

# Justus-Liebig Universität Gießen

Fachbereich 08 Chemie und Biologie

Heinrich-Buff Ring 17-19  
35392 Gießen



Etablierung eines funktionalen (Meta-)Genomansatzes zur  
Identifizierung von Klonen mit antibakterieller Aktivität

Dissertation vorgelegt von  
Jochen Fracowiak

Gutachter:

1. PD Dr. Jens Glaeser
2. Prof. Peter Kämpfer

## Selbstständigkeitserklärung

Ich erkläre: Ich habe die vorgelegte Dissertation selbstständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Ich stimme einer evtl. Überprüfung meiner Dissertation durch eine Antiplagiat-Software zu. 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.

Bonn, 20.01.2019

Ort, Datum

---

Jochen Fracowiak

## Zusammenfassung

Die Entwicklung von multiresistenten Bakterien als Auslöser von Infektionskrankheiten, die nicht mehr durch Antibiotika behandelbar sind, nimmt stetig zu. Eine weitere Problematik besteht darin, neue Antibiotika zu identifizieren wie zum Beispiel aus Bakterien, weil diese oft unter Standard-Laborbedingungen nicht kultivierbar sind und somit den klassischen Methoden zur Identifizierung von Wirkstoffen nicht zur Verfügung stehen. Daher sind neue Ansätze gefragt um dieses Problem zu lösen. Diese Forschungsarbeit setzt mit einem kombinierten Ansatz aus bioinformatischer Identifizierung von Biosynthese-Genclustern mit Bezug zu Antibiotika, aus sequenzierten bakteriellen Genomen sowie der Erstellung von Klonen als kultur-unabhängige Methode an dieser Schlüsselstelle an. Zunächst wurden bioinformatisch, unerforschte Phyla in Bezug auf die Antibiotika-Forschung untersucht, um gezielter Stämme als Wirkstoffproduzenten zu identifizieren. Diese Analyse umfasste in der Summe 361 Stämme der Phyla *Acidobacteria*, *Bacteroidetes*, *Chloroflexi* sowie des Superphylums *Planctomycetes*, *Verrucomicrobia* und *Chlamydiaeae*. Mit einem statistischen Ansatz wurden Stämme untersucht und festgestellt, dass 80 % aller Acidobakterien eine überdurchschnittliche Gencluster-Dichte aufweisen, von denen ein Großteil relevante Gencluster für eine potentielle Antibiotika-Biosynthese aufzeigen. Dieses Ergebnis wird durch die Identifizierung zahlreicher Resistenzgene unterstützt. Darauf aufbauend wurde das Acidobakterium *Terracidiphilus* sp. S55 für die Erstellung einer Klonbibliothek ausgewählt. Als zweiter Stamm dieses Phylums wurde *Acidicapsa borealis* untersucht. Das Genom wurde sequenziert-, durch eine Hybridassemblierung vollständig geschlossen und annotiert. Desweiteren wurden auch in diesem Stamm überdurchschnittlich viele Gencluster identifiziert. Die Etablierung eines kultur-unabhängigen Ansatzes umfasste die Gewinnung von Bakterien aus dem Termitennest von *Coptotermis niger*, welche zusätzlich taxonomisch klassifiziert wurden. Aus diesem metagenomischen Ansatz wurden 80 Klone für eine weiterführende Untersuchungen erstellt. In der Summe dieser Forschungsarbeit sind 960 Klone aus zwei unterschiedlichen Ressourcen weiter untersucht worden. Dies umfasste eine PCR-basierte Detektion von Adenyldomänen innerhalb des Gencluster-Typs für nicht-ribosomale Peptidsynthasen und zwei direkte Untersuchungsverfahren auf antibakterielle Aktivität. Dadurch konnten neun Klone mit antibakterieller Aktivität identifiziert werden, welche aus der genomischen DNS von *Acidicapsa borealis* stammen. Extrakte aktiver Klone wurden im weiteren Verlauf analytisch durch UHPLC-MS/MS untersucht, um Produkte zu identifizieren. Im Extrakt des Klons JF 7 C3 wurden die Tripeptide Valylprolylleucin und Isoleucylprolylisoleucin als heterolog exprimierte Produkte identifiziert. Die Etablierung dieses kultur-unabhängigen Ansatzes aus priorisierten Stämmen der Acidobakterien wurde in dieser Forschungsarbeit erfolgreich realisiert und stellt eine Grundlage für weitere funktionale (Meta-)Genomansätze zur Wirkstoffsuche dar.

## Abstract

The rising number of multidrug-resistant bacteria causing untreatable infectious diseases is getting a serious issue. This problem is exacerbated due to the difficulty of discovering new antibiotics such as from bacteria. Thus, classical approaches relying on cultivation in natural product discovery are limited. This PhD-Thesis picks up this bottleneck due to a combined approach consisting of a bioinformatics-driven discovery for biosynthetic gene clusters regarding antibiotics and a protocol for preparing clone libraries as an culture-independent approach. In total, 361 underexplored strains concerning antibiotics research were taken into account. Those belong either to the bacterial phylum *Acidobacteria*, *Bacteroidetes*, *Chloroflexi* or the superphylum *Planctomycetes*, *Verrucomicrobia*, *Chlamydia* (*PVC*). It could be found out that 80 % of *Acidobacteria* have high gene cluster densities above the calculated average. In addition, several resistance genes of potentially relevant gene clusters for antibiotics biosynthesis could be identified. Thus *Acidobacteria* tend to be the most promising resource of this dataset. As a consequence, the strain *Terracidiphilus* sp. S55 and *Acidicapsa borealis* were chosen for clone library establishment. The latter was furthermore genome sequenced by using a hybrid assembly approach. Finally the genome was annotated and investigated according to outstanding traits. It could be found out that this strain also has an abnormal number of biosynthetic gene clusters compared to the dataset. In addition to that, a metagenomics-driven approach was established by gaining bacteria from the termite nest of *Coptotermis niger*, which were furthermore taxonomically classified. Subsequently 80 clones could be prepared from this sample. In sum 960 clones from two resources were screened by means of a PCR-driven analysis for non-ribosomal peptide synthases and by direct antibiotics production, using a bioluminescence-based and an extract-based screening approach. Nine clones from *Acidicapsa borealis* DNA revealed antibiotic activity and were introduced to UHPLC-MS/MS analytics for compound detection. As a result, the two tripeptides Valylprolylleucin and Isoleucylprolyl isoleucin, could be identified within the extract of clone JF 7 C3, as products of heterologous expression. In summary, this research project established methods to conduct functional (meta-)genomic experiments, which serves as basis and alternative approach to discover active natural products.

## **Publikationen im Rahmen dieser Arbeit**

### Poster

Markus Oberpaul, Marius Spohn, Jochen Fracowiak, Rüdiger Plarre, Dino Peter McMahon, Peter Hammann, Andreas Vilcinskas, Jens Glaeser: Exploiting termite nest microbiomes for antibiotic drug discovery by using ultra-high throughput Microfluidics/FACS driven pipeline, combined with a microtiter plate based cultivation strategy, Annual Conference of the Society for General and Applied Microbiology (VAAM), 2018 Wolfsburg

Markus Oberpaul, Marius Spohn, Stephan Brinkmann, Maria Patras, Jochen Fracowiak, Rüdiger Plarre, Dino Peter McMahon, Peter Hammann, Andreas Vilcinskas, Jens Glaeser: Combined microfluidics/FACS platform for antibiotic drug discovery of insect-associated microorganisms, INSECTA, 2018 Giessen

Jens Glaeser, Markus Oberpaul, Mai Ngyuen, Tanja Stadler, Jochen Fracowiak, Rüdiger Plarre, Peter Hammann, Andreas Vilcinskas: Exploiting termite nest microbiomes for antibiotic drug discovery by using an ultra-high throughput Microfluidics/FACS driven pipeline, Microbiology and Infection - 5th VAAM Conference, 2017 Würzburg

Markus Oberpaul, Mai Ngyuen, Jochen Fracowiak, Florian Zubeil, Rüdiger Plarre, Peter Hammann, Andreas Vilcinskas, Jens Glaeser: Exploiting termite nest microbiomes for antibiotic drug discovery using an integrated high throughput Microfluidics/FACS pipeline, Biology of Bacteria Producing Natural Products - International VAAM Workshop, 2017 Tübingen

Jochen Fracowiak, Jens Glaeser, Andreas Vilcinskas: Genome-wide secondary metabolite prediction of underexplored bacterial phyla, de.NBI Summer School in Microbial Bioinformatics, 2015 Gießen

# Inhaltsverzeichnis

<b>Zusammenfassung</b>	<b>iii</b>
<b>Abstract</b>	<b>iv</b>
<b>Publikationen</b>	<b>v</b>
<b>Abbildungsverzeichnis</b>	<b>ix</b>
<b>Tabellenverzeichnis</b>	<b>xiii</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Hintergrund . . . . .	1
1.2 Beispiele bekannter Biosynthesewege von Antibiotika in Bakterien . . . . .	4
1.2.1 Polyketid-Biosynthese . . . . .	4
1.2.2 Nicht-ribosomale Peptidsynthese . . . . .	6
1.2.3 Ribosomal-synthetisierte und post-translational modifizierte Peptide . . . . .	7
1.3 Bakterien als Produzenten für neuartige Antibiotika . . . . .	8
1.3.1 Acidobakterien als neue Quelle für Antibiotika . . . . .	9
1.4 <i>state-of-the-art</i> Genomsequenzierung und nachfolgende Sequenzanalysen . . . . .	9
1.4.1 Das <i>Oxford Nanopore MinION</i> -System zur Genomsequenzierung . . . . .	10
1.4.2 Genomassemblierung und Annotation . . . . .	11
1.4.3 <i>In silico</i> -Analysen zur Identifizierung von Biosynthese-Genclustern . . . . .	14
1.5 Heterologe Genexpression zur Produktion von Naturstoffen . . . . .	16
1.5.1 Das <i>Escherichia coli</i> BAC-System zur Genexpression . . . . .	18
1.6 Ziele dieser Doktorarbeit . . . . .	20
<b>2 Material und Methoden</b>	<b>21</b>
2.1 Bioinformatische Untersuchung von Biosynthese-Genclustern in Bakteriengruppen . . . . .	21
2.1.1 Statistische Analysemethoden . . . . .	22
2.1.2 Funktionale Annotation ausgewählter Biosynthese-Genclustern . . . . .	23
2.1.3 Taxonomische Analyse von Stämmen der <i>Bacteroidetes</i> mit Biosynthese-Genclustern . . . . .	23
2.2 Genomsequenzierung von <i>Acidicapsa borealis</i> . . . . .	24
2.2.1 Sequenzierung durch das <i>Oxford Nanopore MinION</i> -System . . . . .	24
2.2.2 Sequenzierung durch das Illumina-HiSeq System . . . . .	24
2.2.3 <i>De-novo</i> Hybridassemblierung . . . . .	24

## Inhaltsverzeichnis

2.2.4	Sequenzanalyse des Genoms von <i>Acidicapsa borealis</i> . . . . .	25
2.3	Verwendete Labormaterialien . . . . .	26
2.3.1	Verwendete Medien zur Kultivierung . . . . .	26
2.3.2	Verwendete Verbrauchsmaterialien . . . . .	27
2.3.3	Verwendete Stämme . . . . .	28
2.3.4	Liste an verwendeten Kits . . . . .	29
2.3.5	Komponenten für die Molekularbiologie . . . . .	29
2.3.6	Anfertigung von Puffern und Stammlösungen . . . . .	30
2.4	Bereitstellung von Bakterien für die Anfertigung von Klonbibliotheken . .	32
2.4.1	Anzucht von <i>Acidicapsa borealis</i> und <i>Terracidiphilus</i> sp. S55 . . . .	32
2.4.2	Gewinnung von Bakterien aus dem Termitennest . . . . .	32
2.4.3	Zellzahlbestimmung gewonnener Bakterien . . . . .	32
2.4.4	Taxonomische Klassifizierung gewonnener Bakterien aus dem Termitennest . . . . .	33
2.5	Erstellung von Klonbibliotheken mit <i>Escherichia coli</i> als Wirt . . . . .	35
2.6	Selektion und Anzucht erstellter Klone . . . . .	39
2.7	Untersuchung erstellter Klonbibliotheken . . . . .	40
2.7.1	BAC-Isolation und Restriktion zur Größenkontrolle der Inserts . .	40
2.7.2	Markergen-spezifische Untersuchung der Klonbibliotheken . . . . .	40
2.7.3	Bioinformatische Determinierung NRPS-positiver BAC-Inserts . .	41
2.7.4	Untersuchung auf antibakterielle Aktivität . . . . .	42
2.7.5	Aktivitätsverifizierung durch <i>scale-up</i> Kultivierung . . . . .	45
<b>3</b>	<b>Ergebnisse und Diskussion</b>	<b>47</b>
3.1	Bioinformatische Analyse von Bakterien hinsichtlich Biosynthese-Gencluster	47
3.1.1	Quantifizierung von Genclustern in alternativen Bakteriengruppen	48
3.1.2	Annotation der Gencluster-Typen <i>other</i> und <i>terpene</i> . . . . .	56
3.1.3	Stammauswahl anhand von Resistenzgenen und Sequenzähnlichkeit	58
3.1.4	Taxonomische Zuordnung von Bakterien mit Genclustern . . . . .	64
3.1.5	Charakterisierung von <i>trans</i> -AT PKS in Acidobakterien . . . . .	66
3.1.6	Die Genom Assemblierung von <i>Acidicapsa borealis</i> . . . . .	71
3.1.7	Die funktionale Annotation von <i>Acidicapsa borealis</i> . . . . .	72
3.1.8	Acidobakterien sind assoziiert mit <i>Coptotermis niger</i> . . . . .	79
3.2	Erstellung einer Klonbibliothek mit DNS aus Acidobakterien . . . . .	84
3.2.1	Molekularbiologische Schritte der Erstellung von Klonbibliotheken	84
3.3	Funktionale Untersuchung der Klonbibliotheken . . . . .	89
3.3.1	Restriktionsverdau zur Insertkontrolle . . . . .	89
3.3.2	Detektion von NRPS-Genfragmenten durch PCR . . . . .	91
3.3.3	Bestimmung der Insertgröße von BACs mit NRPS-Signal . . . . .	95
3.3.4	Untersuchung von Klonbibliotheken hinsichtlich antibakterieller Aktivität . . . . .	101
3.3.5	Peptide: mögliche Verbindungen mit antibakterieller Aktivität . .	105
<b>4</b>	<b>Ausblick</b>	<b>110</b>

<b>5 Annex</b>	<b>112</b>
5.1 Bioinformatik . . . . .	112
5.2 Klonbibliothek . . . . .	118
5.3 Technische Daten der chemischen Analytik . . . . .	120
5.4 Verwendete Chemikalien . . . . .	121
5.5 Verwendete Geräte . . . . .	123
5.6 Abkürzungsverzeichnis . . . . .	124
5.7 Extraktanalytik . . . . .	127
<b>Literaturverzeichnis</b>	<b>160</b>
<b>Danksagung</b>	<b>185</b>

# Abbildungsverzeichnis

1.1	Organisation der Polyketidsynthese-Typen I, II und III, in Anlehnung an Lim <i>et al.</i> 2016 [31]. Dargestellt sind die einzelnen katalytischen Domänen. AT: Acyltransferase; ACP: Acyl-Carrier-Protein; KS: Ketosynthase; DH: Dehydrogenase; ER: Enoylreduktase; KR: Ketoreduktase; TE: Thioesterase; CoA: Coenzym A . . . . .	5
1.2	Modulare Grundstruktur der katalytischen Domänen einer nicht-ribosomalen Peptidsynthase, in Anlehnung an Donadio <i>et al.</i> (2007) [45]. Gezeigt sind folgende Domänen, A: Adenylation; T: Peptidyl-Carrier-Protein, C: Condensation und TE: Thioesterase . . . . .	7
1.3	Illustration der <i>Oxford Nanopore</i> -Sequenzierung basierend auf einer Nukleotidsequenz als Einzelstrang, die durch eine Pore geleitet, eine detektierbare Spannungsänderung erzeugt. Abbildung modifiziert nach Göpfrich <i>et al.</i> 2018 [89] . . . . .	11
1.4	Darstellung einer Genomassemblierung durch einen de-Bruijn Graphen in Anlehnung an Baichoo <i>et al.</i> (2017) [91]. Durch Basenüberlappung von K-meren werden die einzelnen <i>sub strings</i> miteinander verbunden . . . . .	12
1.5	Entwicklung eines <i>profile Hidden Markov</i> Modells. (A) Ausgehend von einem MSA wird eine Konsensussequenz erstellt (B). Das <i>profile</i> HMM enthält zusätzliche Wahrscheinlichkeiten zur Berechnung von Insertionen und Deletionen (C). Abbildung in Anlehnung an Yoon (2009) [99] . . . . .	14
1.6	Prinzipielle Schritte eines klassischen Ansatzes in der funktionalen Metagenomik. Ausgangspunkt ist die Ressource, gefolgt von der DNS-Isolation und der Erstellung von Klonbibliotheken. Diese werden funktional untersucht und selektiert. Ausgewählte Klone werden sequenziert und deren Metabolite charakterisiert. Abbildung in Anlehnung an Trindade <i>et al.</i> (2015) [120] . . . . .	17
2.1	Ablauf der bioinformatischen Detektion und Analyse der Gencluster. Ausgangspunkt war die Datenbank-Akquise, gefolgt von der Identifizierung der Gencluster. Es folgte die statistische, funktionale und taxonomische Analyse . . . . .	22
2.2	Ablauf der Erstellung von Klonbibliotheken mit <i>E. coli</i> MegaX DH10B T1 Electrocomp als Wirt . . . . .	35
2.3	Plasmidkarte von pBeloBac11 für die Erstellung von Klonbibliotheken. Die BamHI-Schnittstelle wurde für die Klonierung genutzt . . . . .	36

## Abbildungsverzeichnis

2.4	Plasmidkarte von pFU166 mit Lumineszenzkassette ( <i>luxA</i> , <i>luxB</i> , <i>luxC</i> , <i>luxD</i> , <i>luxE</i> ) und Selektionsmarker <i>bla</i> gegen $\beta$ -Lactam-Antibiotika, eingesetzt für die primäre Untersuchung als <i>E. coli</i> DH5 $\alpha$ [pFU166] [173] . . .	43
3.1	Verteilung der Genomgröße von Stämmen alternativer Bakterien die Biosynthese-Gencluster (BGC) enthalten oder nicht enthalten . . . . .	49
3.2	Korrelationsanalyse der Actinobakterien . . . . .	49
3.3	Korrelationsanalyse wenig erforschter Bakteriengruppen . . . . .	49
3.4	Gruppierung des Datensatzes anhand der Steigungsgeraden von 0,88 . . .	50
3.5	Gruppierung des Datensatzes nach Gencluster-Dichte je Phylum . . . . .	50
3.6	Taxonomische Zuordnung von Stämmen der <i>Bacteroidetes</i> die Biosynthese-Gencluster enthalten . . . . .	65
3.7	Gencluster mit <i>trans</i> -AT Polyketidsynthasen des Typs I aus drei Acidobakterien, <i>Acidobacterium</i> sp. S8, <i>Terracidiphilus</i> sp. S55 und <i>Bryocella</i> sp. S190. Die entsprechenden Gene sind hervorgehoben . . . . .	68
3.8	Sequenzvergleich der Ketosynthasen mit <i>trans</i> -AT <i>docking</i> Domänen aus <i>Acidobacterium</i> sp. S8, <i>Terracidiphilus</i> sp. S55 und <i>Bryocella</i> sp. S190. Herausgestellt ist jeweils der konservierte Bereich um die katalytische Triade (CHH) und der charakteristische C-Terminus . . . . .	69
3.9	Sequenzlängen-Verteilung der gefilterten <i>reads</i> aus der <i>Oxford Nanopore</i> Sequenzierung für die Genom Assemblierung von <i>A. borealis</i> . . . . .	71
3.10	Quantifizierung biologischer Prozesse anhand von Genontologien des <i>A. borealis</i> Genoms . . . . .	72
3.11	Identifizierte Gene (in grüner Farbe dargestellt) des Stärke- und Saccharose-Metabolismus in <i>A. borealis</i> . . . . .	74
3.12	Gewinnung von Bakterien aus dem turaxierten Termitennest nach Zentrifugation mit Nykodenzzusatz . . . . .	80
3.13	Zellzählung der mittels Nykodenzen gewonnenen Zellen aus dem Termitennest. FL1: Fluoreszenzintensität, FSC: <i>forward scatter</i> beziehungsweise Signalgröße . . . . .	81
3.14	Ergebnis der taxonomischen Klassifizierung auf Klassen-Ebene der gewonnenen Bakterien aus dem <i>C. niger</i> Termitennest . . . . .	82
3.15	Kulturen der zwei Acidobakterien <i>Terracidiphilus</i> sp. S55 (links) und <i>A. borealis</i> (rechts) vor der Zellernte . . . . .	84
3.16	Beispiel der Auftrennung von partial verdauter DNS mittels PFGE. Tache 1: Marker, 2: <i>A. borealis</i> DNS unverdaut, Tache 3 - 10 <i>A. borealis</i> DNS partial verdaut und weiterverarbeitet für die Klonierung (Der Größenbereich zwischen 50 - 150 Kb wurde zuvor ausgeschnitten und anschließend gefärbt) . . . . .	86
3.17	Blau-Weiß-Selektion anhand eines Ausstrichs von 100 $\mu$ L Transformationsvolumen (1 zu 5 verdünnt) auf eine Agarplatte (12 cm im Durchmesser), 24 Stunden nach der Transformation von klonierter DNS aus <i>A. borealis</i> .	87

## Abbildungsverzeichnis

3.18	Restriktionsverdau zur Insertlängenkontrolle. Tasche 1: Marker, Tasche 2: pBeloBac11 zirkulär, Tasche 3: pBeloBac11 verdaut, Tasche 4: JF 3 A5, Tasche 5: Klon mit <i>Terracidiphilus</i> sp. S55 DNS . . . . .	89
3.19	Platte JF 7: Markergen-spezifische Identifizierung der Adenyldomäne von nicht-ribosomalen Peptidsynthasen aus der DNS von <i>A. borealis</i> . Es wurden zwei Signale identifiziert: JF 7 A7 und JF 7 E10. Auf Grund der Plattenkonfiguration beginnt die Nummerierung ab der Ziffer 3 . . . . .	92
3.20	Platte JF 8: Markergen-spezifische Identifizierung der Adenyldomäne von nicht-ribosomalen Peptidsynthasen aus der DNS von <i>A. borealis</i> . Es wurden ein Signal identifiziert: JF 8 B12. Auf Grund der Plattenkonfiguration beginnt die Nummerierung ab der Ziffer 3 . . . . .	93
3.21	Markergen-spezifische Identifizierung der Adenyldomäne von nicht-ribosomalen Peptidsynthasen aus der DNS von <i>A. borealis</i> (Platte 12). Auf Grund der Plattenkonfiguration beginnt die Nummerierung ab der Ziffer 3 . . . . .	94
3.22	Lokalisation der klonierten Inserts ausgewählter BACs mit NRPS-Signalen aus der Klonbibliothek von <i>A. borealis</i> . Gelb: NRPS-Gencluster 5 aus <i>A. borealis</i> , Blau: weiterer genomischer Bereich . . . . .	95
3.23	Chromatogramme als Liste Klons JF 7 C3 im Vergleich zur Referenz <i>E. coli</i> MegaX DH10B T1 [pBeloBAC11] . . . . .	106
3.24	Signale Klons JF 7 C3 herausgestellt. Signal 1: Valylprolylleucin, Signal 2: Isoleucylprolyliso-leucin . . . . .	107
3.25	MS/MS-Spektrenvergleich von Diprotin A (oben) und Isoleucylprolyliso-leucin (unten) . . . . .	108
5.1	Hochrechnung des Gencluster-Potentials anhand des Datensatzes für jedes Phylum. ACI: <i>Acidobacteria</i> , ACT: <i>Actinobacteria</i> , BAC: <i>Bacteroidetes</i> , CHL: <i>Chloroflexi</i> , PVC: <i>Planctomycetes</i> , Verrucomicrobia, <i>Chlamydiae</i> . . . . .	112
5.2	Markergen-spezifische Identifizierung der Adenyldomäne von nicht-ribosomalen Peptidsynthasen der Termitennest-Klonbibliothek . . . . .	118
5.3	Markergen-spezifische Identifizierung von NRPS-Gencluster des isolierten BACs aus Klon JF 7 C3 . . . . .	119
5.4	Technische Daten der verwendeten UHPLC-Methode . . . . .	120
5.5	Chromatogramme als Liste des Klons JF 7 C4 . . . . .	127
5.6	Signale des Klons JF 7 C4 herausgestellt . . . . .	128
5.7	Massenspektrum 1 des Klons JF 7 C4 . . . . .	129
5.8	Massenspektrum 2 des Klons JF 7 C4 . . . . .	130
5.9	Chromatogramme als Liste des Klons JF 8 A9 . . . . .	131
5.10	Signal 1 des Klons JF 8 A9 herausgestellt . . . . .	132
5.11	Massenspektrum 1 des Klons JF 8 A9 . . . . .	133
5.12	Chromatogramme als Liste des Klons JF 9 C12 . . . . .	134
5.13	Signal 1 des Klons JF 9 C12 herausgestellt . . . . .	135
5.14	Massenspektrum 1 des Klons JF 9 C12 . . . . .	136
5.15	Chromatogramme als Liste des Klons JF 9 H5 . . . . .	137

