as eukaryotic organisms [1,2]. Among these, universal stress protein A (UspA) of *Escherichia coli* K-12 is best characterized and found to be highly expressed in response to heat, substrate starvation, exposure to antimicrobial agents and oxidative stress [1].

Subsequently, additional Usps have been described for *E. coli* K-12 and several other bacterial species, including *Haemophilus influenza* [3,4], *Mycobacterium smegmatis* [5], *Mycobacterium tuberculosis* [2,6], *Pseudomonas aeruginosa* [7 9], *Porphyromonas gingivialis* [10], *Shigella sonnei* [11], *Salmonella typhimurium* [12 14] and *Lactobacillus plantarum* [15,16].

The majority of *usp* genes are monocistronically expressed, and different transcription factors, such as σ^{32}, σ^{70} and σ^{E} promote transcription of Usps [1].

The significance of Usps in the model pathogen *Listeria monocytogenes* is presently unknown. *L. monocytogenes* is a ubiquitously occuring gram-positive, facultative intracellular bacterium that causes food-borne infections, which mainly affect pregnant women, newborns, elderly and immunocompromised patients [17,18]. *L. monocytogenes* possesses the remarkable ability to grow under a wide range of temperatures, pH conditions and high osmolarity, allowing the pathogen to survive in nature, food preservation methods as well as in infected host cells [18].

Here we present first evidence that UspA domain containing stress proteins are of importance for *L. monocytogenes* to survive under different stress conditions both *in vitro* and *in vivo*.

Materials and Methods

Bacterial strains, plasmids and culture conditions

All bacterial strains and plasmids used in this study are listed in Table 1.

Appendix

Table 1. Bacterial strains and plasmids used in this study.

Strain or plasmid	Description	Reference or source
L. monocytogenes EGD-e	wild-type	[19]
DH10β	electrocompetent	Life technologies
Δ*lmo*0515	*lmo*0515 deletion (379 nt) strain of EGD-e	This study
Δ*lmo*1580	*lmo*1580 deletion (409 nt) strain of EGD-e	This study
Δ*lmo*2673	*lmo*2673 deletion (471 nt) strain of EGD-e	This study
Δ*lmo*0515Δ*lmo*1580Δ*lmo*2673	*lmo*0515, *lmo*1580 and *lmo*2673 deletion strain of EGD-e	This study
Δ*lmo*0515+pPL2*lmo*0515	complementation of *lmo0515* deletion strain of EGD-e	This study
Δ*lmo*1580+pPL2*lmo*1580	complementation of *lmo1580* deletion strain of EGD-e	This study
Δ*lmo*2673+pPL2*lmo*2673	complementation of *lmo2673* deletion strain of EGD-e	This study
pPL2	site specific phage integration vector	[23]
pCR2.1-TOPO®	single 3′-thymidine (T) overhangs for TA Cloning® 3.9 kb	Life technologies
pAUL-A	temperature sensitive shuttle vector 9.2 kb	[21]

doi:10.1371/journal.pone.0024965.t001

The wild type strain *L. monocytogenes* EGD-e serotype 1/2a [19] and its isogenic *usp* deletion mutants or complemented strains were cultivated aerobically in Brain Heart Infusion medium (BHI; Difco) at 37°C or on BHI agar plates at 37°C.

E. coli strain (DH10β^{M}) was grown in Luria-Bertani broth or LB agar plates at 37°C. When required, antibiotics were added to the following concentrations: erythromycin (Sigma), 300 µg/ml for *E. coli* and 5 or 10 µg/ml for *Listeria*; ampicillin (Sigma), 50 µg/ml for *E. coli* and 200 µg/ml for *Listeria*.

Construction of chromosomal *usp* deletion mutants

Generation of the *usp* in frame deletion mutants was done as previously described [20 22]. Briefly, the flanking regions of each *usp* gene were amplified by PCR (*Taq* polymerase; Fermentas) using primer pairs 1 and 2b for the 5′ flanking region and 3b and 4 for the 3′ flanking region (Table 2), respectively. The resulting PCR products were fused to each other by using primers 1 and 4 in a second PCR reaction with Taq polymerase and the product was cloned into pCR2.1-TOPO®.

For the generation of Δ*lmo0515* and Δ*lmo2673* the pCR2.1-TOPO® vector containing the flanking region of the respective gene was digested with *BamH*I, *Xho*I and *Nco*I (Fermentas) and inserted into pAUL-A [21], that was previously digested with *BamH*I and *Sal*I (Fermentas). The pCR2.1-TOPO® vector containing the flanking region of Δ*lmo1580* was digested with *Xba*I and *Sac*I (Fermentas) and ligated with pAUL-A, also previously digested with the same restriction endonucleases.

We designated resulting pAUL-A vectors containing the flanking regions of the respective *usp* genes, pCS1 for Δ*lmo0515*, pCS2 for Δ*lmo1580* and pCS3 for Δ*lmo2673*. These vectors were transformed into *E. coli* and were isolated, sequenced and subsequently electroporated into *L. monocytogenes* wild-type strain. Gene replacement was performed as previously described by Schaeferkordt et al. [21,22]. To create a triple mutant (Δ*lmo0515*Δ*lmo1580*Δ*lmo2673*) for all three *usp* genes, gene replacement were carried out as described above starting with the electroporation of pCS2 into Δ*lmo0515*. Finally, pCS3 was electroporated into the resulting double mutant to generate the isogenic triple mutant (Δ*lmo0515*Δ*lmo1580*Δ*lmo2673*). The chromosomal deletion was confirmed by DNA sequencing of PCR products using primer 7 and 8 (Table 2).

Table 2. Primers used in this study.

Name	5′-3′ Sequence
lmo0515-1	GCATTGCCACAAACTGGTGA
lmo0515-2b	AAGTGTCGCTGCTACAAGAATGCGATG
lmo0515-3b	CTTGTAGCAGCGACACTTTAATAGCTA
lmo0515-4	TTTTGGGGCTGATCCTACGC
lmo0515-5	ATTATTGTGAATTCATTAG
lmo0515-6	TACTTCGATCCATTTTGAT
lmo0515-7	CGACAGTGGACATGTTGATC
lmo0515-8	ACCCAATTTGGTCATGCGAT
lmo1580-1	GGGTCATTGCCACCCTATT
lmo1580-2b	ACATCGCAAGAATAATTCCTCCAATCAAAA
lmo1580-3b	AGGAATTATTCTTGCGATGTTCTTGTAGTTCG
lmo1580-4	GTTTTCGTTGACCGTATCCA
lmo1580-5	CTAAATCACTCTCCTCGTTA
lmo1580-6b	GCACTTAACCAAGTGCGCG
lmo1580-7	TCTACAATTTTTGCTCCCGC
lmo1580-8	GTAAAACTGCAGAACAAGTA
lmo2673-1	TAAGAGCTGCACTCGGTAGA
lmo2673-2b	TATACATGAATTTGTATCACCCTCTCAAAAGTTTTC
lmo2673-3b	GTGATACAAATTCATGTATAATGAAGGTATTG
lmo2673-4	CCAACAAGCGCGCAACGAATACC
lmo2673-5	AAAGAGACCCCCTTTTCCTC
lmo2673-6	TTTTTCTCCTCCTGCCGTAT
lmo2673-7	AGCATCCGTCACTAGCCCTG
lmo2673-8	GCTATCTTCGTAAGCAGTGA
lmo0515-forward	GATGGTTCAGAACCAGCAAA
lmo0515-reverse	GCTTTTTCTTCTAAGCTGCCAT
lmo1580-forward	GCAGTTGATGGATCCAAAGAA
lmo1580-reverse	TTTATCCGCCATGCTTGTATC
lmo2673-forward	ATAACGCGAAATTTTGAACCCG
lmo2673-reverse	CCAAGGTTTTGCGTGAACAA

doi:10.1371/journal.pone.0024965.t002

The complementation of the Δ*lmo0515*, Δ*lmo1580* and Δ*lmo*2673 deletion mutants was carried out using the *L. monocytogenes* site specific phage integration vector pPL2 [23]. PCR products were generated using primer pairs 5 and 6 with Taq polymerase. The product was cloned into pCR 2.1-TOPO®. The pCR 2.1-TOPO® vector containing the *usp* gene region of *lmo0515* or *lmo2673* was digested with *BamH*I and *Xho*I (Fermentas) and for *lmo1580 with Xho*I and *Sac*I (Fermentas), respectively. The insert was ligated with pPL2, also previously digested with the same restriction endonucleases. Primer sequences used for the complementation of the deleted genes are listed in table 2.

Acid stress assay

Single colonies of the wild type *L. monocytogenes* and its *usp* deletion mutants grown on BHI agar plates were used to inoculate 100 ml Erlenmeyer flasks containing 10 ml of BHI broth, followed by overnight incubation at 37°C with shaking (180 rpm, Unitron Infors). The overnight cultures were then subcultured (1:50) in 40 ml of BHI and grown to an optical density at 600 nm (OD_{600}) of 0.4 prior to our experiments. Then each culture was divided into two aliquots, 20 ml each and cells were harvested by centrifugation at 6,000×*g* for 15 min. at 37°C. Furthermore the supernatant was removed and cells were resuspended in a BHI broth that was acidified with a 6 N HCL solution (Sigma) to pH 2.5. These tubes were incubated with shaking (180 rpm) at 37°C. At time point zero (t = 0) and after 10, 20 and 30 minutes intervals samples were taken from each culture, serially diluted with phosphate-buffered saline (1×PBS) and plated on BHI agar plates. The colony forming units (CFU) were counted after overnight incubation at 37°C. The percentage of survivors was calculated by comparing the number of survivors in BHI (pH 2.5) after 10, 20 and 30 minutes with the number of cells in BHI (pH 2.5) at t = 0, respectively. Results are reported as means of data collected from three independent biological replicates.

Hydrogen peroxide sensitivity assay

BHI broth was used to dilute a 3% hydrogen peroxide (H_2O_2) solution (Ottmar Fischer, Germany) to obtain 0.045% H_2O_2 final concentration.

Overnight cultures in 10 ml BHI broth (in 100 ml Erlenmeyer flasks) were prepared and were diluted 1:50 in 20 ml fresh BHI broth or in 20 ml of different fresh BHI supplemented with 0.045% H_2O_2 (in 100 ml Erlenmeyer flasks) as mentioned before. The optical density (OD_{600}) of the bacterial culture was measured at the indicated times after incubation at 37°C (180 rpm, Unitron, Infors).

Macrophage survival assay

P388D1 murine macrophages, cultured in 24-well plates in RPMI (Life Technologies) supplemented with 10% fetal calf serum (FCS) and 1% nonessential amino acids (NEAs), were infected with 1×10^7 bacteria to obtain an MOI of 10. After 30 min. RPMI was removed, and cells were washed twice with 1×PBS and incubated for 1 h in RPMI medium containing 50 µg/ml gentamicin. Macrophages were then washed three times with 1×PBS and lysed with ice-cold 0.2% Triton X-100 in H_2O. The released bacteria were plated on BHI agar plates in appropriate dilutions and quantified after overnight incubation at 37°C.

For the determination of macrophage survival rates the CFU numbers in the inoculum of the wild-type and the mutants used to infect P388D1 were compared and a factor was calculated. The wild-type factor was set to 1 and depending on the CFU numbers of the mutant inoculums their factor was <1 or >1. For the calculation of the survival in % we set the CFU number of wild-type to 100% and percentage survival of the mutants was set in relation to wild-type.

cDNA synthesis and quantitative Real-Time Polymerase Chain Reaction

Messenger RNA levels of *lmo0515*, *lmo1580* and *lmo2673* isolated from extracellular grown bacteria in BHI and intracellular replicating bacteria in murine macrophages as previously described [24] were assessed in *L. monocytogenes* wild-type strain using quantitative Real-Time-PCR (qRT-PCR). Complementary DNA (cDNA) was obtained from bacteria from each setting by reverese transcription of RNA isolated and purified using the RNeasy Mini kit (Qiagen). Triplicates were performed for each forward and reverse primer pair combination. Primers were purchased from Qiagen (Quantitect Primer Assay) (see Table 2).

Primers were diluted to 1 pmol/µl for further procedure and qRT-PCR was run (7900 HT Fast Real Time System, Applied Biosystems). 16sRNA was used as for normalization and calculation of relative expression.

Threshold cycle values (CT) of the tested genes were determined and normalized expression of each target gene was given as the ΔCT between the log2 transformed CT of the target gene and the log2 transformed CT of the internal control (ACTB). Log2 transformed gene expression levels (ΔCT) of each target gene for i5ntra- and extracellular bacteria were expressed as log2 differences from control (= log2 ΔΔCT method). Data was acquired and analyzed with the SDS 2.3 and RQ-Manager 1.2, respectively.

Animals and insects

Six to eight week-old female BALB/c mice, purchased from Harlan Winkelmann (Borchen, Germany), were kept at our breeding facilities in specific-pathogen-free conditions and used in all experiments. *Galleria mellonella* larvae, purchased from fauna topics (Marbach, Germany) were reared at 32°C in darkness and on an artificial diet (22% maize meal, 22% wheat germ, 11% dry yeast, 17.5% bee wax, 11% honey and 11% glycerine) prior to use. Last instar larvae, each weighing between 250 and 350 mg, were used in all experiments [20].

Insect and animal models of infection

In all experiments, fresh cultures of bacteria, prepared from an overnight culture, were used. Briefly, bacteria were grown in Brain heart Infusion (BHI) at 37°C, harvested in the exponential growth phase and washed twice with 1×PBS. The pellet was resuspended in 1×PBS and the bacterial concentration was calibrated by optical absorption. Further dilutions were prepared in 1×PBS to obtain required numbers of bacteria for infection. The infection of *G. mellonella* was performed as previously described by Mukkerjee et al [20]. Briefly, the human pathogenic *L. monocytogenes* and its isogenic *usp* deletion mutants or complemeted strains were separately injected (10^6 CFU/larva) into the hemocoel of the last instar larvae and the infection was monitored at 37°C.

In all experiments, fresh cultures of bacteria, prepared from an overnight culture, were used. Briefly, bacteria were grown in Brain heart Infusion (BHI) at 37°C, harvested in the exponential growth phase and washed twice with 1×PBS. The pellet was resuspended in 1×PBS and the bacterial concentration was calibrated by optical absorption. Further dilutions were prepared in 1×PBS to obtain required numbers of bacteria for infection.

Primary mice infections *in vivo* infection with *L. monocytogenes* wild-type, Δ*lmo0515*, Δ*lmo1580*, Δ*lmo2673* or Δ*lmo0515*Δ*l-*

*mo1580*Δ*lmo2673* were performed by an intravenous injection of viable bacteria in a volume of 0.2 ml of 1×PBS. Spleens and livers were harvested three days after infection. Bacterial growth in spleens and livers was determined by plating 10-fold serial dilutions of the organ homogenates on BHI agar plates. Colonies were counted after 24 h of incubation at 37°C.

Ethics statement

This study was carried out in strict accordance with the regulation of the National Protection Animal Act (§7-9a Tierschutzgesetz). The protocol was approved by the local Committee on the Ethics of Animal Experiments (Regierungsbezirk Mittelhessen) and permission was given by the local authority (Regierungspraesidium Giessen, Permit Number: GI 15/5-Nr.63/2007).

Statistical analysis

All experimental work was performed for a minimum of three times. Significant differences between two values were compared with a paired Student's *t*-test. Values were considered significantly different when $P<0.05$.

Results

Comparative genomic analysis and promoter prediction

All three Usps of *L. monocytogenes* harbor an UspA domain based on the Pfam analysis (data not shown), and thus represent paralogues of the UspA family of *E. coli* [1]. The structure and function of UspA is highly conserved among several bacteria and eukaryotic organisms. In line with this, comparative analysis in 19 listerial strains including all serotypes using GECO [25] (cut-off protein identity 80%, 90%) revealed a common chromosomal location for all three *usp* genes (Figure S1, S2, S3). Previously identified regulatory RNAs [26] were not located in the flanking regions of the *usp* genes. Furthermore, promoter box analysis of the upstream regions indicated that all *usp* genes habour a σ^B box upstream of their trancriptional start sites (Figure 1).

Universal stress proteins contribute to extracellular survival of *L. monocytogenes* at low pH condition

L. monocytogenes is able to resist and survive at low pH conditions, which may occur in acidic food, in the gastric milieu and within the macrophage phagosome [27,28]. In order to assess the role of listerial Usps at low pH environment we challenged exponentially grown *L. monocytogenes* and its isogenic *usp* deletion mutants (Δ*lmo0515*, Δ*lmo1580* and Δ*lmo2673*) with acidic BHI-medium (pH = 2.5) and determined the percentage of CFU at 10, 20 and 30 min after inoculation (Figure 2).

The CFU counts of both, Δ*lmo0515* and Δ*lmo1580* were significantly lower compared to wild-type bacteria at each time point. We observed no significant difference in survival rate between wild-type and Δ*lmo2673*. Δ*lmo1580* was most susceptible to an increase in acid over the time course which was reflected by a more than 100 fold lower survival rate of the culture as compared to the wild-type strain.

Universal stress proteins are crucial for oxidative stress resistance of *L. monocytogenes*

L. monocytogenes is known to be resistant to oxidative stress, occurring in the macrophage phagosome, where reactive oxygen species (ROS), such as H_2O_2 are produced to kill engulfed pathogens [29].

To elucidate the role of listerial Usps in response to oxidative stress, we determined the survival of Δ*lmo0515*, Δ*lmo1580* and Δ*lmo2673* using optical density measurement in BHI medium supplemented with H_2O_2 following a 3 h incubation period (Figure 3).

Exposure to H_2O_2 significantly impaired the growth of the mutants compared to the wild-type strain ($OD_{600} = 0.55$). While growth attenuation was moderate in Δ*lmo0515* ($OD_{600} = 0.3$), a dramatic effect was seen for the *lmo2673* mutant ($OD_{600} = 0.16$). The strongest effect was observed in Δ*lmo1580* with barely any growth detectable at any observation point. These results indicate a gradual order in susceptibility to oxidative stress in the three *usp* mutants as follows: Δ*lmo1580*>Δ*lmo2673*>Δ*lmo0515*.

Survival attenuation of Δ*lmo1580* and Δ*lmo2673* in murine macrophages

Macrophages play a key role in the phagocytosis and intracellular killing of *L. monocytogenes* by generating ROS and acidic conditions in the phagosome. To examine the impact of Usps for intracellular survival, we determined the number of CFU isolated from P388D1 murine macrophages upon infection with either wild-type strain or the *usp* deletion mutants [30].

In agreement with our results at low pH condition and the oxidative stress response, the Δ*lmo1580* deletion had the strongest effect on intracellular listerial survival. Compared to the wild-type strain, only 72% of Δ*lmo1580* were able to resist the hostile macrophage environment. Δ*lmo2673* exhibited similar survival attenuation with 76% viable bacteria compared to wild-type strain. No significant effect on intracellular survival was detected in the Δ*lmo0515* mutant (Figure 4). To confirm that the *usp* genes are intracellular induced we appllied quantitative RT-PCR. The analysis showed that *usp* transcripts obtained from intracellular versus extracellular grown bacteria displayed a strong induction of *usp* genes (Figure 5).

Decreased virulence of *usp* mutants in an invertebrate *in vivo* model

G. mellonella is currently a well established model host for studying the virulence attributes of the human pathogenic *L. monocytogenes* [20]. In this work, we used this invertebrate model to evaluate the pathogenicity of listerial *usp* deletion mutants. *G. mellonella* larvae were infected with pathogenic EGD-e and its isogenic *usp* mutants namely Δ*lmo1580*, Δ*lmo2673* and Δ*lmo0515* respectively. Survival rate of the infected larvae was determined at 37°C for over a period of 7 days.

Attenuated virulence associated with higher survival rates of larvae (Figure 6) were observed in Δ*lmo1580*, Δ*lmo2673* and Δ*lmo0515* (75%, 65% and 50% respectively) as compared to wild-type infection (37%). The difference between Δ*lmo0515* and wild-type did however not reach statistical significance. As observed previously with intracellular growth in macrophages, the Δ*lmo1580* mutation had the strongest effect on the pathogenicity, exhibiting nearly avirulent characteristics. A similar result was obtained for the triple mutant Δ*lmo0515*Δ*lmo1580*Δ*lmo2673* (Figure S4). To confirm that the deletion of *usp* genes is responsible for the effects in invertebrate infections we complemented the single *usp* mutants. Complementation restored the ability to kill insect larvae to levels comparable to the wild-type bacteria (Figure S5).

We observed no differences in growth of *usp* mutants and wild-type strain when cultured in BHI (data not shown) which excludes the possibility that varying numbers of inoculum CFUs of *usp* deletion mutants affected the survival of insect larvae.

Growth and survival of *Listeria usp* deletion mutants is profoundly impaired in a murine infection model

Although several studies have investigated the role of Usps in different bacteria, an important role of Usps *in vivo* has only been

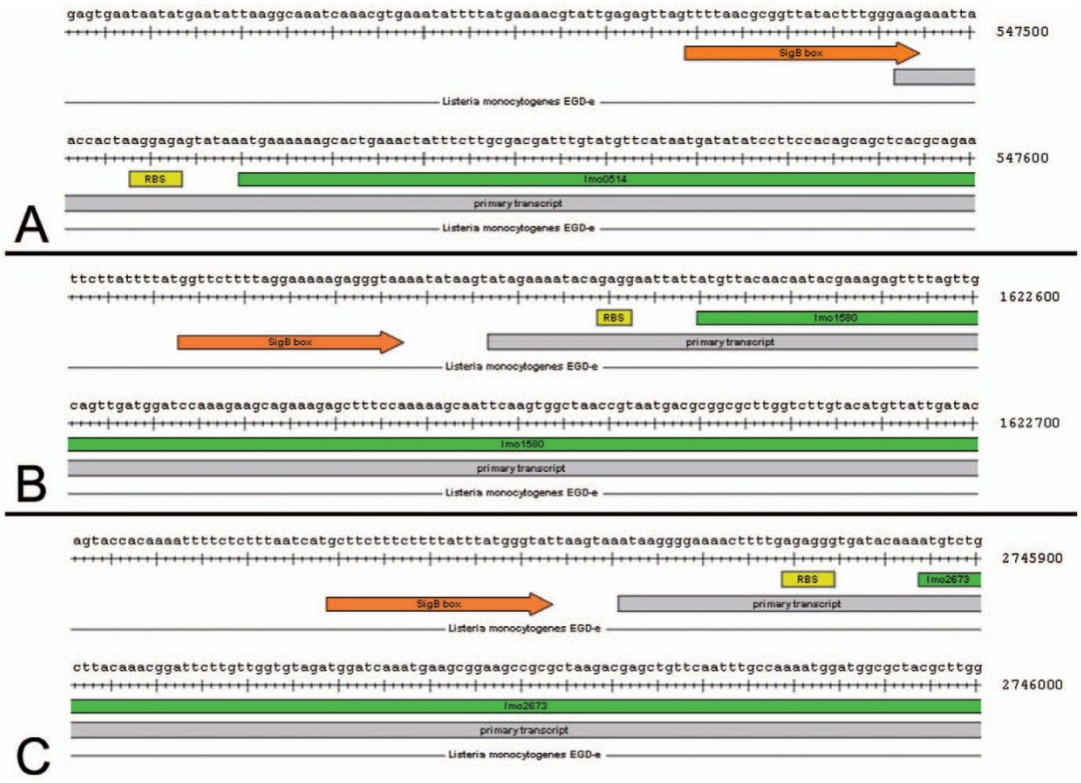


Figure 1. Transcriptional overview of chromosomal loci of *lmo0515* (A), *lmo1580* (B), and *lmo2673* (C) derived from sequencing data of *L. monocytogenes* (data not shown). σ^B boxes are depicted as orange arrows, the primary transcripts from their start sites by grey bars, the ribosome binding site as a yellow box and the respective coding sequence from their start sites by green bars.
doi:10.1371/journal.pone.0024965.g001

shown for the regulation of growth in *S. typhimurium* [12] and *M. tuberculosis* [6]. To elucidate the influence of listerial Usps on the growth and survival *in vivo*, we injected BALB/c mice with three different *usp* deletion mutants or wild-type and determined the number of bacteria isolated from the liver.

Remarkably, the number of CFU of all mutants compared to wild-type was dramatically diminished in this mouse model (Figure 7). The number of viable bacteria lacking the respective *usp* gene was significantly lower in livers and spleens of infected mice. In this setting, growth of Δ*lmo0515* was reduced to 22% in spleen and 32% in liver as compared to wild-type strain. The number of Δ*lmo2673* was reduced to 68% and 60%, and the Δ*lmo1580* count declined to 38% and 68%, respectively. The isogenic *usp* triple mutant showed comparable results as the single *usp* mutants.

Discussion

L. monocytogenes is capable of withstanding a variety of stressors encountered in nature and within infected host cells [18]. The ability to rapidly adjust to the changing environment is essential for its pathogenic lifestyle [18]. In this study, we provide strong evidence that universal stress proteins (Usps) are highly conserved among *Listeriae* and are important for resistance and growth of *L. monocytogenes* by investigating *usp* deletion mutants exposed to different stress conditions both *in vitro* and *in vivo*.

Universal stress proteins were shown to be important in response to stress in gram-negative bacteria, including *E. coli* [31], *S. typhimurium* [12] and *Azospirillum brasilense* [32]. However, the role of these proteins in gram-positive bacteria is not as well delineated. We have previously revealed an induction of three genes, *lmo0515*, *lmo2673* and *lmo1580* in *L. monocytogenes* in response to heat shock and acid stress [33,34]. In accordance with effects described for *E. coli* lacking the *uspA* gene, *L. monocytogenes usp* deletion mutations investigated in this study displayed impaired capability to resist exposure to H_2O_2 and low pH conditions. Variable susceptibility to stress conditions between deletion mutants may represent a non-redundant role of Usps in stress responses. Similar observations were made for *E. coli* and addressed in a recent study testing several *usp* deletion mutants [31]. Nachin et al. uncovered that the UspA proteins differ in their responses to oxidative stress and DNA damaging agents and defined similarities as well as differences in expression pattern based on their biological role in stress response [31]. Thus, certain environmental changes may not require the induction of all *usp* genes, but allow adjustment depending on the specialized role of a particular Usp.

Generally, *usp* genes are expressed in monocistronic units and the production of UspA seems to be primarily regulated at the

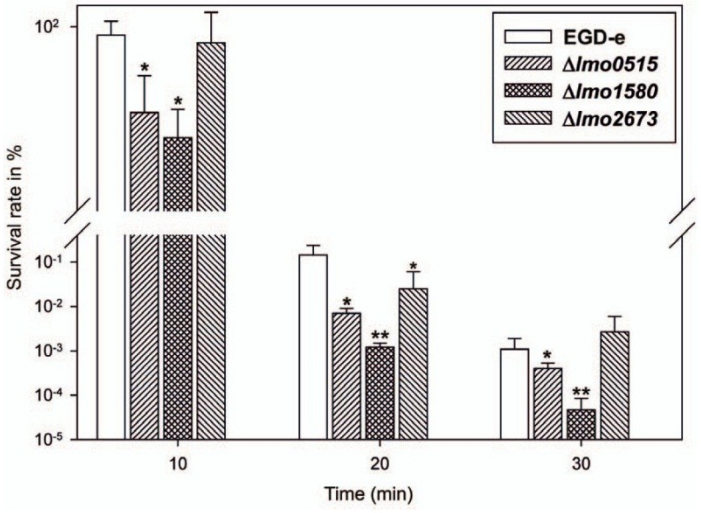


Figure 2. Effect of low acidic condition (pH = 2.5) on survival of *usp* deletion mutants compared to *L. monocytogenes* EGD-e wild-type. Δ*lmo0515* and Δ*lmo1580* display decreased growth at all observations, while impaired resistance in Δ*lmo2673* was visible only at 20 min post challenge. Statistically significant differences were identified using a two-tailed Student *t* test. (*, $p<0.05$ and **, $p<0.005$).
doi:10.1371/journal.pone.0024965.g002

level of transcription initiation [1]. We and others have previously reported an important role for σ^B, a transcription factor that controls an inducible regulon in *L. monocytogenes* induced following challenge to different stress conditions, such as acids, high osmolarity or carbohydrate starvation [17,34,35]. As previously reported, a single σ^B promoter has been identified for *lmo1580*; and promoter analysis revealed σ^B boxes for *lmo2673* and *lmo0515*. Thus, it appears likely that σ^B possesses a major role in stress induced transcriptional regulation of *usp* genes in *L. monocytogenes*. Furthermore, phenolic acids and fructose-6-phosphate were shown to influence *usp* gene expression in *Lactobacillus plantarum* and *E* coli, respectively. *L. monocytogenes* mobilizes a large set of metabolic genes in response to stress [24]. Especially, genes of the carbohydrate metabolism are strongly expressed. Possible effects

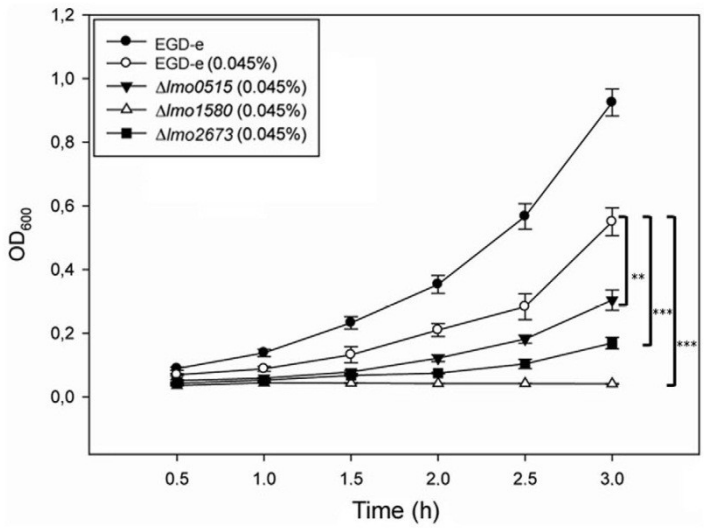


Figure 3. Survival of *L. monocytogenes* EGD-e wild-type and *usp* deletion mutants after exposure to H_2O_2. As appreciable, resistance of all deletion mutants was strongly impaired, resulting in decreased growth as compared to wild-type strain. Statistically significant differences were identified using a two-tailed Student *t* test. (*, $p<0.05$; **, $p<0.005$, and ***, $p<0.0005$).
doi:10.1371/journal.pone.0024965.g003

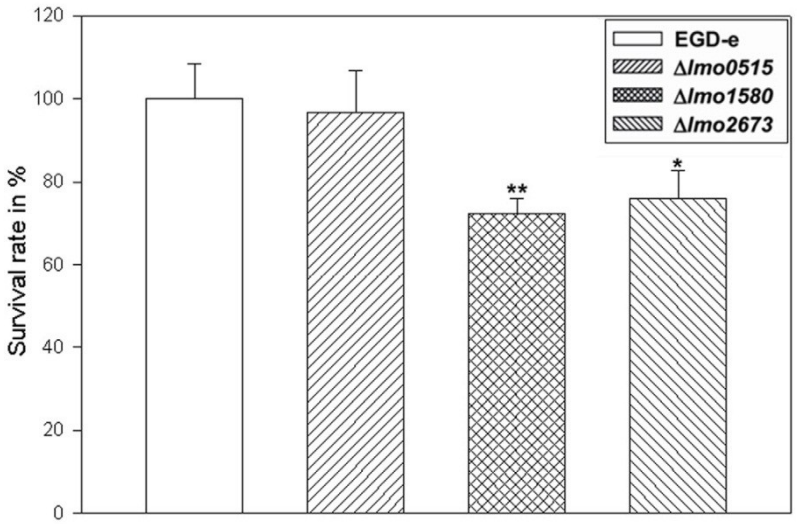


Figure 4. The intracellular survival of *L. monocytogenes* EGD-e wild-type and *usp* deletion mutants was assessed in murine macrophages according to [30]. Intracellular bacteria burden was decreased in macrophages inoculated with Δ*lmo1580* and Δ*lmo2673*, but no effect was visible for Δ*lmo0515*. Statistically significant differences were identified using a two-tailed Student *t* test. (*, $p<0.05$ and **, $p<0.005$).
doi:10.1371/journal.pone.0024965.g004

of metabolic constituents on *usp* gene regulation may therefore not be excluded and require further investigation. [15,36].

In this regard, the temporal and coordinated activities of different Usps of *L. monocytogenes* may enable the bacteria to survive in acid-rich niches such as in the upper intestinal tract or mediate resistance to oxidative stress and bile acids in the spleen and liver.

Following ingestion by phagocytic cells, *Listeriae* are faced with the hostile environment in the phagosome [18]. As shown previously, virulence and intracellular growth adaption of *L. monocytogenes* depends on a concerted gene expression program including the up-regulation of universal stress proteins [24,37]. In this study, we demonstrate that survival of *Listeriae* in murine macrophages is dependent on the presence of Usps. This was reflected by decreased virulence, resulting in a significant reduction of bacteria lacking the respective *usp* genes. In addition we observed intracellular transcriptional induction of all three *usp* genes which supports the

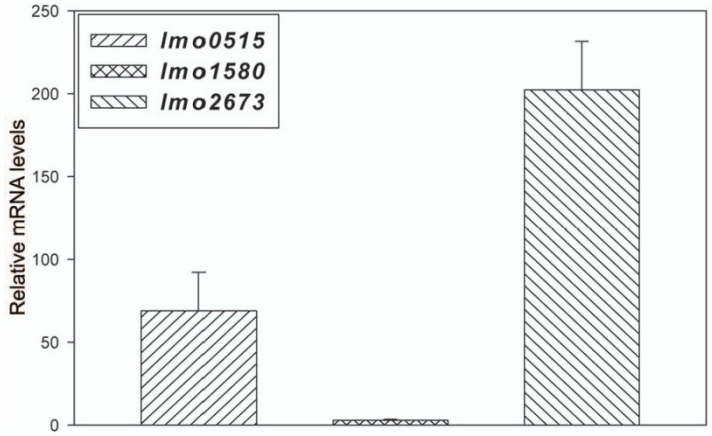


Figure 5. qRT-PCR analysis of *lmo0515*, *lmo1580* and *lmo2673* expressed in *L. monocytogenes* EGD-e. The graph displays the relative expression of the respective *usp* gene in the intracellular vs. extracellular setting. High fold changes indicate that the particular gene was higher expressed in intracellularly localized bacteria as compared to expression in bacteria that were grown in BHI. The expression of *lmo0515* and *lmo2673* is strongly enhanced in intracellular pathogens, while levels of *lmo1580* remain relatively constant. This is consistent with the observation made in this study, which shows a strong dependence on *lmo2673* and *lmo0515* in the murine infection model experiments as compared to extracellular *in vitro* challenges, while *lmo1580* seems to play an important role in both conditions.
doi:10.1371/journal.pone.0024965.g005

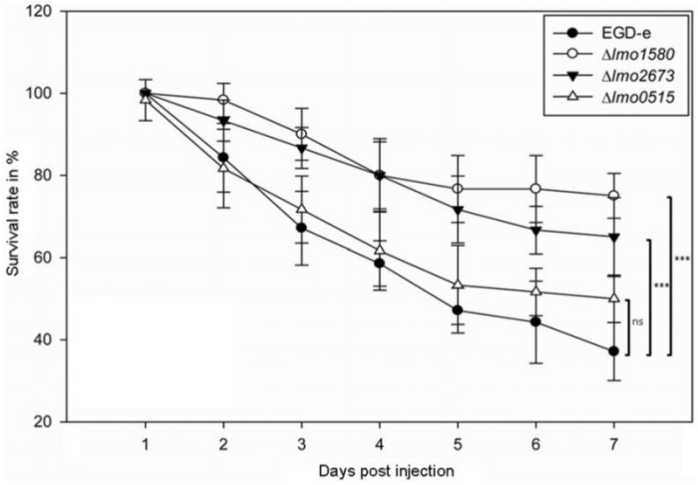


Figure 6. Survival of *G. mellonella* larvae after inoculation with *usp* deletion mutants. As compared to *L. monocytogenes* EGD-e wild-type, all *usp* deletion mutants exhibited impaired virulence, resulting in higher survival rates of *G. mellonella* larvae. Strongest effects were observed for *Δlmo1580 and Δlmo2673*, while deletion of *lmo0515* was not associated with significant increase in survival of *G. mellonella*. Results represent mean values of at least three independent experiments with a total of 80 larvae per treatment. Statistically significant differences were identified using a two-tailed Student *t* test. (***, $p<0.0005$ and ns-not significant).
doi:10.1371/journal.pone.0024965.g006

evidence that the particular stress proteins are involved in the intracellular pathogenic lifestyle of *L. monocytogenes*.

The impact of Usps in infection has only been examined for a few bacteria and their role, if any, in gram-positive pathogens has not been described [6,12,31,37]. For this reason, we further characterized the effect of listerial Usps in *in vivo* infection models. First we tested their effects on the viability of *G. mellonella* larvae, a recently described alternative infection model for *L. monocytogenes*

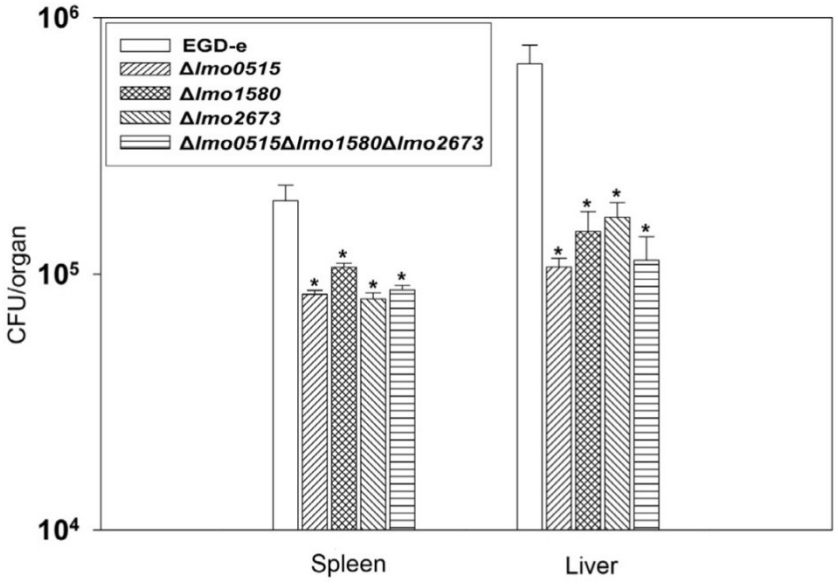


Figure 7. Survival rates of intravenously inoculated *L. monocytogenes* EGD-e or *usp* deletion mutants in mice livers and spleens. Survival of mutant bacteria are impaired in both organs, reflecting a role of Usps in listerial resistance and growth *in vivo*. Statistically significant differences were identified using a two-tailed Student *t* test. (*, $p<0.05$).
doi:10.1371/journal.pone.0024965.g007

[20]. Lack of *usp* genes resulted in improved survival of infected larvae as compared to wild-type infection. In accordance with our *in vitro* results, Δ*lmo1580* was nearly avirulent demonstrating its importance in pathogenicity of *L. monocytogenes*. As also demonstrated here, Usps contribute strongly to the survival and proliferation of these bacteria in livers and spleens of infected mice, while no major differences were apparent when comparing both organs. We have also constructed a triple mutant to investigate the cooperative nature of the three *usp* genes. We could not observe any synergistic effect for the *usp* triple mutant in mice experiments, but in the invertebrate infection model the isogenic triple mutant indicated a similar survival rate for larvae infected with Δ*lmo1580*. Thus, we suspect that Δ*lmo1580* might be the major player among the three *usp* genes involved in pathogenicity.

Previously, Usps were implicated in the resistance to antibiotics [38], biofilm formation by *P. gingivalis* [10] and survival of *P. aeruginosa* under anaerobic conditions [7]. In this study however, we did not identify a significant impact of *usp* gene deletion on biofilm formation or altered resistance under anaerobic/microaerophilic conditions and in response to treatment with several antibiotics (data not shown).

In conclusion, we have identified an important role of three universal stress proteins in resistance and survival of *L. monocytogenes* in response to low pH conditions and oxidative stress as well as intracellular survival within macrophages. We subsequently confirmed our observations in two *in vivo* infection models and show for the first time that universal stress proteins are important for survival and growth in the vertebrate and invertebrate hosts. Finally, we propose the designation of genes for *lmo0515*, *lmo1580* and *lmo2673 as uspL-1* (universal stress protein *Listeria*), *uspL-2* and *uspL-3*, respectively in the species *L. monocytogenes*.

Supporting Information

Figure S1 Comparative analysis of flanking regions of listerial universal stress protein *lmo0515* among 19 *L. monocytogenes* species including all known serotypes.
(PDF)

Figure S2 Comparative analysis of flanking regions of listerial universal stress protein *lmo1580* among 19 *L. monocytogenes* species including all known serotypes.
(PDF)

Figure S3 Comparative analysis of flanking regions of listerial universal stress protein *lmo2673* among 19 *L. monocytogenes* species including all known serotypes.
(PDF)

Figure S4 Survival of *G. mellonella* larvae after inoculation with *usp* deletion mutants. As compared to *L. monocytogenes* EGD-e wild-type, triple deletion mutant of *lmo1580*, *lmo2673* and *lmo0515* exhibited impaired virulence similar to the single gene deletion mutant Δ*lmo1580*, resulting in significant higher survival rates of *G. mellonella* larvae. Results represent mean values of at least three independent experiments and each repetition contained 30 larvae per treatment. Statistically significant differences were identified using a two-tailed Student *t* test. (***, $p<0.005$).
(TIF)

Figure S5 Complemetation of *lmo1580*, *lmo2673* and *lmo0515* into their respective isogenic deletion mutants Δ*lmo1580*, Δ*lmo2673* and Δ*lmo0515* resulted in induced virulence. Artificial introduction of the individual *usp* genes *lmo1580*, *lmo2673* and *lmo0515* into their respective isogenic deletion mutants Δ*lmo1580*, Δ*lmo2673* and Δ*lmo0515* resulted in an increase of virulence similar to the wild-type EGD-e. Results represent means of at least three independent determinations ± standard deviations. Each repetition contained 20 larvae per treatment.
(TIF)

Acknowledgments

We would like to thank Alexandra Amend, Juri and Nelli Schklarenko, Silke Silva and Martina Hudel for excellent advice and technical assistance.

Author Contributions

Conceived and designed the experiments: CSG BI FP WM MAM KM YA TC TH. Performed the experiments: CSG BI FP WM MAM KM. Analyzed the data: CSG BI MAM KM YA TC TH. Wrote the paper: BI FP TH TC. IT support and contributed to the manuscript: AB.

References

1. Kvint K, Nachin L, Diez A, Nystrom T (2003) The bacterial universal stress protein: function and regulation. Curr Opin Microbiol 6: 140 145.
2. O'Toole R, Williams HD (2003) Universal stress proteins and *Mycobacterium tuberculosis*. Res Microbiol 154: 387 392.
3. Sousa MC, McKay DB (2001) Structure of the universal stress protein of *Haemophilus influenzae*. Structure 9: 1135 1141.
4. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269: 496 512.
5. O'Toole R, Smeulders MJ, Blokpoel MC, Kay EJ, Lougheed K, et al. (2003) A two-component regulator of universal stress protein expression and adaptation to oxygen starvation in *Mycobacterium smegmatis*. J Bacteriol 185: 1543 1554.
6. Drumm JE, Mi K, Bilder P, Sun M, Lim J, et al. (2009) *Mycobacterium tuberculosis* universal stress protein Rv2623 regulates bacillary growth by ATP-Binding: requirement for establishing chronic persistent infection. PLoS Pathog 5: e1000460.
7. Schreiber K, Boes N, Eschbach M, Jaensch L, Wehland J, et al. (2006) Anaerobic survival of *Pseudomonas aeruginosa* by pyruvate fermentation requires an Usp-type stress protein. J Bacteriol 188: 659 668.
8. Boes N, Schreiber K, Hartig E, Jaensch L, Schobert M (2006) The *Pseudomonas aeruginosa* universal stress protein PA4352 is essential for surviving anaerobic energy stress. J Bacteriol 188: 6529 6538.
9. Boes N, Schreiber K, Schobert M (2008) SpoT-triggered stringent response controls *usp* gene expression in *Pseudomonas aeruginosa*. J Bacteriol 190: 7189 7199.
10. Chen W, Honma K, Sharma A, Kuramitsu HK (2006) A universal stress protein of *Porphyromonas gingivalis* is involved in stress responses and biofilm formation. FEMS Microbiol Lett 264: 15 21.
11. Chen J (2007) *uspA* of *Shigella sonnei*. J Food Prot 70: 2392 2395.
12. Liu WT, Karavolos MH, Bulmer DM, Allaoui A, Hormaeche RD, et al. (2007) Role of the universal stress protein UspA of *Salmonella* in growth arrest, stress and virulence. Microb Pathog 42: 2 10.
13. Fink RC, Evans MR, Porwollik S, Vazquez-Torres A, Jones-Carson J, et al. (2007) FNR is a global regulator of virulence and anaerobic metabolism in *Salmonella enterica* serovar Typhimurium (ATCC 14028s). J Bacteriol 189: 2262 2273.
14. Saguruthi SR, Panigrahi RR, Gowda G, Savithri HS, Murthy MR (2007) Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of universal stress protein F (YnaF) from *Salmonella typhimurium*. Acta Crystallogr Sect F Struct Biol Cryst Commun 63: 957 960.
15. Licandro-Seraut H, Gury J, Tran NP, Barthelmebs L, Cavin JF (2008) Kinetics and intensity of the expression of genes involved in the stress response tightly induced by phenolic acids in *Lactobacillus plantarum*. J Mol Microbiol Biotechnol 14: 41 47.
16. Gury J, Seraut H, Tran NP, Barthelmebs L, Weidmann S, et al. (2009) Inactivation of PadR, the repressor of the phenolic acid stress response, by molecular interaction with Usp1, a universal stress protein from *Lactobacillus plantarum*, in *Escherichia coli*. Appl Environ Microbiol 75: 5273 5283.
17. Hain T, Hossain H, Chatterjee SS, Machata S, Volk U, et al. (2008) Temporal transcriptomic analysis of the *Listeria monocytogenes* EGD-e sigma B regulon. BMC Microbiol 8: 20.
18. Vazquez-Boland JA, Kuhn M, Berche P, Chakraborty T, Dominguez-Bernal G, et al. (2001) *Listeria* pathogenesis and molecular virulence determinants. Clin Microbiol Rev 14: 584 640.
19. Glaser P, Frangeul L, Buchrieser C, Rusniok C, Amend A, et al. (2001) Comparative genomics of *Listeria* species. Science 294: 849 852.
20. Mukherjee K, Altincicek B, Hain T, Domann E, Vilcinskas A, et al. (2010) *Galleria mellonella* as a model system for studying *Listeria* pathogenesis. Appl Environ Microbiol 76: 310 317.