5.16	Signal 1 des Klons JF 9 H5 herausgestellt . . . . .	138
5.17	Massenspektrum 1 des Klons JF 9 H5 . . . . .	139
5.18	Massenspektrum 1 des Klons JF 9 H5 . . . . .	140
5.19	Massenspektrum 2 des Klons JF 9 H5 . . . . .	141
5.20	Chromatogramme als Liste des Klons JF 11 B9 . . . . .	142
5.21	Signal 1 des Klons JF 11 B9 herausgestellt . . . . .	143
5.22	Massenspektrum 1 des Klons JF 11 B9 . . . . .	144
5.23	Chromatogramme als Liste des Klons JF 13 C6 . . . . .	145
5.24	Signal 1 des Klons JF 13 C6 herausgestellt . . . . .	146
5.25	Massenspektrum 1 des Klons JF 13 C6 . . . . .	147
5.26	Signal 2 des Klons JF 13 C6 herausgestellt . . . . .	148
5.27	Massenspektrum 2 des Klons JF 13 C6 . . . . .	149
5.28	Chromatogramme als Liste des Klons JF 13 G11 . . . . .	150
5.29	Signal 1 des Klons JF 13 G11 herausgestellt . . . . .	151
5.30	Massenspektrum 1 des Klons JF 13 G11 . . . . .	152
5.31	Signal 2 des Klons JF 13 G11 herausgestellt . . . . .	153
5.32	Massenspektrum 2 Klons JF 13 G11 . . . . .	154
5.33	Chromatogramme als Liste Klons JF 13 G5 . . . . .	155
5.34	Signal 1 des Klons JF 13 G5 herausgestellt . . . . .	156
5.35	Massenspektrum 1 des Klons JF 13 G5 . . . . .	157
5.36	Signal 2 Klons JF 13 G5 herausgestellt . . . . .	158
5.37	Massenspektrum 2 als Klons JF 13 G5 . . . . .	159

# Tabellenverzeichnis

2.1	Komponenten für LB-Medium . . . . .	26
2.2	Komponenten für R2A . . . . .	26
2.3	Verwendete Verbrauchsmaterialien . . . . .	27
2.4	Verwendete Stämme in dieser Arbeit . . . . .	28
2.5	Verwendete Kits für die Molekularbiologie . . . . .	29
2.6	Verwendete Enzyme für die Molekularbiologie . . . . .	29
2.7	Verwendete <i>primer</i> in dieser Arbeit . . . . .	29
2.8	Komponenten für 50x TAE . . . . .	30
2.9	Komponenten für 10x TBE . . . . .	30
2.10	Komponenten für den Partialverdau-Puffer . . . . .	30
2.11	Komponenten der Lysozymlösung . . . . .	31
2.12	Komponenten der Achromopeptidase-Lösung . . . . .	31
2.13	Komponenten der Zellyse-Lösung . . . . .	31
2.14	Stopplösung I . . . . .	31
2.15	Sequenzanalyse erstellter <i>reads</i> anhand der SILVAngs 1.3 <i>pipeline</i> [169] . .	34
2.16	Komponenten für die Linearisierung von pBeloBAC11 als Vorbereitung auf die Klonierung . . . . .	37
2.17	Komponenten für die Ligation hochmolekularer DNS in pBeloBAC11 . . .	39
2.18	Komponenten für die Restriktion und Insertlängen-Kontrolle ausgewählter Klone . . . . .	40
2.19	Komponenten für die Detektion von Adenyldomänen in nicht-ribosomalen Peptidsynthesen . . . . .	41
2.20	PCR-Protokoll zur Markergen-spezifischen Detektion von Adenyldomänen in nicht-ribosomalen Peptidsynthesen . . . . .	41
2.21	Klonbibliotheken für die <i>scale-up</i> Extraktion und chemische Analytik . . .	45
2.22	Protokoll für die chromatographische Auftrennung generierter Extrakte .	46
3.1	Datensätze der bioinformatischen Auswertung auf DVD . . . . .	47
3.2	Statistische Erhebung von Genomgröße und BGC-Anzahl von wenig er- forschter Bakteriengruppen. Extremwerte sind hervorgehoben . . . . .	51
3.3	Statistische Erhebung von Genomgröße und BGC-Anzahl des Referenz- phylums <i>Actinobacteria</i> . Extremwert ist hervorgehoben . . . . .	51
3.4	Anzahl an detektierten Genclustern im Datensatz. ACI: <i>Aciobacteria</i> ; BAC: <i>Bacteroidetes</i> ; CHL: <i>Chloroflexi</i> ; PVC: <i>Planctomycetes</i> , <i>Verrucomi- crobia</i> , <i>Chlamydiae</i> ; ACT: <i>Actinobacteria</i> . Gencluster mit <i>trans</i> -AT Poly- ketidsynthase sind hervorgehoben . . . . .	52

3.5	Die fünf häufigsten Proteine aus dem Gencluster-Typ <i>terpene</i> . . . . .	57
3.6	Die fünf häufigsten Proteine aus dem Gencluster-Typ <i>other</i> . . . . .	57
3.7	Stammauswahl des Phylums <i>Acidobacteria</i> anhand von Resistenzgenen und Sequenzähnlichkeit . . . . .	60
3.8	Stammauswahl des Phylums <i>Bacteroidetes</i> anhand von Resistenzgenen und Sequenzähnlichkeit, Teil 1 . . . . .	61
3.9	Stammauswahl des Phylums <i>Bacteroidetes</i> anhand von Resistenzgenen und Sequenzähnlichkeit, Teil 2 . . . . .	62
3.10	Stammauswahl des Phylums <i>Chloroflexi</i> und <i>PVC</i> anhand von Resistenzgenen und Sequenzähnlichkeit . . . . .	63
3.11	Die 15 häufigsten Protein- oder Enzymfamilien des Carbohydrat-Stoffwechsels von <i>A. borealis</i> . . . . .	76
3.12	Identifizierte BGC im assemblierten Genom von <i>A. borealis</i> . . . . .	77
3.13	Identifizierte Markergene der Biosynthese-Gencluster in <i>Acidicapsa borealis</i> . . . . .	78
3.14	Zusammenfassung der Erstellung von Klonbibliotheken aus <i>Terracidiphilus</i> sp. S55, <i>A. borealis</i> und <i>C. niger</i> Termitennest . . . . .	88
3.15	Auflistung annotierter Proteine des NRPS-Genclusters fünf aus <i>A. borealis</i> , Teil 1 . . . . .	97
3.16	Auflistung annotierter Proteine des NRPS-Genclusters fünf aus <i>A. borealis</i> , Teil 2 . . . . .	99
3.17	Ergebnisse der Inhibierung auf Basis der Biolumineszenz (primäre Untersuchung) und der Extrakt-basierten Inhibierung auf Basis der optischen Dichte (sekundäre Untersuchung). Die Ergebnisse sind als Duplikat dargestellt . . . . .	103
3.18	Ergebnisse der wiederholten Extrakt-basierten Untersuchung als 10x Konzentrat aus 50 mL Kultur ( <i>scale-up</i> ) . . . . .	104
3.19	Identifizierte Massen des Klons JF 7 C3 im Duplikat . . . . .	108
3.20	Massenspuren weiterer Klone mit antibakterieller Aktivität . . . . .	109
5.1	Klassifikation identifizierter Resistenzen in Gencluster, Teil 1 . . . . .	113
5.2	Klassifikation identifizierter Resistenzen in Gencluster, Teil 2 . . . . .	114
5.3	Klassifikation identifizierter Resistenzen in Gencluster, Teil 3 . . . . .	115
5.4	Stämme mit Resistenzgenen und Genclusterähnlichkeit, Teil 1 . . . . .	116
5.5	Stämme mit Resistenzgenen und Genclusterähnlichkeit, Teil 2 . . . . .	117
5.6	Verwendete Antibiotika . . . . .	121
5.7	Verwendete Chemikalien . . . . .	122
5.8	Verwendete Geräte . . . . .	123
5.9	Aminosäuren und Abkürzungen . . . . .	124
5.10	Fachtermini, Teil 1 . . . . .	125
5.11	Fachtermini, Teil 2 . . . . .	126

# 1 Einleitung

## 1.1 Hintergrund

### Der Beginn der Antibiotikaforschung

Die erste dokumentierte Anwendung einer antibiotisch wirksamen Substanz für medizinische Zwecke ist auf das Jahr 1909 datiert. Es handelte sich um die synthetische Verbindung Salvarsan, welche in der Forschergruppe um Paul Ehrlich entwickelt worden ist [1]. Die zusätzliche Entdeckung von Penicillin im Jahr 1928 durch Alexander Fleming [2] eröffnete erstmals die Anwendung von Antibiotika aus dem Pilz *Penicillium fungi*. Heutzutage werden Antibiotika in der Human- und Veterinärmedizin eingesetzt. Antibiotika dienen zur Bekämpfung von bakteriellen Infektionen, wobei zwischen einer hemmenden Wirkung (bakteriostatisch) und einer abtötenden Wirkung (bakteriozid) unterschieden wird [3]. Aus chemischer Perspektive lassen sich Antibiotika in zwei Kategorien einteilen, zum einen in synthetische Verbindungen und zum anderen aus der Natur abgeleitete Verbindungen, die generell als Naturstoffe bezeichnet werden.

Viele Naturstoffe mit antibiotischer Wirkung entstammen aus Mikroorganismen, wie Pilzen oder Bakterien und dienen der eigenen Verteidigung oder als Signalmolekül [4]. Unter dem taxonomischen Gesichtspunkt zählt die Ordnung *Actinomycetales* und deren Genus *Streptomyces* zu den bekanntesten Bakterien, die in der Forschung untersucht worden sind [5].

Etwa die Hälfte der heute auf dem Markt befindlichen Antibiotika geht auf die Entdeckungen zwischen 1950 bis 1960 zurück, weshalb dieser Zeitraum als goldene Ära der Antibiotika-Forschung bezeichnet wird [6].

### Resistenzentwicklung als akutes Problem

Durch den unsachgemäßen Gebrauch, beispielsweise ohne hinreichende medizinische Indikation bei einer bakteriellen Infektion, ist die natürliche Resistenzentwicklung gegen Antibiotika in Bakterien gefördert worden, die auch als Selektionsdruck bezeichnet wird [7]. Im Zuge der Resistenzentwicklung wird zwischen drei Kategorien unterschieden, der intrinsischen Resistenz, der adaptiven Resistenz und der akquirierten Resistenz [8]. Die intrinsische Resistenz wird durch die Eigenschaften der Zellhülle vermittelt, die aus der Zellwand sowie der Zellmembran und deren Polysacchariden besteht. Im Falle von Gram-negativen Bakterien besitzt die Zellhülle zwei Membranen, von denen die äußere Schicht eine zusätzliche Permeabilitäts-Barriere darstellt. Dies betrifft sowohl die Diffusion durch die Lipidschicht von hydrophoben Antibiotika, als auch deren Transport von

## 1.1 Hintergrund

hydrophilen Antibiotika durch Porine [9] [10].

Die adaptive Resistenz ist eine transiente Eigenschaft, welche durch einen konzentrationsabhängigen Umweltreiz ausgelöst wird und eine Änderung der Genexpression, beispielsweise von Transportern zur Folge hat.

Die akquirierte Form der Resistenz basiert auf Mutation oder durch Weitergabe von Resistenzgenen, zum Beispiel in Form des horizontalen Gentransfers [11]. Bekannt sind Mutationen an Transportern, welche den Influx von Antibiotika verhindern oder spezialisierte Enzyme, wie  $\beta$ -Lactamasen, welche das Antibiotikum inaktivieren [7].

Die Weitergabe von stammspezifischen Resistenzgenen in Folge unsachgemäßer Antibiotika-Anwendung und des ausgeübten Selektionsdrucks, führt darüber hinaus nicht nur dazu, dass obligat-pathogene Bakterien, wie *Mycobacterium tuberculosis*, resistenter werden, sondern auch dazu, dass weitere Kommensalen des Menschen, wie zum Beispiel *Escherichia coli* oder diverse Staphylokokken, diese Resistenzen erwerben. Als Resultat können auch bakterielle Infektionen von fakultativ pathogenen Bakterien schwer behandelbar werden [12].

Bakterien von besonderer klinisch-pathologischer Relevanz werden unter dem Akronym ESKAPE zusammengefasst, dass die Stämme *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* und *Enterobacter* sp. bezeichnet [13]. Im Zusammenhang dieser Entwicklung wird die zunehmende Anzahl Gram-negativer Bakterien als besonders kritisch angesehen. Diese sind *Klebsiella pneumoniae*, *Acinetobacter baumannii* und *Pseudomonas aeruginosa*. Eine Statistik innerhalb des europäischen Raums für das Jahr 2015 ergab eine Anzahl von 33.110 Infektionen mit Todesfolge durch nicht mehr behandelbare Bakterien. Für dieses Ergebnis sind insbesondere Cephalosporin- und Meticillin-resistente Bakterien verantwortlich [14].

### Neue Ansätze in der Antibiotikaforschung

Die vermehrte Resistenzentwicklung gefährdet zunehmend den Erfolg einer antibiotischen Behandlung. Diese problematische Entwicklung soll durch internationale- sowie nationale Bestrebungen, wie beispielsweise der Deutschen Antibiotika Resistenzstrategie (DART) entgegen gewirkt werden [15]. Neben dem kontrollierten Einsatz von Antibiotika bedarf es zukünftig alternativer Strategien zur Entdeckung neuer antibiotischer Metabolite [16]. Der Bedarf einer Intensivierung in diesem Forschungsbereich wird unterstrichen durch die niedrige Anzahl von nur zwei neuartigen Substanzen seit 1990, wie Linezolid und Daptomycin, welche gegen Gram-positive Bakterien wirken [16]. Ferner lässt sich diese Entwicklung in der Forschung an vier Hemmnissen festmachen, die durch einen Bericht der *Boston Consulting Group* genannt werden [17], wie zum Beispiel Mangel an entsprechenden Experten und chemischen Leitstrukturen sowie schlechte Gewinnaussichten und ein hohes Ausfallrisiko in der Entwicklungsphase.

Bisweilen beträgt der Anteil an entwickelten Antibiotika aus Naturstoffen 73 % [15] und entstammt aus Pilzen oder Bakterien welche durch Kultivierung identifiziert worden sind. Schätzungen gehen davon aus, dass etwa nur 1 % von Bakterien unter Standard-

## *1.1 Hintergrund*

Laborbedingungen kultivierbar sind [18], was die konventionelle Wirkstoffsuche stark limitiert. Daher bedarf es alternative Methoden welche die Wirkstoffsuche erweitern wie beispielsweise aus der Bioinformatik und Molekularbiologie. Diese bieten den Vorteil, die genetische Grundlage der Biosynthese von Naturstoffen in Form der Desoxyribonukleinsäure (DNS), unabhängig von der Kultivierung zu nutzen. Diesbezüglich konnten zum Beispiel erste Erfolge durch die Erstellung von Klonbibliotheken aus Umwelt-DNS gezeigt werden [19]. Um ein Verständnis der Biosynthese von Antibiotika zu erhalten, sind im nachfolgenden Kapitel die genetischen Grundlagen und bekannte Stoffwechselwege beschrieben, auf die Bezug genommen wird.

## 1.2 Beispiele bekannter Biosynthesewege von Antibiotika in Bakterien

Die Biosynthese von Antibiotika in Mikroorganismen zählt zu den spezialisierten Stoffwechselwegen. Deshalb werden diese enzymatischen Prozesse in den Sekundärmetabolismus eingeordnet, der Regulationsmechanismen unterliegt und nur unter bestimmten Bedingungen aktiviert wird, wie etwa zur Verteidigung [20] [21] oder während der Zelldifferenzierung bei einigen Bakterien [22]. Der Sekundärmetabolismus unterscheidet sich daher vom Primärmetabolismus, welcher essentiell für die Energiegewinnung und das Wachstum ist [23]. Dennoch können einzelne Metabolite für die Antibiotika-Biosynthese aus dem Primärmetabolismus genutzt werden [23].

Bekannte Beispiele sind Manoyl-CoA für die Polyketid-Biosynthese [24], oder Phenylalanin aus der Shikimat-Biosynthese [25], der aromatische Aminosäuren für die nicht-ribosomale Peptidsynthese bereitstellt [26].

### 1.2.1 Polyketid-Biosynthese

Polyketidsynthasen (PKS) werden enzymatisch in drei Typen (I, II und III) eingeteilt, die einen Multienzym-Komplex ausbilden und aus essentiellen und fakultativen Domänen bestehen. Diese Enzyme haben sich aus Fettsäuresynthasen (FAS) entwickelt [27] und können zum Teil die gleichen kurzkettigen Fettsäurederivate, wie Acetyl-CoA oder Propionyl-CoA in ihrer aktivierten Form verwenden. Jedoch ist das PKS-Substratspektrum durch weitere kurzkettige Carbonsäuren gegenüber Fettsäuresynthasen erweitert [28].

Der PKS Typ I gliedert sich in zwei Subtypen, die sich vom Ablauf der Biosynthese unterscheiden (nicht-iterativ oder iterativ), aber vom modularen Aufbau stark ähneln. Grundsätzlich bestehen Polyketidsynthasen des Typs I aus Modulen mit mindestens drei essentiellen katalytischen Domänen. Diese sind das Acyl-Carrier-Protein (ACP), welches den Transport der wachsenden Polyketidkette durch eine Thioester-Verbindung vermittelt, die Acyltransferasedomäne (AT), welche weitere Einheiten für den Einbau bereitstellt und der Ketosynthasedomäne (KS), die durch Claisenkondensation den Einbau dieser Einheiten in die bestehende Polyketidkette über eine Thioesterbindung vermittelt. Diese drei Domänen bilden das Minimum für ein Modul, welche hintereinander geschaltet, die wachsende Polyketidkette durch sequenziellen Einbau verlängert (Abbildung 1.1).

Der Unterschied zwischen dem nicht-iterativen und dem iterativen Typ besteht in dem Ablauf der Biosynthese. Diese verläuft bei dem nicht-iterativen Subtyp linear von links nach rechts, jeweils ein Mal pro Modul. Dieses Prinzip wird auch als Co-Linearitätsprinzip beschrieben [29]. Bei dem iterativen Subtyp verläuft die Biosynthese über ein Modul, welches mehrmals aktiv sein kann (Abbildung 1.1).

Bevor die Polyketidkette synthetisiert werden kann, muss ein Aktivierungsschritt vom apo- zum holo-Enzym durch die Phosphopantetheinyl-transferase (PPtase) erfolgen, welche das 4'-Phosphopantetheinyl auf den Serinrest der ACP-Domäne(n) überträgt [30].



### 1.2.2 Nicht-ribosomale Peptidsynthese

einem nicht-kovalenten Komplex zusammenlagern, aber die gleichen Starter- und Extender-einheiten wie Typ I PKS nutzen können. Eine weitere Besonderheit, wie in Abbildung 1.1 gezeigt, stellen die zwei unterschiedlichen Ketosynthase-Subtypen Alpha und Beta dar. Der Beta-Subtyp zeichnet sich durch eine andere Aminosäure-Sequenz (Glutamin anstatt Cystein) im katalytischen Zentrum aus und ist nicht wie der Alpha-Subtyp an der Claisen-kondensation beteiligt. Dennoch ist der Beta-Subtyp nicht inaktiv und kontrolliert die Länge der wachsenden Polyketidkette [36]. Ein Beispiel für ein PKS-Typ II synthetisiertes Antibiotikum ist das Oxytetracyclin [28] [37].

Typ III Polyketidsynthasen verfügen über eine katalytische Funktion als diskretes Enzym (Abbildung 1.1). Weitere Domänen sind nicht vorhanden, dennoch ist eine Reaktion zur Kettenverlängerung und Cyclisierung möglich [38]. Polyketidsynthasen dieses Typs produzieren eine Reihe unterschiedlicher Sekundärmetabolite, wie zum Beispiel Pyrone, Acridone oder Stilbene. Die Biosynthese von Chalcone in Pflanzen gilt als bekanntester Weg [39] [31]. In Bezug zur Antibiotika-Produktion ist dieser Typ jedoch wenig erforscht und wurde 1999 erstmals in Bakterien entdeckt [40]. Dennoch gibt es Untersuchungen, dass dieser Typ als Hybrid an der Biosynthese von beispielsweise Glykopeptid-Antibiotika involviert sein kann [41] [42].

### 1.2.2 Nicht-ribosomale Peptidsynthese

Ähnlich wie Polyketidsynthasen bestehen nicht-ribosomale Peptidsynthasen (NRPS) aus multimodular aufgebauten katalytischen Domänen, die sowohl proteinogene als auch nicht-proteinogene Aminosäuren als Bausteine zu einer Kette katalysieren. Neben den 20 proteinogenen Aminosäuren sind 500 nicht-proteinogene Aminosäuren bekannt, die zur Produktdiversität beitragen [43]. Die Biosynthese erfolgt modular aus mindestens drei katalytischen Domänen, wie der Adenylierungsdomäne (A-Domäne) für die Substraterkennung, der Peptidyl-Carrier-Protein Domäne (PCP oder als T abgekürzt), die das Substrat mit dem Intermediat kovalent bindet und der Kondensationsdomäne (C), welche die Extender-einheit mit dem Intermediat über eine Amid- oder Esterbindung kondensiert (Abbildung 1.2). Die Biosynthese der Aminosäurenkette kann, ähnlich der Polyketid-Biosynthese, über eine Thioesterase (TE) abgeschlossen werden. Neben den Basismodulen können auch weitere Module enthalten sein, beispielsweise zur Epimerisierung, Methylierung oder zur Reduzierung des Moleküls. Als Produkte können Makrozyklen und linear verzweigte Verbindungen entstehen, die zusätzlich acyliert oder glykosyliert werden können [44].

### 1.2.3 Ribosomal-synthetisierte und post-translational modifizierte Peptide

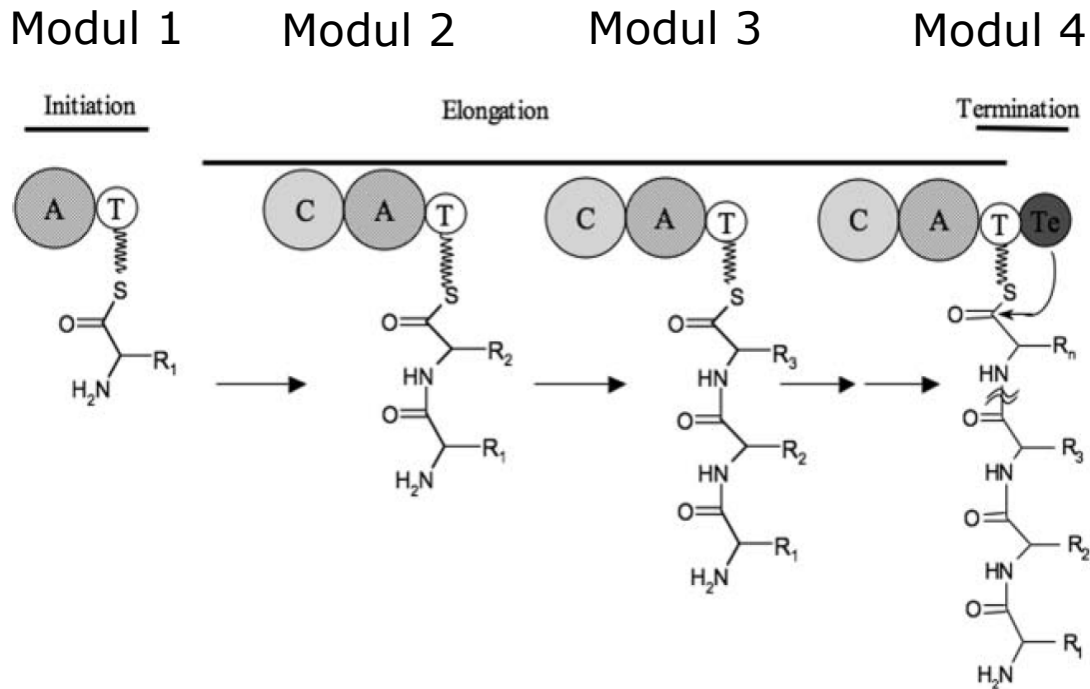


Abbildung 1.2: Modulare Grundstruktur der katalytischen Domänen einer nicht-ribosomalen Peptidsynthase, in Anlehnung an Donadio *et al.* (2007) [45]. Gezeigt sind folgende Domänen, A: Adenylation; T: Peptidyl-Carrier-Protein, C: Condensation und TE: Thioesterase

Antibiotika, die aus diesem Biosyntheseweg hervorgehen, sind beispielsweise  $\beta$ -Lactam-Antibiotika, wie etwa die Gruppe der Penicilline. Weitere wichtige Substanzen sind Daptomycin (Lipopeptid) [46] und Vancomycin (Glykopeptid) [31]. Neben den eigenständigen Genclustern existieren auch PKS/NRPS-Hybridformen [47] [48]. Ein Beispiel für ein Hybridmolekül ist das Bacillaen aus *Bacillus subtilis* [49].