Appendix

21. Schaferkordt S, Chakraborty T (1995) Vector plasmid for insertional mutagenesis and directional cloning in *Listeria* spp. Biotechniques 19: 720 725.
22. Gaillard JL, Berche P, Mounier J, Richard S, Sansonetti P (1987) *In vitro* model of penetration and intracellular growth of *Listeria monocytogenes* in the human enterocyte-like cell line Caco-2. Infect Immun 55: 2822 2829.
23. Lauer P, Chow MY, Loessner MJ, Portnoy DA, Calendar R (2002) Construction, characterization, and use of two *Listeria monocytogenes* site-specific phage integration vectors. J Bacteriol 184: 4177 4186.
24. Chatterjee SS, Hossain H, Otten S, Kuenne C, Kuchmina K, et al. (2006) Intracellular gene expression profile of *Listeria monocytogenes*. Infect Immun 74: 1323 1338.
25. Kuenne CT, Ghai R, Chakraborty T, Hain T (2007) GECO linear visualization for comparative genomics. Bioinformatics 23: 125 126.
26. Mraheil MA, Billion A, Mohamed W, Mukherjee K, Kuenne C, et al. (2011) The intracellular sRNA transcriptome of *Listeria monocytogenes* during growth in macrophages. Nucleic Acids Res 39: 4235 4248.
27. Cotter PD, Gahan CG, Hill C (2000) Analysis of the role of the *Listeria monocytogenes* F0F1 -ATPase operon in the acid tolerance response. Int J Food Microbiol 60: 137 146.
28. Ferreira A, Sue D, O'Byrne CP, Boor KJ (2003) Role of *Listeria monocytogenes* sigma B in survival of lethal acidic conditions and in the acquired acid tolerance response. Appl Environ Microbiol 69: 2692 2698.
29. Rea R, Hill C, Gahan CG (2005) *Listeria monocytogenes* PerR mutants display a small-colony phenotype, increased sensitivity to hydrogen peroxide, and significantly reduced murine virulence. Appl Environ Microbiol 71: 8314 8322.
30. Machata S, Tchatalbachev S, Mohamed W, Jansch L, Hain T, et al. (2008) Lipoproteins of *Listeria monocytogenes* are critical for virulence and TLR2-mediated immune activation. J Immunol 181: 2028 2035.
31. Nachin L, Nannmark U, Nystrom T (2005) Differential roles of the universal stress proteins of *Escherichia coli* in oxidative stress resistance, adhesion, and motility. J Bacteriol 187: 6265 6272.
32. Galindo Blaha CA, Schrank IS (2003) An *Azospirillum brasilense* Tn5 mutant with modified stress response and impaired in flocculation. Antonie Van Leeuwenhoek 83: 35 43.
33. van d V, Hain T, Wouters JA, Hossain H, de Vos WM, et al. (2007) The heat-shock response of *Listeria monocytogenes* comprises genes involved in heat shock, cell division, cell wall synthesis, and the SOS response. Microbiology 153: 3593 3607.
34. Wemekamp-Kamphuis HH, Wouters JA, de Leeuw PP, Hain T, Chakraborty T, et al. (2004) Identification of sigma factor sigma B-controlled genes and their impact on acid stress, high hydrostatic pressure, and freeze survival in *Listeria monocytogenes* EGD-e. Appl Environ Microbiol 70: 3457 3466.
35. Kazmierczak MJ, Mithoe SC, Boor KJ, Wiedmann M (2003) *Listeria monocytogenes* sigma B regulates stress response and virulence functions. J Bacteriol 185: 5722 5734.
36. Persson O, Valadi A, Nystrom T, Farewell A (2007) Metabolic control of the *Escherichia coli* universal stress protein response through fructose-6-phosphate. Mol Microbiol 65: 968 978.
37. Camejo A, Buchrieser C, Couve E, Carvalho F, Reis O, et al. (2009) *In vivo* transcriptional profiling of *Listeria monocytogenes* and mutagenesis identify new virulence factors involved in infection. PLoS Pathog 5: e1000449.
38. Schobert M, Jahn D (2010) Anaerobic physiology of *Pseudomonas aeruginosa* in the cystic fibrosis lung. Int J Med Microbiol 300: 549 556.

13.9 Appendix publication 8

JOURNAL OF BACTERIOLOGY, Mar. 2010, p. 1473–1474 Vol. 192, No. 5
0021-9193/10/$12.00 doi:10.1128/JB.01415-09
Copyright © 2010, American Society for Microbiology. All Rights Reserved.

Complete Genome Sequence of *Listeria seeligeri*, a Nonpathogenic Member of the Genus *Listeria*▽

Christiane Steinweg,[1]§ Carsten T. Kuenne,[1]§ André Billion,[1] Mobarak A. Mraheil,[1] Eugen Domann,[1] Rohit Ghai,[1] Sukhadeo B. Barbuddhe,[1] Uwe Kärst,[2] Alexander Goesmann,[3] Alfred Pühler,[4] Bernd Weisshaar,[5] Jürgen Wehland,[2] Robert Lampidis,[6] Jürgen Kreft,[6] Werner Goebel,[6] Trinad Chakraborty,[1]* and Torsten Hain[1]*

Institute of Medical Microbiology, Justus-Liebig-University, Frankfurter Strasse 107, D-35392 Giessen, Germany[1]; Helmholtz-Centre for Infection Research, Inhoffenstrasse 7, D-38124 Braunschweig, Germany[2]; Bioinformatics Resource Facility, Centrum für Biotechnologie, University of Bielefeld, Universitätsstrasse 25, D-33615 Bielefeld, Germany[3]; Lehrstuhl für Genetik, University of Bielefeld, Universitätsstrasse 25, D-33615 Bielefeld, Germany[4]; Lehrstuhl für Genomforschung, University of Bielefeld, Universitätsstrasse 25, D-33615 Bielefeld, Germany[5]; and Lehrstuhl für Mikrobiologie, University of Würzburg, Am Hubland/Biozentrum, D-97074 Würzburg, Germany[6]

Received 29 October 2009/Accepted 23 December 2009

We report the complete and annotated genome sequence of the nonpathogenic *Listeria seeligeri* SLCC3954 serovar 1/2b type strain harboring the smallest completely sequenced genome of the genus *Listeria*.

Listeria seeligeri is one of seven species of the genus *Listeria*, a group of Gram-positive, motile, facultative anaerobic, low-GC-content, nonsporulating rods (3, 10, 12). To obtain a better understanding of the evolution of this nonpathogenic *Listeria* species, the type strain, SLCC 3954 (serovar 1/2b), a soil isolate from Germany (11), was sequenced using Sanger technology. Two small (1.5- to 2.5-kb) insert plasmid libraries were constructed with the TOPO Shotgun Subcloning Kit (Invitrogen) as described previously (9). Additionally, a small insert plasmid library (~1.0 to 1.5 kb) and a medium insert plasmid library (~5 kb) were constructed in the pUC19 cloning vector (New England Biolabs) by Qiagen (Hilden, Germany). A fosmid library harboring fragments of around 40 kb was created using the CopyControl Fosmid Library Production Kit (Epicentre) as published before (9).

Sequencing was performed by Agowa (Berlin, Germany), Qiagen (Hilden, Germany), and the Max-Planck-Institut (Köln, Germany) by using ABI Big Dye Terminator technology. A total of 30,961 and 1,305 high-quality reads from the shotgun, and fosmid libraries were used to generate a draft assembly with an overall coverage of ~7-fold by using the Phred/Phrap/Consed assembly package (4, 5, 7). Contigs were linked by primer walking on shotgun clones and fosmids as well as by PCR gap closure followed by sequencing of the PCR product. Genome annotation was performed as described previously (9).

The genome of *L. seeligeri* consists of a circular chromosome of 2,797,636 bp and hence is slightly smaller than those of previously sequenced listerial strains (6, 9). Functional classification of genes obtained by mapping against clusters of orthologous groups (14) was predicted by Augur (1). Comparative analysis of these clusters indicates gene loss in categories such as amino acid/carbohydrate transport and metabolism as well as in transcription, thus confirming a general trend toward genome shrinkage. *L. seeligeri* harbors no plasmid and carries only a single copy of a prophage and no transposon in its genome. The mean G+C content of the *L. seeligeri* genome is 37.4%, which is close to the average value of all known *Listeria* strains (6, 9). G/C skew analysis revealed a bidirectional replication mechanism, and the origin of replication (oriC) is located close to the *dnaA* gene, which is positioned diametrically opposite to the replication terminus. We identified six 16S-23S-5S rRNA operons, all of which are located on the leading strand, two on the right and four on the left replichore. Additionally, a total of 67 tRNA genes were detected. We used MAVID (2) for phylogenetic analysis of the genomes of *L. monocytogenes*, *L. innocua*, *L. welshimeri*, and *L. seeligeri* and conclude that the "phylogenomic" relationship corresponds exactly to phylogenetic analysis based on either 16S rRNA genes (8, 15) or other additional specific marker genes (8, 13).

The genome sequence of *L. seeligeri* is the fourth species of genus *Listeria* to be reported.

Nucleotide sequence accession number. The genome sequence of *L. seeligeri* serovar 1/2b (SLCC3954) reported here has been deposited in the EMBL database under accession number FN557490.

* Corresponding author. Mailing address: Institute of Medical Microbiology, Justus-Liebig-University, Frankfurter Strasse 107, D-35392 Giessen, Germany. Phone: 49-641 99 46400. Fax: 49-641 99 46409. E-mail for T. Hain: Torsten.Hain@mikrobio.med.uni-giessen.de. E-mail for T. Chakraborty: Trinad.Chakraborty@mikrobio.med.uni-giessen.de.
§ Both coauthors contributed equally to the work.
▽ Published ahead of print on 8 January 2010.

We thank Alexandra Amend, Claudia Zörb, Nelli and Juri Schklarenko, and Prisca Viehoever for excellent technical assistance.
This work was supported by the fund obtained from the BMBF through the Competence Network PathoGenoMik (031U213B) and the ERA-NET PathoGenoMics program grant SPATELIS to T.C. and T.H.

REFERENCES

1. **Billion, A., R. Ghai, T. Chakraborty, and T. Hain.** 2006. Augur—a computational pipeline for whole genome microbial surface protein prediction and classification. Bioinformatics **22**:2819–2820.

Appendix

2. **Bray, N., and L. Pachter.** 2004. MAVID: constrained ancestral alignment of multiple sequences. Genome Res. **14:**693–699.
3. **Collins, M. D., S. Wallbanks, D. J. Lane, J. Shah, R. Nietupski, J. Smida, M. Dorsch, and E. Stackebrandt.** 1991. Phylogenetic analysis of the genus *Listeria* based on reverse transcriptase sequencing of 16S rRNA. Int. J. Syst. Bacteriol. **41:**240–246.
4. **Ewing, B., and P. Green.** 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8:**186–194.
5. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8:**175–185.
6. **Glaser, P., L. Frangeul, C. Buchrieser, C. Rusniok, A. Amend, F. Baquero, P. Berche, H. Bloecker, P. Brandt, T. Chakraborty, A. Charbit, F. Chetouani, E. Couve, A. de Daruvar, P. Dehoux, E. Domann, G. Dominguez-Bernal, E. Duchaud, L. Durant, O. Dussurget, K. D. Entian, H. Fsihi, F. Garcia-del Portillo, P. Garrido, L. Gautier, W. Goebel, N. Gomez-Lopez, T. Hain, J. Hauf, D. Jackson, L. M. Jones, U. Kaerst, J. Kreft, M. Kuhn, F. Kunst, G. Kurapkat, E. Madueno, A. Maitournam, J. M. Vicente, E. Ng, H. Nedjari, G. Nordsiek, S. Novella, B. de Pablos, J. C. Perez-Diaz, R. Purcell, B. Remmel, M. Rose, T. Schlueter, N. Simoes, A. Tierrez, J. A. Vazquez-Boland, H. Voss, J. Wehland, and P. Cossart.** 2001. Comparative genomics of *Listeria* species. Science **294:**849–852.
7. **Gordon, D., C. Abajian, and P. Green.** 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8:**195–202.
8. **Graves, L. M., L. O. Helsel, A. G. Steigerwalt, R. E. Morey, M. I. Daneshvar, S. E. Roof, R. H. Orsi, E. D. Fortes, S. R. Milillo, H. C. den Bakker, M. Wiedmann, B. Swaminathan, and B. D. Sauders.** 2009. *Listeria marthii* sp. nov., isolated from the natural environment, Finger Lakes National Forest. Int. J. Syst. Evol. Microbiol. doi 10.1099/ijs.0.014118–0.
9. **Hain, T., C. Steinweg, C. T. Kuenne, A. Billion, R. Ghai, S. S. Chatterjee, E. Domann, U. Karst, A. Goesmann, T. Bekel, D. Bartels, O. Kaiser, F. Meyer, A. Puhler, B. Weisshaar, J. Wehland, C. Liang, T. Dandekar, R. Lampidis, J. Kreft, W. Goebel, and T. Chakraborty.** 2006. Whole-genome sequence of *Listeria welshimeri* reveals common steps in genome reduction with *Listeria innocua* as compared to *Listeria monocytogenes*. J. Bacteriol. **188:**7405–7415.
10. **Rocourt, J.** 1999. The genus *Listeria* and *Listeria monocytogenes*: phylogenetic positions, taxonomy, and identification, p. 1–20. *In* E. T. Ryser and E. Marth (ed.), *Listeria*, listeriosis, and food safety. Macel Dekker Inc., New York, NY.
11. **Rocourt, J., and P. Grimont.** 1983. *Listeria welshimeri* sp. nov. and *Listeria seeligeri* sp. nov. Int. J. Syst. Bacteriol. **33:**866–869.
12. **Sallen, B., A. Rajoharison, S. Desvarenne, F. Quinn, and C. Mabilat.** 1996. Comparative analysis of 16S and 23S rRNA sequences of *Listeria* species. Int. J. Syst. Bacteriol. **46:**669–674.
13. **Schmid, M. W., E. Y. Ng, R. Lampidis, M. Emmerth, M. Walcher, J. Kreft, W. Goebel, M. Wagner, and K. H. Schleifer.** 2005. Evolutionary history of the genus *Listeria* and its virulence genes. Syst. Appl. Microbiol. **28:**1–18.
14. **Tatusov, R. L., D. A. Natale, I. V. Garkavtsev, T. A. Tatusova, U. T. Shankavaram, B. S. Rao, B. Kiryutin, M. Y. Galperin, N. D. Fedorova, and E. V. Koonin.** 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res. **29:**22–28.
15. **Vaneechoutte, M., P. Boerlin, H. V. Tichy, E. Bannerman, B. Jager, and J. Bille.** 1998. Comparison of PCR-based DNA fingerprinting techniques for the identification of *Listeria* species and their use for atypical *Listeria* isolates. Int. J. Syst. Bacteriol. **48:**127–139.

13.10 Appendix publication 9

Pischimarov et al. BMC Genomics 2012, 13:384
http://www.biomedcentral.com/1471-2164/13/384

BMC Genomics

DATABASE Open Access

sRNAdb: A small non-coding RNA database for gram-positive bacteria

Jordan Pischimarov[1†], Carsten Kuenne[1†], André Billion[1], Jüergen Hemberger[2], Franz Cemič[2], Trinad Chakraborty[1] and Torsten Hain[1*]

Abstract

Background: The class of small non-coding RNA molecules (sRNA) regulates gene expression by different mechanisms and enables bacteria to mount a physiological response due to adaptation to the environment or infection. Over the last decades the number of sRNAs has been increasing rapidly. Several databases like Rfam or fRNAdb were extended to include sRNAs as a class of its own. Furthermore new specialized databases like sRNAMap (gram-negative bacteria only) and sRNATarBase (target prediction) were established. To the best of the authors' knowledge no database focusing on sRNAs from gram-positive bacteria is publicly available so far.

Description: In order to understand sRNA's functional and phylogenetic relationships we have developed sRNAdb and provide tools for data analysis and visualization. The data compiled in our database is assembled from experiments as well as from bioinformatics analyses. The software enables comparison and visualization of gene loci surrounding the sRNAs of interest. To accomplish this, we use a client–server based approach. Offline versions of the database including analyses and visualization tools can easily be installed locally on the user's computer. This feature facilitates customized local addition of unpublished sRNA candidates and related information such as promoters or terminators using tab-delimited files.

Conclusion: sRNAdb allows a user-friendly and comprehensive comparative analysis of sRNAs from available sequenced gram-positive prokaryotic replicons. Offline versions including analysis and visualization tools facilitate complex user specific bioinformatics analyses.

Background

In recent years numerous small non-coding RNAs (sRNAs) were discovered in bacteria. This class of RNAs is crucial to prokaryotic life, modulating transcription or translation leading to either activation or repression of important physiological processes. sRNAs enable bacteria to trigger rapid physiological responses in order to adapt to the environment or infectious processes [1-3].

To cope with the increasing number of identified sRNAs, databases such as fRNAdb, Rfam, sRNAMap and sRNATarBase were developed [4-9]. All of these approaches have certain drawbacks. fRNAdb contains all classes of RNAs, but allows no further analysis. Rfam is one of the most informative data collections, allowing detailed analyses via a web front-end. sRNAMap is a webserver-based application for gram-negative bacteria only. sRNATarBase compiles experimental data and allows the prediction of sRNA targets. But all databases available to date limit the analysis to published data only. Therefore bioinformatics analyses of candidate sRNAs in combination with genomes, terminators and other relevant information that has not yet been published is still a very complicated task.

In an attempt to overcome some of the aforementioned drawbacks, we have developed sRNAdb. Our database is a locally installable web-suite, permitting the comparative analysis of sRNAs of gram-positive bacteria including their flanking genes. User modified files in GenBank format and gram-negative bacterial genomes, pooled sRNA candidates or further features of interest can be included in locally installed databases. Furthermore all integrated analysis tools can also be used locally.

* Correspondence: Torsten.Hain@mikrobio.med.uni-giessen.de
†Equal contributors
[1]Institute of Medical Microbiology, Justus-Liebig-University, Schubertstrasse 81, Giessen D-35392, Germany
Full list of author information is available at the end of the article

BioMed Central

© 2012 Pischimarov et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Construction and content

A database scheme of unique keys and entities, combined with corresponding relations and connections is given in Figure 1. Optional user defined extensions to locally installed versions of the database are indicated with a lighter background color than the boxes representing database entities.

Input data

To the best authors' knowledge, no general nomenclature convention for sRNAs exists to date. Therefore sRNAs imported into our database from the literature cannot always be unambiguously distinguished by name, locus or annotation only. Furthermore a large number of published sRNAs is currently annotated as predicted or putative. This leads to a myriad of sRNAs bearing indistinct names, positions or ambiguous annotations. To cope with this difficulty, sRNAdb contains a unique key composed of information about the authors, experimental conditions and sRNA properties as shown in the table termed snrax of Figure 1. Annotated sequences of organisms or plasmids downloaded from NCBI's RefSeq database [10] represent the replicons in the database. Information annotated in GenBank-formatted files such as sequences, or genes filtered from these files are automatically inserted into sRNAdb. When sRNAdb is installed locally, users can furthermore modify the local database by adding customized features such as terminators, promoters and other additional data. Terminators predicted by TransTermHP [11] serve as examples for this option, as described on the official sRNAdb server homepage.

Architecture and design

Our public sRNAdb server is implemented in Java 1.6 on a Debian Linux PC. It facilitates a client–server architecture using Java Server Pages (JSPs), Java Servlets, and Cascading Stylesheets (CSS). Apache Tomcat and MySQL serve as webserver and database, respectively.

Related sRNAs are determined using BLASTN [12], while protein homologies are established by a combination

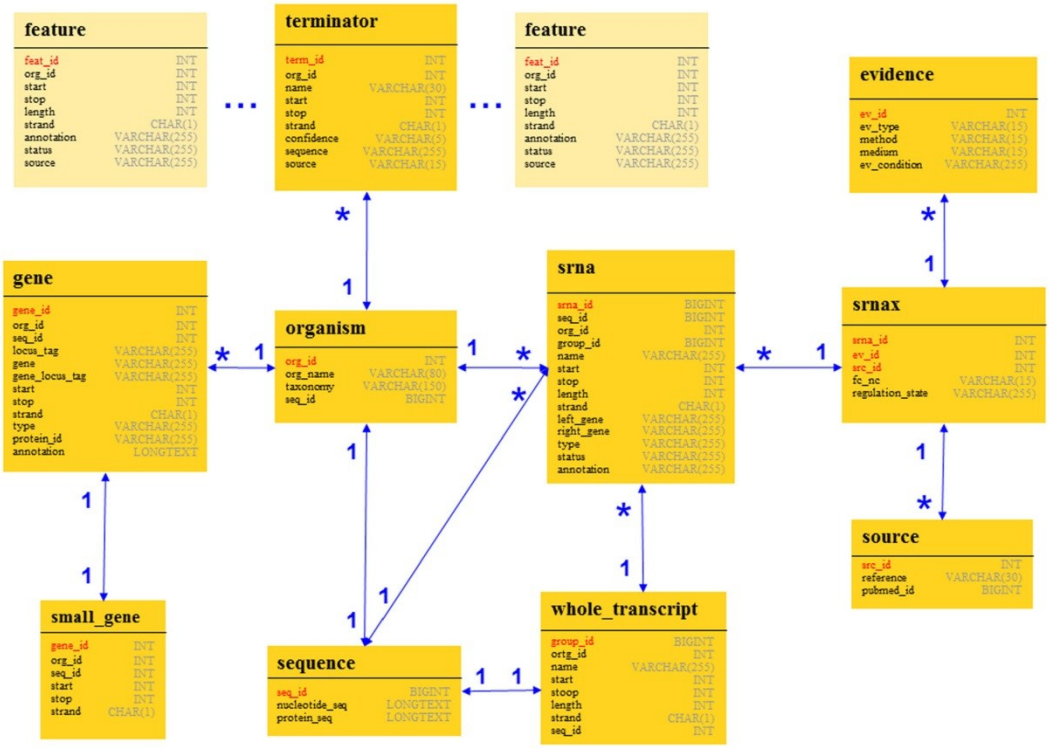


Figure 1 Database schema. The whole database with connections between tables and specific attributes are shown in UML-Notation. Unique and foreign keys of each table are given in bold letters while relations between entities are stated above the connection arrows. Optional features which can be inserted by the user into local versions of the database, are indicated using a lighter background color than employed for boxes representing entities.

of BLASTCLUST and BLASTP [12]. The addition of new data (replicons, sRNAs, terminators, promoters, RBS, etc.) to a local installation of sRNAdb is a simple process based on GenBank and tab-delimited flat-files.

Currently, the public sRNAdb server contains 558 gram-positive genomes and plasmids as well as 9993 automatically predicted and 671 experimentally verified sRNAs. An overview is given in Table 1.

Utility and discussion

The sRNAdb web-database aims to collect all published and predicted sRNAs of gram-positive bacteria for comparative analysis. sRNAs featuring an environmental condition-depending range of sizes can optionally be joined to a combined transcript. The public version of sRNAdb contains terminators predicted by TransTermHP [11]. Three web-interfaces are provided for retrieval and analysis of the data. The first module is called *search* and offers a rich query interface for the database, as shown in Figure 2A. Properties of sRNAs can be selected and filters can be defined to create task-specific queries resulting in a tabular output (Figure 2B). Related or customized data can also be collated to the query, based on the up- or downstream distance to an sRNA of interest. Furthermore, a secondary structure prediction of selected sRNA sequences by energy minimization can be performed using RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi).

Another interface named *blast* (Figure 3A) was created to enable homology searches of sRNAs versus either public or proprietary sRNAs or whole chromosomes/plasmids using BLASTN [12]. This can be used for initial screening of potential genomic regions. Concise matrix outputs for comparative analysis purposes as shown in Figure 3B and Figure 3C, are implemented. Complete BLAST alignments are displayed in Figure 3D. Sequences from the BLAST output table can be easily selected by setting checkmarks to extract data into a multifasta-formatted file, ready to serve as input to multiple sequence alignment programs such as CLUSTALW (http://www.ebi.ac.uk/Tools/msa/clustalw2/). The resulting output can be used to predict structurally conserved and thermodynamically stable RNA secondary structures using e.g., RNAz (http://rna.tbi.univie.ac.at/cgi-bin/RNAz.cgi), facilitating screens for sRNA-homologs across genomes.

For comprehensive visual assessment the *vision* servlet (Figure 4A) was developed. This allows for a comparative analysis of multiple, related chromosome/plasmid loci of the genomic neighborhood of a single sRNA of interest (single mode) as displayed in Figure 4B. The results are translated into an image (.png-formatted) whereby homologous genes (CDS, RNA) of the sRNA locus are identified by BLASTP [12] and presented with an identical colour code. Terminators and any number of additional features previously defined can be included as desired. Each object in the image is associated with a popup-box, displaying further information and linked to corresponding database entries. The width of the resulting image can be varied to compensate for different screen resolutions. Thus one sRNA locus can be

Table 1 The table shows an overview of the current database entries. These are compiled from experiments or from bioinformatic analyses

Reference	sRNAs	Organism	Pubmed_id
Arnvig et al. 2009 [13]	9	*Mycobacterium tuberculosis* H37Rv	19555452
Bohn et al. 2010 [14]	28	*Staphylococcus aureus subsp. aureus* N315	20511587
Christiansen et al. 2006 [15]	3	*Listeria monocytogenes* EGD-e	16682563
D'Hérouel et al. 2011 [16]	22	*Enterococcus faecalis* V583	21266481
Geissmann et al. 2009 [17]	11	*Staphylococcus aureus* subsp. *aureus* N315	19786493
Irnov et al. 2010 [18]	90	*Bacillus subtilis* subsp. *subtilis* str. 168	20525796
Kumar et al. 2010 [19]	50	*Streptococcus pneumonia* TIGR4	20525227
Livny et al. 2008 [20]	9993	Gram-positive bacteria	18787707
Mandin et al. 2007 [21]	12	*Listeria monocytogenes* EGD-e	17259222
Mraheil et al. 2011 [22]	150	*Listeria monocytogenes* EGD-e	21278422
Nielsen et al. 2008 [23]	1	*Listeria monocytogenes* EGD-e	18621897
Perez et al. 2009 [24]	33	*Streptococcus pyogenes* MGAS5005	19888332
Rasmussen et al. 2009 [25]	84	*Bacillus subtilis* subsp. *subtilis* str. 168	19682248
Tezuka et al. 2009 [26]	12	*Streptomyces griseus* subsp. *griseus* NBRC 13350	19465662
Toledo-Arana et al. 2009 [27]	103	*Listeria monocytogenes* EGD-e	19448609
Vockenhuber et al. 2010 [28]	63	*Streptomyces coelicolor*	21521948

The organisms for which sRNAs are listed in the database, including references, the number of identified sRNAs for the specific organisms and their relevant pumed identification number are listed.

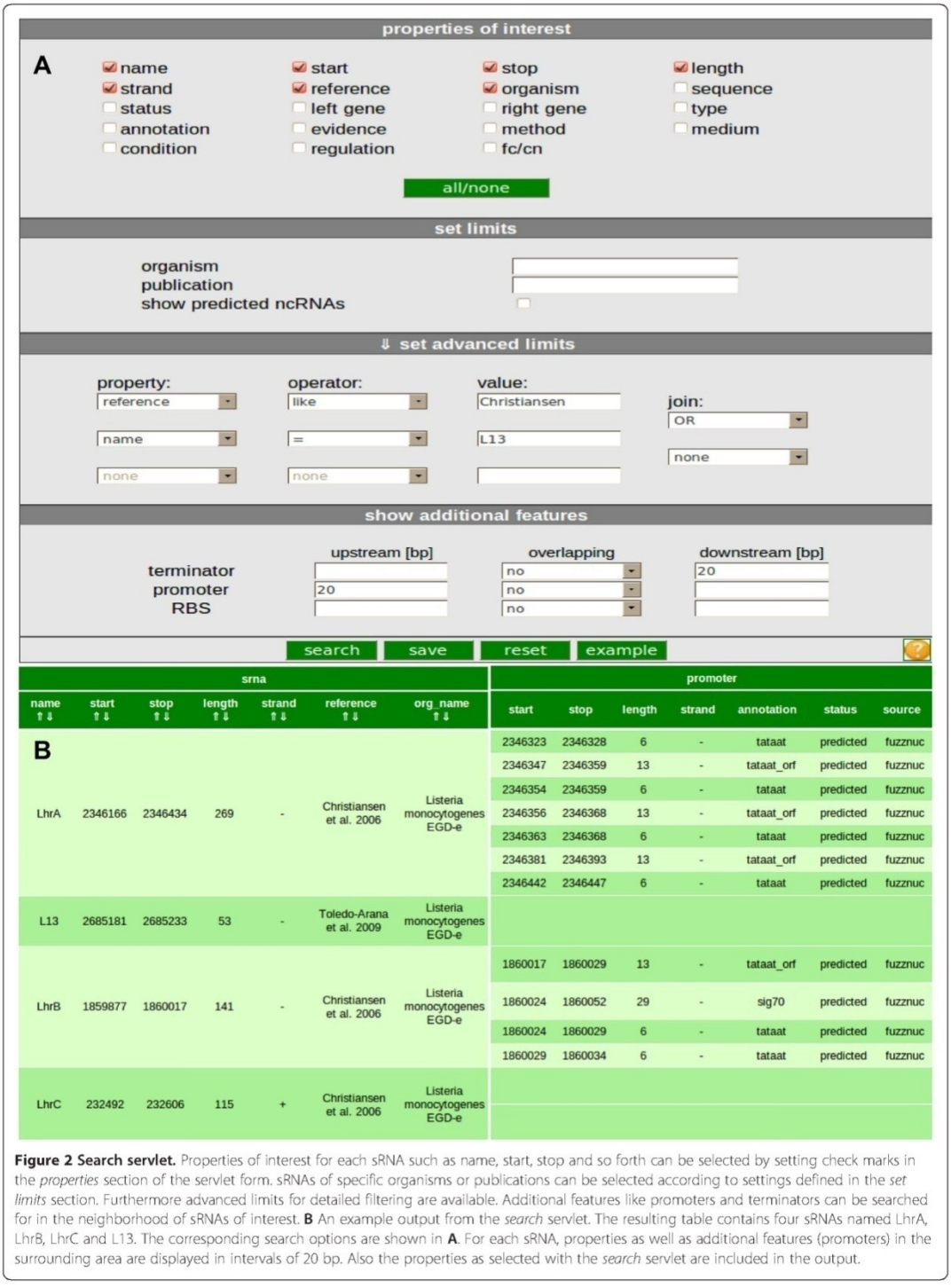


Figure 2 Search servlet. Properties of interest for each sRNA such as name, start, stop and so forth can be selected by setting check marks in the *properties* section of the servlet form. sRNAs of specific organisms or publications can be selected according to settings defined in the *set limits* section. Furthermore advanced limits for detailed filtering are available. Additional features like promoters and terminators can be searched for in the neighborhood of sRNAs of interest. **B** An example output from the *search* servlet. The resulting table contains four sRNAs named LhrA, LhrB, LhrC and L13. The corresponding search options are shown in **A**. For each sRNA, properties as well as additional features (promoters) in the surrounding area are displayed in intervals of 20 bp. Also the properties as selected with the *search* servlet are included in the output.

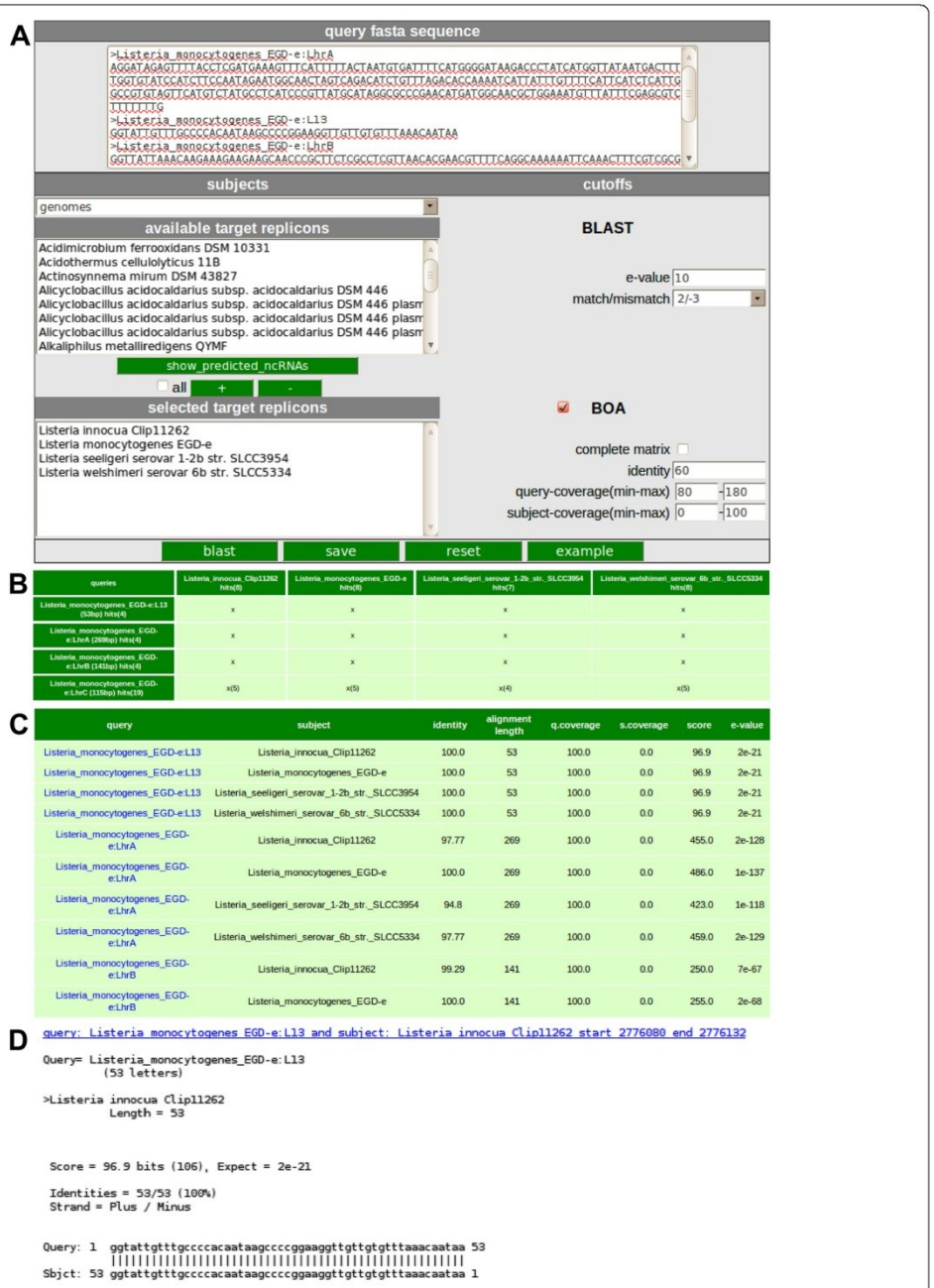


Figure 3 Blast servlet form and corresponding output. A FASTA formatted sRNA sequences can be inserted into the query box. Also target genomes or sRNAs have to be selected for multiple alignment using BLAST. For a detailed BLAST analysis the BLAST output analysis (BOA) options has to be selected. In this example four sRNAs resulting from a search with parameters shown in Figure 1 were selected as input. Genomes of the genus *Listeria* were set as targets and the BOA options were enabled. **B** The number of sRNAs detected in the target organism is displayed in a comparative matrix form. **C** All hits listed in a table and are linked to their corresponding alignment. **D** A detailed BLAST alignment of all results can also be plotted.

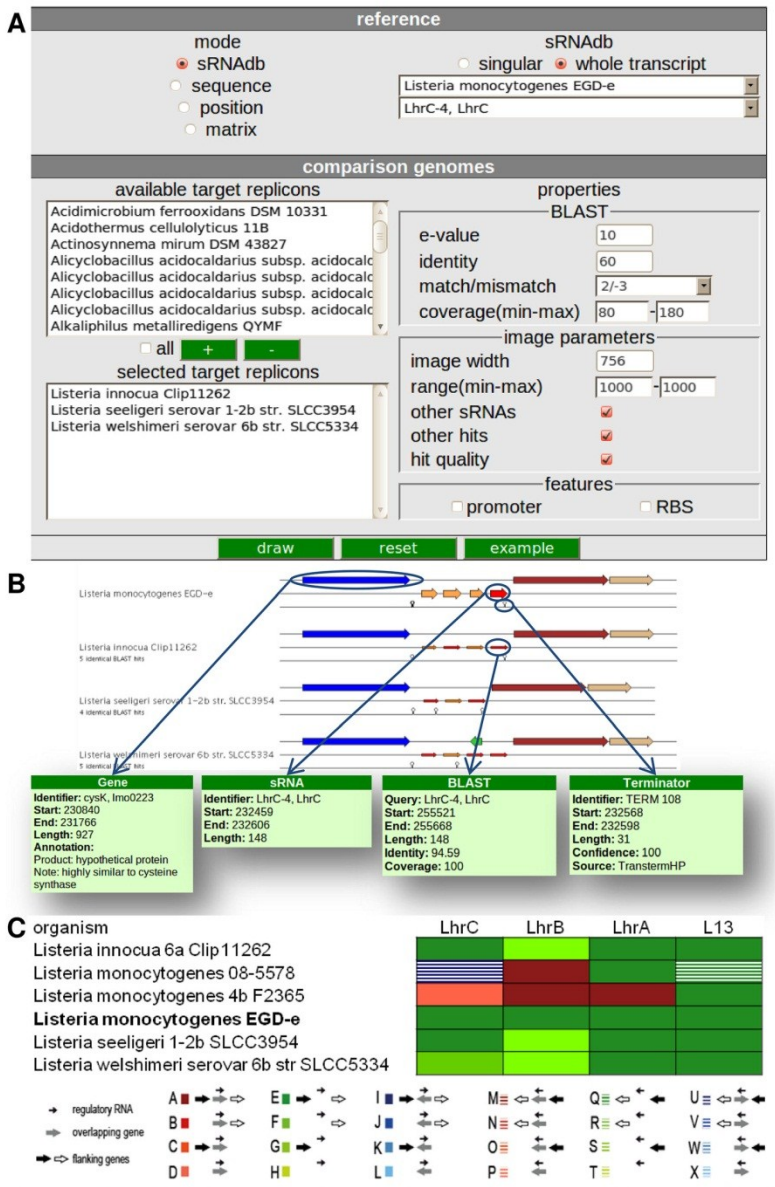


Figure 4 Vision servlet forms and result of single and batch mode. Different input options are available. After selecting the sRNA of interest, replicons can be selected for visualization. Options for further analyses based on BLAST, as well as properties relating to the image output can be set. **A** An example relating to the LhrC transcript is displayed. **B** Single mode: the resulting image shows a comparative representation of a single sRNA candidate and flanking genes in selected organisms. Moving the mouse pointer over these, the corresponding properties of each object is shown in a separate popup window. **C** Batch mode: sRNAs displayed in Figure 1 are used as input in this example. The output-matrix indicates occurrence of the sRNA candidates in selected organisms and their directional relationships with respect to their surrounding genes.

compared to different chromosomes/plasmids in a concise image output.

For the genome wide analysis of multiple sRNA loci an additional batch mode is available. Results from an application of this batch mode have already been published by Mraheil and collaborators [22]. In order to permit this global analysis an option was implemented that enables export of the data to an Excel sheet. This contains a visualization matrix (Figure 4C) which indicates the occurrence of the sRNA of interest in the target organism together with its directional relationships of the flanking genes.

The software tool presented here is a valuable extension to existing solutions and will assist in the rapid analysis of large volumes of data to understand the distribution and evolution of sRNAs in bacteria. Compared to other databases the comparative batch mode of sRNAdb's *vision* servlet facilitates analyses such as *in silico* screening for phylogenetic markers, or identification of drug targets related to bacterial sRNAs. As exemplified by Mraheil and colleagues [22] a grouping of sRNAs from pathogenic, apathogenic or non-pathogenic bacterial strains based on the *vision* servlet´s result matrix, allows the user to identify sRNAs as putative phylogenetic markers. Specifically, sRNAs found exclusively in pathogenic strains can be identified as drug target candidates. Furthermore after download and local installation of sRNAdb, both the database and the dedicated software tools are available to the user. Since proprietary replicons or putative sRNAs can easily be included into locally installed versions of the database, these may be analysed making use of the full power of sRNAdb's software tools, simplifying detailed analyses of unpublished bacterial replicons or sRNA candidates. To the best of the author's knowledge, this functionality is currently not supported by any other publicly available sRNA database.

Conclusion

sRNAdb offers biologists an easy access and analysis to both proprietary and public data and allows the identification of a core set of sRNAs which can be used as putative drug targets in antimicrobial therapeutic approaches as well as specific sRNAs for potential diagnostic markers for the detection of gram-positive bacteria.

Availability and requirements

The database including documentation and tools for analysis are available free of charge at http://bioinfo.mikrobio.med.uni-giessen.de/sRNAdb.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Designed and implemented the database and related software tools: JP, CK, AB, TH. Analyzed data: JP, CK, AB, FC, TH. Wrote the paper: JP, CK, JH, FC, TC, TH. All authors read and approved the final manuscript.

Funding

This work was supported by grants from the German Federal Ministry of Education and Research (BMBF ERA-NET) Pathogenomics Network to the sncRNAomics project (62080061) to T.H. and the German Centre for Infection Research, Justus-Liebig University Giessen.

Author details

[1]Institute of Medical Microbiology, Justus-Liebig-University, Schubertstrasse 81, Giessen D-35392, Germany. [2]Institute for Biochemical Engineering and Analytics, University of Applied Sciences Giessen-Friedberg, Wiesenstrasse 14, Giessen D-35390, Germany.