### 1.2.3 Ribosomal-synthetisierte und post-translational modifizierte Peptide

Weitere Metabolite, die zur Verteidigung von Bakterien gebildet werden, sind ribosomal-synthetisierte und post-translational modifizierte Peptide (Ripps) [50]. Ripps besitzen durch ihre funktionale Seitenkette(n) eine elektrische Ladung, meist kationisch oder amphipatisch. Mit dieser Seitenkette können sie von außen die Membran penetrieren, bis zur Disruption der ganzen Zelle [51]. Verbindungsklassen von Ripps aus Bakterien sind zum Beispiel Bacteriocine [52] und Defensine [53]. Ein Ripp, welches in der Lebensmittelindustrie zur Konservierung eingesetzt wird, ist Nisin, produziert von *Lactococcus lactis* [54]. Ripps, wie Macroline die durch ihre bioaktive Eigenschaft häufig im Kontext von Antibiotika genannt werden, tragen auch die Bezeichnung antimikrobielle Peptide (AMPs). Auf genetischer Ebene bestehen Ripp-Gencluster im Vergleich zu PKS oder

### 1.3 Bakterien als Produzenten für neuartige Antibiotika

NRPS-Genclustern aus wenigen Genen [55]. Die Biosynthese von Ripps erfolgt über drei Schritte, wie der Transkription des Gens und der Translation von *messenger RNA*. Im letzten Schritt erfolgt die post-translationale Modifikation, durch die der N-terminale Teil des Peptids proteolytisch abgetrennt wird. Der C-Terminus bleibt vorhanden und kann weiter zum finalen Ripp posttranslational modifiziert werden, wodurch eine große strukturelle Diversität entsteht, die in 22 unterschiedliche Subtypen gegliedert werden kann [56].

### 1.3 Bakterien als Produzenten für neuartige Antibiotika

Die Ordnung *Actinomycetales*, insbesondere deren Genus *Streptomyces*, waren in den vergangenen Jahrzehnten die Hauptquelle für Antibiotika. Bisher wurden nur vereinzelt andere Bakterien entdeckt, wie beispielsweise Cyanobakterien, sowie Stämme der Genera *Burkholderia* [57] oder *Bacillus* [58], die Naturstoffe mit antibiotischer Wirkung produzierten [59]. Durch molekularbiologische Ansätze, wie der Polymerasekettenreaktion (PCR) auf konservierte Bereiche von PKS und NRPS Genclustern werden auch anderen Bakteriengruppen die potentielle Fähigkeit zur Biosynthese von Antibiotika zugeschrieben, wie etwa dem Phylum *Planctomycetes* [60]. Weitere genom-basierte Untersuchungen zeigten, dass marine *Chloroflexi* und *Bacteroidetes* Spezies in komplexen Symbiosen zu Schwämmen stehen können, die als neue Quelle für antibakterielle Sekundärmetabolite dienen [61] [62] [63] [64]. Eine der Hauptursachen für die geringe Anzahl an neuen Bakterien ist die eingeschränkte Methodik, Bakterien unter Standard-Laborbedingungen zu kultivieren, weil das natürliche Habitat nicht vollständig imitiert werden kann [65]. Beispiele sind die unzureichende Fähigkeit komplexe Symbiosen aufrechtzuerhalten [66], oder dormante Bakterien zu reaktivieren [67] [65] [68] [69].

Daher kann nur ein sehr kleiner Anteil an Bakterien aus Habitaten mit hoher mikrobieller Diversität in Kultur gebracht werden.

Diese Problematik konnte mit Hilfe der Metagenomik und der Weiterentwicklung in der Sequenzierung von DNS durch sogenannte *next generation sequencing* Technologien (NGS) reduziert werden, wodurch es nun möglich geworden ist, Genome zu entschlüsseln [70]. Als Resultat können unbekannte Biosynthese-Wege für potentielle Antibiotika unabhängig von der Kultivierung von Bakterien identifizieren werden [71] [72] [70].

Cimermanic *et al.* (2014) [73] zeigten in einer Phyla-übergreifenden Analyse hohe Abundanzen von Biosynthese-Genclustern des Sekundärmetabolismus aus dem Phylum der *Proteobacteria*. Weitere Phyla, wie *Chloroflexi*, *Bacteroidetes* und das Superphylum *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae* (*PVC*) [74] wurden ebenfalls hinsichtlich ihres Biosynthese-Potentials untersucht und sind aufgrund neuer Genominformationen auch Bestandteil dieser Forschungsarbeit.

In Bezug auf das Phylum *Acidobacteria* gibt es Hinweise auf ein hohes antibiotisches Potential, welches in der Publikation von Cimermanic *et al.* (2014) nicht berücksichtigt worden ist. Durch den stetigen Zuwachs von Genominformationen wird dieses Phylum nachfolgend ausführlich thematisiert und ein Einblick in die bisherigen Forschungsergebnisse gegeben.

#### 1.3.1 Acidobakterien als neue Quelle für Antibiotika

Das Phylum *Acidobacteria* wurde taxonomisch durch diverse Metagenomanalysen untersucht und weist eine große Verbreitung sowohl terrestrisch [75], als auch aquatisch [76] auf. Nach aktuellem Kenntnisstand spaltet sich das Phylum der Acidobakterien in 26 taxonomisch anerkannten *sub divisions* auf [77]. Das erste kultivierbare Bakterium aus diesem Phylum ist *Acidobacterium capsulatum*, isoliert aus Sediment mit saurem pH-Wert im Jahr 1991 [78]. Laut der Publikation von Ward *et al.* (2009) [79] besitzen einige Vertreter des Phylums eine ausgeprägte Carbohydrathydrat-Verwertung für Cellulose oder Xylose.

Im Kontext der Antibiotika-Forschung gilt das Phylum bisher als wenig wissenschaftlich beachtet und ist nur durch einzelne Metagenomanalysen sowie Klonbibliotheken beschrieben. Parsley *et al.* (2011) [80] charakterisierte eine Klonbibliothek aus einem Metagenom von Acidobakterien, in der homologe Gene von *mtaD* entdeckt worden sind. Für dieses Gen ist eine Funktion in der Biosynthese von Myxothiazol beschrieben.

Eine Studie von Crits-Christoph *et al.* (2018) [81] identifizierte 149 neue Acidobakterien anhand von *draft* Genomsequenzen aus Bodenproben durch einen kombinierten Ansatz von Metagenomik und Metatranskriptomik. Anhand dieser Untersuchung wurde eine Vielzahl unbekannter Polyketidsynthasen und nicht-ribosomale Peptidsynthasen entdeckt, welche durch eine Netzwerkanalyse mit diversen Regulatoren für zum Beispiel Zwei-Komponentensystemen verknüpft wurden, die relevant für die Aktivierung der Biosynthese sind.

Diese ersten Untersuchungen hinsichtlich schwer kultivierbarer Acidobakterien demonstrieren die Möglichkeiten der modernen Molekularbiologie, welche zunehmend in der Antibiotika-Forschung angewendet werden. Im Rahmen dieser Forschungsarbeit wurde diesem vielversprechenden Ansatz nachgegangen und umfassend ergänzt durch eine bioinformatische Untersuchung vorhandener Daten bezüglich Gencluster die im Kontext der Antibiotika-Produktion stehen können und einer vollständigen Sequenzierung von *Acidicapsa borealis*. Das nachfolgende Kapitel gibt einen Überblick über die verwendeten bioinformatischen Programme und Algorithmen zu diesem Ansatz.

## 1.4 *state-of-the-art* Genomsequenzierung und nachfolgende Sequenzanalysen

Jeglicher bioinformatischer Analyse geht die Sequenzierung von DNS voraus. Diese wird im Routinebetrieb zum Beispiel durch Sequenzierapparate der zweiten Generation durchgeführt. Als weit verbreitet gilt die Sequenzierung durch Illumina, die nach dem Prinzip Sequenzierung durch Synthese agiert [82]. Dabei dient ein DNS-Strang als Matrize, während der komplementäre Strang durch Fluorophor-markierte Nucleotide generiert wird. Die Sequenz des komplementären Strangs wird durch die Fluorophoremmission nach Nucleotideinbau optisch detektiert. Die Vorteile dieser Sequenzierung liegen im sehr hohen Durchsatz von kurzen Sequenzen von 200 - 500 Basenpaaren als *paired-end*, die eine sehr niedrige Fehlerrate pro Base haben (1 %) [83]. Durch repetitive Basenabfolgen im

#### 1.4.1 Das Oxford Nanopore MinION-System zur Genomsequenzierung

Genom ist die vollständige Genomschließung ohne weitere Methoden, wie *genome finishing* durch *primer walking* kompliziert [84]. Eine Methode diesen Aufwand zu umgehen, bieten Sequenzierapparate der dritten Generation wie *Pacific Biosciences* SMRT oder die *Oxford Nanopore* Sequenzierung, die eine Sequenzlänge von rund 40 Kilobasen generieren können [85] [86] und somit komplexe Genombereiche nicht extra assembliert werden müssen. Im nachfolgenden Kapitel ist die Sequenzierung durch das *Oxford Nanopore MinION*-System und die bioinformatische Auswertung der Sequenzdaten beschrieben.

#### 1.4.1 Das Oxford Nanopore MinION-System zur Genomsequenzierung

Zur Genomsequenzierung wurde das *Oxford Nanopore MinION* System verwendet. Das Sequenzierprinzip basiert auf einem porenbildenden Protein, dem  $\alpha$ -Hemolysin aus dem Bakterium *Staphylococcus aureus*, welches in eine Lipid-Doppelschicht integriert ist (Abbildung 1.3). Das  $\alpha$ -Hemolysin ist an der Oberseite mit einer Helikase gekoppelt, welche die DNS als Doppelhelix zum Einzelstrang entwindet und verlangsamt in die Pore einführt. Die Porenprotein-Matrix ist umgeben von einer Kaliumchlorid-Elektrolyt-Lösung, welche Bestandteil eines Stromkreises ist. Beim Eintritt der DNS in die Pore kommt es zu einer Änderung der Spannung (Abbildung 1.3). Im weiteren Sequenzierungsvorgang erzeugt jede der vier unterschiedlichen Basen (Adenin, Thymin, Cytosin und Guanin) eine typische Spannung, mit welcher die Nukleotidsequenz bestimmt wird [87]. Nach derzeitigem Stand der Technik beträgt die *alignment accuracy* für die Einzelstrang-Sequenzierung (1D) 94 % [88].

### 1.4.2 Genomassemblierung und Annotation

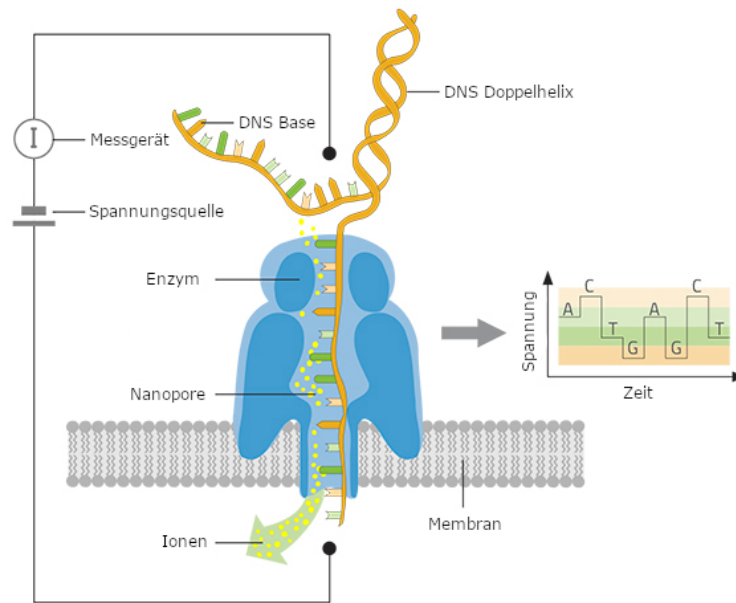


Abbildung 1.3: Illustration der *Oxford Nanopore*-Sequenzierung basierend auf einer Nukleotidsequenz als Einzelstrang, die durch eine Pore geleitet, eine detektierbare Spannungsänderung erzeugt. Abbildung modifiziert nach Göpfrich *et al.* 2018 [89]

### 1.4.2 Genomassemblierung und Annotation

Die Verwendung von Sequenzinformationen aus einem Illumina HiSeq-Lauf in Kombination mit dem *Oxford Nanopore MinION*-System für eine Hybridassemblierung bietet zwei Vorteile, die für eine vollständige *de novo* Rekonstruktion eines Genoms wichtig sind. Erstens können durch das *MinION*-System Sequenzfragmente relativ einfach zu sehr langen Sequenzabschnitten (*contigs*) assembliert werden, die durch Illumina *reads* zu einem fehlerarmen Sequenzkonstrukt (*scaffold*) oder Genom geschlossen werden können. Die durchgeführte Assemblierung besteht aus einer Abfolge von sieben Schritten, die durch das Programm *Unicycler* durchgeführt wurde [90]. Sie umfasst folgende Schritte: *short read assembly*, *multiplicity*, *short read bridging*, *long read bridging*, *bridge application*, *contig merging* sowie das *polishing*, welche nachfolgend kurz beschrieben werden.

Der erste Schritt, das *short read assembly*, basiert auf dem de-Bruijn Graph-Algorithmus, der durch *SPAdes* angewendet wird. In diesem Schritt werden zunächst die *reads*, welche in ihrer Länge variieren, zu kürzeren Sequenzen einer bestimmten Länge  $K$  definiert. Jede Sequenz (*string*) wird erneut in zwei kürzere Fragmente unterteilt, die als  $K$ -mere bezeichnet werden. Daraus resultiert, dass es für ein *string* zwei  $K$ -mere gibt, die sich jeweils am Präfix des einen und am Suffix des anderen um eine Base überlappen. Dieser Zustand wird als  $K$ -1-mer bezeichnet. Als Graph ausgedrückt repräsentiert ein  $K$ -mer einen Brückenpunkt (*node*) (Abbildung 1.4). Zwei  $K$ -mere werden durch ihre überlappende Base in Form eines Pfeils zusammengeführt (*edge*), der den vollständigen

### 1.4.2 Genomassemblierung und Annotation

*string* repräsentiert. Eindeutige Verbindungen werden anhand des Graphen als *contigs* zusammengefasst. Bei idealer Durchführung resultiert daraus ein langer *string*, bei dem jeder Brückenpunkt nur einmal durchlaufen wird und den kürzesten Weg repräsentiert, um das Genom vollständig zu assemblieren.

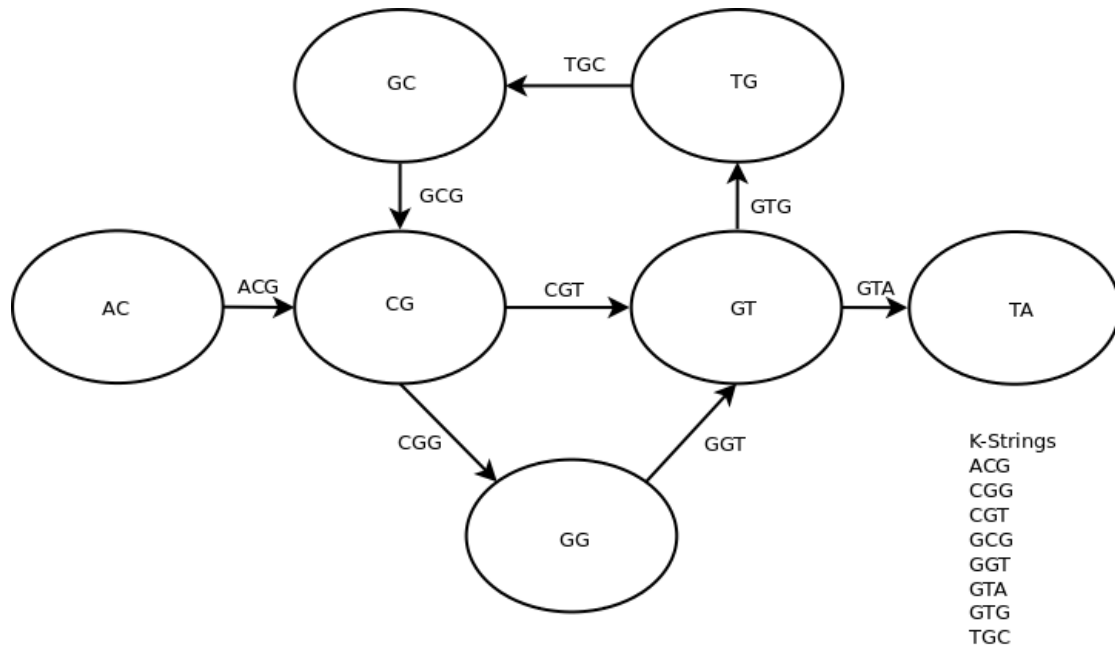


Abbildung 1.4: Darstellung einer Genomassemblierung durch einen de-Bruijn Graphen in Anlehnung an Baichoo *et al.* (2017) [91]. Durch Basenüberlappung von K-meren werden die einzelnen *sub strings* miteinander verbunden

In der Praxis wird dieser ideale Zustand selten erreicht, da Sequenzdaten aus vielfältigen Gründen nicht absolut fehlerfrei sind. Eine Fehlerquelle besteht darin, dass nicht jeder Genomabschnitt tatsächlich in Form eines *reads* vorhanden ist oder dass jede Base korrekt vom *base-caller* Algorithmus übersetzt wurde. Daher erhöhen Sequenzfehler bei der de-Bruijn Graph Assemblierung die Wahrscheinlichkeit zur Bildung von Abzweigungen, (Abbildung 1.4), die bei einer massiven Anzahl zu einem Abbruch der Assemblierung führen können.

Im zweiten Schritt der Assemblierung wird die Anzahl an gleichen K-meren berechnet, die als K-mer *multiplicity* bezeichnet wird. Diese Anzahl wird dazu verwendet, um falsche Abzweigungen, die durch mehrere Überlappungsmöglichkeiten entstanden sind, anhand der Anzahl an *reads* zu korrigieren.

Im dritten und vierten Schritt, dem *short- und long read bridging*, erfolgt die erste Korrektur des erstellten Graphen. Es werden anhand der *multiplicity* uneindeutige Verzweigungen durch die Berechnung von alternativen Übergängen vereinfacht, die in Folge einzelner *reads* entstanden sind. Dies erfolgt zunächst durch die Verwendung kurzer

### 1.4.2 Genomassemblierung und Annotation

*reads*. Im Anschluss werden lange *reads* verwendet, um voneinander getrennte *contigs* durch Brücken zu verbinden.

Im fünften Schritt, der *bridge application*, werden die berechneten Brücken anhand von Qualitätsparametern, wie der *read coverage* und der Qualität eines Sequenzvergleichs (*alignment*) zwischen Graphensequenz und der Brücken-Konsensussequenz dazu verwendet, den Graphen weiter zu simplifizieren.

Im abschließenden Schritt, dem *contig merging* und dem *polishing*, wird die Graphensequenz durch überlappende Sequenzen zusammengelegt, um unnötige Sequenzinformation zu löschen. Replikationsgene wie *dnaA* oder *repA* werden gesucht, um den Replikationsstart und die Orientierung innerhalb des Genoms zu determinieren.

#### Genvorhersage und Annotation

Die Genvorhersage und die Genannotation sind die weiteren Schritte nach der Genomassemblierung. Die Genvorhersage basiert auf statistischen Modellen [92] [93], welche anhand von konservierten Sequenzmotiven, beispielsweise Start- oder Stoppcodons, der *Shine-Dalgarno* Sequenz, *transfer RNA (tRNA)* und Nucleotidhäufigkeiten, Gene vorhersagt [94]. Im Folgeschritt werden die putativen Gene durch ein *alignment* mit einer Datenbank verglichen und bei Signifikanz annotiert.

Dies geschieht in der Regel auf Ebene der Aminosäuren und besagt, dass bei einer signifikanten Sequenzähnlichkeit zu einer Proteinfamilie, dem Protein die gleiche Funktion zugesprochen wird. Für den paarweisen Sequenzvergleich werden beispielsweise Algorithmen wie der *Smith-Waterman* Algorithmus, oder das effizientere *Basic Local Alignment Search Tool* (BLAST) verwendet [95] [96].

Eine weitere Methode, Sequenzen einer Funktion zuzuordnen, ist mit Hilfe von *Hidden Markov* Modellen (HMMs). Diese probabilistischen Modelle klassifizieren Sequenzen anhand ihrer Ähnlichkeit zu hinterlegten Proteinfamilien [97]. In der Bioinformatik werden *profile* HMMs für die Annotation als Standard verwendet und im Folgenden erläutert. Jede Nukleotid- oder Proteinsequenz kann bioinformatisch als ein *string* von unterschiedlichen, aber repetitiven Zeichen angesehen werden. Ein *profile* HMM wird aus einem multiplen *alignment* von Sequenzen (MSA) der gleichen Funktion generiert [94] [98]. An jeder Position im *alignment* werden die relativen Häufigkeiten der Zeichen berechnet, welche als *match* (M) bezeichnet werden und zunächst das einfache HMM repräsentieren (Abbildung 1.5).

### 1.4.3 *In silico*-Analysen zur Identifizierung von Biosynthese-Genclustern

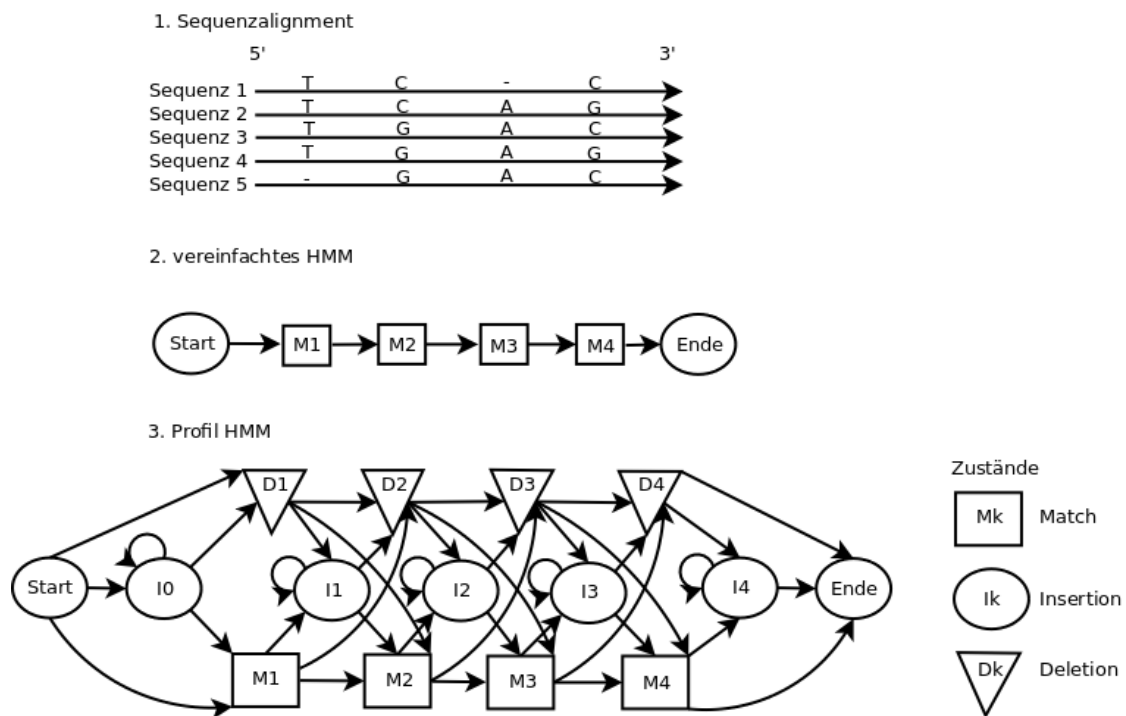


Abbildung 1.5: Entwicklung eines *profile Hidden Markov* Modells. (A) Ausgehend von einem MSA wird eine Konsensussequenz erstellt (B). Das *profile HMM* enthält zusätzliche Wahrscheinlichkeiten zur Berechnung von Insertionen und Deletionen (C). Abbildung in Anlehnung an Yoon (2009) [99]

In einem *alignment* muss nicht jede Sequenz an jeder Position identisch sein. Deshalb werden weitere Wahrscheinlichkeiten in das Modell einkalkuliert, welche zum Beispiel Insertionen (I) oder Deletionen (D) beschreiben. Desweiteren besitzt jeder Übergang von Zustand zum nächsten Zustand eine Wahrscheinlichkeit, die als Transition bezeichnet wird und in Abbildung 1.5 als Pfeil gekennzeichnet ist. Das mathematische Ziel des Algorithmus für die Generierung des *profile HMMs* ist, die maximale Wahrscheinlichkeit aus den möglichen Sequenzabfolgen zu berechnen und in Form einer Häufigkeitsverteilung zu beschreiben. Eine häufig angewendete Methode für diesen Zweck ist der Viterbi Algorithmus [100] [101].