Received: 3 April 2012 Accepted: 1 July 2012
Published: 10 August 2012

References

1. Frohlich KS, Vogel J: **Activation of gene expression by small RNA.** *Curr Opin Microbiol* 2009, **12**:674–682.
2. Mraheil MA, Billion A, Kuenne C, Pischimarov J, Kreikemeyer B, Engelmann S, Hartke A, Giard JC, Rupnik M, Vorwerk S, Beier M, Retey J, Hartsch T, Jacob A, Cernic F, Hemberger J, Chakraborty T, Hain T: **Comparative genome-wide analysis of small RNAs of major Gram-positive pathogens: from identification to application.** *Microb Biotechnol* 2010, **3**:658–676.
3. Waters LS, Storz G: **Regulatory RNAs in bacteria.** *Cell* 2009, **136**:615–628.
4. Cao Y, Wu J, Liu Q, Zhao Y, Ying X, Cha L, Wang L, Li W: **sRNATarBase: a comprehensive database of bacterial sRNA targets verified by experiments.** *RNA* 2010, **16**:2051–2057.
5. Kin T, Yamada K, Terai G, Okida H, Yoshinari Y, Ono Y, Kojima A, Kimura Y, Komori T, Asai K: **fRNAdb: a platform for mining/annotating functional RNA candidates from non-coding RNA sequences.** *Nucleic Acids Res* 2007, **35**:D145–D148.
6. Mituyama T, Yamada K, Hattori E, Okida H, Ono Y, Terai G, Yoshizawa A, Komori T, Asai K: **The Functional RNA Database 3.0: databases to support mining and annotation of functional RNAs.** *Nucleic Acids Res* 2009, **37**:D89–D92.
7. Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, Bateman A: **Rfam: updates to the RNA families database.** *Nucleic Acids Res* 2009, **37**:D136–D140.
8. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR: **Rfam: an RNA family database.** *Nucleic Acids Res* 2003, **31**:439–441.
9. Huang HY, Chang HY, Chou CH, Tseng CP, Ho SY, Yang CD, Ju YW, Huang HD: **sRNAMap: genomic maps for small non-coding RNAs, their regulators and their targets in microbial genomes.** *Nucleic Acids Res* 2009, **37**:D150–D154.
10. Pruitt KD, Katz KS, Sicotte H, Maglott DR: **Introducing RefSeq and LocusLink: curated human genome resources at the NCBI.** *Trends Genet* 2000, **16**:44–47.
11. Kingsford CL, Ayanbule K, Salzberg SL: **Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake.** *Genome Biol* 2007, **8**:R22.
12. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403–410.
13. Arnvig KB, Young DB: **Identification of small RNAs in *Mycobacterium tuberculosis*.** *Mol Microbiol* 2009, **73**:397–408.
14. Bohn C, Rigoulay C, Chabelskaya S, Sharma CM, Marchais A, Skorski P, Borezee-Durant E, Barbet R, Jacquet E, Jacq A, Gautheret D, Felden B, Vogel J, Bouloc P: **Experimental discovery of small RNAs in *Staphylococcus aureus* reveals a riboregulator of central metabolism.** *Nucleic Acids Res* 2010, **38**:6620–6636.
15. Christiansen JK, Nielsen JS, Ebersbach T, Valentin-Hansen P, Sogaard-Andersen L, Kallipolitis BH: **Identification of small Hfq-binding RNAs in *Listeria monocytogenes*.** *RNA* 2006, **12**:1383–1396.
16. Fouquier DA, Wessner F, Halpern D, Ly-Vu J, Kennedy SP, Serror P, Aurell E, Repoila F: **A simple and efficient method to search for selected primary transcripts: non-coding and antisense RNAs in the human pathogen *Enterococcus faecalis*.** *Nucleic Acids Res* 2011, **39**:e46.

17. Geissmann T, Chevalier C, Cros MJ, Boisset S, Fechter P, Noirot C, Schrenzel J, Francois P, Vandenesch F, Gaspin C, Romby P: **A search for small noncoding RNAs in *Staphylococcus aureus* reveals a conserved sequence motif for regulation.** *Nucleic Acids Res* 2009, **37**:7239–7257.
18. Irnov I, Sharma CM, Vogel J, Winkler WC: **Identification of regulatory RNAs in *Bacillus subtilis*.** *Nucleic Acids Res* 2010, **38**:6637–6651.
19. Kumar R, Shah P, Swiatlo E, Burgess SC, Lawrence ML, Nanduri B: **Identification of novel non-coding small RNAs from *Streptococcus pneumoniae* TIGR4 using high-resolution genome tiling arrays.** *BMC Genomics* 2010, 11:350.
20. Livny J, Teonadi H, Livny M, Waldor MK: **High-throughput, kingdom-wide prediction and annotation of bacterial non-coding RNAs.** *PLoS One* 2008, **3**:e3197.
21. Mandin P, Repoila F, Vergassola M, Geissmann T, Cossart P: **Identification of new noncoding RNAs in *Listeria monocytogenes* and prediction of mRNA targets.** *Nucleic Acids Res* 2007, **35**:962–974.
22. Mraheil MA, Billion A, Mohamed W, Mukherjee K, Kuenne C, Pischimarov J, Krawitz C, Retey J, Hartsch T, Chakraborty T, Hain T: **The intracellular sRNA transcriptome of *Listeria monocytogenes* during growth in macrophages.** *Nucleic Acids Res* 2011, **39**:4235–4248.
23. Nielsen JS, Olsen AS, Bonde M, Valentin-Hansen P, Kallipolitis BH: **Identification of a sigma B-dependent small noncoding RNA in *Listeria monocytogenes*.** *J Bacteriol* 2008, **190**:6264–6270.
24. Perez N, Trevino J, Liu Z, Ho SC, Babitzke P, Sumby P: **A genome-wide analysis of small regulatory RNAs in the human pathogen group A *Streptococcus*.** *PLoS One* 2009, **4**:e7668.
25. Rasmussen S, Nielsen HB, Jarmer H: **The transcriptionally active regions in the genome of *Bacillus subtilis*.** *Mol Microbiol* 2009, **73**:1043–1057.
26. Tezuka T, Hara H, Ohnishi Y, Horinouchi S: **Identification and gene disruption of small noncoding RNAs in *Streptomyces griseus*.** *J Bacteriol* 2009, **191**:4896–4904.
27. Toledo-Arana A, Dussurget O, Nikitas G, Sesto N, Guet-Revillet H, Balestrino D, Loh E, Gripenland J, Tiensuu T, Vaitkevicius K, Barthelemy M, Vergassola M, Nahori MA, Soubigou G, Regnault B, Coppee JY, Lecuit M, Johansson J, Cossart P: **The *Listeria* transcriptional landscape from saprophytism to virulence.** *Nature* 2009, **459**:950–956.
28. Vockenhuber MP, Sharma CM, Statt MG, Schmidt D, Xu Z, Dietrich S, Liesegang H, Mathews DH, Suess B: **Deep sequencing-based identification of small non-coding RNAs in *Streptomyces coelicolor*.** *RNA Biol* 2011, **8**:468–477.

doi:10.1186/1471-2164-13-384
Cite this article as: Pischimarov *et al.*: sRNAdb: A small non-coding RNA database for gram-positive bacteria. *BMC Genomics* 2012 **13**:384.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central

13.11 Appendix publication 10

Dynamic integration hotspots and mobile genetic elements shape the genome structure of the species *Listeria monocytogenes*

Carsten Kuenne[1], André Billion[1], Mobarak Abu Mraheil[1], Axel Strittmatter[2], Rolf Daniel[2], Alexander Goesmann[3], Sukhadeo Barbuddhe[4], Torsten Hain[1§], Trinad Chakraborty[1§]

[1]Institute of Medical Microbiology, Justus-Liebig University, Giessen, Germany and German Centre for Infection Research
[2]Department of Genomic and Applied Microbiology and Goettingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University Goettingen, Goettingen, Germany
[3]Bioinformatics Resource Facility, Center for Biotechnology, Bielefeld University, Bielefeld, Germany
[4]ICAR Research Complex for Goa, Ela, Old Goa 403402, India

[§]Corresponding authors

Email addresses:
CK: Carsten.Kuenne@mikrobio.med.uni-giessen.de
AB: Andre.Billion@mikrobio.med.uni-giessen.de
MM: Mobarak.Mraheil@mikrobio.med.uni-giessen.de
AS: AxelStrittmatter@eurofins.com
RD: rdaniel@gwdg.de
AG: agoesman@cebitec.uni-bielefeld.de
SB: barbuddhe@icargoa.res.in
TH: Torsten.Hain@mikrobio.med.uni-giessen.de
TC: Trinad.Chakraborty@mikrobio.med.uni-giessen.de

Abstract

Background

To explore the genetic and phenotypic diversity of the facultative pathogenic species *Listeria monocytogenes*, we genome sequenced an additional 11 strains thereby completing coverage of all known serotypes and performed comparative analysis on them in conjunction with publicly available data.

Results

The species pan-genome of *L. monocytogenes* is highly stable but open, suggesting an ability to adapt to new niches by generating or including new genetic information. The majority of gene-scale differences resulted from nine hyper variable hotspots, a similar number of different prophages, three transposons (Tn916, Tn554, IS3-like), and two mobilizable islands. Only a subset of strains showed CRISPR/Cas bacteriophage resistance systems of different subtypes, suggesting a supplementary function in maintenance of chromosomal stability. Multiple phylogenetic branches of the genus *Listeria* imply long common histories of strains of each lineage as revealed by a SNP-based core genome tree highlighting the impact of small mutations for the

evolution of species *L. monocytogenes*. Frequent loss or truncation of genes described to be vital for virulence or pathogenicity was confirmed as a recurring pattern, especially for strains belonging to lineages III and II. New candidate genes implicated in virulence function were predicted based on functional domains and phylogenetic distribution. A comparative analysis of small regulatory RNA candidates supports observations of a differential distribution of *trans*-encoded RNA, hinting at a diverse range of adaptations unlike the commonly conserved *cis*-regulatory elements that were putatively present in the last common ancestor of *L. monocytogenes*.

Conclusions

This study determined commonly occurring hyper variable hotspots and mobile elements as primary effectors of gene-scale evolution of species *L. monocytogenes* based on the most complete pan-genomic distribution available to date. The discovery of common and disparately distributed genes considering lineages, serogroups, serotypes and strains of species *Listeria monocytogenes* will be helpful for further diagnostic, phylogenetic and functional research, assisted by the comparative genomic GECO-LisDB analysis server (http://bioinfo.mikrobio.med.uni-giessen.de/geco2lisdb).

Background

The genus *Listeria* consists of eight species being *L. monocytogenes*, *L. innocua*, *L. welshimeri*, *L. seeligeri*, *L. ivanovii*, *L. grayi*, *L. marthii* and *L. rocourtiae* [1-3]. *Listeria* are saprotrophic with *L. monocytogenes* and *L. ivanovii* considered facultative pathogens, the latter predominantly causing infections in ruminants [4]. *L. monocytogenes* represents the species most commonly associated with listeriosis in humans which primarily affects immunocompromised individuals [5]. The majority of infections are thought to be foodborne and results in high mortality rates [6]. Virulence of the bacterium is heavily dependent on the virulence gene cluster (VGC, LIPI-1) which promotes cytosolic replication as well as intra- and intercellular movement [7]. A second cluster required for virulence contains an operon of two genes (*inlA/B*) that encode internalins necessary for the attachment to and invasion of non-phagocytic host cells [8]. The species *L. ivanovii* displays a specific island with virulence factors called LIPI-2, comprising of multiple internalins and *smcL* sphingomyelinase hemolysis gene [9]. A subset of strains of lineage I carry an additional hemolysin called listeriolysin S (LIPI-3) which contributes to virulence *in vitro* [10]. Other genes involved in the infectious process modulate the bacterial metabolism and stress response [11,12].

A variety of cell wall components are important for the survival of strains of species *Listeria monocytogenes* in the environment and the infected host. They serve as docking sites crucial for processes including detection, adhesion or invasion and can be grouped according to their anchoring domains. This study will focus on surface proteins covalently linked to peptidoglycan (LPXTG), those with domains facilitating non-covalent interactions (GW, P60, LysM), and membrane-bound lipoproteins [13,14]. Internalins belong to a family of leucine-rich repeat (LRR) containing surface-associated proteins implicated in facilitating protein-protein interactions vital for adhesion and invasion [15]. LRR domains were also found to be vital parts of a lymphocyte-based immune system of jawless vertebrates attesting to the wide-spread utility of these diverse sequences [16,17].

To protect from bacteriophage activity, some Archaea and bacteria have developed an adaptive immune system (CRISPR: clustered regularly interspaced short palindromic repeats) based on a variable module of repeats, spacers and protein coding genes (Cas: CRISPR associated) [18]. By integrating target DNA (spacer) between repeats the system can recognize and silence the expression of these sequences. Recently it was shown that CRISPR spacers can bear sequences homologous to chromosomal genes which may represent a form of autoimmunity or regulatory mechanism [19,20]. Multiple types or guilds of Cas genes were differentiated bearing a similar functionality split among a variable number of frequently non-homologous genes [21]. Some CRISPR/Cas subtypes lacking endoribonucleases necessary for the maturation of crRNAs were shown to appropriate a *trans*-encoded small RNA (*tracrRNA*) in combination with a host factor (Rnase III) in order to facilitate the silencing of foreign nucleic acids [22]. Apart from silencing expression *cas* genes were also described to have a function in DNA repair and chromosome segregation [23].

Small non-coding regulatory RNAs have emerged as a further layer of gene expression regulation in prokaryotes. They regulate transcription by pairing with other RNAs, forming parts of RNA-protein complexes, or adopting regulatory secondary structures [24]. Small non-coding RNAs have been implicated in responses to iron limitation, oxidative stress, low temperature and intracellular growth [25,57].

Strains of *L. monocytogenes* can be grouped into four evolutionary lineages and 12 serotypes representing distinct phylogenetic, ecologic and phenotypic characteristics [26,27]. Lineage I was found to be overrepresented among human clinical isolates and epidemic outbreaks in most studies while lineage II is typically sporadically isolated from both humans and animals. Lineage III and IV are rare and predominantly identified in animals. These associations show frequent regional differences, thus rendering the definition of a natural environment difficult. Lineages II, III and IV show higher recombination rates and a lower degree of sequence similarity than lineage I. This observation was proposed to result from less diverse lifestyles for the latter and may denote strains of lineage I as descendants of a recently emerged highly virulent clone [28,29]. Plasmids are more prevalent in lineage II and include a multitude of resistance genes dealing with toxic metals, horizontal gene transfer, oxidative stress and small toxic peptides [30]. Furthermore, strains of this lineage often show virulence attenuated phenotypes due to deletions inside important virulence genes [31]. Approximately half of the strains compared in this study were found to be virulence attenuated as assessed by low invasion rates of epithelial cells [32]. About 98% of human cases of listeriosis are caused by strains of serotypes 4b, 1/2a, 1/2b and 1/2c [33].

This study is the first one to base its evolutionary analyses on a set of 16 completely sequenced genomes of species *L. monocytogenes* including strains of all serotypes. We determined common and distinct genetic elements to understand the diversity of forces shaping the species down to the level of strains. Considering the scope of the underlying data, this work is intended to serve as a framework to support future analyses for the *Listeria* research community.

Results and Discussion

Common features of strains selected among all serotypes of *L. monocytogenes*

In order to analyze the evolution and pan-genomic potential of the species, strains of *L. monocytogenes* spanning all known serotypes originating from various sources were selected for comparison (Table 1). These include five completely sequenced public strains being 4a L99, 4b F2365, 1/2a EGD-e, 1/2a 08-5578 and 1/2a 08-5923 as well as the eleven newly sequenced genomes [34-37]. The chromosome of *L. monocytogenes* 7 SLCC2482 harbours one gap which is located at 2125011 bp and is estimated to have a size of approximately 10000 bp. Four strains harbored plasmids which were described previously [30]. All strains were classified according to known sequence types and chromosomal complexes using the BIGSdb software [26,38].
The chromosomes compared show a similar size, G+C content, and average length of protein coding genes and percentage of protein coding DNA (Table 2). The number of coding sequences ranged from 2755 (SLCC2376) to 3010 (08-5578). We identified six 16S-23S-5S-rRNA operons in most strains with the exception of 1/2a 08-5578 and 1/2a 08-5923 which lack one rRNA module each. Similarly, the majority of strains contain 67 tRNA genes apart from 4e SLCC2378 and 3c SLCC2479, which harbor 66 and 65 tRNA genes respectively, as well as 1/2a 08-5578 and 08-5923 that were found to encode only 58 tRNA genes.

Pan-genome model predicts a highly conserved species

The pan-genome of 16 chromosomes of *Listeria monocytogenes* was found to contain 4387 genes including 114 paralogues based on a similarity cutoff of 60% amino acid identity and 80% coverage of protein alignments (Figure 1). Approximately 78% of coding sequences per strain consist of mutually conserved core genes (2354 / species) indicating a highly stable species backbone with relatively few accessory genes (2033 / species) (Additional File 1). More than half of the species accessory genes (1161) furthermore displayed homologues in only one or two strains implying relatively recent insertions that are rarely fixed in the population. A power law regression analysis predicting a future pan-genomic distribution after further sequencing resulted in a mean power law fitting for new genes of $n=397.4N^{-0.7279}$ ($\alpha=0.7279$). This indicates a conserved but open pan-genome that permits limited integration of foreign DNA or generation of genetic diversity by other evolutionary forces such as mutation, duplication and recombination as previously described [39]. Regression curves predict the presence of ca. 6000 different genes in the pan-genome of *L. monocytogenes* after 100 strains have been completely sequenced.
Other studies relying on the hybridization of eight lineage III strains on a microarray based on 20 strains (two complete, 18 draft chromosomes) found a closed species pan-genome [40,41]. These likely represent an underestimation of true sequence diversity of the species because they lack multiple serotypes (e.g. 3a, 3b, 3c, 4e, 7), less stringent similarity cutoffs and a lower number of fully sequenced strains. The pan-genome of genus *Listeria* based on chromosomes of 13 strains (six complete, seven draft) was determined to be open [42].
Studies of species of other genera found varying results reflecting differences in habitat, evolutionary pressure and gene pool. Only around one half of the genes of each strain were considered to be mutually conserved in species *Bacillus anthracis*, *Streptococcus pneumoniae* and *Escherichia coli* leading to diverse pan-genomes [43-

45]. Species *Streptococcus pyogenes*, *Streptococcus agalactiae* and *Haemophilus influenzae* seemed much more stable and displayed approximately three-quarters of core genes similar to species *L. monocytogenes* [43,46,47]. It should be noted that these analyses relied on disparate measures of homology and included between 10 and 20 strains of various qualities and sources, thus turning a comparison into a non-trivial task.

Summarily, our research shows a conserved species, which tolerates low but vital levels of horizontal gene transfer, overrepresented in highly diverse areas on the chromosome.

Hyper variable hotspots contain one fourth of the accessory genes and permitted the insertion of major pathogenicity determinants

Chromosomes of the strains compared contain a number of regions, which display at least three non-homologous insertions between mutually conserved core genes. These hotspots are suggested to be the result of a founder effect [45]. Once a permissive region has acquired a substantial integration without reducing the fitness of the respective strain, future insertions become more likely since the locus now offers a larger target for neutral insertion. We identified nine high-flux hotspots that were located between mutually conserved anchor genes. The latter showed no overrepresentation among any particular functional or genetic category. Hotspots contained 55 (4a L99) to 111 (4b F2365) genes per strain (Additional File 2, Additional File 3). Interestingly, strains of lineage III displayed an average of 56 genes inside these loci, while strains of lineage I and II contained nearly twice as many (80-90), indicating either stronger deleterious forces in the former or an increased number of insertions in the latter. Taken together, every tenth gene of the pan-genome (454 genes) is located in such a highly variable region. This translates to 22% of the non-core accessory genes. One third of these were accounted for by strain-specific insertions leading to a low average conservation of hotspot genes in only three strains. The majority of these genes have no known function (298), 35 are part of restriction modification systems, and 13 are involved in genetic mobilization.

We could only identify a small number of genes located inside hotspot loci which exhibit an obvious adaptive value for the host genome, including the previously described pathogenicity determinants *inlA/B* and LIPI-3 [8,10]. Transposon Tn916 introduced additional cadmium resistance genes into its host strain 1/2a EGD-e. Two variants of an IS3-like transposon were inserted in different hotspot integration sites of the epidemic lineage I and found to bear multiple surface-associated proteins. The latter are implied in attachment, invasion, and other interactions with the environment and were identified in most hotspots resulting in the presence of a total of 40 genes of this category.

Hyper variable regions in species *L. monocytogenes* may be considered evolutionary test areas that attract new genetic information by frequent insertions, deletions and other differentiating forces, rarely leading to fixation of genes in the population. Interestingly, all but one of these hotspots are located on the right replichore, which thus represents an area of increased genomic plasticity. Only half of the variable regions displayed identifiable mobilization genes indicating either unidentified mobilization genes, decay or other means to facilitate insertions putatively including also mechanisms for homologous recombination.

Chromosomal mobile genetic elements are major sources of diversity – prophages, transposons and genetic islands

In order to find large insertions in the chromosomes of the respective strains we plotted all coding sequences, which were not conserved in all strains, resulting in the identification of between one and five mobile genetic elements (MGE) such as prophages, transposons, insertion sequences and genomic islands per chromosome (Figure 2). These introduced 6 to 235 protein coding genes per strain included in 15 different MGE insertions into 13 distinct chromosomal loci (Additional file 4). This translates into 703 genes of the pan-genome (15%) or one third of the accessory genes.

Among these are 8 different prophages which are typically inserted by site-specific recombination into chromosomal loci adjacent to tRNA genes as previously observed [48]. We also found two different bacteriophages (A006 and A118) which targeted the *comK* gene. Most prophages belong to the class of listeriaphages (B025, A118, and A006) or show a high similarity to unnamed prophages also found in the genera *Bacillus, Enterococcus, Clostridium* and *Staphylococcus*. The large majority of prophage genes code for hypothetical and structural proteins. Multiple *Listeriae* were described to up-regulate phage gene expression including the mutually conserved *lma* operon (part of monocin prophage-rest) and surrounding genes to assist intracellular survival [34]. It should be noted that the only strains without apparently complete prophages are both strains of serotype 4b as well as 4c SLCC2376.

Three putative transposons were identified in the strains studied. Two of them are located between homologues of genes *lmo1096-lmo1115* in strain EGD-e (ICELm1, TN916-like) and *lmo2676-lmo2677* in 3c SLCC2479 and 1/2c SLCC2372 (TN554-like), respectively (Additional File 5) [49]. ICELm1 contains two genes involved in cadmium resistance and a fibrinogen-binding protein with an LPXTG domain which is implied in host cell attachment in *Staphylococcus epidermidis* [50]. The Tn554-like transposon introduced an arsenate resistance operon (*arsCBADR*) also found in *Enterococcus faecalis* (ca. 70% amino acid identity) into its host chromosomes. The third putative transposon consists of 15 genes including two insertion elements bearing two IS3-type transposases as found in its complete form in strain 3b SLCC2540 (Additional File 6). It contains a module consisting of a transcriptional regulator and four homologues of a lipoprotein. The latter was predicted by previous studies to furthermore contain an RGD motif implied in integrin binding and a weak homology to leucine-rich-repeat domains, indicating a putative function in host-pathogen interaction [14,36]. Deletion versions of this transposon, which have lost one insertion element, can be found at the same relative position at approximately 2.1 Mb (e.g. *LMOf2365_2051-9*) in all strains of lineage I and another variant at ca. 0.5 Mb (e.g. *LMOf2365_0493-500)* in a subset of strains of all lineages. Interestingly, indels of the complete transposon and the lipoprotein itself have led to a distribution of 4-7 instances of the lipoprotein in epidemic lineage I in comparison to 0-1 in lineages II and III, which further indicates these two modules as potential targets for research regarding pathogenicity determinants. All but one transposon were found in a hyper variable hotspot suggesting either relaxed deleterious forces in these areas or an enrichment of repeats targeted by the respective transposases.