### 1.4.3 *In silico*-Analysen zur Identifizierung von Biosynthese-Genclustern

Biosynthese-Wege für die Produktion von Antibiotika sind genetisch determiniert und in Bakterien sehr häufig in Form eines Genclusters organisiert [102]. Diese Organisation besteht aus einer Co-Lokalisation von essentiellen Genen zur Biosynthese, sowie weiterer Gene, zum Beispiel für die intrinsische Resistenzvermittlung, gegen das produzierte Antibiotikum [103]. Aufgrund der stetig wachsenden Genominformationen hat sich im Bereich der Bioinformatik eine Strategie für die Detektion dieser Gencluster entwickelt, die

### 1.4.3 *In silico*-Analysen zur Identifizierung von Biosynthese-Genclustern

als *genome mining* bezeichnet wird. Im Kontext der genomischen Naturstoff-Forschung sind Programme entwickelt worden, die spezialisierte *profile* HMMs für die Annotation nutzen [104], wie zum Beispiel NP.searcher [105], NaPDos [106] oder PKMiner [107]. Der Unterschied zwischen den Programmen besteht in der Funktionalität. PKMiner ist ausschließlich auf PKS-Typ II Synthasen spezialisiert, NP.searcher und NaPDos können PKS- und NRPS- Gencluster-Typen detektieren. Neben der Strategie, Gencluster zu klassifizieren, gibt es einen weiteren Ansatz, der Gencluster in einem phylogenetischen Kontext untersucht. Dieser Ansatz ist als *evomining* bekannt und findet Anwendung in den Programmen NaPDos sowie dem *Antibiotic Resistant Target Seeker* (ARTS) [103]. Aufgrund von zunehmenden Sequenzinformationen, wird die Strategie der bioinformatischen Identifizierung von Genclustern stetig erweitert. In diesem Zusammenhang wird die *antibiotics & secondary metabolite analysis shell* (*antiSMASH*), in der Version 3, an die Entwicklungen kontinuierlich angepasst und kann derzeit 44 Gencluster-Typen detektieren [108] [109]. Zudem bieten zusätzliche Algorithmen, wie der *knownclusterblast*, die Möglichkeit, detektierte Gencluster hinsichtlich Sequenzähnlichkeit zu vergleichen oder durch den *clusterFinder* Algorithmus [110] [111] putativ weitere Gencluster-Typen in Genomen vorzuschlagen.

## 1.5 Heterologe Genexpression zur Produktion von Naturstoffen

Es ist bekannt, dass nur ein geringer Anteil an Bakterien, von schätzungsweise einem Prozent, unter Standard-Laborbedingungen kultivierbar ist [112]. Deshalb bleibt der Großteil an neuen bakteriellen Ressourcen für die Antibiotika-Forschung verborgen [113] [114]. Um diese Limitierung zu umgehen, ist die heterologe Genexpression als molekularbiologische Methode geeignet und wird in zahlreichen bakteriellen Wirtssystemen angewendet, um durch die gezielte Expression fremder DNS in einem Wirt, Produkte, wie Enzyme und Naturstoffe, zu generieren. Nachgeschaltete Assays zur Identifikation einer bestimmten Eigenschaft, wie beispielsweise eine Hemmhofbildung von Kulturüberständen aus einer Klonbibliothek, sind Bestandteile des funktionalen Ansatzes [115] [116] [117] [118]. Dieses Vorgehen, unter Verwendung von DNS, die nicht aus kultivierbaren Organismen stammt, wird auch als funktionale Metagenomik bezeichnet [119]. Die Abbildung 1.6 illustriert die wesentlichen Schritte in diesem Prozess.

## 1.5 Heterologe Genexpression zur Produktion von Naturstoffen

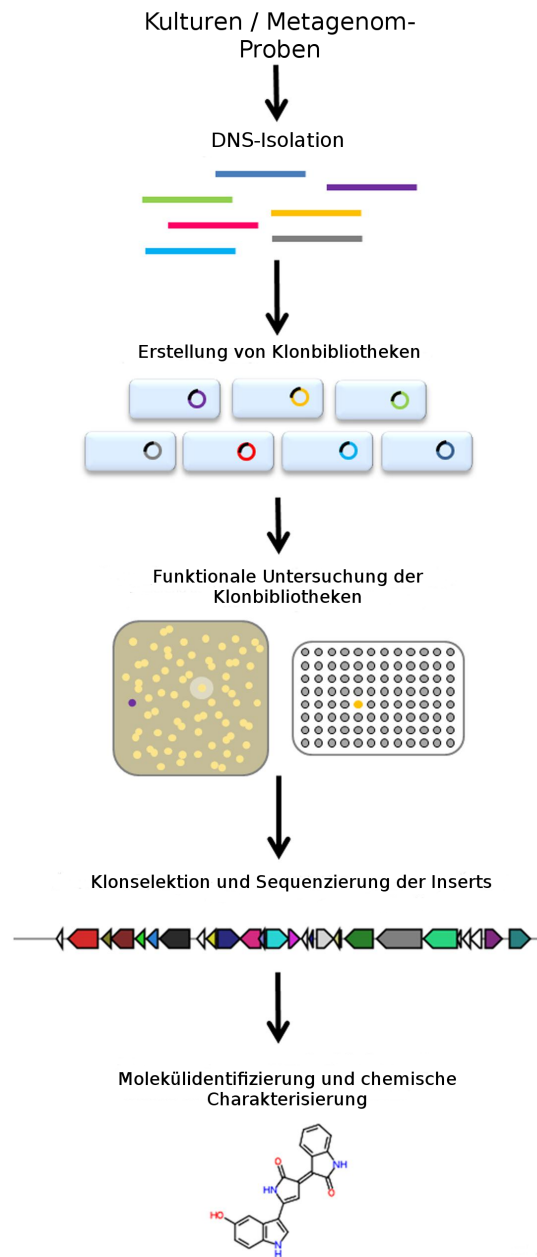


Abbildung 1.6: Prinzipielle Schritte eines klassischen Ansatzes in der funktionalen Metagenomik. Ausgangspunkt ist die Ressource, gefolgt von der DNS-Isolation und der Erstellung von Klonbibliotheken. Diese werden funktional untersucht und selektiert. Ausgewählte Klone werden sequenziert und deren Metabolite charakterisiert. Abbildung in Anlehnung an Trindade *et al.* (2015) [120]

### 1.5.1 Das *Escherichia coli* BAC-System zur Genexpression

Die heterologe Genexpression gilt als Standardmethodik für die Expression einzelner Gene im Größenbereich unter zehn Kilobasen. Für die Genexpression ganzer Biosynthesegencluster zur Produktion eines komplexen Naturstoffs, ist die Methodik jedoch auf bestimmte Wirts-Vector-Systeme beschränkt. Die Gründe hierfür umfassen die Größe von mehreren Kilobasen einzelner Biosynthesegencluster, den einsetzbaren Träger der DNS in Form des Plasmids und deren Wirt. Plasmide sind die Träger der zu exprimierenden DNS und liegen in der Regel episomal im Wirt vor, können aber auch über homologe DNS-Bereiche chromosomal oder durch andere Techniken wie Crispr/Cas9 in das Wirtsgenom integriert werden [121]. Die fremde DNS wird für gewöhnlich durch eine *multiple cloning site* (MCS) in das Plasmid integriert. Plasmide können in unterschiedlicher Anzahl im Wirt vorliegen und werden daher als *low-copy* (1 – 12 Kopien), *medium-copy* (15 – 20 Kopien) oder *high-copy* Plasmide mit >20 Kopien pro Zelle klassifiziert [122]. Die Anzahl richtet sich nach der Art des *origin of replication* (ori) und den zugehörigen Regulatorgenen. Mit Zunahme der Plasmidanzahl wird die Produktion durch gesteigerte Gendosis erhöht, die zusätzlich durch einen starken Promoter reguliert werden kann. Allerdings steigt dadurch auch die metabolische Belastung durch erhöhten Biosyntheseaufwand, wodurch sich das Zellwachstum verlangsamt und der Druck steigt, unnötige DNS abzubauen (Plasmidsegregation). Aus diesem Grund werden für Genexpressions-Experimente, Selektionsmarker verwendet und Vektoren durch eine stringente Kontrolle des Replikationsmechanismus reguliert. Daher liegen Vektoren als Träger von DNS im mehreren Kilobasenbereich häufig in einfacher bis zweifacher Kopie im Wirt vor. Vektoren welche für diese Anforderungen konstruiert worden sind, sind Cosmide, Fosmide, oder *Bacterial Artificial Chromosomes* (BACs). Cosmide sind für die Klonierung von bis zu 48 Kilobasen konzipiert und besitzen *cohesive-end sites* (cos) aus der DNS des Lambdaphagen, geeignet für die *in vitro* Verpackung zur Transfektion von *Escherichia coli* [123]. Durch die geringe Größe des Capsids, ist die Insertlänge für die Transfektion begrenzt. Die Entwicklung von BACs erfolgte im Rahmen des Humangenom-Projekts in den 1990iger Jahren. *Bacterial Artificial Chromosomes* basieren auf dem F-Replicon aus *E. coli* und werden in funktionalen Metagenomstudien verwendet [124]. BACs können hochmolekulare DNS von bis zu 300 Kilobasen tragen [125] [126]. Fosmide bieten die dritte Möglichkeit, Klonbibliotheken mit *E. coli* herzustellen. Diese können, ähnliche wie Cosmide, eine Insertgröße von circa 48 Kilobasen aufnehmen und besitzen wie BACs das F-Replicon von *E. coli*. Da sie aber ursprünglich durch Transfektion und nicht wie BACs mittels Elektroporation in die Zelle transportiert wurden, ist die maximale Insertgröße ähnlich die der Cosmide [127] [128].

### 1.5.1 Das *Escherichia coli* BAC-System zur Genexpression

*Escherichia coli* ist ein Gram-negatives Enterobakterium aus dem Phylum  $\gamma$ -*Proteobacteria*. Das Bakterium wurde ursprünglich aus dem menschlichen Darm isoliert und zu diversen unbedenklichen Laborstämmen wie *E. coli* DH5alpha domestiziert. Das Genom ist vollständig sequenziert und gilt als sehr gut erforscht. Daher werden diverse Stämme sowohl für den Labor-Routinebetrieb, als auch für spezielle Anwendungen, wie etwa als Wirt für große Klonbibliotheken, auf Grund seiner kurzen Generationszeit und der ge-

### 1.5.1 Das *Escherichia coli* BAC-System zur Genexpression

netischen Zugänglichkeit eingesetzt [129] [130]. Bezüglich der genetischen Eigenschaften besitzt *E. coli* sieben Sigmafaktoren für die Aktivierung der Genexpression. Diese lassen sich in die Familie Sigma 70, verantwortlich für die Aktivierung von Haushaltsgenen und der alternativen Sigmafaktorfamilie 54 einteilen [131]. Für die erfolgreiche Genexpression fremder DNS sind weitere genetische Eigenschaften für einen Wirt essentiell. Desweiteren spielt die Fähigkeit des Wirts, fremde DNS in eine richtige Aminosäuresequenz zu übersetzen, eine entscheidende Rolle. In diesem Prozess, der als Translation bezeichnet wird, werden jeweils drei Nukleotide der DNS zu einem Codon zusammengefasst. Durch *tRNAs* werden Codons erkannt und die korrespondierende Aminosäure bereitgestellt [132]. Die Auswahl von *tRNAs* zur Übersetzung der DNS in eine Aminosäuresequenz, wird als *codon usage* bezeichnet [133]. Häufig steht die *codon usage* im Zusammenhang mit dem GC-Gehalt der DNS. Das Genom von *E. coli* besitzt einen GC-Gehalt von 50 % [134]. In Studien zur Genexpression konnte festgestellt werden, dass ein relativ ähnlicher GC-Gehalt von fremder DNS zum Wirt, die Wahrscheinlichkeit der Genexpression erhöht [135] [136]. Auf Ebene der Aminosäuren ist zusätzlich die korrekte Faltung des Proteins entscheidend. Dies kann durch Faltungshelfer (*chaperone*), wie zum Beispiel das DnaK-DnaJ-GrpE oder das GroEL-GroES System, unterstützt werden, um ausgefallene, funktionslose Proteine (*inclusion bodies*) zu vermeiden [137]. Weitere wirtseigene Enzyme, welche die Produktivität heterologer Proteine beeinflussen, sind Proteasen, die die Halbwertszeit der Produkte durch proteolytische Aktivität verringern. Ein Beispiel ist das Stressprotein Lon [138]. Auch die molekulare Masse des gebildeten Proteins spielt eine Rolle hinsichtlich Stabilität. Ein weiterer Faktor ist die Verfügbarkeit von Enzymen für die Aktivierung von Polyketidsynthasen und nicht ribosomale Peptidsynthasen durch Phosphopantethein als posttranslationale Proteinmodifikation. Dieser Schritt erfolgt durch die Phosphopantetheinyl Transferase (PPtase), welche auf Basis ihrer Substrataffinitäten in drei Typen unterschieden wird. Aus dem Stamm *Escherichia coli* K-12 sind drei solcher Enzyme bekannt. In diesem Zusammenhang ist die PPtase EntD von besonderer Bedeutung, welche als Typ II, Polyketidsynthasen und nicht ribosomale Peptidsynthasen aktivieren kann [139]. Weitere wichtige Enzyme sind Glykosyltransferasen, die beispielsweise das Aglykon zu einem bioaktiven Naturstoff modifizieren, wie am Beispiel der Erythromycin-Biosynthese bekannt ist [140]. Viele der oben beschriebenen Faktoren zur heterologen Genexpression wurden in *E. coli* erstmalig untersucht, wodurch dieses Bakterium als weitreichend erforscht gilt. Dies lässt sich auch an der Anzahl kommerziell erhältlicher Stämme für unterschiedlichste Zwecke festmachen, wie beispielsweise der elektrokompetente Stamm *E. coli* DH10B, optimiert für die Erstellung von Klonbibliotheken [141]. Aus diesem Grund wurde auch in dieser Forschungsarbeit zur die Methodenetablierung auf *E. coli* zurückgegriffen.

## 1.6 Ziele dieser Doktorarbeit

Diese Arbeit ist innerhalb der Forschung für die Fraunhofer-Sanofi Kooperation *Natural Product Center of Excellence* entstanden mit dem Ziel, einen (meta-)genombasierten Ansatz für die Produktion neuer antibiotisch-wirksamer Naturstoffe zu etablieren.

Dieser Ansatz erschloss sich aus der Hintergrund-Problematik neuartige chemische Leitstrukturen für die Entwicklung von Antibiotika zu finden, die nicht durch konventionelle Methoden wie der Kultivierung entdeckt werden können. Daraus ergab sich die erste wissenschaftliche Fragestellung, alternative Bakterien als neue Ressource gegenüber Actinobakterien, für die Biosynthese von Antibiotika zu identifizieren. Hierzu wurden Genomdaten von wenig erforschten Bakterien verwendet und diese bioinformatisch nach ausschlaggebenden Kriterien untersucht. Kriterien für eine Priorisierung waren Anzahl an Genclustern, Gencluster-Typ, Resistenzgene und Sequenzähnlichkeit. Dabei wurde sowohl auf publizierte Daten, als auch auf eine selbstständig durchgeführte Sequenzierung und funktionale Annotation des Acidobakteriums *Acidicapsa borealis* zurückgegriffen. Als zweite Fragestellung an diese Forschungsarbeit schließt sich die molekularbiologische Methodik an. Aufgrund der Problemstellung, der schwierigen Kultivierung neuer Bakterien unter Standard-Laborbedingungen, wurde in dieser Arbeit ein Protokoll zur Konservierung und heterologen Genexpression etabliert, um neue genetische Ressourcen kulturunabhängig zu erschließen.

Gegenstand der bioinformatischen Untersuchung waren Stämme bakterieller Phyla, die in der aktuellen Literatur, beispielsweise von Cimermancic *et al.* (2014) [73] zur Antibiotikaforschung unterrepräsentiert sind, wie *Acidobacteria*, *Bacteroidetes*, *Chloroflexi* und des Superphylums *Planctomycetes*, *Verrucomicrobia*, und *Clamydiae* (PVC). Für die Analyse wurden bestehende *state of the art* Programme wie *antibiotics & secondary metabolite analysis shell* (*antiSMASH*) und *Antibiotic Resistant Target Seeker* (ARTS) verwendet. Als Ausgangspunkt für die Etablierung der Methodik von Klonbibliotheken wurden die bioinformatischen Ergebnisse zur Stamms Selektion berücksichtigt. Neben der Verwendung von zwei Kulturen, wurde die Etablierung eines Metagenomansatzes aus einem Termitennest als kultur-unabhängige Methode realisiert.

Um die Klonbibliotheken funktional hinsichtlich Konservierung von Genclustern und antimikrobieller Metabolite zu untersuchen, war neben der Etablierung die Integrierung der Klonbibliotheken in einen bestehenden Routinebetrieb zur Detektion von antibakterieller Aktivität ein weiteres Ziel. Dieser Prozess bestand aus Methoden, wie der PCR zur DNS-basierten Identifikation von Gencluster-Fragmenten und einer Biolumineszenz-basierter Untersuchung. Als dritte Methode erfolgte die Extrakterstellung erstellter Klonbibliotheken, mit dem weitere Bakterien auf antibakterielle Aktivität getestet wurden.

Der gesamte Prozess, beginnend bei der bioinformatischen Identifizierung von Bakterien, über die Erstellung der Klonbibliotheken bis zur Untersuchung mit chemischen Analytik, ist in dieser Arbeit als Methodenetablierung anzusehen, die in der Gesamtheit eine Arbeitsgrundlage für weitere funktionale (Meta-)Genomansätze bietet.

## 2 Material und Methoden

### 2.1 Bioinformatische Untersuchung von Biosynthese-Genclustern in Bakteriengruppen

Die bioinformatische Untersuchung wurde durchgeführt, um alternative Bakteriengruppen hinsichtlich Biosynthese-Gencluster zu identifizieren und zu priorisieren. Abbildung 2.1 zeigt den schematischen Ablauf der bioinformatischen Analyse. Diese basiert auf vollständigen und unvollständigen (*draft*) Genomsequenzen. Es wurden sieben Genomsequenzen aus Acidobakterien der Stämme *Bryocella elongata*, *Terriglobus* sp. L56, *Granulicella* sp. S137, *Terracidiphilus* sp. S55, *Granulicella* sp. L46, *Silvibacterium* sp. S15 und *Acidobacterium* sp. S8 verwendet, die nicht öffentlich zugänglich sind und auf der DVD zur Dissertation hinterlegt sind. Die öffentlichen Genomdaten wurden über das *Portal National Center for Biotechnology Information* (NCBI) bezogen. Für den Bezug dieser Sequenzdaten wurde das Script *Datenbankabfrage.py* sowie *Datenbankabfrage-draft.py* geschrieben, welche mit Python Version 2.7 ausgeführt wurden. Die Skripte enthalten eine Suchzeile (*search\_term*) zur Eingabe des Suchbegriffs, der durch weitere Operatoren wie *[Title]* und *[prop]* eingegrenzt wurde (Quellcode).

```
from Bio import Entrez
from Bio import SeqIO
import urllib , urllib2
from urllib2 import HTTPError

search_term = '(actinobacteria [Organism] _\ \
AND_complete_genome [ Title ] _\ \
AND_srcdb_ddbj/embl/genbank [ prop ] _\ \
NOT_plasmid [ Title ] _NOT_segment [ Title ] ) '
```

Als Sucheinträge wurden eingegeben: *Acidobacteria*, *Bacteroidetes*, *Chloroflexi*, *Verrucomicrobia*, *Planctomycetes*, *Chlamydiae* und als Referenzphylum *Actinobacteria*. Durch den Operator *srcdb* wurden die Datenbanken *DNA data bank of Japan* [142], *EMBL Nucleotide Sequence Database* [143] und die Datenbank *GenBank* [144] abgefragt. Über das Suchinterface *Entrez* verlief die Kommunikation von Suchanfrage und NCBI [145]. Bei erfolgreicher Suchanfrage erfolgte der Datenaustausch von DNS-Sequenzen welche in eine Fastadatei geschrieben wurden.

Im nächsten Schritt erfolgte die Formatierung der Fastadateien durch das Skript *makeDB.py*. Durch diesen Schritt wurde die Genomgröße pro Stamm berechnet, sowie weitere Informationen wie Stammbezeichnung und NCBI-*accession* in eine separate Datei

### 2.1.1 Statistische Analysemethoden

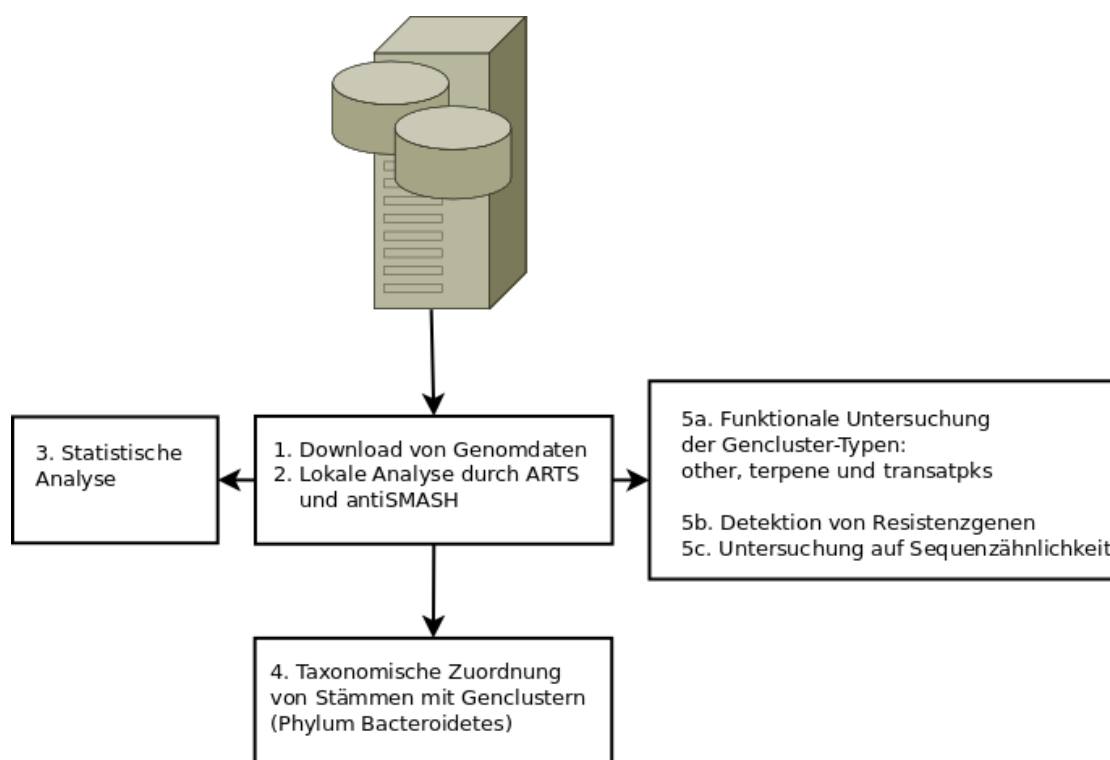


Abbildung 2.1: Ablauf der bioinformatischen Detektion und Analyse der Gencluster. Ausgangspunkt war die Datenbank-Akquise, gefolgt von der Identifizierung der Gencluster. Es folgte die statistische, funktionale und taxonomische Analyse

geschrieben.

Die anschließende Detektion von Genclustern erfolgte mit *antiSMASH* Version 3.0.5 [146]. Die Informationen der Gencluster wurden anschließend in zwei Dateien vereint und tragen die Bezeichnungen *Datenbank-genome* und *Datenbank-draft*. Die Phyla *Planctomycetes*, *Verrucomicrobia* und *Chlamydiae* wurden nachträglich als das Superphylum *PVC* zusammengefasst [147]. Die Dateien wurden als Datengrundlage zur statistischen und taxonomischen Analyse verwendet (Abbildung 2.1). Zur Detektion von Resistenzgenen wurde das Programm *Antibiotic Resistant Target Seeker* verwendet (ARTS) [103].

#### 2.1.1 Statistische Analysemethoden

Die Grundlage der statistischen Analysen basierte auf den zuvor erstellten Dateien *Datenbank-genome* und *Datenbank-draft*. Die Visualisierung der Boxplots sowie die Korrelationsanalyse nach *Pearson* erfolgte mit Python [148] [149]. Zur Einteilung der kategorischen Variablen überdurchschnittlich und unterdurchschnittlich diente die Regressionsanalyse und die errechnete Steigung der Geraden von 0,88. Zur Untersuchung auf statistische Signifikanz wurde R (Version 3.3.1) verwendet. Die *rarefaction* Analyse wur-

### 2.1.2 Funktionale Annotation ausgewählter Biosynthese-Gencluster

de nach dem Ansatz von Cao *et al.* berechnet und basierte auf den Gencluster-Typen und deren Abundanzen [150].

### 2.1.2 Funktionale Annotation ausgewählter Biosynthese-Gencluster

Nach der Detektion der Gencluster-Typen wurden, gezielt die Typen *other* und *terpene* für eine ausführliche Genannotation aus dem Datensatz herausgefiltert und in eine separate Fastadatei umgeschrieben. Die Annotation der Gencluster-Typen *terpene* und *other* erfolgte durch die mikrobielle Annotations-Plattform (RAST) [151].

Die Gencluster des Typs *trans*-AT PKS I aus *Acidobacterium* sp. S8 (*t1pks-transatpks-nrps*), *Terracidiphilus* S55 (*t3pks-nrps-transatpks-t1pks*) und *Bryocella* sp. S190 (*transatpks*), wurden durch Prokka annotiert [152]. Die Gene mit *trans*-AT Domänen wurden hervorgehoben und durch einen Proteinblast auf Sequenzähnlichkeit untersucht. Im Anschluss erfolgte die Sequenzextraktion der korrespondierenden Domänen. Diese wurden durch ein *alignment* mittels *Muscle* [153] verglichen und konservierte Sequenzmotive durch den *MViewer* [154] und Weblogo 3 [155] visualisiert.

### 2.1.3 Taxonomische Analyse von Stämmen der *Bacteroidetes* mit Biosynthese-Gencluster

Zur taxonomischen Analyse wurde das Phylum *Bacteroidetes* ausgewählt und alle Stämme die Gencluster enthalten verwendet. Die NCBI-*accession* diente zur Datenbankabfrage durch das R-Paket *taxonomizr*, wodurch die entsprechende Taxonomie bestimmt wurde. Die Ergebnisse dieser Untersuchung wurden quantifiziert und hierarchisch dargestellt. Eine tabellarische Darstellung dieser Untersuchung befindet sich in der Datei *Bacteroidetes\_gencluter.csv*.

## 2.2 Genomsequenzierung von *Acidicapsa borealis*

Zur Sequenzierung des Genoms wurde eine 500 mL Kultur gemäß Abschnitt 2.4.1 angefertigt, aliquotiert und zur DNS-Isolation herangezogen. Die Isolierung hochmolekularer DNS erfolgte gemäß des VWR peqGOLD Protokolls. Es wurde eine DNS-Gesamtmenge von 5 µg isoliert.

### 2.2.1 Sequenzierung durch das *Oxford Nanopore MinION*-System

Zur Präparation für die Sequenzierung wurde die DNS mittels Covaris *g-TUBE* größenfragmentiert und durch den *NEBNext FFPE Repair Mix* zu *blunt-end* DNS repariert. Im Anschluss erfolgte eine Adenylierung, wodurch die Ligation von Sequenzieradapter mittels *Ligation sequencing Kit 1D* ermöglicht wurde. Zur Sequenzierung wurde 3,5 µg DNS auf die *SpotON Flow Cell* eingesetzt und die Sequenzierung dokumentiert durch das *MinKNOW* Programm Version 1.7.7.

Das Programm *Poretools* [156] wurde ausgeführt, um das FAST5-Format in das FASTQ-Standardformat zu konvertieren. Aus der FASTQ-Datei wurden zur Assemblierung des Genoms alle *reads* mit einer Länge von weniger als 1.000 Basen entfernt. Abbildung 3.9 gibt die Größenverteilung der restlichen *reads* wieder.

### 2.2.2 Sequenzierung durch das *Illumina-HiSeq* System

Die Sequenzierung mittels Illumina erfolgte durch die Firma Macrogen, Amsterdam. Dazu wurde eine Illumina *TruSeq DNA PCR-free library* erstellt.

Der HiSeq-Sequenzierungslauf erfolgte von 2x 100 Basenpaaren unter Generierung von 50 Millionen *reads*. Um die anschließende Assemblierung effizient durchzuführen, wurden zuvor die *paired-end reads* durch das Skript *bbnorm* (*bbtools*) <https://jgi.doe.gov/data-and-tools/bbtools/> auf eine 40x Sequenzabdeckung normalisiert.

### 2.2.3 *De-novo* Hybridassemblierung

Die *de novo* Genomassemblierung von *Acidicapsa borealis* erfolgte unter Verwendung von *reads* aus der *Oxford nanopore* und der Illumina Sequenzierung, durch das Programm *Unicycler* [90]. Die Assemblierung durch die Konstruktion eines de-Bruijn Graphen, erfolgte automatisch durch das Programm *SPAdes* [157]. Nachfolgende Schritte, wie das *polishing* des Graphen, wurden durch die Programme *Pilon* [158] und *Racon* [159] automatisch durchgeführt.

## 2.2.4 Sequenzanalyse des Genoms von *Acidicapsa borealis*

Zur Berechnung der Sequenzabdeckung (*coverage*) des assemblierten Genoms wurde die Formel 2.1 angewendet [160].

$$coverage = \frac{(Readlaenge \times Readanzahl)}{Genomgroe\beta e} \quad (2.1)$$

### Funktionale Annotation

Die vollständige Annotation des Genoms erfolgte durch die Plattform *Genix* [161]. Diese umfasste eine Genvorhersage durch Prodigal, sowie eine Vorhersage von *tRNA* durch Aragon. Die anschließende Annotation auf Basis der Aminosäuresequenzen erfolgte anhand von Proteinen der UniProt *KnowledgeBase* (UniprotKB)-Datenbank [162]. Es wurden alle verfügbaren Proteine des Phylums *Acidobacteria* (Uniprot-ID: 57723) verwendet. Für den Sequenzvergleich wurde ein BLASTp *eval* kleiner als 0.0005 gewählt.

Die nachfolgende Annotation durch die *Kyoto encyclopedia of genes and genomes* (KEGG) [163] und der Genontologien, erfolgte auf Basis der Uniprot-Einträge (IDs). Diese wurden für die jeweilige Analyse durch Uniprot umgewandelt [164]. Für die KEGG Analyse ergaben sich aus 4.560 Uniprot-IDs, 7.595 KEGG-IDs sowie 7.658 GO-IDs für die Untersuchung der Genontologien. Anhand der KEGG-IDs erfolgte das *mapping* auf den Stärke- und Saccharose-Metabolismus. Für die Untersuchung der Genontologien wurde die Plattform REVIGO verwendet [165]. Es erfolgte eine Normalisierung der Einträge unter Berücksichtigung der Semantik (Resnik).

Eine weitere Untersuchung auf Enzymfamilien des Carbohydrat Metabolismus erfolgte durch die Datenbank *Carbohydrate-Active enZymes Database* (CAZy), Version CAZyDB.07202017 und dem HMMER-Paket Version 3.1 [166] [167].

Zur Identifizierung von Genclustern, Sequenzähnlichkeit sowie Resistenzgenen wurden die Programme *antiSMASH* [146] und *Antibiotic Resistant Target Seeker* (ARTS) [103], jeweils in der Version 3 angewendet.

## 2.3 **Verwendete Labormaterialien**

Sofern nicht anders erwähnt, wurden für alle Puffer und Lösungen autoklaviertes MilliQ-Wasser verwendet. Medien wurden mit MilliQ-Wasser angesetzt und anschließend autoklaviert.

### 2.3.1 **Verwendete Medien zur Kultivierung**

Tabelle 2.1: Komponenten für LB-Medium

Komponente	Menge
Hefeextrakt	5 g/L
Trypton	10 g/L
Natriumchlorid	5 g/L
Agar (optional)	15 g/L

Tabelle 2.2: Komponenten für R2A

Komponente	Menge
Hefeextrakt	0,5 g/L
Pepton	0,5 g/L
Casaminosäure	0,5 g/L
Glukose	0,5 g/L
Lösliche Stärke	0,5 g/L
Na-Pyruvat	0,3 g/L
Di-Kaliumhydrogenphosphat	0,3 g/L
Magnesiumsulfat x H <sub>2</sub> O	0,05 g/L
MES-Hydrat	1,95 g/L
Agar (optional)	15 g/L

### 2.3.2 Verwendete Verbrauchsmaterialien

### 2.3.2 Verwendete Verbrauchsmaterialien

Tabelle 2.3: Verwendete Verbrauchsmaterialien

Bezeichnung	Hersteller	Artikelnr.
Elektroporations-Küvetten	<i>Thermo Fisher</i>	12338212
Membranfilter 0,1 mm	<i>Fisher Scientific</i>	10526902
<i>Hard-Shell</i> 96-well Platten	Bio-Rad	HSP-9631
96-well Mikrotiter-Platten <i>clear/w</i>	<i>Fisher Scientific</i>	10601442
384-well Mikrotiter-Platten	<i>Fisher Scientific</i>	10497092
96-well Verschlussmatten	<i>Fisher Scientific</i>	10433382
2,2 mL <i>deep-well</i> Platten	<i>Fisher Scientific</i>	10060072
<i>sandwich cover</i> für 96 <i>deep-well</i> Platten	<i>Kuhner Shaker</i>	SMCR1296b
50 mL Zentrifugengefäße	<i>Fisher Scientific</i>	10788561
96-well Versiegelungsfolie	<i>Fisher Scientific</i>	10595001
2 mL <i>safe-lock</i> Zentrifugengefäße	<i>Fisher Scientific</i>	10318661
<i>g-Tube</i>	Covaris	010145
Glasware	Schott	-
0,2 µm Spritzen-Filter	Corning	14831

### 2.3.3 Verwendete Stämme

#### 2.3.3 Verwendete Stämme

Tabelle 2.4: Verwendete Stämme in dieser Arbeit

Stamm	Genotyp	Phylum / Herkunft	Bezug
<i>Acidicapsa borealis</i>	Wildtyp	<i>Acidobacteria</i>	DSMZ
<i>Terracidiphilus</i> sp. S55	Wildtyp	<i>Acidobacteria</i>	Petr Baldrian Kooperation
<i>Escherichia coli</i> MegaX DH10B™ T1R Electrocomp	F- <i>mcrA</i> $\Delta(mrr-hsdRMS-mcrBCmrr-hsdRMS-mcrBC)$ $\Phi 80lacZ \Delta M15 \Delta lacX74$ <i>recA1 endA1 araD139 \Delta(ara, leu)7697</i> <i>galU galK \Delta-rpsL nupG tonA</i>	<i>Proteobacteria</i>	Thermo Fisher
<i>Pseudomonas aeruginosa</i> PAO 750	$\Delta mexAB-oprM, \Delta mexCD-oprJ,$ $\Delta mexEF-OprN$ $\Delta mexJK, \Delta mexXY, \Delta oprM$	<i>Proteobacteria</i>	Sanofi
<i>Escherichia coli</i> ATCC 25922	Wildtyp	<i>Proteobacteria</i>	ATCC
<i>Escherichia coli</i> $\Delta tolC$	$\Delta tolC$	<i>Proteobacteria</i>	Sanofi
Methicillin-sensitiver <i>Staphylococcus aureus</i> ATCC 25923	Wildtyp	<i>Proteobacteria</i>	ATCC

### 2.3.4 Liste an verwendeten Kits

#### 2.3.4 Liste an verwendeten Kits

Tabelle 2.5: Verwendete Kits für die Molekularbiologie

Bezeichnung	Hersteller	Artikelnr.
<i>Bacteria cell counting kit</i>	<i>Thermo Fisher</i>	B7277
<i>NucleoSpin Soil</i>	Macherey-Nagel	740780.10
<i>Large-Construct Kit</i>	Qiagen	12462
Plasmid Maxi Kit	Qiagen	12162

#### 2.3.5 Komponenten für die Molekularbiologie

Tabelle 2.6: Verwendete Enzyme für die Molekularbiologie

Bezeichnung	Hersteller	Artikelnr. oder CAS
<i>T4 DNA Ligase</i>	<i>Thermo Fisher</i>	EL0014
<i>Dream Taq DNAP</i>	<i>Thermo Fisher</i>	EP0701
dNTPs	<i>Thermo Fisher</i>	10297018
Lysozym	Sigma-Aldrich	12650-88-3
Achromopeptidase	Sigma-Aldrich	123175-82-6
Proteinase K-Lösung	Qiagen	19133
BamHI	<i>Thermo Fisher</i>	ER0051
<i>FastDigest</i> BamHI	<i>Thermo Fisher</i>	FD0055
Alkalische Phosphatase	<i>Thermo Fisher</i>	EF0652
Agarase	<i>Thermo Fisher</i>	EO0461
<i>FFPE DNA Repair Mix</i>	NEB	M6630S
<i>5' DNA Adenylation Kit</i>	NEB	E2610S
Pyrogenfreies Wasser	Roth	3255.1
<i>Ligation Sequencing Kit 1D</i>	<i>Oxford Nanopore</i>	SQK-LSK108

Tabelle 2.7: Verwendete *primer* in dieser Arbeit

Bezeichnung	Orientierung	Nukleotidsequenz	Quelle
A3-NRPS	vorwärts	GCSTACSYSATSTACACSTCSGG	[168]
A7R-NRPS	rückwärts	SASGTCVCCSGTSCGGTAS	[168]
M13rev2	vorwärts	GAGTTAGCTCACTCATTAGG	LGC
pBeloR	rückwärts	TGTGCTGCAAGGCGATTAAG	LGC
U341	vorwärts	CCTACGGGRSGCAGCAG	LGC
U805	rückwärts	GGACTACVSGGGTATCTAAT	LGC

## 2.3.6 Anfertigung von Puffern und Stammlösungen

### Herstellung von Antibiotika-Stammlösungen

Die Herstellung von antibiotischen Stammlösungen umfasste die Einwaage von Ampicillin, Chloramphenicol und Gentamycin in pulverform. Als Ampicillin-Stocklösung wurden 100 mg eingewogen und in 1 mL H<sub>2</sub>O gelöst. Die Herstellung der Gentamycin-Stammlösung erfolgte analog. Für die Herstellung der Chloramphenicol-Stocklösung wurde 10 mg Chloramphenicol eingewogen und in Ethanol gelöst. Alle Lösungen wurden durch einen 0,2 µm Spritzenfilter sterilfiltriert und bei -20 °C für maximal zwei Wochen gelagert.

### Anfertigung von Puffern

Tabelle 2.8: Komponenten für 50x TAE

Komponente	Menge
100 % Essigsäure	57,1 mL/L
0,5 M EDTA	100 mL/L
Tris-Base	242 g/L

Tabelle 2.9: Komponenten für 10x TBE

Komponente	Menge
Tris-Base	108 g/L
Borsäure	55 g/L
EDTA	7,5 g/L

Tabelle 2.10: Komponenten für den Partialverdau-Puffer

Komponente	Menge oder Konzentration
10x Tango Puffer oder BamHI	1,6 mL
0,1 M Spermidinlösung	600 µL
H <sub>2</sub> O	14,4 mL
BamHI	25 <i>Units</i> /500 µL Probe

### 2.3.6 Anfertigung von Puffern und Stammlösungen

#### Anfertigung sonstiger Lösungen

Tabelle 2.11: Komponenten der Lysozymlösung

Komponente	Menge oder Konzentration
Kaliumdihydrogen-Phosphat, pH 6,2	66 mM
Lysozym	1 mg/mL

Tabelle 2.12: Komponenten der Achromopeptidase-Lösung

Komponente	Menge oder Konzentration
Tris-HCl, pH 8	10 mM
NaCl	10 mM
Achromopeptidase	1 mg/mL

Tabelle 2.13: Komponenten der Zelllyse-Lösung

Komponente	Menge oder Konzentration
10 % (w/v) SDS	800 µL
10 % (w/v) Sarkosyl	800 µL
H <sub>2</sub> O	14,4 mL
Qiagen Proteinase-K Lösung	160 µL

Tabelle 2.14: Stopplösung I

Komponente	Menge oder Konzentration
10x Tango Puffer oder BamHI	1,6 mL
0,1 M PMSF	200 µL
Isopronanol	19,8 mL

## 2.4 Bereitstellung von Bakterien für die Anfertigung von Klonbibliotheken

### 2.4.1 Anzucht von *Acidicapsa borealis* und *Terracidiphilus* sp. S55

Der Stamm *Acidicapsa borealis* wurde aus einer eingefrorenen (-80 °C) Glycerinlösung (50 % (v/v)) entnommen und in 500 mL R2A (pH 5,5 eingestellt mit 5 M HCl), bei 80 rpm und Raumtemperatur (RT) kultiviert. Die Zellzahl wurde mit Hilfe einer Thoma-kammer von 0,1 mm Tiefe und 0,0025 mm<sup>2</sup> je Kleinquadrat berechnet. Um das Volumen eines Großquadrats ( $V_G$ ) zu ermitteln, wurde das Volumen eines Kleinquadrats mit 16 multipliziert. Für beide Stämme wurden fünf Großquadrate ausgezählt und der Mittelwert berechnet ( $\bar{x}$ ). Anschließend wurde nach Formel 2.2 die finale Zellzahl pro Milliliter berechnet.

$$\left(\frac{\bar{x}}{V_G}\right) \times 1000 = \text{Zellzahl/mL} \quad (2.2)$$

Für *Acidicapsa borealis* wurde eine Zellzahl von  $1,52 \times 10^7$  Zellen/mL bei Ernte berechnet. Analog dazu erfolgte die Kultivierung von *Terracidiphilus* sp. S55, die Zellzahl betrug bei der Ernte  $3,94 \times 10^6$  Zellen/mL.

### 2.4.2 Gewinnung von Bakterien aus dem Termitennest

Für die Gewinnung der Bakterien aus dem Termitennest von *Coptotermis niger* wurde eine Nykoden-Dichtezentrifugation angewendet, die aus folgenden Schritten bestand:

- Einwaage von 2 g Termitennest und lösen in 10 mL H<sub>2</sub>O
- Dreimal für zwei Sekunden turaxieren (9.500 1/min)
- Termitennest-Suspension unterschichten mit 60 % (w/v) Nykoden-Lösung
- Zentrifugieren bei 4.000 x g für 60 Minuten mit einem A-4-81 Ausschwingrotor (Eppendorf Zentrifuge 5810R), bei 4 °C
- Entnahme der Zellschicht

### 2.4.3 Zellzahlbestimmung gewonnener Bakterien

Um die Zellzahl der mittels Nykoden-Lösung getrennten Zellschicht zu bestimmen, wurde eine 1 zu 80 Verdünnung angefertigt und die Zellen mittels *Bacteria cell counting Kit* (*Thermo Fisher*) nach Herstellerangaben gefärbt. Die Messung wurde mit einem BD FACScalibur durchgeführt. Als Referenz wurden *beads* auf eine Zellzahl von  $1 \times 10^6$  Zellen/mL verdünnt und zur Probe hinzugeführt. Im BD *Cell Quest* Programm wurde eine Grafik erstellt und die Probe durch ein *gate* anhand der Fluoreszenz-Intensität (FL1) und der Signalgröße im *forward side scatter* (FSC) definiert. Es wurden 15 µL Probe vermessen und die *events* jeweils für extrahierte Bakterien und *beads* gezählt. Anhand der Formel 2.3 und unter Berücksichtigung der Verdünnung wurde die finale Zellzahl

#### 2.4.4 Taxonomische Klassifizierung gewonnener Bakterien aus dem Termitennest

berechnet. Anschließend wurde das Material bei 9.933 x g für zwei Minuten pelletiert und für die DNS-Isolation eingesetzt.

$$\frac{Events_{Zellen}}{Events_{Beads}} \times 10^6 \frac{Zellen}{mL} = \text{gewonnene Zellen} \frac{Zellen}{mL} \quad (2.3)$$

#### 2.4.4 Taxonomische Klassifizierung gewonnener Bakterien aus dem Termitennest

Im Anschluss an die Gewinnung der Zellen aus dem Termitennest erfolgte die DNS-Isolation für die taxonomische Klassifizierung. Dazu wurde das MN *Nucleospin-Kit Soil* verwendet und nach Herstellerangaben gearbeitet. Es wurde 294 mg Zellmaterial für die DNS-Isolation verwendet. Die Sequenzierung der hypervariablen Region 3 des 16S rRNA Gens erfolgte durch die *primer* U341/U806 (Tabelle 2.7) und wurde von der Firma LGC durchgeführt. Es erfolgte eine Illumina-Sequenzierung von 2,5 Millionen *paired-end reads* durch den MiSeq V3.

Ausgangspunkt der Analyse waren 120.358 *paired-end reads*. Die Auswertung erfolgte durch die Firma Ribocon. Die Schritte der Auswertung sind in Tabelle 2.15 zusammengefasst. Im ersten Schritt der Auswertung wurden alle *reads* hinsichtlich einer validen 16S rRNA-Gensequenz durch ein *alignment* verifiziert. Sequenzen, welche als uneindeutig klassifiziert worden sind, wurden entfernt. Im zweiten Schritt der Prozessierung erfolgte eine weitere Filterung der *reads* hinsichtlich Qualität und Anzahl an Homopolymeren. Im dritten Schritt wurden die verbliebenen *reads* quantifiziert auf Basis einer exakten Sequenzübereinstimmung. Im nächsten Schritt wurden alle *reads* auf Basis einer 98 prozentigen Sequenzübereinstimmung zu *operational taxonomic units* (OTUs) zusammengefasst. Die längste Sequenz diente jeweils als Referenz und wurde für die taxonomische Klassifizierung verwendet. Diese erfolgte durch einen kombinierten Ansatz aus BLAST und der SILVA SSU oder LSU Ref-Datenbank. Sequenzen mit einer Übereinstimmung weniger als 93 % wurden als unklassifiziert vermerkt. Durch die taxonomisch klassifizierten Referenzsequenzen wurden die korrespondierenden OTUs ebenfalls klassifiziert.

#### 2.4.4 Taxonomische Klassifizierung gewonnener Bakterien aus dem Termitennest

Tabelle 2.15: Sequenzanalyse erstellter *reads* anhand der SILVAngs 1.3 *pipeline* [169]

Modul	Programm	Aufgabe
<i>align</i>	Sina Sina v1.2.10	<i>alignment</i> gegen die SILVA SSU rRNA SEED
<i>quality</i>	Sina Sina v1.2.10	Filterung von <i>reads</i> mit >2 % Homopolymeren, >2 % Uneindeutigkeit, <300 bp
<i>dereplication</i>	cd-hit-est [170]	Suche nach identischen <i>reads</i>
<i>clustering</i>	cd-hit-est	Zusammenfassung von Sequenzen mit mind. 98 % Identität
<i>classification</i>	blastn + Silva SSU Datenbank Version 128	Taxonomische Annotation
<i>visualisation</i>	Krona [171]	Taxonomische Zuordnung der Daten mit entsprechender Abundanz

## 2.5 Erstellung von Klonbibliotheken mit *Escherichia coli* als Wirt

Für die Erstellung der Klonbibliotheken mit *E. coli* wurde der Stamm MegaX DH10B T1 Electrocomp verwendet (Tabelle 2.4). Als BAC wurde pBeloBAC11 von *New England Biolabs* bezogen und ist in Abbildung 2.3 dargestellt. Abbildung 2.2 stellt die Schritte zur Erstellung der Klonbibliothek vor.