Another type of MGE is designated genomic island and denotes a mobilizable element frequently composed of modules derived from chromosomes, plasmids or bacteriophages. One of these was called Listeria genomic island 1 (LGI1) and putatively introduced by serine recombinases into 1/2a 08-5923 and 08-5578 [37]. It was described to include genes involved in secretion, protein-protein interaction,

adhesion, multidrug efflux, signal transduction and restriction modification. We identified a second genomic island named LGI2, which has not yet been described in the literature. It spans approximately 35000 bp in strains 4e SLCC2378 and 4d ATCC19117 and integrated into genes orthologous to *lmo2224* (1/2a EGD-e). This mobile element consists of 36 genes and putatively inserted by means of a bacteriophage integrase (*LMOSLCC2378_2256*) distantly related to temperate *Lactococcus lactis* bacteriophage phiLC3 [51]. Additionally, a putative operon of eight genes coding for arsenate resistance proteins (*LMOSLCC2378_2263-70*) was found to be homologous to a region of *Listeria innocua* Clip11262 plasmid pLI100, indicating recombination between phages, plasmids and chromosomes which resulted in the formation of this mobile element. Other genes of this locus code for ATP transporters, a putative anti-restriction protein, a secreted and a cell wall surface anchor protein.

In summary, a contribution to pathogenicity or environmental adaptation due to genes inserted into the chromosome of *L. monocytogenes* by identifiable MGEs seems to be rare. Phage-related genes of other genera were previously described to contribute to virulence and pathogenicity due to toxins and resistances versus antibiotics and osmotic, oxidative and acid stresses [52,53]. We were unable to identify prophage-related genes in the strains compared which bear homology to published pathogenicity modulators, with the exception of integrases that were demonstrated to assist biofilm formation in *E. coli* [54]. The intracellular up-regulation of prophage-genes nonetheless indicates a function for intracellular survival of *Listeriae* that has still to be uncovered [34].

The distribution of most MGEs is heterogeneous indicating either recent insertions and/or frequent deletion of these sequences. MGEs occur in similar proportions among all lineages, with lineage I, II and III containing a mean of 3.7, 3.9, and 3 mobile genetic elements per strain, including 1.6, 2.7, and 2.5 prophage remnants, respectively. The majority of the compared strains of lineages I and III (6/9) do belong to serogroup 4 and contain either one or two prophage remnants with the exception of 4a L99 which contains four of these elements. All serogroup 4 strains of this study also lack 12 out of 16 genes of an operon encoding a rhamnose pathway for teichoic acid biosynthesis conserved in all other strains (*lmo1076-91*) as well as several glycosyl transferases (*lmo0497, lmo0933, lmo2550*) that are responsible for differences in teichoic acid composition, which may inhibit recognition by certain phages [55]. The general rarity of mobile genetic elements in the compared strains supposes mechanisms to limit inclusion of foreign DNA.

CRISPR/Cas systems represent supplementary bacteriophage defense mechanisms for the species *L. monocytogenes*

Chromosomes of *L. monocytogenes* contain parts of a CRISPR/Cas-system implied in defense versus bacteriophages at three different loci (Additional File 7). These were identified by a combination of PILER-CR 1.02, CRT 1.1 and manual correction using BLASTN leading to slightly higher counts of repeat/spacer modules than previously published for strains 4a L99 and 1/2a EGD-e [34] (Additional File 8).

All strains bear a putative remnant of a CRISPR-system at ca. 0.5 Mb in strain 1/2a EGD-e. The array contains 4-11 repeats with a consensus sequence of 29 bp and is not associated with any *cas* genes. Spacer content shows a clear separation along lineage borders with a considerable amount of similarity between strains inside lineage I and II, respectively. No correlation could be found between spacers of strains of lineage

III, indicating that ancestors of lineage I and II have lost the *cas* genes necessary to create new spacers inside locus 1, leading to a relatively homogenous distribution, while strains of lineage III maintained this ability for a period long enough to completely differentiate their spacer sequences. This locus is also conserved in strains of other species such as *L. innocua* Clip11262 and *L. seeligeri* SLCC3954, but not in *L. welshimeri* SLCC5334.

Locus 2 is located ca. 10kb adjacent to locus 1 and resembles the *Thermotoga neapolitana* (Tneap) subtype which consists of *cas6*, *cst1*, *cst2*, *cas5t*, *cas3* and *cas2* [21]. Homologues of this system exist in 4a L99, 7 SLCC2482 and 1/2b SLCC2755 at the same relative chromosomal position and no sequence remnants could be identified in other chromosomes, suggesting the insertion of this locus in a common ancestor of these strains. It contains arrays with 21-28 repeats and displays a consensus sequence with only one mismatch in comparison to repeats of locus 1 indicating a close relation. Spacers are identical in strains 7 SLCC2482 and 1/2b SLCC2755, while 4a L99 shows a completely different content. Locus 2 is also conserved in other strains of *L. monocytogenes* (1/2a: F6854, F6900, J0161, J2818, 1/2b: R2-503, 4a: HCC23, 4b: N1-017) and *L. seeligeri* SLCC3954.

Locus 3 is inserted into homologues of a lipoprotein gene (*lmo2595*) located at ~2.7 Mb relative to the chromosome of reference strain 1/2a EGD-e. It was found to be present in 1/2a SLCC5850, 7 SLCC2482, 1/2b SLCC2755 and 3b SLCC2540 without any local sequence homologies in other strains, implying insertion into a common ancestor of the former strains. Strain 4a L99 does not contain CRISPR/Cas locus 3 but listeriaphage A118 is located at the same relative position here, highlighting a role for this region as a common insertion point for horizontal gene transfer. This locus was found to contain *csn2*, *cas2*, *cas1* and *csn1* and thus classified as subtype *Neisseria meningitidis* (Nmeni). The array spans 25-41 repeats, each with a length of ca. 36 bp. These were found to exhibit a considerable global homology to the 29 bp repeats of locus 1 and 2 (17 identical nucleotides) implying a relation between the *cas* genes of all loci. Spacer content of locus 3 is clonal for strains 7 SLCC2482 and 1/2b SLCC2755 while 1/2a SLCC5850 and 3b SLCC3540 display mostly unique spacers, including a number of duplicates versus listeriaphages A500 and A118. We found a range of further mostly pathogenic strains of *L. monocytogenes,* which also display this locus (1/2a: 10403S, F6854, F6900, N3-165, J0161, J2818, 1/2b: J1.194, R2-503, 4b: N1-017). Strain *L. innocua* Clip11262 contains the only chromosome outside of species *L. monocytogenes* which harbours a partial homologue. Locus 3 belongs to subtype Nmeni which was previously described to rely on a *trans*-encoded sRNA (tracrRNA) located upstream of *csn1* and host factor RNase III in order to compensate for a missing endoribonuclease gene [22]. We could exclusively identify perfect matches of the 94 bp tracrRNA variant as expressed by *L. innocua* Clip11262 in all compared strains of *L. monocytogenes* bearing locus 3 at a position upstream of *csn1*. We thus hypothesize, that this locus functions according to the former principles and may only be able to silence foreign nucleic acids inside a host which is able to supply an RNase III enzyme.

All identifiable spacers (81/276) are directed versus known listeriaphages or related composite prophages. Nearly two thirds (50/81) of these displayed a perfect match while the rest contained a maximum of one mismatch. Non-perfect matches are indicative of either decay, rendering the spacer ineffective, or a related but different target, that has not yet been sequenced. We also encountered multiple different spacers homologous to sequences of the same phage in the same array, as well as identical duplications of one spacer. It is tempting to speculate that inclusion of

redundant spacer sequences increases the likelihood of a successful defense against the respective bacteriophage (ex.: A118, A500, B025). We never observed identical spacers to be present in multiple arrays, indicating a clear separation of all loci.
In conclusion, we hypothesize that an ancestor of genus *Listeria* contained CRISPR locus 1 (*lmo0519-lmo0520*) that lost its associated *cas* genes during early evolutionary events. Interestingly, this locus was previously described as trans-acting small non-coding RNA RliB in strain 1/2a EGD-e indicated in control of virulence [56,57]. Thus, this remnant CRISPR array may have been adapted for regulation in 1/2a EGD-e and possibly other strains of the species.
Five of 16 strains compared in this work still contain at least one of two types of putatively functional CRISPR/Cas systems indicating an ongoing selective pressure by bacteriophages. On the other hand, presence or lack of such a system does not correlate with number or type of prophages identified per strain and 11 strains neither bear a functional CRISPR/Cas system nor an increase of other defense mechanisms such as restriction modification systems (data not shown). We suggest that CRISPR/Cas represents an additional line of defense directed against bacteriophage attacks that can be gained by horizontal gene transfer and seems to be effective only for a subset of strains of genus *Listeria*.
We could not identify any spacer targeting chromosomal or plasmid sequences of species *L. monocytogenes* apart from integrated prophages, indicating that CRISPR/Cas does not serve further regulatory roles. The variable nature of CRISPR-arrays suggests their future use in differentiating strains or lineages by typing procedures. Further research will now be necessary to determine the operational capability of locus 2 and 3 in the environment or host.

Phylogenies compared – relationships between lineages, serogroups, serotypes and strains according to genomic and genetic content

This analysis used the complete genomic sequences of 19 strains of genus *Listeria* including those of related species being *L. innocua* 6a Clip11262, *L. welshimeri* 6b SLCC5334 and *L. seeligeri* 1/2b SLCC3954 to identify phylogenetic relationships. In order to enable phylogenetic clustering we created a well-supported (bootstrap >80%) core-genome tree based on an alignment of all concatenated core genes (2018) of 19 strains using Mugsy [58] (Figure 3A/B). This tree shows distances between strains based on small adaptations inside mutually conserved genes, which translate into an approximate timeline when assuming consistent rates of evolution. We found that strains of species *L. monocytogenes* clustered inside three clearly separated lineages in support of previous observations. Lineage III contains serotypes 4a and 4c, lineage II includes 1/2a, 1/2c, 3a and 3c and lineage I bears strains of serotypes 1/2b, 3b, 4b, 4d, 4e and 7. Differentiation leading to separate serotypes apparently had little impact on the placement of the branches apart from the general lineage. We identified the closest relationships between strains of different serotypes being 1/2b SLCC2755, 7 SLCC2482 (termed phylogenomic group 1 or PG1) and 4e SLCC2378, 4b F2365 (PG2) in lineage I as well as 1/2a EGD-e, 1/2c SLCC2372, 3c SLCC2479 (PG3) in lineage II, with the exception of clonal strains 08-5578 and 08-5923 which both belong to serotype 1/2a. Strains of serotypes 4e and 4d were found on a branch displaying strain 4b L312 as its oldest ancestor in support of a previous hypothesis indicating serotype 4b as ancestral state for serotypes 4e and 4d [26].
We additionally clustered all strains based on the accessory gene content (presence/absence of 2953 genes) to identify the impact of gene-scale indels, which

includes most horizontal gene transfer events [59] (Figure 3C). This methodology was shown to be biased towards a tree topology that parallels convergence in lifestyle and thus displays a phenotypical relationship among the compared strains [60]. The resulting tree was found to be well supported (>80% bootstrap) with the exception of the placement of branches neighboring the central *L. monocytogenes* junction, implying early indels and recombination that lead to inconsistent topologies.

If only gene gain and loss are taken into account, lineages of *L. monocytogenes* are closely related to other listerial species, indicating that large evolutionary timeframes shown by the SNP-based core-genome tree resulted in a low number of conserved gene-scale indels.

The opposite is apparent when considering phylogenomic groups, which were found to be closely related in the core-genome tree but to a much lesser degree considering gene content, implying a number of young indels. Interestingly, phylogenomic groups are located at the end of shorter common branches in the gene content tree, which is due to a small number of exclusively conserved genes (PG1: 28, PG2: 20, PG3: 22, primarily hypothetical and truncated genes) (Additional File 1). Thus, phylogenomic groups can be considered closely related but do not necessarily share the same niche or phenotype. Other branches are supported by a varying number of conserved and predominantly hypothetical genes (ex. 4b L312, 4b F2365, 4d ATCC19117, 4e SLCC2378: 18 genes; 3b SLCC2540, 1/2b SLCC2755, 7 SLCC2482: 5 genes) that are distributed along the chromosomes in small modules.

We identified three topological changes between core-genome and gene content tree hinting at shared indels that run contrary to the phylogenomic signal of core-genome SNPs. Removal of genes related to mobile genetic elements (34% of accessory genes) from the gene content matrix resulted in a topology very similar to the core-genome tree. Thus, large-scale insertions, which resulted mainly from bacteriophage integration, run contrary to the "true" phylogenetic signal by inserting many genes in one event as well as by putative parallel insertions into different strains. The only remaining difference was observed considering a common branch for strains of lineage III and apathogenic species, highlighting small-scale indels as causative force. This supports a previous hypothesis suggesting lineage III as a possible deleterious intermediate state between lineages I/II and apathogenic species [61].

Interestingly, the majority of accessory genes of species *L. monocytogenes* were either scattered along the chromosomes (46%) or found inside hyper variable regions (20% when excluding MGE) and thus likely originated from a wide range of diversifying forces. Gradual change seems to be a superior factor for the evolution of gene content of *Listeriae* when compared to large-scale insertions of multiple genes by mobile elements.

In summary, tree topologies based on a core-genome alignment and gene content were found to be highly similar despite the obfuscating influence of mobile genetic elements. Other studies on *Rickettsia/Orienta* species and *E.coli/Shigella* found considerable differences in the respective phylogenies indicating more distinct evolutionary histories for the gene repertoires involved [45,62]. The relative correspondence of SNPs and gene-scale indels in genus *Listeria* could be a result of differential acquisition and loss of genes in accordance to various evolutionary descents as previously described considering other genera [63,64].

Frequent loss and disruption of known virulence-associated genes may explain observed phenotypic attenuations

About one third of the genes which displayed compelling evidence for involvement in the infectious process were found to be absent or to code for a truncated protein in at least one of the strains studied, putatively impacting the disease phenotype (Additional file 9) [10,11,65,66]. Thus, most of the strains compared here showed similar growth rates in BHI medium at 37°C with the exception of strain 1/2a SLCC5850 which grew at a considerably lowered rate (Additional File 10). Rates of mortality of larvae in the *Galleria mellonella* model system indicative of pathogenicity showed that strains of serotype 4b killed most larvae, followed by 1/2c SLCC2372, 3a SLCC7179, 1/2a EGD-e, 1/2b SLCC2755, 3b SLCC2540 and 3c SLCC2479 (Figure 4). The remaining strains showed a low degree of pathogenicity in this model.

In order to correlate phenotypes with genomic differences we performed detailed analyses of virulence-associated genes that allow us to present hypotheses on the evolutionary descent of these changes (Additional File 12). In short, deletions affecting primary virulence genes *prfA* (1/2a SLCC5850), *plcA* (3a SLCC7179), *inlA* (3c SLCC2479), and *inlB* (4b F2365) [67] were identified in four strains. A number of surface-associated genes were found to be absent from strains of lineage III and especially from strain 4a L99. Further deletions which putatively interfere with regulation of the SigB regulon during stress are related to genes *rsbS* (1/2c SLCC2372), *rsbV* (4d ATCC19117) and *rsbU* (3c SLCC2379) [68-70].

In conclusion, strains of *L. monocytogenes* frequently lose determinants of pathogenicity leading to virulence-attenuated phenotypes, which may be advantageous in some environments, especially considering lineage III. Interestingly, highly invasive and/or pathogenic strains of serotypes 4b, 1/2a, 1/2b, and 1/2c also displayed a range of deletions here, indicating a certain amount of redundancy of these functions.

Distribution of surface-associated genes displays conserved lineage-backbones with strain-specific adaptations

A detailed examination was undertaken to spot relevant patterns of presence or absence of surface-associated genes mediating interaction with the environment and the infected host, and to invoke evolutionary explanations (Additional File 13, Additional File 14).

To conclude, genes bearing P60 or LysM domains showed little variation among the strains studied (Additional File 15). Between 6 and 16 non-core lipoprotein coding genes were identified, indicating some differentiation. These were frequently located in chromosomal hotspots of horizontal gene transfer and found inside or adjacent to prophage insertions, hinting at putative methods of transmission. Interestingly, all strains of epidemic lineage I show an exclusive gene (*LMOf2365_1974*) with both LPXTG and GW domains, which may become a future research target when considering the role of cell wall anchored modulators of virulence or pathogenicity.

Internalins are implied in cell adhesion and invasion of host cells and contain a leucine-rich repeat (LRR) domain indicated in protein-protein interaction (Additional File 17). InlB B-repeats represent a hallmark of previously described virulence-associated internalins [71], and were identified in 15 clusters, thus increasing the

probability of the respective genes to be involved in host-pathogen interaction. The distribution of putative internalins revealed that only four of 42 homology clusters are mutually conserved, confirming previous observations of diversity, especially considering lineages II and III [72]. A number of known virulence-associated internalins were absent in a subset of strains, putatively resulting in a reduced number of infectable cell types (lineage III: *inlC* and *inlF*, 4a L99: *inlGHE*, *inlI* and *inlJ*, 3c SLCC2479: *inlA*, 4b F2365: *inlB*). The absence of *inlC* in strains of lineage III may have been caused by a deleterious transposition moving two adjacent lipoprotein coding genes (*lmo1264-5*) by approximately 600kb to replace the internalin (Additional File 16). Interestingly, we identified two different versions of *inlF* and *inlJ* in lineage I as compared to lineages II/III, putatively resulting in different adhesion properties. Only one internalin was found to be specific and mutually conserved for lineage I (*LMOf2365_0805*), indicating this gene for further research regarding virulence.

Taken together, we found that most surface-associated genes are either mutually conserved or were likely present in an early ancestor of a lineage, implying a fixed core-functionality that is rarely complemented by strain-specific additions. Of all the classes surveyed, internalins seemed to be the most diverse, driven by duplication, recombination and transposition. We furthermore identified a large number of novel surface-associated genes, including their distribution among all serotypes of species *L. monocytogenes*, thereby presenting a pool of candidates for future analysis considering virulence and pathogenicity.

Ancestral genes of serotypes, serogroups and lineages reveal new marker and virulence-associated genes

In order to identify those ancestral genes important for the differentiation of lineages, we collected genes, which were found in all strains of a lineage (>60% amino acid identity, >80% coverage) and lacking in all strains of other lineages. These genes were correlated with data from a previous analysis examining intracellular gene expression, monitored by microarrays, inside macrophages in order to identify a possible role in intracellular survival [34] (Additional file 18).

Thus, 33 lineage-III-specific, 22 lineage-II-specific and 14 lineage-I-specific core genes could be identified, which are supported by previous microarray-based studies[41,73]. We were able to predict the largest number of specific core genes for lineage III, a result that is influenced by the lower number of compared strains in this lineage. Interestingly, the distinct lineage core repertoire of lineage II predominantly includes PTS systems and ABC transporters involved in carbohydrate metabolism, while those of lineages I and III mainly consist of hypothetical and surface-associated proteins. Furthermore, the majority of specific lineage-core genes of lineage II are largely organized in three operon-like islands, while those of lineages I and III are generally spread discretely along the chromosomes. The former organization is thought to largely follow from deletions of these modules in lineages I and III, based on the lack of deviation from average G/C content and codon usage frequently associated with horizontal gene transfer, as well as from the presence of sequence remnants of some of these genes in strains of lineages I and III (*lmo0734-9, lmo1060-3*, ~60bp with >75% nucleotide identity). Neighborhood and sequence of specific core genes of lineages I and III show more ambiguous patterns including putative insertions, especially considering surface-related proteins (data not shown). We thus hypothesize that ancestral strains of lineage I and III diverged from lineage II by loss

of genes related to carbohydrate metabolism and gain of surface-associated genes serving different needs of interaction with the environment.

In order to determine new candidate genes, which may be involved in resistance to food preservation measures, pathogenicity, or virulence we also collected genes found to be mutually conserved in predominantly human listeriosis-related lineages I and II while being absent from lineage III. Among these are 45 genes which are conserved in 13 out of 14 strains of lineage I and II while being absent from both lineage III strains (Additional File 19). These comprise genes coding for 16 hypothetical proteins, 14 metabolic enzymes, 6 surface-associated proteins and 4 transcriptional regulators. Affected metabolic pathways include non-mevalonate isoprenoid, fructose and arginine biosynthesis. The former is affected by the deletion of *gcpE* and *lytB* in lineage III and strains of apathogenic strains such as *L. innocua* Clip11262, *L. seeligeri* SLCC3954 and *L. welshimeri* SLCC5334, bearing implications for intraperitoneal infection [74]. A fructose IIABC PTS system and its associated regulator (*lmo0630-34*) were also found to be absent from lineage III as well as strain 4b L312. Considering the clonality of the junction, this deletion likely happened in a common ancestor. Furthermore, the complete arginine metabolic pathway (*lmo0036-41*) is absent from lineage III [75]. A nitroreductase and a hydrolase that were each associated with a LysR family regulator and only present in lineages I and II may assist survival in the gastrointestinal tract. Other genes related to stress resistance exclusively conserved in these lineages include the intracellularly up-regulated A118-like prophage rest also known as monocin or *lma*-operon. Furthermore, lineage III does not contain genes coding for multiple internalins and amidases associated with invasion (*inlF, inlC, lmo0129, lmo0849*). Remnant genes of amidases *lmo0129* and *lmo0849* in strains of lineage III display identical C-terminal sequence deletions, suggesting loss in a common ancestor.

Summarizing, strains of less virulent and pathogenic lineage III mainly differ from the other two lineages by loss of genes involved in metabolism, stress resistance and surface-associated functions [34]. These genes are implied in adaptation to the complex inter- and intracellular environment inside the host with regard to nutrients, survival in the gastrointestinal tract, attachment and invasion, as well as resistance towards food preservation measures.