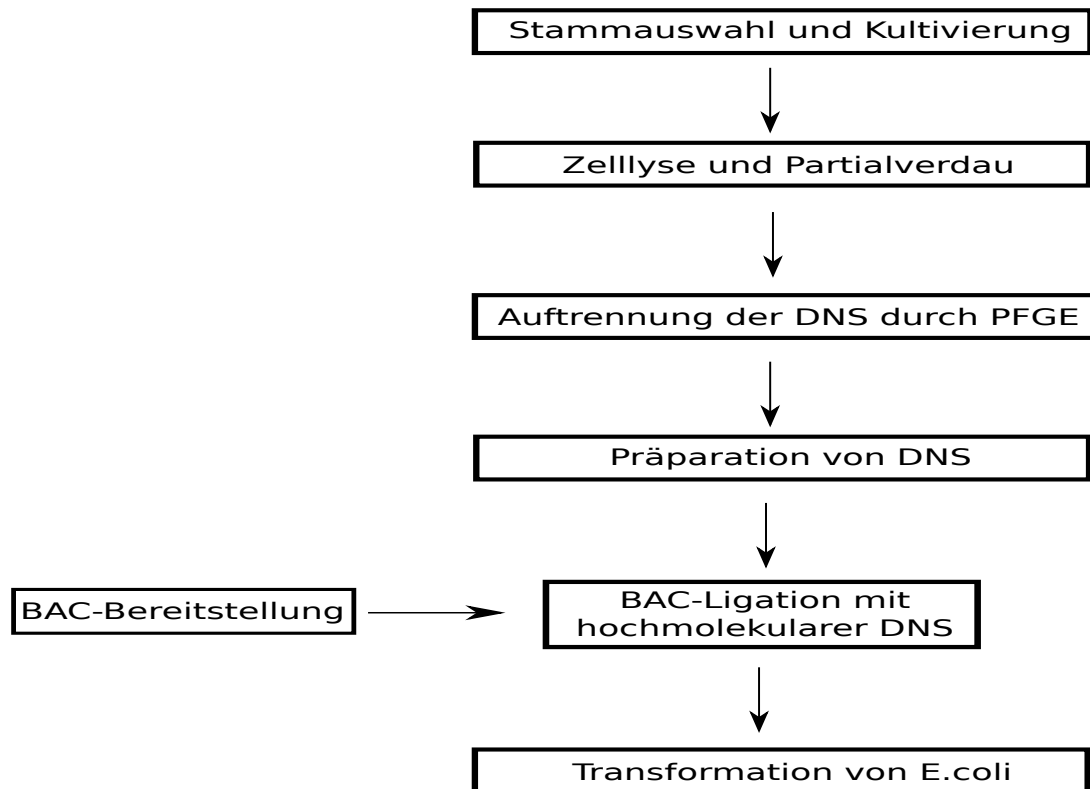


Abbildung 2.2: Ablauf der Erstellung von Klonbibliotheken mit *E. coli* MegaX DH10B T1 Electrocomp als Wirt

## 2.5 Erstellung von Klonbibliotheken mit *Escherichia coli* als Wirt

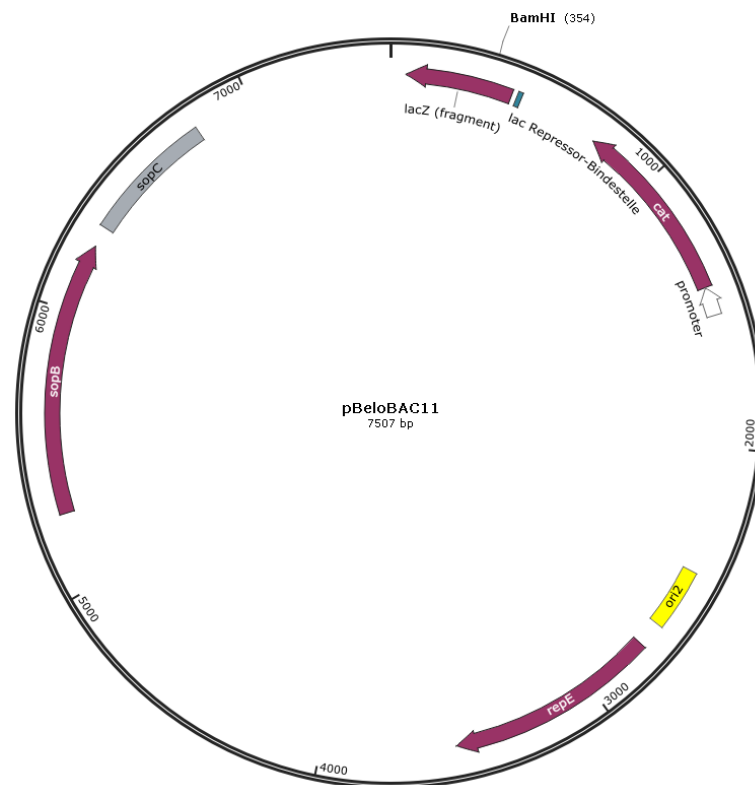


Abbildung 2.3: Plasmidkarte von pBeloBac11 für die Erstellung von Klonbibliotheken. Die BamHI-Schnittstelle wurde für die Klonierung genutzt. Gene; *lacZ*-*pha*: Markergen für die Blau-Weiß-Selektion; *cat*: Chloramphenicol-Acetyltransferase, vermittelt die Resistenz; *repE*: kodiert für das Replikationsprotein RepE; *sopB*: kodiert für das Plasmidpartitions-Protein; *sopC*: Partitionsregion, *ori*: Replikations-Initiierung

Die Isolation von pBeloBAC11 aus *E. coli* K12 ER2420 [pBeloBAC11] erfolgte durch das Qiagen Maxi Prep Kit nach Herstellerangaben. Dazu wurde eine 500 mL LB-Kultur + 12,5 µg/mL Chloramphenicol mit einer Kolonie angeimpft und für 16 Stunden bei 37 °C und 220 rpm kultiviert. Für die Klonierung des BACs wurde die BamHI-Schnittstelle gewählt. Zum Verdau wurde die alkaline Phosphatase hinzugegeben um die Rezirkularisierung zu vermeiden (Tabelle 2.16). Der Verdau erfolgte über vier Stunden bei 37 °C. Zur Kontrolle der Linearisierung wurde der BAC nach jedem Verdau mittels Gel-Elektrophorese (1 % (w/v) LE-Agarose gelöst in 1x TAE, 60 V, 1 Stunde, Standard: 5 µL GeneRuler 1 Kb Marker, 0,13 µL Ladepuffer je µL Probe) + Gelextraktion aufgereinigt.

## 2.5 Erstellung von Klonbibliotheken mit *Escherichia coli* als Wirt

Tabelle 2.16: Komponenten für die Linearisierung von pBeloBAC11 als Vorbereitung auf die Klonierung

Komponente	Menge
1 µg pBeloBAC11	3,33 µL
FD-BamHI	1 µL
10x FD-Puffer	1 µL
Alkaline Phosphatase	1 µL
pyrogenfreies Wasser	add 20 µL

### Isolierung hochmolekularer DNS

Für die Gewinnung hochmolekularer DNS aus den Zellpellets wurde das folgende Isolationsprotokoll etabliert.

Nach der Anzucht von *Acidicapsa borealis* und *Terracidiphilus* sp. S55 wurden diese bei 10.000 x g für 5 Minuten pelletiert. Der Überstand wurde verworfen. Die gewonnenen Bakterien aus dem Termitennest wurden analog präpariert. Um die zu präparierende DNS vor intrinsischer DNase-Aktivität zu schützen, wurde das Pellet für eine Minute bei 95 °C inkubiert.

Nach der Hitzebehandlung erfolgte der Verdau der Zellwand, indem das Pellet in 1 mL Lysozymlösung resuspendiert und für 2 Stunden bei 37 °C inkubiert wurde. Um hinreichend DNS aus den Bakterien zu isolieren, wurde eine spezielle Gießform mit einem Fassungsvermögen von 200 µL je Gelstück konstruiert in denen der nächste Schritt erfolgte. Um im weiteren Verlauf die DNS nicht zu scheren, wurden die unverdauten Zellen in Agarose eingegossen. Dazu wurde die Zellsuspension bei 3.900 x g für 15 Sekunden zentrifugiert und der Überstand entfernt. Das Pellet wurde vorsichtig in 1 mL 1,2 % (w/v) *low melting*-Agarose resuspendiert und in die Gießform transferiert.

Nachdem die Gelstücke ausgehärtet waren, wurde der Verdau der Zellen fortgeführt. Die Gelstücke wurden in ein 50 mL Falcon-Gefäß überführt und pro Gelstück 1 mL Achromopeptidase-Lösung hinzugefügt. Die Inkubation erfolgte für drei Stunden bei Raumtemperatur. Im letzten Schritt des Zellverdaus wurde die Lösung mit 1 mL Zellyse-Lösung pro Gelstück ausgetauscht und bei Raumtemperatur für 16 Stunden inkubiert.

Die Gelstücke wurden im weiteren Verlauf für den Partialverdau vorbereitet. Es erfolgten mehrere Waschstufen, um die Gelstücke von inhibierendem Detergenz und restlicher Proteinase K zu befreien. Dazu wurde die Zellyse-Lösung mit 40 mL TE-Puffer ausgetauscht und für zwei Stunden inkubiert. Danach erfolgte der Austausch des TE-Puffers mit 1 mL Stopplösung pro Gelstück und Inkubation für 2 Stunden. Dieser Schritt wurde wiederholt. Abschließend wurde die Stopplösung mit 40 mL TE-Puffer ausgetauscht und erneut für 2 Stunden inkubiert.

Im Anschluss wurden die Schritte für den Partialverdau der immobilisierten DNS durchgeführt. Zunächst wurden die Gelstücke in ein neues 50 mL Falcon-Gefäß überführt. Da-

## 2.5 Erstellung von Klonbibliotheken mit *Escherichia coli* als Wirt

mit das Restriktionsenzym BamHI zur DNS diffundieren kann, erfolgte eine vierstündige Inkubation der Gelstücke auf Eis mit 500  $\mu\text{L}$ /Gelstück Partialverdau-Puffer. Anschließend wurde das Gefäß in ein 37 °C warmes Wasserbad überführt und für eine Stunde unter leichtem Schütteln inkubiert. Nach diesem Schritt erfolgte der sofortige Stopp der Restriktion durch Hinzugabe von 500  $\mu\text{L}$  Stopplösung II pro Gelstück und Transfer der Proben auf Eis.

Um die DNS von Proteinen zu befreien und die partial-verdaute DNS im Größenbereich von 50 - 150 Kilobasen zu selektieren, wurde die DNS mittels Pulsfeld-Gelelektrophorese aufgetrennt. Dafür wurde zunächst das präparative Gel vorbereitet, indem 1,2 % (w/v) *low melting*-Agarose in  $\text{H}_2\text{O}$  gelöst und anschließend in die Gießform überführt wurde. Die Verwendung von autoklavierten MilliQ-Wasser erfolgte, um in der DNS-Aufbereitung keine Rückstände von inhibierendem Tris-Boratpuffer zu erhalten. Die Auftrennung der DNS wurde optimiert, indem das Gel für zusätzlich 30 Minuten bei 4 °C gelagert wurde. Anschließend wurden die Gelstücke in das Gel präpariert und mit 1,2 % (w/v) *low melting*-Agarose versiegelt. Als Marker wurde der *CHEF DNA Size Standard* (Bio-Rad) verwendet. Für die Auftrennung wurde das Steuergerät auf 6 V/cm, 1 - 45 Sekunden, 20 Stunden, 10 °C und 120 ° Winkel eingestellt. Als Laufpuffer wurden 3 Liter 0,5x TBE-Puffer verwendet.

Nach der Auftrennung wurde das Gel entnommen und die ersten drei Geltaschen vom Rest abgetrennt. Es erfolgte eine Färbung mit 0,5  $\mu\text{g}/\text{mL}$  Ethidiumbromid für 30 Minuten. Zunächst wurde die zweite Geltasche (DNS unverdaut) mit der dritten Tasche (DNS partiell verdaut) verglichen, um die Wirkung des Partialverdaus zu überprüfen. Anschließend wurde anhand des Markers der Bereich von 50 - 150 Kilobasen im Gel unter UV-Licht markiert. Die Präparation der zu klonierenden DNS erfolgte ohne UV-Licht anhand der Markierung.

Für die Extraktion der DNS aus dem Gel wurde ein Verdau mittels Agarase durchgeführt. Dazu wurden zunächst Gelstücke von maximal einen Zentimeter ausgeschnitten und in 2 mL Zentrifugengefäße überführt. Jede Probe wurde mit 1 mL TE-Puffer für eine Stunde gewaschen. Anschließend wurde die Probe bei 70 °C für 15 Minuten aufgeschmolzen und dann in einen 42 °C warmen Thermoblock transferiert. Der Agaroseverdau erfolgte durch Hinzugabe von 1  $\mu\text{L}$  Agarase und einstündiger Inkubation. Alle Proben aus einem Gellauf wurden anschließend vereinigt.

### **Klonierung und Transformation**

Die Ligation der extrahierten DNS mit dem linearisierten BAC erfolgte nach Tabelle 2.17. Um die Ligationseffizienz durch mehr DNS zu erhöhen, wurde der Reaktionsansatz von 20 auf 100  $\mu\text{L}$  erhöht.

## 2.6 Selektion und Anzucht erstellter Klone

Tabelle 2.17: Komponenten für die Ligation hochmolekularer DNS in pBeloBAC11

Komponente	Menge
20 ng linearisierter pBeloBAC11	2,3 $\mu\text{L}$
10x T4-Ligasepuffer	10 $\mu\text{L}$
T4-Ligase	2 $\mu\text{L}$
Extrahierte DNS, Konzentration (Tabelle 3.14)	85,7 $\mu\text{L}$

Der Reaktionsansatz für die Ligation wurde auf Eis angefertigt und für 16 Stunden inkubiert. Um die Effizienz der Transformation zu optimieren, erfolgte nach der Ligation eine Enzymdeaktivierung durch Hitze bei 65 °C für 30 Minuten.

Eine weitere Optimierung erfolgte durch die anschließende Aufkonzentrierung des Ligationsansatzes. Dazu wurden in einer Petrischale 30 % (w/v) Polyethylenglykol 8000 vorgelegt. Der Ligationsansatz wurde dann auf ein zuvor eingelegtes Filterpapier (0,0025 mm Membran) aufgetragen und für 30 Minuten inkubiert.

Die Transformation von *E. coli* MegaX DH10B T1 Electrocomp erfolgte nach Herstellerangaben und jeweils im Triplikat. Dazu wurden 2,5  $\mu\text{L}$  konzentrierter Ligationsansatz auf 20  $\mu\text{L}$  Zellen pipettiert und zusammen in eine 0,1 mm vorgekühlte Elektroporations-Küvette gegeben. Der Impuls erfolgte mit 2 kV (Bio-Rad Micropulser). Anschließend wurde der Ansatz sofort in 250  $\mu\text{L}$  *super optimal broth* (SOC) überführt und für eine Stunde bei 37 °C, 220 rpm inkubiert.

## 2.6 Selektion und Anzucht erstellter Klone

Die Selektion der transformierten Stämme erfolgte durch Ausplattieren von 100  $\mu\text{L}$  einer 1 zu 5 verdünnten Zellsuspension auf LB-Agarplatten mit 50  $\mu\text{g}/\text{mL}$  X-Gal, 25  $\mu\text{g}/\text{mL}$  IPTG und 12,5  $\mu\text{g}/\text{mL}$  Chloramphenicol. Diese wurde bei 16 Stunden für 37 °C inkubiert und anschließend auf Färbung untersucht.

Für die Anzucht im Milliliter-Maßstab wurden 2 mL *deep-well* Masterblöcke im 96-*well* Format mit Sauerstoff-durchlässigen Deckel (*sandwich cover*) verwendet. Das Arbeitsvolumen bestand aus 1,4 mL Müller Hinton II Medium. Durch die Blau-Weiß-Selektion mittels X-Gal wurden ausschließlich weiße Kolonien ausgewählt und in die Masterblöcke überführt. Die Kultivierung erfolgte über 24 Stunden bei 37 °C, 220 rpm mit 85 % Luftfeuchtigkeit. Zur Konservierung der Klonbibliotheken wurde jeweils 50  $\mu\text{L}$  Kultur entnommen und 1 zu 1 mit einer 50 % (v/v) Glycerinlösung verdünnt und bei -80 °C eingefroren.

## 2.7 Untersuchung erstellter Klonbibliotheken

### 2.7.1 BAC-Isolation und Restriktion zur Größenkontrolle der Inserts

Die BAC-Isolation erfolgte anhand des Qiagen *Large-Construct Kits*. Dazu wurde jeweils eine 500 mL LB-Kultur mit 12,5 µg/mL Chloramphenicol durch eine Kolonie angeimpft und bei 37 °C, 220 rpm für 24 Stunden kultiviert. Die isolierten BACs wurden nach Tabelle 2.18 verdaut und mittels PFGE größen-kontrolliert (Abbildung 3.18). Es wurden folgende Volumina für den Restriktionsverdau eingesetzt: 6 µL pBeloBAC11, 9 µL JF 3 A5, 7,6 µL S55 A3 und 2,2 µL JF 14 A4.

Tabelle 2.18: Komponenten für die Restriktion und Insertlängen-Kontrolle ausgewählter Klone

Komponente	Menge
500 ng BAC-DNS	siehe Text
10x FD-Puffer	2 µL
FD-NotI	1 µL
pyrogenfreies Wasser	add 20 µL

Für die Auftrennung dieses Restriktionsverdau wurde erneut eine Pulsfeld-Gel-elektrophorese durchgeführt. Es wurde 1,2 % LM-Agarosegel gelöst in 0,5 x TBE verwendet. Als Standard wurden 2 µL NEB *Lambda Monocut-Mix*, verdünnt mit 8 µL H<sub>2</sub>O und 2 µL Ladepuffer verwendet. Vor dem Pipettieren wurde der Mix 10 Minuten bei 65 °C erhitzt. Die Auftrennung erfolgte bei 6 V/cm, 120 ° Winkel, 12 °C und 12 Stunden.

### 2.7.2 Markergen-spezifische Untersuchung der Klonbibliotheken

Die Markergen-spezifische Untersuchung erfolgte durch eine PCR. Es wurden die Klone mit der DNS aus *Acidicapsa borealis* und der DNS aus gewonnenen Bakterien des Termitennests (*Coptotermes niger*) verwendet. Für die PCR wurden die *primer* A3-NRPS und A7R-NRPS (Tabelle 2.7) verwendet. Diese binden an der Adenylierungsdomäne von nicht-ribosomalen Peptidsynthasen und generieren ein Amplifikat von 700 Basen.

Die PCR erfolgte im 96-*well* Format. Dazu wurde nach der Kultivierung aus dem Masterblock 0,5 µL Kulturvolumen entnommen und direkt in den vorgelegten PCR-Mix, erstellt nach Tabelle 2.19, gegeben. Für diese PCR wurde zusätzlich 5 µL 10 % (v/v) Dimethylsulfoxid (DMSO) verwendet. Nach der PCR wurde der Ansatz 1 zu 10 verdünnt und auf einen *DNA 5K/HT LabChip* aufgetragen. Die Dokumentation erfolgte durch den *LabChip GX Touch HT* (PerkinElmer).

### 2.7.3 Bioinformatische Determinierung NRPS-positiver BAC-Inserts

Tabelle 2.19: Komponenten für die Detektion von Adenyldomänen in nicht-ribosomalen Peptidsynthasen

Komponente	Menge
Kulturvolumen	0,5 µL
10 mM A3-NRPS-Primer	2 µL
10 mM A7R-NRPS-Primer	2 µL
10x <i>DreamTaq</i> -Puffer	2,5 µL
<i>DreamTaq</i> DNAP	0,5 µL
10 mM dNTPs	0,5 µL
10 % DMSO	2,5 µL
pyrogenfreies Wasser	14,5 µL

Tabelle 2.20: PCR-Protokoll zur Markergen-spezifischen Detektion von Adenyldomänen in nicht-ribosomalen Peptidsynthasen

Programm	Temperatur	Zeit	Zyklen
Initiale Denaturierung	95 °C	5 min.	-
Denaturierung	95 °C	30 sek.	35
Annealing	58 °C	2 min.	
Extention	72 °C	4 min.	
Finale Extention	72 °C	10 min.	-

### 2.7.3 Bioinformatische Determinierung NRPS-positiver BAC-Inserts

Zur Bestimmung der klonierten Fragmentlängen sowie des Genomlokus wurden die vier Klone JF 7 A7, JF 7 E10, JF 8 B12 und JF 12 C9 untersucht. Diese wurden als Kolonien der Firma LGC Genomics übergeben und ansequenziert. Es wurden die *primer* M13rev2 und pBeloR verwendet (Tabelle 2.7). Die Sequenzierung erfolgte nach der Kettenabbruch-Methode von Sanger.

Die Bestimmung der Insertlängen wurden durch ein Sequenzvergleich mit dem Genom von *A. borealis* in dem Programm GenDB [172] durchgeführt. Die Darstellung der klonierten Bereiche erfolgte mit Python. Dafür wurde der zuvor mit Prokka annotierte Genomabschnitt von 2.700.127 bis 2.793.896 Basen extrahiert und mit den Sequenzpositionen aus GenDB abgeglichen.

Als zusätzliche Klassifikation wurden die Proteinsequenzen des Gencluster-Bereichs durch eine HMM-Analyse mit der Pfam-Datenbank untersucht [98]. Als Datenbank diente dazu eine *profile*-HMM basierte Version von Pfam (Pfam-A), die mittels HMMER-Paket (Version 3.1) ausgeführt wurde.

### 2.7.4 Untersuchung auf antibakterielle Aktivität

Nach der funktionalen Untersuchung mittels PCR und Konservierung der Klonbibliotheken, wurde das verbliebene Volumen von 1.350  $\mu\text{L}$  für die Untersuchung auf antibakterielle Aktivität eingesetzt. Dazu wurde zunächst ein zellfreier Überstand erzeugt. Die Masterblöcke wurden dazu bei 3.900 x g für 30 Minuten, 4 °C zentrifugiert. Die funktionale Untersuchung auf antibakterielle Aktivität wurde mit zwei Methoden im Mikrotiterplatten-Format durchgeführt und sind nachfolgend beschrieben.

#### Primäruntersuchung durch Biolumineszenz

In der primären Untersuchung wurden alle erstellten Klonbibliotheken auf anti Gram-negative Aktivität gegen *E. coli* DH5 $\alpha$  [pFU166] (Abbildung 2.4) untersucht. *E. coli* DH5 $\alpha$  [pFU166] eingesetzt. Das Plasmid pFU166 besitzt einen vorgeschalteten Promoter für das Haushaltsgen Glyceraldehyde-3-Phosphat- dehydrogenase (*gapA*), wodurch die Biolumineszenz-Kassette konstitutiv exprimiert wird. Als Basis für diesen Test auf Biolumineszenz wurden 50  $\mu\text{L}$  zellfreier Kulturüberstand in weiße, *clear botton 96-well* Mikrotiter-Platten eingesetzt und mittels Vakuumzentrifugation bei Raumtemperatur eingetrocknet.

## 2.7.4 Untersuchung auf antibakterielle Aktivität

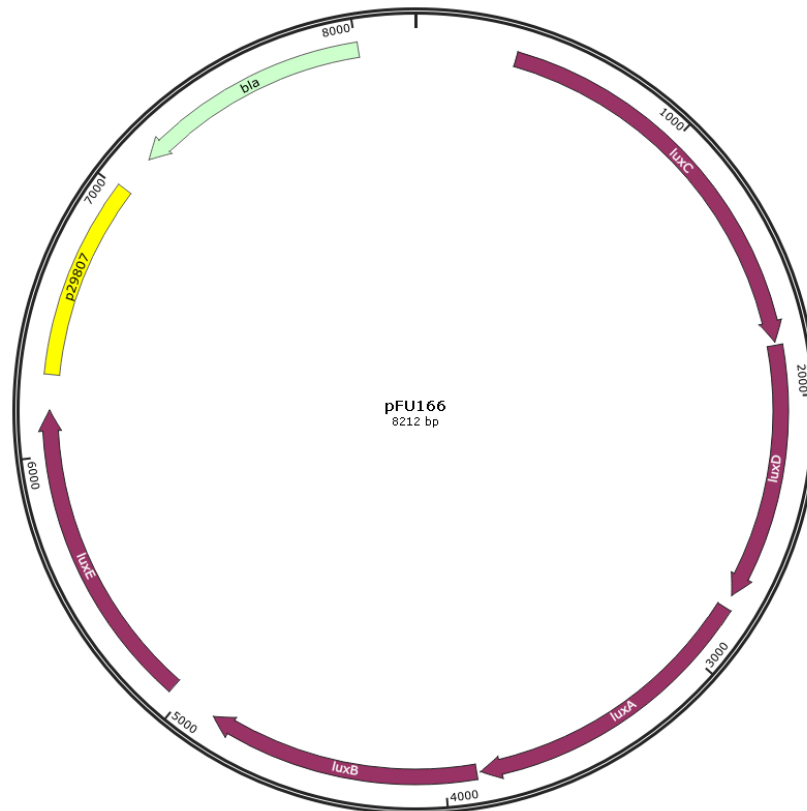


Abbildung 2.4: Plasmidkarte von pFU166 mit Lumineszenzkassette (*luxA*, *luxB*, *luxC*, *luxD*, *luxE*) und Selektionsmarker *bla* gegen  $\beta$ -Lactam-Antibiotika, eingesetzt für die primäre Untersuchung als *E. coli* DH5 $\alpha$  [pFU166] [173]

Die Bereitstellung des Testorganismus *E. coli* DH5 $\alpha$  [pFU166] erfolgte aus einem eingefrorenen Glycerin-Bakteriengemisch. Dazu wurde eine 30 mL Vorkultur mit Müller Hinton II plus 100  $\mu\text{g}/\text{mL}$  Ampicillin angefertigt und durch eine Impföse angeimpft. Die Kultivierung erfolgte bei 37  $^{\circ}\text{C}$ , 180 rpm für 16 Stunden kultiviert. Vor jedem Experiment wurde die Testkultur auf Biolumineszenz (LU) durch das Gerät Wallac Victor<sup>2</sup> 1420 überprüft. Es wurde je eingedampfter Zellüberstand das gleiche Volumen an Testorganismus hinzugegeben. Als Negativkontrolle wurde 50  $\mu\text{L}$  Müller Hinton II, als Positivkontrolle eine Gentamycin-Verdünnungsreihe in den Konzentrationen 160, 80, 40, 20, 10, 5 und 2,5  $\mu\text{g}/\text{mL}$  eingesetzt. Die Biolumineszenz wurde zu zwei Zeitpunkten ( $t =$  Stunden) bei  $t_0$  und  $t_6$  bestimmt. Zwischen den Zeitpunkten wurden die Platten bei 37  $^{\circ}\text{C}$ , 180 rpm und 85 % Luftfeuchtigkeit kultiviert. Die Berechnung der Inhibition (I in %) erfolgte nach folgender Formel:

$$I = 100 \times \left( 1 - \frac{(LU_{\text{Probe-t6}} - LU_{\text{Probe-t0}})}{(LU_{\text{Blank-t6}} - LU_{\text{Blank-t0}})} \right) \quad (2.4)$$

#### 2.7.4 Untersuchung auf antibakterielle Aktivität

##### Extraktion in Mikrotiterplatten und Sekundäruntersuchung

In Vorbereitung auf diese Untersuchung, wurden Extrakte aus den 1.300  $\mu\text{L}$  Kulturvolumina angefertigt. Die Herstellung dieser Extrakte sowie das Testverfahren ist nachfolgend beschrieben.

Als Vorbereitung auf die Extraktion wurden die Masterblöcke bei  $-80\text{ }^{\circ}\text{C}$  eingefroren und anschließend lyophilisiert bis keine Flüssigkeit mehr vorhanden war. Die Extraktion erfolgte für die sekundäre Untersuchung wie folgt:

- Hinzugabe von 1,2 mL 75 % (v/v) MeOH auf das Lyophilisat
- Sonifikation für 5 Minuten
- Masterblock für 1 Stunde schütteln, bei 180 rpm und RT
- Masterblock bei 3.900 x g für 10 Minuten,  $4\text{ }^{\circ}\text{C}$  zentrifugieren und 1 mL in einen neuen Masterblock (Extraktionsplatte) überführen
- Vakuumzentrifugation des Masterblocks für eine Stunde bei RT
- Vorherige Schritte mit dem ursprünglichen Masterblock wiederholen
- Nach Hinzugabe des zweiten Milliliters in die Extraktionsplatte, das Volumen mittels Vakuumzentrifugation eindampfen für 24 Stunden bei RT
- Die getrockneten Extrakte mit 130  $\mu\text{L}$  75 % (v/v) MeOH rüchlösen
- Sonifikation für 5 Minuten
- Versiegeln der Extraktionsplatte und Schütteln mit Qiagen *TissueLyser*, 1 Minute bei 20 Hertz
- Zentrifugieren bei 3.900 x g für 5 Minuten,  $4\text{ }^{\circ}\text{C}$
- Überführen von 100  $\mu\text{L}$  10x konzentrierten Extrakt in eine neue 96-well Platte

Für diese Untersuchung wurde das 10x konzentrierte Exktrakt aus den Klonbibliothek-Kulturen verwendet. Die Verteilung von 5  $\mu\text{L}$  Extrakt erfolgte als Duplikat in transparenten 384-well Mikrotiter-Platten und wurde automatisch durch einen *CyBio Well vario* Roboter (Analytik Jena) durchgeführt. Als Testbakterien wurden *P. aeruginosa* PAO 750, *E. coli* ATCC 25922, *E. coli*  $\Delta\text{tolC}$  und MSSA ATCC 25923 für die Untersuchung eingesetzt (Tabelle 2.4). Alle Stämme wurden in Müller Hinton II bei  $37\text{ }^{\circ}\text{C}$ , 240 rpm kultiviert. Um eine zweifache Nettoextrakt-Konzentration zu erhalten, wurden auf den eingedampften Extrakt 25  $\mu\text{L}$  Testbakterium mit einer zuvor bestimmten *McFarland*-Einheit von 1 hinzupipettiert, was etwa  $20^8$  Zellen/mL entspricht [174]. Die Inkubation erfolgte bei  $37\text{ }^{\circ}\text{C}$ , 220 rpm für 24 Stunden bei 85 % Luftfeuchtigkeit. Nach der Kultivierung wurde die optische Dichte bei 600 Nanometern durch den LUMIstar Omega (BMG Labtech) bestimmt. Die Berechnung der Inhibition (I in %) dieser Endpunktmessung erfolgte nach folgender Formel:

$$I = 100 x \left( 1 - \frac{(OD_{\text{Probe}} - OD_{\text{Blank}})}{(OD_{\text{Kontrolle}} - OD_{\text{Blank}})} \right) \quad (2.5)$$

### 2.7.5 Aktivitätsverifizierung durch *scale-up* Kultivierung

Klone mit detektierter antibakterieller Aktivität aus der Untersuchung mittels Biolumineszenz oder festgestellt anhand der Extrakt-basierten Untersuchung, wurden zur Bestätigung wiederholt. Dazu wurden die jeweiligen Klone in einem größeren Maßstab (50 mL) über 24 Stunden bei 37 °C und 220 rpm kultiviert. Zur Überprüfung wurden im Duplikat die Mediumskontrolle Müller Hinton II sowie der Wirt *E. coli* MegaX DH10B T1 Electrocomp [pBeloBAC11] mitbearbeitet. Die Kulturen wurden in einen 250 mL Rundkolben überführt, eingefroren und lyophilisiert. Tabelle 2.21 stellt die Klone dar welche für die *scale-up* Extraktion herangezogen wurden. Die Extraktion in diesem Maßstab erfolgte wie folgt:

- Hinzugabe von 20 mL entsprechendem Lösemittel auf das Lyophilisat (Tabelle 2.21)
- Sonifikation für 10 Minuten
- Überführen in ein 50 mL Röhrchen und zentrifugieren für 5 Minuten bei 3.900 x g
- Überstand überführen in ein neues 50 mL Röhrchen und Pellet nach Schritt 1 - 3 behandeln
- 40 mL Gesamtlösemittel durch Vakuumzentrifugation eindampfen bei RT
- Das getrocknete Extrakt in 1 mL 100 % MeOH rüchlösen (50x generiertes Extrakt)
- Extrakt bei 19.000 x g für 30 Minuten zentrifugieren und Überstand für Messung entnehmen

Tabelle 2.21: Klonbibliotheken für die *scale-up* Extraktion und chemische Analytik

Klon	Extraktionsmittel
JF 7 C3	100 % (v/v) Aceton oder EtAc
JF 7 C4	100 % (v/v) Aceton oder EtAc
JF 8 A9	75 % (v/v) MeOH
JF 9 H5	75 % (v/v) MeOH
JF 9 C12	75 % (v/v) MeOH
JF 11 B9	75 % (v/v) MeOH
JF 13 C6	75 % (v/v) MeOH
JF 13 G5	75 % (v/v) MeOH
JF 13 G11	75 % (v/v) MeOH

Mit den angefertigten Extrakten wurde die sekundäre Untersuchung in den Nettoextrakt-Konzentrationen 1x, 2x und 10x analog wie zum ersten Durchlauf durchgeführt. Zusätzlich wurden die Extrakte der chemischen Analytik übergeben.

### 2.7.5 Aktivitätsverifizierung durch scale-up Kultivierung

#### Chemische Analytik als aktiv eingestufte Klone

Die Extraktanalytik der Klone nach Tabelle 2.21 wurden mittels UHPLC-MS/MS untersucht. Dazu wurden jeweils 2  $\mu\text{L}$  des 50x Extrakts injiziert und die Probe über eine  $\text{C}_{18}$ -Säule (Firma ACQuity) mit einer Fließrate von 0,6 mL/min chromatographisch durch das *Agilent 1290 Infinity LC* System aufgetrennt. Die Säulenofen-Temperatur betrug 40 °C. Als Laufmittel wurden zwei Medien verwendet und in unterschiedlichen Verhältnissen kombiniert: (A: 99 %  $\text{H}_2\text{O}$ , 0.1 % Ameisensäure (FA); B: 99 % Acetonitril (ACN), 0.1 % Ameisensäure (FA))

Tabelle 2.22: Protokoll für die chromatographische Auftrennung generierter Extrakte

Zeit in Minuten	Komponente A	Komponente B
0	95 %	5 %
0,3	95 %	5 %
18	4,75 %	95,25 %
22,5	-	100 %
22,6	95 %	5 %
25	95 %	5 %

Der Analytik wurde anschließend mittels einer Elektrospray-Ionisation in das Massenspektrometer (maXis IITM, Bruker Daltronics) überführt und die Ionen der Masse nach aufgetrennt. Zur Auswertung der Ergebnisse wurden die Chromatogramme in dem *Compass DataAnalysis* Programm auf Unterschiede verglichen. Zur weiteren Analyse der Massen wurde die Datenbank *antiBase* verwendet.

Für den Vergleich der Massenfragmente aus Probe JF C3-Aceton und Diprotin A wurde zusätzlich eine Tandem-MS/MS-Analyse durchgeführt. Diprotin A wurde dazu in einer geringen Menge in 100 % (v/v) MeOH gelöst.

## 3 Ergebnisse und Diskussion

Im nachfolgenden Kapitel werden die Ergebnisse der durchgeführten Untersuchungen dargestellt und diskutiert. Diese gliedern sich in einen bioinformatischen Teil (Kapitel 3.1), sowie in einen experimentellen Teil zur Erstellung und funktionalen Untersuchung der Klonbibliotheken (Kapitel 3.2). Aufgrund der großen Datenmengen sind die Ergebnisse aus beiden Ergebnisteilen in der Druckversion auf das Wesentliche beschränkt. Es wird an dieser Stelle auf die jeweilige Datei verwiesen, welche auf der DVD zur Dissertation hinterlegt ist (Tabelle 3.1).

Tabelle 3.1: Datensätze der bioinformatischen Auswertung auf DVD

Dateiname	Kapitel
Datenbank-genome.csv	3.1.1
Datenbank-draft.csv	3.1.1
Resistenzgene.csv	3.1.3

### 3.1 Bioinformatische Analyse von Bakterien hinsichtlich Biosynthese-Gencluster

In diesem Kapitel werden die bioinformatischen Ergebnisse zu der statistischen und funktionalen Genomauswertung dargestellt. Diese basieren auf Genclustern ausgewählter Bakterien, worauf ein Stamm zur Methodenetablierung ausgewählt wurde. Kapitel 3.1.1 untersucht den Zusammenhang zwischen Genomgröße und Biosynthese-Gencluster (BGC). Weiterführend werden in dem Kapitel 3.1.2 die Gencluster-Typen *terpene* und *other* auf besondere Gene überprüft. Das Kapitel 3.1.3 stellt die Bakterien mit Resistenzgenen sowie Gencluster ohne Sequenzähnlichkeit als zentrales Ergebnis der bioinformatischen Untersuchung vor. Das Kapitel 3.1.4 illustriert die Identifizierung von Flavobakterien als abundanteste Gruppe mit Genclustern innerhalb des Phylums *Bacteroidetes*. Das Kapitel 3.1.5 charakterisiert *trans*-AT PKS Gencluster im Zusammenhang neuer Metabolite. Das Kapitel 3.1.6 behandelt die Genomsequenzierung von *A. borealis* sowie die funktionale Annotation. Zuletzt wird im Kapitel 3.1.8 die taxonomische Klassifizierung gewonnener Bakterien aus dem Termitennest *Coptotermes niger* als alternative Bioresource für die Antibiotikaforschung untersucht.

### 3.1.1 Quantifizierung von Genclustern in alternativen Bakteriengruppen

Das Phylum *Actinobacteria* galt jahrzehntlang als die Quelle aus der am meisten Antibiotika entwickelt worden sind [22]. Es wurden zahlreiche Produzenten dieses Phylums genomsequenziert und lieferten Erkenntnisse bezüglich Gencluster und Abundanz pro Genom, die miteinander korrelieren [73]. Auf Basis dieses Zusammenhangs wurde der akquirierte Datensatz alternativer Phyla bestehend aus *Acidobacteria*, *Bacteroidetes*, *Chloroflexi* sowie des Superphylums *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae* ebenfalls untersucht, um sich auf Bakterien mit ausgewählten Genclustern für die Erstellung von Klonbibliotheken zu fokussieren.

Ausgangspunkt der statistischen Analyse war die Identifizierung von Genclustern aus 361 Genomsequenzen. Auf dieser Basis konnte bereits eine erste Selektion hinsichtlich Identifizierung von Genclustern erfolgen. Es konnten 202 Genome mit Gencluster identifiziert werden und 159 Genome, in denen keine Gencluster identifiziert wurden. Die berechneten Genomgrößen aller Bakterien wurden darauffolgend gruppiert und in Form eines Boxplots dargestellt (Abbildung 3.1). Die Boxplots zeigen die Genomgrößen in Form einer Verteilung als Quartile, die durch den Median separiert werden. Die Abstände zwischen Boxrahmen und *Whiskers* (Minimum beziehungsweise Maximum) stellen die Außenbereiche der Verteilung dar und beinhalten jeweils 25 % aller Werte. Punkte außerhalb der *Whiskers* werden als Ausreißer bezeichnet. Die Box markiert 25 - 75 % der Werte um den Median.

Zu erkennen ist, dass sowohl die Gruppe mit Genclustern als auch die Gruppe ohne Gencluster, Ausreißer beinhalten, die jeweils außergewöhnlich große Genome haben. Die gesamte Werteverteilung beider Gruppen scheint dennoch ausgewogen. Aus der Abbildung der Genom-Größenverteilung ist ersichtlich, dass Genome mit Gencluster ( $\sim 4,1$  Megabasen) mehr als doppelt so groß sind, als Genome die keine Gencluster ( $\sim 1,5$  Megabasen) aufweisen. Um das Ergebnis dieser Verteilung statistisch zu belegen, wurde der Welch t-test durchgeführt. Dieser parametrische Test errechnet die statistische Signifikanz auf Basis des Mittelwerts für zwei Gruppen [175]. Als wichtige Voraussetzung zu diesem Test wurde der Datensatz auf Normalverteilung durch den Shapiro-Wilk t-test überprüft und ergab eine Normalverteilung anhand des p-Werts von  $>0,05$  für beide Gruppen. Die zu testenden Hypothesen für den Welch t-test wurden wie folgt formuliert;  $H_0$ : Es liegt kein signifikanter Unterschied zwischen beiden Gruppen vor (p-Wert  $>0,05$ ) und  $H_1$ : Es liegt ein signifikanter Unterschied vor (p-Wert  $<0,05$ ). Der errechnete p-Wert von  $2,2 \cdot 10^{-16}$  ergab den Ausschluß von  $H_0$  und die Annahme von  $H_1$ . Demnach besteht ein Zusammenhang zwischen Genomgröße und dem Auftreten von Biosynthese Genclustern (BGC) in diesem Datensatz.

### 3.1.1 Quantifizierung von Genclustern in alternativen Bakteriengruppen

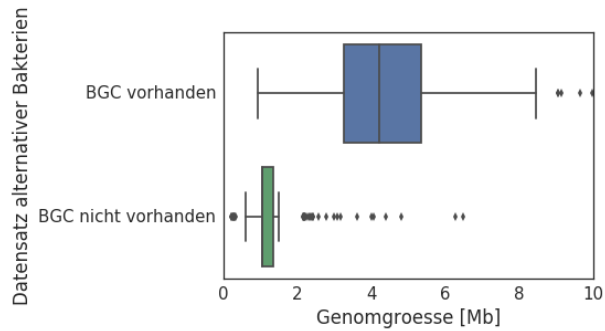


Abbildung 3.1: Verteilung der Genomgröße von Stämmen alternativer Bakterien die Biosynthese-Gencluster (BGC) enthalten oder nicht enthalten

Im weiteren Verlauf der Analyse wurde mit den 202 Genomen gearbeitet, in denen Gencluster identifiziert worden sind. Um den Zusammenhang zwischen Genomgröße und Gencluster-Anzahl zu untersuchen, wurde eine lineare Regression erstellt und der Korrelationskoeffizient berechnet (Abbildungen 3.2 und 3.3). Die Korrelation gibt grundsätzlich einen Wert zwischen -1 und +1 wieder. Werte über 0 implizieren eine positive Korrelation. Als Ergebnis wurde eine starke Korrelation für Actinobakterien von 0,89 und eine Korrelation von 0,54 für alternative Bakterien errechnet.

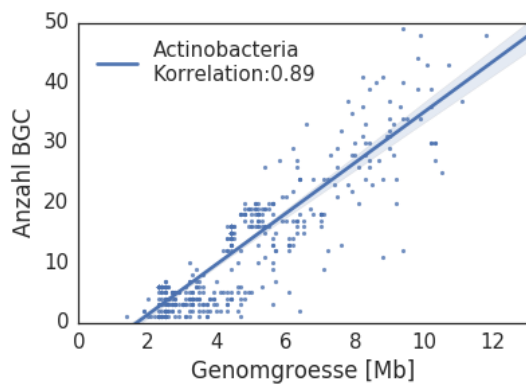


Abbildung 3.2: Korrelationsanalyse der Actinobakterien

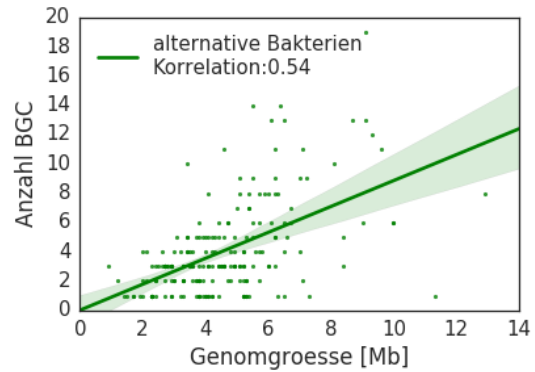


Abbildung 3.3: Korrelationsanalyse weniger erforschter Bakteriengruppen

### 3.1.1 Quantifizierung von Genclustern in alternativen Bakteriengruppen

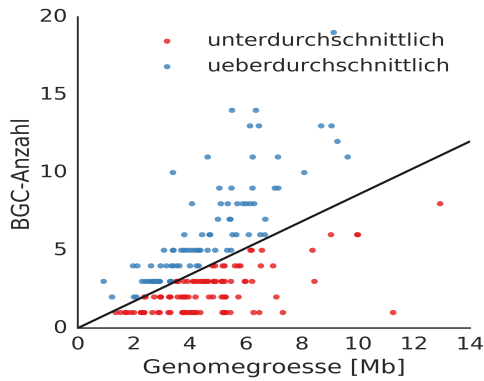


Abbildung 3.4: Gruppierung des Datensatzes anhand der Steigungsgeraden von 0,88

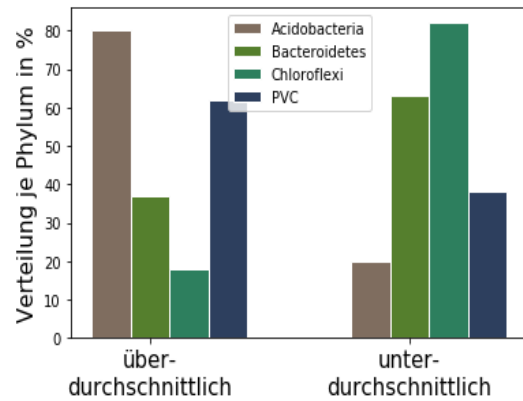


Abbildung 3.5: Gruppierung des Datensatzes nach Genclusterdichte je Phylum

Die Regressionsanalyse zeigt einen starken Zusammenhang zwischen Anzahl von Genclustern und der Genomgröße beider Gruppen. Die getesteten Parameter BGC-Anzahl und Genomgröße wurden individuell für jedes Phylum berechnet und sind in Tabelle 3.2 gezeigt. Als Vergleich dazu sind in Tabelle 3.3 die Werte für Actinobakterien angegeben. Um zielführender Stämme mit besonders hoher Dichte an Genclustern aus alternativen Bakterien zu identifizieren, wurde anhand der Steigung der Regressionsgeraden von 0,88 der Datensatz unterteilt (Abbildung 3.4). Stämme mit einem Quotienten  $>0,88$  wurden in die Kategorie überdurchschnittlich, Stämme  $<0,88$  in die Kategorie unterdurchschnittlich eingestuft. Die Verteilung der Stämme in die jeweilige Kategorie ist phylumspezifisch in Abbildung 3.5 dargestellt. Sie zeigt, dass für die Kategorie überdurchschnittlich 80 % der Acidobakterien diese Eigenschaft zutrifft. Für die weiteren Phyla ergab diese Einteilung 62 % *PVC*, 37 % *Bacteroidetes* und 18 % *Chloroflexi*.

### 3.1.1 Quantifizierung von Genclustern in alternativen Bakteriengruppen

Tabelle 3.2: Statistische Erhebung von Genomgröße und BGC-Anzahl von wenig erforschter Bakteriengruppen. Extremwerte sind hervorgehoben

Phylum	<i>Acidobacteria</i>		<i>Bacteroidetes</i>	
Genome	25		139	
Parameter	BGC-Anzahl	Genomgröße [Mb]	BGC-Anzahl	Genomgröße [Mb]
Mittelwert	7	5,6	3	4,2
STABW	3	2,1	3	1,6
Minimum	1	2	1	1,2
Maximum	13	<b>13</b>	<b>19</b>	11,25

Phylum	<i>Chloroflexi</i>		<i>PVC</i>	
Genome	11		27	
Parameter	BGC-Anzahl	Genomgröße [Mb]	BGC-Anzahl	Genomgröße [Mb]
Mittelwert	3	4,2	5	5,5
STABW	4	1,7	3,6	2,6
Minimum	1	1,7	1	<b>0,9</b>
Maximum	<b>14</b>	6,1	13	10

Tabelle 3.3: Statistische Erhebung von Genomgröße und BGC-Anzahl des Referenzphyllums *Actinobacteria*. Extremwert ist hervorgehoben

Phylum	<i>Actinobacteria</i>	
Genome	486	
Parameter	BGC-Anzahl	Genomgröße [Mb]
Mittelwert	12	4,6
STABW	10	2,2
Minimum	1	1,43
Maximum	<b>51</b>	12,1

Die Analyse von Genomen mit Genclustern wurde fortgeführt und in einem komparativen Ansatz mit denen von Actinobakterien verglichen. Tabelle 3.4 stellt die absoluten Werte aller 43 Biosynthese-Genclustertypen dar, die in den alternativen Phyla detektiert wurden. Um die Übersicht zu wahren, sind die Gencluster in sechs Kategorien (PKS, NRPS, Ripp, Terpene, Sonstige und Hybrid) eingeteilt. Die Daten in der Tabelle zeigen, dass vereinzelt Gencluster-Typen für jedes Phylum, mit Ausnahme der *Chloroflexi* gefunden wurden. Für *Acidobacteria* ist dies der Hybridcluster *t3pks-nrps-transatpks-t1pks* aus *Terracidiphilus* sp. S55, sowie *otherks-t1pks-oligosaccharide* aus *Solibacter usitatus* Ellin 6076. Im Phylum *Bacteroidetes* wurden die Gencluster *microviridin* und *ladderaneterpene* beispielsweise aus *Chryseobacterium* sp. StRB126, sowie *transatpks-otherks* und *lanthipeptide-nrps-transatpks* aus *Chitinophaga pinensis* DSM 2588 gefunden. Für das Superphylum *PVC* wurde das *otherks-t3pks* als einziges Gencluster aus *Planctomyces limnophilus* DSM 3776 gefunden.

### 3.1.1 Quantifizierung von Genclustern in alternativen Bakteriengruppen

Tabelle 3.4: Anzahl an detektierten Genclustern im Datensatz. ACI: *Acidobacteria*; BAC: *Bacteroidetes*; CHL: *Chloroflexi*; PVC: *Planctomycetes, Verrucomicrobia, Chlamydiae*; ACT: *Actinobacteria*. Gencluster mit *trans*-AT Polyketidsynthase sind hervorgehoben

Kategorie	Biosynthese-Gencluster	ACI	BAC	CHL	PVC	ACT
PKS	<i>t1pks</i>	2	3	1	6	834
	<i>t3pks</i>	25	44	1	10	249
	<b><i>transatpks</i></b>	1	0	0	0	3
	<i>resorcinol</i>	2	23	0	2	2
	<i>ladderane</i>	0	1	0	1	10
	<i>aryl polyene</i>	1	51	0	2	26
	<i>otherks</i>	1	0	4	6	32
NRPS	<i>nrps</i>	9	16	4	5	952
Ripp	<i>bacteriocin</i>	27	18	1	18	382
	<i>microviridin</i>	0	5	0	0	0
	<i>lanthipeptide</i>	2	37	0	2	201
	<i>lassopeptide</i>	11	9	0	0	51
	<i>microcin</i>	2	0	0	0	0
	<i>linaridin</i>	1	0	0	0	8
Terpene	<i>terpene</i>	50	149	19	53	727
Sonstige	<i>phenazine</i>	1	0	0	1	5
	<i>oligosaccharide</i>	1	0	0	0	10
	<i>butyrolactone</i>	0	2	0	0	95
	<i>homoserine lactone</i>	0	1	0	0	1
	<i>other</i>	26	22	1	24	861
	<i>siderophore</i>	0	18	0	0	236
	<i>ectoine</i>	0	0	2	2	136
Hybrid	<i>aryl polyene-resorcinol</i>	0	29	0	0	0
	<i>aryl polyene-t3pks</i>	0	7	0	0	0
	<i>t1pks-otherks</i>	0	5	0	9	38
	<i>t3pks-terpene</i>	1	4	0	0	3
	<b><i>transatpks-otherks</i></b>	0	1	0	0	0
	<i>otherks-t3pks</i>	0	0	0	1	0
	<i>ladderane-terpene</i>	0	1	0	0	4
	<i>lanthipeptide-bacteriocin</i>	1	0	0	0	11
	<i>lanthipeptide-terpene</i>	0	2	0	0	6
	<i>nrps-t1pks</i>	4	8	3	6	472
	<i>thiopeptide-lanthipeptide</i>	1	0	1	0	47
	<b><i>t3pks-nrps-transatpks-t1pks</i></b>	1	0	0	0	0
	<i>otherks-t1pks-oligosaccharide</i>	1	0	0	0	0
	<i>t1pks-transatpks-nrps</i>	1	0	0	0	13
	<i>otherks-PUFA-t1pks</i>	0	3	0	0	0
	<i>nrps-t1pks-bacteriocin</i>	1	0	0	0	1
	<i>nrps-t1pks-lanthipeptide</i>	0	0	0	0	11
	<i>nrps-lanthipeptide-t1pks-otherks</i>	0	1	0	0	1
	<i>terpene-nrps-t1pks</i>	1	0	0	0	4
	<b><i>lanthipeptide-nrps-transatpks</i></b>	0	1	0	0	0

### 3.1.1 Quantifizierung von Genclustern in alternativen Bakteriengruppen

Im vorherigen Abschnitt wurden die Ergebnisse der bioinformatischen Analyse zur Stamm-priorisierung, bestehend aus insgesamt 361 genomsequenzierten Bakterien der alternativen Phyla *Acidobacteria*, *Bacteroidetes*, *Chloroflexi* sowie des Superphylums *Planctomycetes*, *Verrucomicrobia* und *Chlamydiae* (PVC), gezeigt. Als Referenz zu dieser Untersuchung wurden 486 genomsequenzierte Stämme aus dem Phylum *Actinobacteria* hinzugezogen.