We also tried to identify exclusive indels for serogroups or –types that are represented by at least two strains in this analysis in order to uncover ancestral sequences (Additional file 20). We found nine genes to be specific for all strains of serogroup 4, while 16 genes are specifically absent. Most of these were already described to be responsible for differences in teichoic acid composition and are located in the same locus (*lmo1068-91*) [35,76]. Serogroups 3 and 1/2 do not show exclusive gene indels, indicating that the respective variable antigens either result from minor changes inside coding genes, from differences located in intergenic regions (ex. promoters, imperfect automatic prediction of ORFs, operon structures, etc.) or from heterogeneous causes. The same is true for strains of serotype 1/2a and 4b.

No genes were exclusively present or absent in strains of serotypes primarily indicated in human listeriosis (4b, 1/2a, 1/2b, 1/2c), suggesting that the virulent phenotype of these strains is not based on uniform gene-scale indels, but resulted from a range of differing virulence factors and minor mutations.

Strain-specific genes rarely represent an obvious extended functionality

Many genes which only exist in one strain represent young insertions or truncations and thus the most recent adaptations. It should be noted that strains 1/2a 08-5923 and 08-5578 are highly similar isolates of one outbreak leading to a very low number of specific genes [37]. In order to alleviate this artifact, we chose to refer to genes exclusively present in either one or both strains as 08-specific. Nonetheless, strain-specific genes identified will involve considerable bias due to the relation of strains selected for this study.

Between 11 (3c SLCC2479) and 177 (4a L99) genes per strain were classified as specific using a homology cutoff of 60% amino acid identity and 80% coverage (Additional file 21). These include 0 (4b L312) to 93 (4a L99) genes inserted by a set of previously determined mobile genetic elements dominated by specific prophages. Up to 37 strain-specific genes were found to be fragments of genes either split or truncated by the insertion of a premature stop-codon ("pseudogenes"). Most of these are transporters, metabolic enzymes or regulators and in many cases associated with virulence or pathogenicity as described previously. Strains 1/2a SLCC5850 and 7 SLCC2482 displayed an overrepresentation of fragmentary CDS, which may mark the recent onset of a reductive adaptation (31 and 35 genes, respectively, versus a species mean of 14). After exclusion of prophage and fragmentary genes, we found between 2 (3c SLCC2479) and 82 (4a L99) genes, which are more likely to constitute an extended functionality. The majority of these have been inserted in modules of 2-53 genes, frequently as a part of transposons and genetic islands and described mostly as hypothetical proteins.

Strain 4b L312 shows a specific insertion of an additional lactose/cellobiose PTS (*LMOL312_2315-20*) putatively facilitated by homologous recombination. Hydrolase *LMOL312_2315* inserted next to hydrolase *LMOL312_2314* (39% nucleotide identity). Since 4b L312 was isolated from cheese, this may represent an adaptation to dairy products. Another exclusive module was found in 4d ATCC19117 (*LMOATCC19117_2375-83*). It contains an additional set of genes putatively involved in the biosynthesis of aromatic amino acids. A specific element found in strain 3b SLCC2540 might constitute a bacteriocin transport and resistance system (*LMOSLCC2540_2733-40*) resembling lantibiotic sublancin 168 (up to 28% amino acid identity at 100% coverage) [77]. We found no homologue to the *sun*A bacteriocin peptide, indicating that either this system exports a different bacteriocin or it serves exclusively to resist these molecules.

Interestingly, eight non-homologous restriction-modification systems were also found to be strain-specific. Half of these are located in cassettes that do not show any mobility genes, indicating a potential supplementary function as mobile genetic elements. No RM system was commonly present in more than three of the compared strains, confirming observations of their "selfish" and competitive nature [78].

Small non-coding RNA candidates of *L. m.* EGD-e are largely conserved within the species

A recent transcriptomic analysis uncovered 150 regulatory sRNA candidates expressed in *L. monocytogenes* 1/2a EGD-e [25]. We identified homologues of these 150 sRNA candidates using a minimum BLASTN cutoff of 60% nucleotide identity and 80% coverage (Additional File 22).

Only 25 of these were found to be differentially distributed including a clear overrepresentation of sRNAs specific for a subset of strains of lineage II (12/25). Approximately half of those differentially distributed sRNAs, that were previously described to be involved in virulence or pathogenicity by either deletion mutants (*rli33-1, rli38, rli50*) or differential regulation in blood or intestinal lumen (*rli24, rli28, rli29, rli38, rli48, rliC*) [25,57] were also exclusively present in a varying subset of strains of lineage II. It should be noted that this subset never included strain 3a SLCC7179, implying that ancestral strains of 3c SLCC2479, 1/2c SLCC2372 and serotype 1/2a contributed a specific range of sRNAs in order to adapt to the environment and to modulate the infectious process. In total, 10 sRNAs were present in at least half of the strains of lineages I/II and absent from at least one strain of lineage III (*rli24, rli75, rli33-1, rli33-2, rli38, rli58, rli98, rli48, rli112, anti1974),* which may be involved in functions related to intracellular survival.
We found only one sRNA to be specific for strain 1/2a EGD-e (*rli62*) and inserted as part of prophage A118 (MGE-13). Transcriptional activation of prophage genes was reported previously, but an impact on phenotype due to prophage-related sRNAs has still to be elucidated in species *L. monocytogenes* [32, 34].
Interestingly, strain 3a SLCC7179 shows a fragmented homologue of *ssrA* (*tmRNA*, 391/500 bp = 78% coverage) necessary for the *trans*-translation of mRNAs that lack a natural stop-codon. In some *Escherichia coli* strains, an alternative sRNA termed *afrA* (*yhdL*) can serve as a possible replacement but was found to be absent from all compared strains [79]. Thus, we speculate that either the shortened *ssrA* gene is still functional or that species *L. monocytogenes* or specifically strain 3a SLCC7179 harbor another yet unknown system to recycle stalled ribosomes and incomplete polypeptides.
In summary, evolution of small non-coding RNAs represents an ongoing process in species *Listeria monocytogenes.* This excludes all riboswitches found to be mutually conserved in all compared chromosomes, strengthening the hypothesis that *cis* acting RNA regulation is an ancient mechanism. Small non-coding RNA transcriptomic analysis of strains of lineages I and III will now be required to uncover their specific regulatory networks on this level.

LisDB – a comparative genomics server for the *Listeria* research community

A large part of the analysis presented in this study is based on the GECO comparative genomics software [80]. We have created a public web-server that includes all published chromosomes and plasmids of genus *Listeria,* as well as a subset of genomes of related genera. The main function of this tool is the identification of homologous genes between replicons to uncover pan-genomic distributions, which can be visualized graphically or exported in the form of tab-delimited lists. Among these are matrices sorted for conservation in selected replicons or for synteny according to a reference strain. Gene gain and loss between two replicons can be identified and nucleotide or amino acid sequences can be exported. GECO-LisDB is accessible at the following address: http://bioinfo.mikrobio.med.uni-giessen.de/geco2lisdb.

Conclusion

Listeria monocytogenes represents a well-characterized pathogen and model system for infection research. Extension of fully sequenced genomes by 11 strains to include

all serotypes of the species allowed evolutionary analyses of unprecedented depth. Comparative examination in conjunction with public data revealed that (i) the species pan-genome is highly stable but not closed, (ii) accessory genes are mainly located in defined chromosomal regions (nine hyper variable hotspots, nine different prophages, three transposons, and two mobilizable islands) constituting primary loci of gene-scale species evolution, (iii) potentially functional CRISPR/Cas systems of different subtypes are infrequent but may shape genome diversity, (iv) evolutionary distances observed between lineages of *L. monocytogenes* and apathogenic species are mostly the result of SNPs rather than gene-scale indels that are rarely commonly inherited, highlighting the potential impact of small-scale mutation on long-term development, (v) frequent loss or truncation of genes described to be vital for virulence or pathogenicity was confirmed as a recurring pattern, especially for lineages II and III. The presence or absence of genes among all serotypes of species *L. monocytogenes* uncovered by this study will be helpful for further diagnostic, phylogenetic and functional research, and is assisted by the comparative genomic GECO-LisDB analysis server (http://bioinfo.mikrobio.med.uni-giessen.de/geco2lisdb).

Materials and Methods

Sequencing

The DNA was purified using Epicentre's MasterPure gram-positive DNA purification kit as recommended by the manufacturer and sequenced on a 454 GS-FLX System. Between 213437 and 297585 reads per strain were *de novo* assembled with the GS Assembler (Newbler) and mapped versus published strains of *L. monocytogenes* for scaffolding purposes. Contigs resembling plasmid DNA were identified in *L. monocytogenes* 7 SLCC2482, 1/2c SLCC2372 and 1/2b SLCC2755. PCR-based techniques followed to close the remaining gaps partially assisted by Minimap (unpublished software, available upon request) to identify specific primer pairs. PCRs were sequenced with Sanger ABI Big Dye technology (Applied Biosystems) by Roche (Germany), Goettingen Genomics Laboratory (Goettingen, Germany) and Agowa (Berlin, Germany). Sanger reads were incorporated into the assembly using the GAP4 software package v4.11 and SeqMan (Lasergene). The completed chromosomes achieved coverages between 16-26x and 99.67-99.93% of the bases carried Q40 or higher quality scores. The final gap in the chromosome of *L. monocytogenes* 7 SLCC2482 was marked with a sequence of 100 Ns. All replicons were deposited in the EMBL database (see Table 1 for accession numbers).

Annotation

The data was annotated using GenDB, which includes steps for the identification of protein coding sequences (CDS), rRNA and tRNA genes as well as similarity searches against major gene and protein databases [81]. The annotation was enriched using a separate bidirectional best BLASTP step to incorporate data from *L. monocytogenes* 4a L99 (EMBL-Bank: FM211688) and the surface protein prediction software Augur [82]. All preliminary information was assembled in GECO [80] for comparative analysis and curated manually.

Comparative analyses

Homologous coding sequences were identified by BLASTCLUST [83] as implemented in the comparative genomics software GECO [80]. The standard

similarity criterion was set to a minimum of 60% amino acid identity and 80% coverage of both proteins. Chromosomal regions were checked manually using the comparative genome browser of GECO in order to find orthologous CDS which satisfied the homology criteria and were located in a syntenic region in comparison to a reference strain. In some cases a stricter analysis based on 80% amino acid identity and 90% coverage was additionally employed to reduce the number of false positives. In order to avoid excessive redundancy, we denote only one gene of a homologous cluster in brackets, which can be further assessed using either the GECO LisDB server (http://bioinfo.mikrobio.med.uni-giessen.de/geco2lisdb) or the supplementary homology matrix (Additional File 1).

Pan-genome analysis

The pan-genome size of *L. monocytogenes* was predicted based on the chromosomes of 16 sequenced strains compared in this study. We employed the standard BLASTCLUST homology cutoff of 60% amino acid identity and 80% coverage for this analysis. Chromosomes were added 10000 times in a randomized order without replacement, and the number of core (mutually conserved), and accessory (found in at least one but not all strains) genes was noted using GECO. Since mean and median values for each step showed little variation, mean numbers of gene classes were plotted. In order to predict a possible future pan-genomic distribution for this species we performed a power law fitting as described previously [39].

Identification of large insertions

The colinearity of chromosomes of *L. monocytogenes* allowed a relatively simple method to identify large insertions. First we masked the sequence inversion surrounding the oriC in strain 08-5923 (LM5923_2737-0270) and 08-5578 (LM5578_2788-0270) by reordering coding sequences to follow the usual chromosomal layout as found in strain 1/2a EGD-e. CDS were then compared in a bidirectional best BlastP analysis using similarity criteria of more than 60% amino acid identity and 80% coverage of both CDS. Core-CDS existing in all compared strains were identified by single linkage clustering (AB + BC = ABC). All core-CDS showing a break in the synteny (translocation, inversion) relative to reference strain 1/2a EGD-e were removed from the pool. Finally, the number of CDS located between syntenic core-CDS was plotted as a bar chart per strain. Exact borders of mobile genetic elements were identified based on annotation, deviation of GC-content and comparative analysis with sequenced phages and strains of genus *Listeria*.

CRISPR/Cas analysis

Spacer/repeat-arrays were identified with PILER-CR 1.02 and CRT 1.1 using standard parameters with the exception of maximum repeat length, which was increased to 40 [84,85]. Resulting arrays were combined and controlled manually leading to the removal of eleven false positives inside LRR- and LPXTG-domain containing coding sequences. Consensus sequences of repeats of remaining loci were employed for a BLASTN search versus chromosomes of all strains resulting in the identification of multiple decaying spacer/repeat modules that had been ignored by Piler and CRT due to repeat sequence mismatches of up to 20%. Spacers were compared to 10 published bacteriophages of genus *Listeria* (A006: NC_009815, A118: NC_003216, A500: NC_003216, A511: NC_009811, B025: NC_009812, B054: NC_009813, P100: NC_009813, P35: NC_009814, P40: EU855793, PSA: NC_003291), 16 chromosomes and 4 plasmids of strains of this study and the NCBI

nt-database using BLASTN. Alignments showing up to 1 mismatch were deemed homologous. Finally, all spacers where compared to each other using BLASTCLUST considering perfect matches only and mapped to mirror the order of spacers inside the respective loci to visualize the degree of relatedness (Additional File 3, software BlastclustToMatrix available upon request). Softening of the homology cutoffs to 80% nucleotide identity at 80% coverage did not result in a meaningful increase of matches. Cas genes were identified by sequence homology to published data found in the NCBI NT database and Pfam [86].

Phylogeny

A phylogenetic core-genome tree was created based on mutually conserved core CDS of all compared strains including out-group strains *L. innocua* 6a Clip11262, *L. welshimeri* 6b SLCC5334 and *L. seeligeri* 1/2b SLCC3954. These were extracted from a GECO homology matrix (amino acid identity >60%, coverage > 80%) (Additional File 1) following removal of all clusters showing paralogues. A total of 2018 protein coding genes were concatenated resulting in approximately 2 Mb of nucleotide sequence information per strain. The data was aligned using Mugsy [58] and resulting locally collinear blocks were joined per strain and imported into MEGA5 and SplitsTree4 [87,88]. Based on the alignment we created multiple phylogenomic trees (maximum parsimony, minimum evolution, neighbor joining) including 100 bootstrap replicates. Since tree topology was identical in all cases and relative branch lengths showed little variation, we only present trees based on the neighbor joining algorithm.

In order to identify the impact of indels on phylogeny we built a second tree based on the presence and absence of 2953 accessory genes using GeneContent [59]. Distance between strains was calculated with the Jaccard coefficient [89] and a tree was inferred using the neighbor joining reconstruction method including 100 bootstrap replicates.

Identification of surface-associated genes and putative internalins

Surface-associated genes were identified based on sequence similarity to known motifs (P60, LysM, GW, LRR, LPXTG, lipo) using various Hidden Markov Models (HMM) and SignalP as implemented by Augur [82]. A domain was considered present if HMM e-value < 10 and HMM score > 5. All surface-associated homology matrices were created using a higher standard cutoff (80% amino acid identity, 90% coverage) in order to achieve a higher degree of resolution and thus identify even small amounts of sequence dissimilarity. Clusters showing paralogous CDS were manually split according to a GECO synteny analysis.

All CDS containing a leucine rich repeat (LRR) domain were assumed to be putative internalins and checked for the presence of a signal peptide. False positives and negatives as revealed by synteny analysis were corrected manually and the homology cutoff was reduced to 50% identity and 40% coverage if necessary. Apprehension of internalin-types based on predicted internalins from a previous study [90] as well as domains identified by Augur completed the analysis.

Measurement of bacterial growth

Bacterial cultures were grown over night at 37°C in brain heart infusion broth (BHI) and diluted 1:200 the next day for fresh cultures. Automated measuring at 37°C was

performed using the Infinite 200 plate reader (Tecan) in 96-well plates with 150 µl volume/well.

Galleria mellonella infection model

Bacterial inoculums were injected dorsolaterally into the hemocoel of last instar larvae using 1 ml disposable syringes and 0.4 x 20mm needles mounted on a microapplicator as described previously [91]. After injection, larvae were incubated at 37°C. Larvae were considered dead when they showed no movement in response to touch. No mortality of *Galleria* larvae were recorded when injected with 0.9% NaCl. The majority of experiments were conducted independently with 20 animals and three biological replicates, while 10 animals per experiment and two biological replicates were used for strains 1/2a 08-5923, 1/2a 08-5578, 1/2a SLCC5850 and 4b F2365.

Authors' contributions

CK carried out bioinformatic tasks related to the sequencing process, performed annotation and comparative analyses and drafted the manuscript. AB participated in the annotation and bioinformatic analyses and assisted in drafting the manuscript. MM performed phenotypic experiments. AS performed the genomic sequencing. AG participated in the design of the study. RD participated in the design of the study. SB assisted in drafting the manuscript. TH conceived of the study, participated in its design and coordination and assisted in drafting the manuscript. TC participated in the design of the study and assisted in drafting the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This project was funded by the German Federal Ministry of Education and Research through ERA-NET program grants sncRNAomics and SPATELIS to T. H and T. C. We thank Sonja Voget, who performed the majority of the sequencing experiments that this project is based on and Anita Hoeland, who constructed data visualizations. We would also like to acknowledge Mark Achtman and Jana Haase for retrieving information about isolates stored in the *Listeria* Culture Collection (SLCC). Jordan Pischimarov, Julian Krauskopf, and Sebastian Oehm assisted in genome annotation and Keith Jolley created a BIGSdb database for MLST classification. Last but not least, we would like to thank Alexandra Amend, Claudia Zörb, Nelli Schklarenko and Burkhard Linke for excellent technical assistance.

Figures

Figure 1 - Pan-genomic distribution

Distribution of CDS based on a homology measure of 60% amino acid identity and 80% coverage. Chromosomes were added 10000 times without replacement in a randomized order and the number of core (mutually conserved) and accessory (found in at least one but not all strains) genes was noted. Since mean and median values for each step showed only little variation the mean numbers of gene classes were plotted. In order to predict a possible future pan-genomic distribution for this species we performed a power law fitting. A) Pan-genomic CDS after each consecutive addition of a strain, B) mutually conserved CDS, C) conservation of CDS and homology clusters.

Figure 2 - Insertions between syntenic core genes

Bar chart of CDS inserted between syntenic core-CDS existing in all strains depicted relative to reference strain 1/2a EGD-e. The oriC inversion of strains 08-5923 and 08-5578 was removed for this analysis. Mobile genetic elements (MGE) are classified as prophage (red triangle), transposon/IS element (blue square), genetic island (green circle). The MGEs were numbered according to their relative position in strain 1/2a EGD-e. Putative anchor genes in the chromosome (ex.: tRNA, comK) are included in square brackets. If different elements inserted at the same chromosomal locus, the strains involved are denoted in round brackets. Multiple designations per element are delimited by a colon. If an element was not described yet, the genus bearing the highest overall nucleotide similarity to the respective region was included instead (e.g. *Bacillus subtilis, Enterococcus faecalis*). Lineages are denoted with roman numbers.

Figure 3 - Phylogenomic and -genetic trees

(A) Neighbor joining tree based on an alignment of 2018 mutually conserved core genes (amino acid identity >60%, coverage >80%) of 19 strains of genus *Listeria*. Bootstrap support of 100 replicates was always found to be above 80% and thus omitted. Lineages of *L. monocytogenes* are marked in roman letters and phylogenomic groups (PG) describe closely related strains. (B) Data of panel A reimagined as cladogram to highlight branching. (C) Neighbor joining tree of gene content (presence/absence). Only bootstrap support values below 80% (100 replicates) are indicated.

Figure 4 – *Galleria mellonella* mortality rates

Mortality rates of *Galleria mellonella* larvae over the course of seven days post injection. Respective standard deviations can be found in the supplementary material (Additional File 11).

Tables

Table 1 - Origin of compared strains of species *L. monocytogenes*.

Table 2 - General features of the chromosomes of compared strains.

Supporting information

Additional File 1 - Species homology matrices (.xls)

General homology matrices showing the distribution of all coding sequences among 16 strains of species *L. monocytogenes* and 19 strains of genus *Listeria* at different cutoffs. This table is sorted for maximum conservation (core genes = top, specific genes = bottom).

Additional File 2 - Insertional hotspot ranges (.xls)

Hotspots showing at least three separate insertions denoted by locustag ranges.

Additional File 3 - Comparative genomic GECO figures of hyper variable hotspots (.pdf)

Comparative GECO depictions of insertional hotspots highlighting extensive mosaicism.

Additional File 4 - Mobile genetic elements (.xls)

Distribution of mobile genetic elements ordered by relative position in the chromosome of *L. monocytogenes* 1/2a EGD-e.

Additional File 5 - Comparative genomic GECO figures of transposons ICELm1 and TN554 (.pdf)

Comparative GECO depiction using a homology measure of 60% amino acid identity and 80% coverage. Displays content and conservation of two transposons.

Additional File 6 - Comparative genomic GECO figures of IS3 elements (.pdf)

Comparative GECO depiction using a homology measure of 60% amino acid identity and 80% coverage. Displays duplication of IS3-like transposon.

Additional File 7 - Comparative genomic GECO figure of CRISPR/Cas loci (.pdf)

Comparative GECO depictions of three CRISPR/Cas loci using a minimum CDS homology measure of 60% amino acid identity and 80% coverage.Cas genes and spacer/repeat arrays are framed. Locus 1 displayed no associated Cas genes. Locus 3 includes a *trans*-acting sRNA called tracrRNA that was described to compensate for a missing endoribonuclease in conjunction with host factor RNase III.

Additional File 8 - CRISPR/Cas loci (.xls)

Homology matrices and positions of CRISPR/Cas genes and associated arrays of three loci. Spacers were additionally mapped versus the NCBI nt database to identify possible target sequences.

Additional File 9 - Known virulence genes (.xls)

Homology matrix of known virulence genes.

Additional File 10 - Growth curves BHI (.pdf)

Growth of *L. monocytogenes* in BHI medium at 37°C.

Additional File 11 – *Galleria* standard deviations (.xls)

Standard deviations calculated for independent experiments considering mortality rates of *Galleria mellonella* larvae over the course of seven days post infection.

Additional File 12 - Detailed analyses of reductive evolution of virulence-associated genes (.pdf)

In-depth information about previously described virulence and pathogenicity indicated genes that are absent or truncated in one of the compared strains.

Additional File 13 - Plot of Surface-associated CDS (.pdf)

Bar plot depicting the distribution of all surface-associated protein coding genes among studied strains.

968 Additional File 14 - Distribution of surface-associated genes displays
969 conserved lineage-backbones with strain-specific adaptations (.pdf)
970 Detailed analysis of presence and absence of surface-associated genes.

971 Additional File 15 - Surface-associated CDS (.xls)
972 Homology matrices of genes containing a surface-associated domain (NLPC/p60,
973 LysM, GW, LRR, LPxTG, Lipobox, signal peptide).

974 Additional File 16 - Putative transposition of lipoproteins *lmo1264-5* in
975 lineage III (.pdf)
976 Comparative GECO depiction using a homology measure of 80% amino acid identity
977 and 90% coverage. Displays the putative transposition of lipoproteins *lmo1264-5* in
978 lineage III into the locus that putatively held *inlC* previously.

979 Additional File 17 – Internalins (.xls)
980 Homology matrix of genes containing a leucine rich repeat domain and an optional
981 signal peptide.

982 Additional File 18 - Lineage-specific CDS (.xls)
983 Homology matrix of coding genes specifically present in one lineage.

984 Additional File 19 - Lineage I/II exclusive CDS (.xls)
985 Homology matrix of genes conserved in 13/14 strains of lineages I and II, while being
986 absent from both strains of lineage III.

987 Additional File 20 - Serogroup and –type ancestral indels (.xls)
988 Homology matrix of CDS found to be commonly present or absent (ancestral indel)
989 for either one or multiple serogroups or -types.

990 Additional File 21 - Strain-specific CDS (.xls)
991 Homology matrix of coding genes specifically present in one strain.

992 Additional File 22 - Small non-coding regulatory RNAs (.xls)
993 Homology matrix of sRNA candidates.
994

References

1. Schmid MW, Ng EY, Lampidis R, Emmerth M, Walcher M, Kreft J, Goebel W, Wagner M, Schleifer KH: **Evolutionary history of the genus *Listeria* and its virulence genes.** *Syst Appl Microbiol* 2005, **28:**1-18.

2. Graves LM, Helsel LO, Steigerwalt AG, Morey RE, Daneshvar MI, Roof SE, Orsi RH, Fortes ED, Milillo SR, den Bakker HC et al.: ***Listeria marthii* sp. nov., isolated from the natural environment, Finger Lakes National Forest.** *Int J Syst Evol Microbiol* 2010, **60:**1280-1288.

3. Leclercq A, Clermont D, Bizet C, Grimont PA, Le Fleche-Mateos A, Roche SM, Buchrieser C, Cadet-Daniel V, Le MA, Lecuit M et al.: ***Listeria rocourtiae* sp. nov.** *Int J Syst Evol Microbiol* 2010, **60:**2210-2214.

4. Vazquez-Boland JA, Kuhn M, Berche P, Chakraborty T, Dominguez-Bernal G, Goebel W, Gonzalez-Zorn B, Wehland J, Kreft J: ***Listeria* pathogenesis and molecular virulence determinants.** *Clin Microbiol Rev* 2001, **14:**584-640.

5. Allerberger F, Wagner M: **Listeriosis: a resurgent foodborne infection.** *Clin Microbiol Infect* 2010, **16:**16-23.

6. Mead PS, Slutsker L, Dietz V, McCaig LF, Bresee JS, Shapiro C, Griffin PM, Tauxe RV: **Food-related illness and death in the United States.** *Emerg Infect Dis* 1999, **5:**607-625.

7. Portnoy DA, Chakraborty T, Goebel W, Cossart P: **Molecular determinants of *Listeria monocytogenes* pathogenesis.** *Infect Immun* 1992, **60:**1263-1267.

8. Hamon M, Bierne H, Cossart P: ***Listeria monocytogenes*: a multifaceted model.** *Nat Rev Microbiol* 2006, **4:**423-434.

9. Dominguez-Bernal G, Muller-Altrock S, Gonzalez-Zorn B, Scortti M, Herrmann P, Monzo HJ, Lacharme L, Kreft J, Vazquez-Boland JA: **A spontaneous genomic deletion in *Listeria ivanovii* identifies LIPI-2, a species-specific pathogenicity island encoding sphingomyelinase and numerous internalins.** *Mol Microbiol* 2006, **59:**415-432.

10. Cotter PD, Draper LA, Lawton EM, Daly KM, Groeger DS, Casey PG, Ross RP, Hill C: **Listeriolysin S, a novel peptide haemolysin associated with a subset of lineage I *Listeria monocytogenes*.** *PLoS Pathog* 2008, **4:**e1000144.

11. Camejo A, Buchrieser C, Couve E, Carvalho F, Reis O, Ferreira P, Sousa S, Cossart P, Cabanes D: **In vivo transcriptional profiling of *Listeria monocytogenes* and mutagenesis identify new virulence factors involved in infection.** *PLoS Pathog* 2009, **5:**e1000449.

12. Mohamed W, Sethi S, Darji A, Mraheil MA, Hain T, Chakraborty T: **Antibody targeting the ferritin-like protein controls *Listeria* infection.** *Infect Immun* 2010, **78:**3306-3314.

13. Cabanes D, Dehoux P, Dussurget O, Frangeul L, Cossart P: **Surface proteins and the pathogenic potential of *Listeria monocytogenes*.** *Trends Microbiol* 2002, **10:**238-245.

14. Bierne H, Cossart P: ***Listeria monocytogenes* surface proteins: from genome predictions to function.** *Microbiol Mol Biol Rev* 2007, **71:**377-397.

15. Milillo SR, Wiedmann M: **Contributions of six lineage-specific internalin-like genes to invasion efficiency of *Listeria monocytogenes*.** *Foodborne Pathog Dis* 2009, **6:**57-70.

16. Kasamatsu J, Suzuki T, Ishijima J, Matsuda Y, Kasahara M: **Two variable lymphocyte receptor genes of the inshore hagfish are located far apart on the same chromosome.** *Immunogenetics* 2007, **59:**329-331.

17. Boehm T: **Design principles of adaptive immune systems.** *Nat Rev Immunol* 2011, **11:**307-317.

18. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF et al.: **Evolution and classification of the CRISPR-Cas systems.** *Nat Rev Microbiol* 2011, **9:**467-477.

19. Aklujkar M, Lovley DR: **Interference with histidyl-tRNA synthetase by a CRISPR spacer sequence as a factor in the evolution of *Pelobacter carbinolicus*.** *BMC Evol Biol* 2010, **10:**230.

20. Stern A, Keren L, Wurtzel O, Amitai G, Sorek R: **Self-targeting by CRISPR: gene regulation or autoimmunity?** *Trends Genet* 2010, **26:**335-340.

21. Haft DH, Selengut J, Mongodin EF, Nelson KE: **A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes.** *PLoS Comput Biol* 2005, **1:**e60.

22. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E: **CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.** *Nature* 2011, **471:**602-607.

23. Babu M, Beloglazova N, Flick R, Graham C, Skarina T, Nocek B, Gagarinova A, Pogoutse O, Brown G, Binkowski A et al.: **A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair.** *Mol Microbiol* 2011, **79:**484-502.

24. Storz G, Opdyke JA, Zhang A: **Controlling mRNA stability and translation with small, noncoding RNAs.** *Curr Opin Microbiol* 2004, **7:**140-144.

25. Mraheil MA, Billion A, Mohamed W, Mukherjee K, Kuenne C, Pischimarov J, Krawitz C, Retey J, Hartsch T, Chakraborty T et al.: **The intracellular sRNA transcriptome of *Listeria monocytogenes* during growth in macrophages.** *Nucleic Acids Res* 2011, **39:**4235-4248.

26. Ragon M, Wirth T, Hollandt F, Lavenir R, Lecuit M, Le MA, Brisse S: **A new perspective on *Listeria monocytogenes* evolution.** *PLoS Pathog* 2008, **4:**e1000146.

27. Orsi RH, den Bakker HC, Wiedmann M: ***Listeria monocytogenes* lineages: Genomics, evolution, ecology, and phenotypic characteristics.** *Int J Med Microbiol* 2011, **301:**79-96.

28. den Bakker HC, Didelot X, Fortes ED, Nightingale KK, Wiedmann M: **Lineage specific recombination rates and microevolution in *Listeria monocytogenes*.** *BMC Evol Biol* 2008, **8:**277.

29. Orsi RH, Sun Q, Wiedmann M: **Genome-wide analyses reveal lineage specific contributions of positive selection and recombination to the evolution of *Listeria monocytogenes*.** *BMC Evol Biol* 2008, **8:**233.

30. Kuenne C, Voget S, Pischimarov J, Oehm S, Goesmann A, Daniel R, Hain T, Chakraborty T: **Comparative analysis of plasmids in the genus *Listeria*.** *PLoS One* 2010, **5**.

31. Roche SM, Gracieux P, Milohanic E, Albert I, Virlogeux-Payant I, Temoin S, Grepinet O, Kerouanton A, Jacquet C, Cossart P et al.: **Investigation of specific substitutions in virulence genes characterizing phenotypic groups of low-virulence field strains of *Listeria monocytogenes*.** *Appl Environ Microbiol* 2005, **71:**6039-6048.

32. Chatterjee SS, Otten S, Hain T, Lingnau A, Carl UD, Wehland J, Domann E, Chakraborty T: **Invasiveness is a variable and heterogeneous phenotype in *Listeria monocytogenes* serotype strains.** *Int J Med Microbiol* 2006, **296:**277-286.

33. Swaminathan B, Gerner-Smidt P: **The epidemiology of human listeriosis.** *Microbes Infect* 2007, **9:**1236-1243.

34. Hain T, Ghai R, Billion A, Kuenne CT, Steinweg C, Izar B, Mohamed W, Mraheil MA, Domann E, Schaffrath S et al.: **Comparative genomics and transcriptomics of lineages I, II, and III strains of *Listeria monocytogenes*.** *BMC Genomics* 2012, **13:**144.

35. Nelson KE, Fouts DE, Mongodin EF, Ravel J, Deboy RT, Kolonay JF, Rasko DA, Angiuoli SV, Gill SR, Paulsen IT et al.: **Whole genome comparisons of serotype 4b and 1/2a strains of the food-borne pathogen *Listeria monocytogenes* reveal new insights into the core genome components of this species.** *Nucleic Acids Res* 2004, **32:**2386-2395.

36. Glaser P, Frangeul L, Buchrieser C, Rusniok C, Amend A, Baquero F, Berche P, Bloecker H, Brandt P, Chakraborty T et al.: **Comparative genomics of *Listeria* species.** *Science* 2001, **294:**849-852.

37. Gilmour MW, Graham M, Van DG, Tyler S, Kent H, Trout-Yakel KM, Larios O, Allen V, Lee B, Nadon C: **High-throughput genome sequencing of two**

***Listeria monocytogenes* clinical isolates during a large foodborne outbreak.** *BMC Genomics* 2010, **11:**120.

38. Jolley KA, Maiden MC: **BIGSdb: Scalable analysis of bacterial genome variation at the population level.** *BMC Bioinformatics* 2010, **11:**595.

39. Tettelin H, Riley D, Cattuto C, Medini D: **Comparative genomics: the bacterial pan-genome.** *Curr Opin Microbiol* 2008, **11:**472-477.

40. Phillippy AM, Deng X, Zhang W, Salzberg SL: **Efficient oligonucleotide probe selection for pan-genomic tiling arrays.** *BMC Bioinformatics* 2009, **10:**293.

41. Deng X, Phillippy AM, Li Z, Salzberg SL, Zhang W: **Probing the pan-genome of *Listeria monocytogenes*: new insights into intraspecific niche expansion and genomic diversification.** *BMC Genomics* 2010, **11:**500.

42. den Bakker HC, Cummings CA, Ferreira V, Vatta P, Orsi RH, Degoricija L, Barker M, Petrauskene O, Furtado MR, Wiedmann M: **Comparative genomics of the bacterial genus *Listeria*: Genome evolution is characterized by limited gene acquisition and limited gene loss.** *BMC Genomics* 2010, **11:**688.

43. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS et al.: **Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome".** *Proc Natl Acad Sci U S A* 2005, **102:**13950-13955.

44. Hiller NL, Janto B, Hogg JS, Boissy R, Yu S, Powell E, Keefe R, Ehrlich NE, Shen K, Hayes J et al.: **Comparative genomic analyses of seventeen *Streptococcus pneumoniae* strains: insights into the pneumococcal supragenome.** *J Bacteriol* 2007, **189:**8186-8195.

45. Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C, Bouvet O et al.: **Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths.** *PLoS Genet* 2009, **5:**e1000344.

46. Lefebure T, Stanhope MJ: **Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition.** *Genome Biol* 2007, **8:**R71.

47. Hogg JS, Hu FZ, Janto B, Boissy R, Hayes J, Keefe R, Post JC, Ehrlich GD: **Characterization and modeling of the *Haemophilus influenzae* core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains.** *Genome Biol* 2007, **8:**R103.

48. Hain T, Steinweg C, Kuenne CT, Billion A, Ghai R, Chatterjee SS, Domann E, Karst U, Goesmann A, Bekel T et al.: **Whole-genome sequence of *Listeria welshimeri* reveals common steps in genome reduction with *Listeria***

***innocua* as compared to *Listeria monocytogenes*.** *J Bacteriol* 2006, **188:**7405-7415.

49. Burrus V, Pavlovic G, Decaris B, Guedon G: **The ICESt1 element of *Streptococcus thermophilus* belongs to a large family of integrative and conjugative elements that exchange modules and change their specificity of integration.** *Plasmid* 2002, **48:**77-97.

50. Pei L, Palma M, Nilsson M, Guss B, Flock JI: **Functional studies of a fibrinogen binding protein from *Staphylococcus epidermidis*.** *Infect Immun* 1999, **67:**4525-4530.

51. Blatny JM, Godager L, Lunde M, Nes IF: **Complete genome sequence of the *Lactococcus lactis* temperate phage phiLC3: comparative analysis of phiLC3 and its relatives in lactococci and streptococci.** *Virology* 2004, **318:**231-244.

52. Novick RP, Christie GE, Penades JR: **The phage-related chromosomal islands of Gram-positive bacteria.** *Nat Rev Microbiol* 2010, **8:**541-551.

53. Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, Wood TK: **Cryptic prophages help bacteria cope with adverse environments.** *Nat Commun* 2010, **1:**147.

54. Wang X, Kim Y, Wood TK: **Control and benefits of CP4-57 prophage excision in *Escherichia coli* biofilms.** *ISME J* 2009, **3:**1164-1179.

55. Promadej N, Fiedler F, Cossart P, Dramsi S, Kathariou S: **Cell wall teichoic acid glycosylation in *Listeria monocytogenes* serotype 4b requires gtcA, a novel, serogroup-specific gene.** *J Bacteriol* 1999, **181:**418-425.

56. Mandin P, Repoila F, Vergassola M, Geissmann T, Cossart P: **Identification of new noncoding RNAs in *Listeria monocytogenes* and prediction of mRNA targets.** *Nucleic Acids Res* 2007, **35:**962-974.

57. Toledo-Arana A, Dussurget O, Nikitas G, Sesto N, Guet-Revillet H, Balestrino D, Loh E, Gripenland J, Tiensuu T, Vaitkevicius K et al.: **The *Listeria* transcriptional landscape from saprophytism to virulence.** *Nature* 2009, **459:**950-956.

58. Angiuoli SV, Salzberg SL: **Mugsy: fast multiple alignment of closely related whole genomes.** *Bioinformatics* 2011, **27:**334-342.

59. Gu X, Zhang H: **Genome phylogenetic analysis based on extended gene contents.** *Mol Biol Evol* 2004, **21:**1401-1408.

60. Dutilh BE, van N, V, van der Heijden RT, Boekhout T, Snel B, Huynen MA: **Assessment of phylogenomic and orthology approaches for phylogenetic inference.** *Bioinformatics* 2007, **23:**815-824.

61. Chen J, Jiang L, Chen X, Luo X, Chen Y, Yu Y, Tian G, Liu D, Fang W: ***Listeria monocytogenes* serovar 4a is a possible evolutionary intermediate**

between *L. monocytogenes* serovars 1/2a and 4b and L. innocua. *J Microbiol Biotechnol* 2009, **19:**238-249.

62. Georgiades K, Merhej V, El KK, Raoult D, Pontarotti P: **Gene gain and loss events in *Rickettsia* and *Orientia* species.** *Biol Direct* 2011, **6:**6.

63. Snel B, Bork P, Huynen MA: **Genome phylogeny based on gene content.** *Nat Genet* 1999, **21:**108-110.

64. van SW, Top J, Riley DR, Boekhorst J, Vrijenhoek JE, Schapendonk CM, Hendrickx AP, Nijman IJ, Bonten MJ, Tettelin H et al.: **Pyrosequencing-based comparative genome analysis of the nosocomial pathogen *Enterococcus faecium* and identification of a large transferable pathogenicity island.** *BMC Genomics* 2010, **11:**239.

65. Begley M, Sleator RD, Gahan CG, Hill C: **Contribution of three bile-associated loci, bsh, pva, and btlB, to gastrointestinal persistence and bile tolerance of *Listeria monocytogenes*.** *Infect Immun* 2005, **73:**894-904.

66. Mraheil MA, Billion A, Mohamed W, Rawool D, Hain T, Chakraborty T: **Adaptation of *Listeria monocytogenes* to oxidative and nitrosative stress in IFN-gamma-activated macrophages.** *Int J Med Microbiol* 2011, **301:**547-555.

67. Chen Y, Ross WH, Whiting RC, Van SA, Nightingale KK, Wiedmann M, Scott VN: **Variation in *Listeria monocytogenes* dose responses in relation to subtypes encoding a full-length or truncated internalin A.** *Appl Environ Microbiol* 2011, **77:**1171-1180.

68. Kang CM, Brody MS, Akbar S, Yang X, Price CW: **Homologous pairs of regulatory proteins control activity of *Bacillus subtilis* transcription factor sigma(b) in response to environmental stress.** *J Bacteriol* 1996, **178:**3846-3853.

69. Palma M, Cheung AL: **sigma(B) activity in *Staphylococcus aureus* is controlled by RsbU and an additional factor(s) during bacterial growth.** *Infect Immun* 2001, **69:**7858-7865.

70. Chaturongakul S, Boor KJ: **RsbT and RsbV contribute to sigmaB-dependent survival under environmental, energy, and intracellular stress conditions in *Listeria monocytogenes*.** *Appl Environ Microbiol* 2004, **70:**5349-5356.

71. Ebbes M, Bleymuller WM, Cernescu M, Nolker R, Brutschy B, Niemann HH: **Fold and function of the InlB B-repeat.** *J Biol Chem* 2011, **286:**15496-15506.

72. Tsai YH, Orsi RH, Nightingale KK, Wiedmann M: ***Listeria monocytogenes* internalins are highly diverse and evolved by recombination and positive selection.** *Infect Genet Evol* 2006, **6:**378-389.

73. Doumith M, Cazalet C, Simoes N, Frangeul L, Jacquet C, Kunst F, Martin P, Cossart P, Glaser P, Buchrieser C: **New aspects regarding evolution and virulence of *Listeria monocytogenes* revealed by comparative genomics and DNA arrays.** *Infect Immun* 2004, **72:**1072-1083.

74. Begley M, Bron PA, Heuston S, Casey PG, Englert N, Wiesner J, Jomaa H, Gahan CG, Hill C: **Analysis of the isoprenoid biosynthesis pathways in *Listeria monocytogenes* reveals a role for the alternative 2-C-methyl-D-erythritol 4-phosphate pathway in murine infection.** *Infect Immun* 2008, **76:**5392-5401.

75. Ryan S, Begley M, Gahan CG, Hill C: **Molecular characterization of the arginine deiminase system in *Listeria monocytogenes*: regulation and role in acid tolerance.** *Environ Microbiol* 2009, **11:**432-445.

76. Zhang C, Zhang M, Ju J, Nietfeldt J, Wise J, Terry PM, Olson M, Kachman SD, Wiedmann M, Samadpour M et al.: **Genome diversification in phylogenetic lineages I and II of *Listeria monocytogenes*: identification of segments unique to lineage II populations.** *J Bacteriol* 2003, **185:**5573-5584.

77. Dubois JY, Kouwen TR, Schurich AK, Reis CR, Ensing HT, Trip EN, Zweers JC, van Dijl JM: **Immunity to the bacteriocin sublancin 168 Is determined by the SunI (YolF) protein of *Bacillus subtilis*.** *Antimicrob Agents Chemother* 2009, **53:**651-661.

78. Kobayashi I: **Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution.** *Nucleic Acids Res* 2001, **29:**3742-3756.

79. Himeno H: **Novel factor rescues ribosomes trapped on non-stop mRNAs.** *Mol Microbiol* 2010, **78:**789-791.

80. Kuenne CT, Ghai R, Chakraborty T, Hain T: **GECO--linear visualization for comparative genomics.** *Bioinformatics* 2007, **23:**125-126.

81. Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J, Kalinowski J, Linke B, Rupp O, Giegerich R et al.: **GenDB--an open source genome annotation system for prokaryote genomes.** *Nucleic Acids Res* 2003, **31:**2187-2195.

82. Billion A, Ghai R, Chakraborty T, Hain T: **Augur--a computational pipeline for whole genome microbial surface protein prediction and classification.** *Bioinformatics* 2006, **22:**2819-2820.

83. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410.

84. Edgar RC: **PILER-CR: fast and accurate identification of CRISPR repeats.** *BMC Bioinformatics* 2007, **8:**18.

1270 85. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, Hugenholtz
1271 P: **CRISPR recognition tool (CRT): a tool for automatic detection of**
1272 **clustered regularly interspaced palindromic repeats.** *BMC Bioinformatics*
1273 2007, **8:**209.

1274 86. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N,
1275 Forslund K, Ceric G, Clements J et al.: **The Pfam protein families database.**
1276 *Nucleic Acids Res* 2012, **40:**D290-D301.

1277 87. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5:**
1278 **molecular evolutionary genetics analysis using maximum likelihood,**
1279 **evolutionary distance, and maximum parsimony methods.** *Mol Biol Evol*
1280 2011, **28:**2731-2739.

1281 88. Huson DH, Bryant D: **Application of phylogenetic networks in**
1282 **evolutionary studies.** *Mol Biol Evol* 2006, **23:**254-267.

1283 89. Wolf YI, Rogozin IB, Grishin NV, Koonin EV: **Genome trees and the tree of**
1284 **life.** *Trends Genet* 2002, **18:**472-479.

1285 90. Bierne H, Sabet C, Personnic N, Cossart P: **Internalins: a complex family of**
1286 **leucine-rich repeat-containing proteins in *Listeria monocytogenes*.**
1287 *Microbes Infect* 2007, **9:**1156-1166.

1288 91. Mukherjee K, Altincicek B, Hain T, Domann E, Vilcinskas A, Chakraborty T:
1289 ***Galleria mellonella* as a model system for studying *Listeria* pathogenesis.**
1290 *Appl Environ Microbiol* 2010, **76:**310-317.
1291
1292