Das Programm *antiSMASH* wurde für die Detektion der Gencluster eingesetzt. In der statistischen Auswertung wurde der Datensatz anschließend in zwei Gruppen von Stämmen mit Gencluster (Anzahl 202) und ohne Gencluster (Anzahl 159) unterteilt. Das Programm *antiSMASH* detektiert grundsätzlich Gencluster, die dem Sekundärmetabolismus zugehörig sind und damit nicht essentiell sein müssen [176]. Um den biologischen Zusammenhang der Gruppe von Bakterien die in diesem Datensatz keine detektierbaren Gencluster aufweisen, zu verstehen, wurden einzelne Vertreter dieser Gruppe näher untersucht.

Diese Bakterien entstammen aus den Phyla *Bacteroidetes*, *Chlamydiae* und *Chloroflexi*. Viele dieser Bakterien haben eine starke Adaption an eine ökologische Nische, was die Ausbreitung vieler Konkurrenten zum Beispiel durch extreme Bedingungen von vornherein begrenzt und sehr wahrscheinlich keine Gencluster zur Bildung von Antibiotika als Verteidigung oder für die Kommunikation notwendig sind. Als Beispiele seien hier Stämme wie *Riemerella anatipestifer* [177] oder *Prevotella intermedia* [178] genannt. Neben diesen Vertretern der *Bacteroidetes* beinhaltet das Phylum *Chlamydiae* hauptsächlich pathogene Bakterien, die intrazellulär im Menschen überleben können [179]. Ebenso weist der Genus *Dehalococcoides* aus dem Phylum *Chloroflexi* viele obligate anaerobe Bakterien auf, die Organohalogenide verstoffwechseln und damit bereits eine spezielle ökologische Nische ausfüllen [180]. Neben dem biologischen Zusammenhang, das spezifische Gene des Sekundärmetabolismus nicht vorkommen müssen, sollte auch die technische Limitierung der Detektionsmodelle mitberücksichtigt werden, die auf bekannte Proteinfamilien beschränkt sind.

Im weiteren Verlauf der Analyse wurde mit der Gruppe von 202 Bakterien gearbeitet, die Gencluster aufweisen. Es wurde ähnlich wie bei den Actinobakterien eine positive Korrelation von 0,54 zwischen Gencluster-Anzahl und Genomgröße festgestellt. Dennoch gibt es auch innerhalb des Phylums *Actinobacteria* Bakterien, die eine deutlich niedrigere Gencluster-Dichte aufweisen, wie zum Beispiel *Propionibacterium* und *Corynebacterium*. Genera mit außergewöhnlich hoher Gencluster-Dichte sind *Streptomyces*, *Mycobacterium* und *Actinoplanes*. Um diese Eigenschaft, welche für viele Antibiotika-Produzenten zutrifft auch in alternativen Bakterien zu untersuchen, wurde die Regressionsanalyse als Maß für eine weitere Selektion der Stämme verwendet. Anhand der Steigung wurde das Datenset gruppiert in überdurchschnittlich und unterdurchschnittlich (Abbildung 3.4). Grundsätzlich konnten in jedem Phylum Stämme identifiziert werden, welche eine außergewöhnlich hohe Gencluster-Dichte aufweisen, wie zum Beispiel *Silvibacterium* sp. S15 (*Acidobacteria*), *Chitinophaga pinensis* DSM 2588 (*Bacteroidetes*), *Herpetosiphon aurantiacus* DSM 785 (*Chloroflexi*) und *Gemmata* sp. SH-PL17 (PVC). Anhand der Einteilung aus Abbildung 3.5 ist erkennbar, dass prozentual viele Stämme des Phylums

### 3.1.1 Quantifizierung von Genclustern in alternativen Bakteriengruppen

*Acidobacteria* diese Eigenschaft besitzen.

Um im weiteren Verlauf qualitativ Stämme anhand des Gencluster-Typs auszuwählen, sind alle Gencluster-Typen nach Kategorie und Phylum in der Tabelle 3.4 aufgelistet. Von besonderer Relevanz sind Gencluster und ihre korrespondierende Naturstoff-Klasse wie zum Beispiel Polyketide, sowie nicht-ribosomale Peptide und ribosomal-synthetisierte und post-translational modifizierte Peptide (Ripps). Aus diesen drei Klassen können potentiell antibiotische Metabolite hervorgehen, die sich in weitere Untergruppen diversifizieren lassen. Bekannte Beispiele sind Typ I PKS synthetisierte Makrolide [181] wie Erythromycin,  $\beta$ -Lactame aus der Biosynthese von nicht-ribosomalen Peptidsynthasen oder Lantibiotika aus der Klasse der Ripps [54]. Die Klassifizierung von Genclustern beruht in *antiSMASH* auf konservierten Sequenzmotiven, welche für die Klasse der Polyketide weitestgehend anhand der Ketosynthase-Domäne durchgeführt werden. Diese lassen sich in fünf Familien mit diversen Untergruppen einteilen und unterscheiden sich mechanistisch und strukturell [182]. In diesem Kontext ist auch der Gencluster-Typ *otherks* von Bedeutung. Dieser ist mit Ausnahme von Stämmen der *Bacteroidetes* vereinzelt vertreten und zeichnet sich durch atypische Reaktionen der Ketosynthasen aus, welche nicht der gewöhnlichen Claisen-Kondensation entsprechen müssen [183] [184]. Detektierte Gencluster des Typs *transatpks* werden in dem Kapitel 3.1.5 näher untersucht und zeichnen sich ebenfalls durch atypische Biosynthesemechanismen aus.

Im Falle der Detektion von nicht-ribosomalen Peptidsynthasen, basiert die Identifizierung anhand der Kondensations-, Adenyl- oder der Thiolations-Domäne. Unter diesem Aspekt sind Siderophore ebenfalls differenziert zu betrachten und im Falle der Detektion durch *antiSMASH* auf Basis einer signifikanten Sequenzähnlichkeit auf nicht-ribosomale Siderophore wie Aerobactin beschränkt. Weitere Erkenntnisse zeigen aber, dass Siderophore auch durch nicht-ribosomale Peptidsynthasen gebildet werden können [185].

Eine ebenfalls große Differenzierung unterschiedlicher Sequenzmotive wird in *antiSMASH* für die Klasse der Ripps gemacht, wie zum Beispiel anhand des N- oder C-Terminus für die Gencluster-Detektion von *bacteriocin* und *lanthipeptide* [186]. Deshalb ist die Wahrscheinlichkeit gegeben diese Gencluster häufiger zu identifizieren, was die relativ hohe Anzahl im Datensatz erklärt.

Weitere Einträge die aus der Tabelle 3.4 hervorgehen, gehören der Kategorie Hybride an. Diese umfasst unterschiedliche Gencluster welche durch den Suchalgorithmus zusammengefasst worden sind. Hybridgencluster sind in Bakterien verbreitet, können jedoch auch Artefakte durch den Suchalgorithmus selbst sein, der ausgehend von einem detektierten Sequenzmotiv, 10 - 15 Kilobasen auf- und abwärts der DNS-Sequenz nach weiteren Motiven sucht (*greedy approach*) [108]. Deshalb ist die *in silico* Längenberechnung der Gencluster nicht zwangsläufig korrekt.

Die Berücksichtigung von Terpene als Naturstoffe mit antibiotischer Wirkung ist auf genetischer Ebene nur unter bestimmten Voraussetzungen bekannt. Diese besondere Einschränkung gilt auch für den Gencluster-Typ mit der Bezeichnung *other* und wird im nachfolgenden Kapitel gesondert diskutiert. Ein weiterer Gencluster-Typ in der Kategorie Sonstige sind *oligosaccharide*. Dieser Gencluster-Typ umfasst bekannte Glykosyltransferasen wie zum Beispiel die UDP-Glykosyltransferase (MGT), welche in Verbin-

### 3.1.1 Quantifizierung von Genclustern in alternativen Bakteriengruppen

dung mit der Aktivierung und Deaktivierung von Makroliden stehen. Weitere Gencluster-Typen die besonders in Verbindung mit Signalprozessen stehen sind *homoserine lactone* und *butyrolactone*. Im Falle von *homoserine lactone* handelt es sich um das Gen der *autoinducer* Synthetase welche an der Biosynthese von Homoserinlactone beteiligt ist, die als *quorum sensing* Moleküle fungieren [187]. Das Gencluster *butyrolactone* steht bezeichnend für die *hotdog domain* der A-Faktor ähnlichen Synthetase. Der A-Faktor agiert als globaler Regulator in der Biosynthese von Streptomycin [188]. Die Gencluster-Typen *phenazine* sowie *ectoine* sind weniger in Bezug zu Antibiotika bekannt und kodieren für Enzyme welche an physiologischen Prozessen beteiligt sind, wie beispielsweise an Redoxreaktionen [189] oder der Aufrechterhaltung der Osmolarität [190].

Als abschließende Untersuchung bezüglich des Gencluster-Potentials, wurde anhand der Daten eine Hochrechnung beziehungsweise *rarefaction* durchgeführt. Diese Analysetechnik ist eine aus der Ökologie stammende Methode, um anhand der Anzahl und Diversität an vorhandenen Gencluster-Typen die noch zu erwartende Diversität abzuschätzen [191]. Auf Grund der relativ geringen Anzahl an Genclustern ist diese Analyse nach aktuellem Stand des Datensatzes nur für Actinobakterien zuverlässig, weshalb sie nur als Zusatzinformation herangezogen werden kann (Abbildung 5.1).

### 3.1.2 Annotation der Gencluster-Typen *other* und *terpene*

Eine hohe Gencluster-Dichte ist eine Eigenschaft, welche prinzipiell die Wahrscheinlichkeit der Biosynthese eines Antibiotikums erhöht. Um das Potential bezüglich der Biosynthese eines Antibiotikums gezielter einzugrenzen, wurde eine funktionale Annotation ausgewählter Gencluster-Typen *transatpks* (Kapitel 3.1.5), *terpene* und *other* durchgeführt.

#### Annotation des Gencluster-Typs: *terpene*

Der Gencluster-Typ *terpene*, ist mit 257 Einträgen am häufigsten in den alternativen Bakterien vertreten (Abbildung 3.4). Diese Beobachtung geht mit der Erkenntnis einher, dass Terpene die höchste Diversität an Naturstoffen aufweisen [192] und *antiSMASH* diesbezüglich diverse Detektionsmodelle besitzt. Allerdings steht die Biosynthese von Terpenen in Bakterien nur selten in Verbindung mit Antibiotika, weshalb eine genauere Untersuchung auf Genebene aller Gencluster notwendig war. Es ist ersichtlich, dass ein Großteil der Proteine keine bekannte Annotation besitzen und als hypothetisch klassifiziert wurden (Tabelle 3.5). Gene, welche für Proteine mit enzymatischer Aktivität kodieren, sind die Phytoen Synthase und Phytoen Dehydrogenase. Diese können zum Beispiel Bestandteil der Carotinoid-Biosynthese sein, indem Phytoen zu zeta-Caroten umgewandelt wird, einer Vorstufe für chromophore Moleküle [180]. Carotinoide dienen zum Beispiel der Vermeidung von *reactive oxygen species* (ROS) [193], daher sind diese Formen der Naturstoffe als Antibiotika irrelevant. Im weiteren Verlauf der Analyse wurden Proteine als Transkriptions-Regulatoren oder als mobile Element-Proteine annotiert. Diese Proteine haben essentielle Bedeutung in Bakterien, da sie für DNS-Rekombinationsereignisse verantwortlich sind und zum Beispiel zur Entwicklung einer Antibiotikaresistenz oder zur Pathogenität beitragen können [194].

Bisherige Studien deuten darauf hin, dass nur dann antibiotisch wirksame Terpene gebildet werden können, wenn eine Diterpencyclase vorhanden ist, die Geranylgeranyldiphosphat (GGDP) metabolisiert um damit Vorstufen für antibiotisch-wirksame Substanzen liefern kann [195]. Dieses Enzym ist jedoch nur vereinzelt in Bakterien wie *Streptomyces melanosporofaciens* zu finden [196]. Eine funktionale Klassifizierung aller Terpen-Gencluster durch die Annotationsplattform RAST ergab keine Treffer hinsichtlich einer Diterpencyclase. Ein Beispiel für einen antibiotisch wirksamen Metabolit ist das Terpentecin, welches aus zwei Geranylgeranyldiphosphat-Gruppen besteht und durch dieses Enzym katalysiert wird [197]. Dieses Enzym ist essentiell für die Ringbildung des Terpentecins und unterscheidet sich von anderen Cyclasen [198]. Generell ist diese enzymatische Reaktion von Cyclasen wichtig für weitere chemische Modifikationen wie Hydroxylierung, Methylierung, und Glykosylierung.

### 3.1.2 Annotation der Gencluster-Typen *other* und *terpene*

Tabelle 3.5: Die fünf häufigsten Proteine aus dem Gencluster-Typ *terpene*

Bezeichnung	Häufigkeit
hypothetisches Protein	4359
Phytoen Synthase (EC 2.5.1.32)	631
Transkriptions-Regulator	443
Phytoen Dehydrogenase (EC 1.14.99.-)	345
mobiles Element Protein	132

#### Annotation des Gencluster-Typs: *other*

Der Gencluster-Typ *other* aus Tabelle 3.4 ist insgesamt 61 Mal identifiziert worden. Jedoch lässt sich anhand der Bezeichnung keine weitere Zuordnung von Enzymen oder weiteren Faktoren machen. Aus den Publikationen von Blin *et al.* (2013) [199], sowie Medema *et al.* (2011) [108] gehen 12 unterschiedliche Detektionsmodelle für diesen Gencluster-Typ hervor. Darunter einzelne Gene die in Zusammenhang mit einer Antibiotika-Biosynthese stehen, wie zum Beispiel die Domäne *vlmB* für die Valanimycin Produktion. Auf Grund dieser Relevanz wurden diese Gencluster mittels einer Annotation untersucht und die Ergebnisse der fünf häufigsten Proteine oder Enzyme in Tabelle 3.6 dargestellt.

Tabelle 3.6: Die fünf häufigsten Proteine aus dem Gencluster-Typ *other*

Bezeichnung	Häufigkeit
hypothetisches Protein	895
Transkriptions-Regulator	27
Fettsäureligase (EC 6.2.1.3)	13
hypothetisches Transmembran-Protein und Signalpeptid-Vorhersage	10
mobiles Element Protein	10

Analog wie in der Annotation zuvor bilden hypothetische Proteine die Mehrheit der Ergebnisse und machen weitere Interpretationen schwierig. Es ist aber erkennbar, dass durch die Einträge wie Transkriptions-Regulatoren und Signalpeptide, dieser Gencluster-Typ sehr wahrscheinlich Regulationsmechanismen besitzt. Dieser Aspekt ist von Bedeutung da bekannt ist, dass die Aktivierung von Genclustern durch Signalmoleküle wie beispielsweise  $\gamma$ -Butyrolactone oder Furane gesteuert werden kann [200] [201]. Zusätzlich lässt der Eintrag von Fettsäureligasen daraufhin deuten, dass es sich verstärkt um den Umbau von langkettigen sowie gesättigten oder ungesättigten Fettsäuren handelt. In dieser ATP-getriebenen Reaktion werden diese Verbindungen zu kürzeren Fragmenten wie Acyl-CoA abgebaut [202].

Der Eintrag mobiles Element Protein ist ebenfalls in der Annotation in den Genclustern der Terpene gefunden worden. Damit erhärtet sich der Verdacht, dass es sich im Zusammenhang von Genclustern allgemein um eine wichtige Komponente handelt, durch die

### 3.1.3 Stammauswahl anhand von Resistenzgenen und Sequenzähnlichkeit

Gencluster rekombiniert beziehungsweise transportiert werden können.

#### 3.1.3 Stammauswahl anhand von Resistenzgenen und Sequenzähnlichkeit

Die Produktion von Antibiotika erfordert einen Resistenzmechanismus als Selbstschutz, der in Form eines Gens kodiert ist. In einem Gencluster sind diese Resistenzgene oft zusammen mit Genen zur Biosynthese des Antibiotikums lokalisiert. Die Identifikation von bekannten Resistenzgenen kann deshalb als eine zielführende Strategie angesehen werden, Stämme als potentielle Antibiotika-Produzenten zu identifizieren. Die durchgeführte Untersuchung bestand aus 202 Stämmen, in denen zuvor Gencluster identifiziert worden sind. Im weiteren Verlauf der Analyse wurde das Programm *Antibiotic Resistant Target Seeker* (ARTS) zur Detektion von Resistenzgenen verwendet. In der Summe wurden 101 Gencluster mit mindestens einem Resistenzgen identifiziert. Die entsprechenden Resistenzgene wurden kategorisiert nach ihrem Mechanismus und sind in den Tabellen 5.1 bis 5.3 des Annex zu finden.

Als zweiter Filter wurde der *knownclusterblast*-Algorithmus aus *antiSMASH* angewendet, um alle identifizierten Gencluster mit Resistenzgen auf Ähnlichkeit zu bekannten Genclustern in der Datenbank *Minimum Information about a Biosynthetic Gene cluster* (MIBiG) zu vergleichen [110]. Diese Datenbank ist speziell für Gencluster konzipiert und dient der offiziellen und standardisierten Dokumentation. Aus den 101 Genclustern mit Resistenzgenen wurden 24 Gencluster mit mindestens einen signifikanten Blasthit zu bekannten Genen identifiziert.

Es konnte festgestellt werden, dass der Großteil an Stämmen mit bekannten Genen aus dem Gencluster-Typ *terpene* stammt, die der Carotinoid-Biosynthese zugeordnet werden (Tabellen des Annex 5.4 und 5.5). Nach der Filterung von diesen 24 Genclustern, mit anteilig bekannten Genen, verblieben 77 Gencluster. Diese sind in Tabellen 3.7, 3.8, 3.9 und 3.10 aufgelistet. Die Tabellen stellen die zentralen Ergebnisse dieser Untersuchung dar, da aufgrund fehlender Sequenzähnlichkeit zu bekannten Genclustern prinzipiell von der Biosynthese eines unbekanntes Naturstoffs ausgegangen werden kann.

Aus den Tabellen zur Stammpriorisierung ist zu entnehmen, dass die anteilige Mehrheit der Gencluster-Typen PKS, NRPS oder Ripps entsprechen. Damit bestätigt diese Untersuchung die empirische Betrachtung dieser Naturstoff-Klassen hinsichtlich einer potentiellen Antibiotika-Biosynthese und ermöglicht mit den Ergebnissen aus Kapitel 3.1.1 eine begründbare Stammselektion.

Neben der Stammselektion erfolgte eine Kategorisierung der identifizierten Resistenzgene nach ihrem bekannten Mechanismus. Als Beispiel wurden diverse Lactamase-Gene identifiziert, welche als Enzym Penicillin-Derivate wie Carbapeneme oder Cephalosporine gezielt zerstören. Jedoch wurde ein Großteil dieser Resistenzgene im PKS-Gencluster-Typ identifiziert, was für diesen Typ nicht plausibel erscheint. In diesen Fällen wäre diese Resistenz im NRPS-Gencluster-Typ erwartet worden. Stämme in denen dieses Ergebnis identifiziert worden ist, sind *Acidobacterium* sp. S8, *Hymenobacter* sp. APR13 sowie *Hymenobacter swuensis* Dy53. Gene die eine Resistenz gegenüber Polyketide wie Tetracycline vermitteln, konnten ebenfalls detektiert werden, wie zum Beispiel der Transporter tet-MFS-efflux aus *Chryseobacterium* sp. StRB126 und *Anaerolinea thermophila* UNI-1.

### 3.1.3 Stammauswahl anhand von Resistenzgenen und Sequenzähnlichkeit

Weitere spezielle Resistenzgene wurden gegen Chloramphenicol in den Stämmen *Sphingobacterium* sp. 21 und *Parachlamydia acanthamoebae* entdeckt, sowie gegen Aminoglykoside in *Parachlamydia acanthamoebae*. In diesen Fällen entspricht das Resistenzgen nicht dem Gencluster-Typ der zu vermuten gewesen wäre, da der Resistenzmechanismus beispielsweise gegenüber Aminoglykoside, eine Aminoglycosid Phosphotransferase (APH3), im Gencluster-Typ t3pks aus *Parachlamydia acanthamoebae* UV-7, identifiziert worden ist.

Weitere Resistenzmechanismen konnten identifiziert werden, darunter diverse Transporter-Familien wie zum Beispiel (*ATP-binding cassette*) oder Transporter der *resistance nodulation devision* (RND), welche ein breites Resistenz-Spektrum zur Vermeidung von Toxizität aufweisen. Desweiteren wurden essentielle Gene zur DNS Transkription oder Replikation sowie Haushaltsgene zur Kodierung der Glycerin-3-Aldehyd Dehydrogenase identifiziert. Solche Gene können in mehrfacher genetischer Kopie vorkommen (Genduplikation), jedoch Mutationen aufweisen und dadurch beispielsweise stärker mit dem Antibiotikum wechselwirken. Dadurch wird der eigentliche Wirkungsort abgeschirmt. Eine weitere Strategie welche ebenfalls der Genduplikations-Theorie angehört, ist die gesteigerte Gendosis und Expression, wodurch der Wirkungsort durch massive Vervielfachung dem Antibiotikum entgegenwirkt [203].

Aus Metagenomstudien von stark konservierten Habitaten wie etwa Permafrost-Sedimenten geht hervor, dass die Entwicklung von Resistenzgenen nicht ausschließlich durch den menschlichen Gebrauch von Antibiotika hervorgeht, sondern Teil des natürlichen bakteriellen Genpools (Resistoms) sind [204]. Es besteht die Evidenz, dass beispielsweise Resistenzgene gegen  $\beta$ -Lactamantibiotika, Tetracycline oder Glykopeptid-Antibiotika seit mehreren zehntausend Jahren existieren [205]. Die Verbreitung und evolutionäre Entwicklung von diesen Genen, wird daher besonders in Habitaten mit hoher mikrobieller Interaktion angenommen, die zum Beispiel in bodenassoziierten Actinobakterien ausgeprägt ist. Diese Beobachtung könnte daher auch im übertragenden Sinn für bodenassoziierte Acidobakterien wie *Terriglobus saanensis* zutreffen und die Tendenz dieses Datensatzes erklären, dass diese Gene nicht zwangsläufig mit einem Gencluster assoziiert sind. Aus der Publikation von Monita *et al.* (2016) [206] ist zu entnehmen, dass insbesondere das Resistenzgen gegen Chloramphenicol global im Chromosom lokalisiert sein kann, was auf eine unabhängige Entwicklung ohne Assoziation zu einem Gencluster hindeutet.

Unter epideminologischen Gesichtspunkten lässt sich dennoch beobachten, dass neben der natürlichen Verbreitung von Resistenzen, durch den menschlichen Gebrauch von Antibiotika, die Akkumulation von Resistenzgenen steigt [207], wie am Beispiel von diversen Enterococcen gezeigt werden konnte [208]. Dabei ist die Verbreitung von Resistenzgenen nicht nur durch Plasmide beschränkt, sondern kann auch durch Phagen oder freiliegende DNS erfolgen [209].

### 3.1.3 Stammauswahl anhand von Resistenzgenen und Sequenzähnlichkeit

Tabelle 3.7: Stammauswahl des Phylums *Acidobacteria* anhand von Resistenzgenen und Sequenzähnlichkeit

Phylum	NCBI- <i>accession</i>	Organismus	Gencluster
<i>Acidobacteria</i>	NC_016025.1	<i>Chloracidobacterium thermophilum</i> B chromosome 2	<i>terpene</i>
	NZ_JIAL01000001.1	<i>Acidobacterium ailaoui</i> strain PMMR2	<i>t3pks</i>
	ATWD01000001.1	<i>Acidobacteriaceae bacterium</i> TAA166 strain TAA 166	<i>lassopeptide</i>
	NZ_CBXX010000007.1	<i>Pyrinomonas methylaliphatogenes</i> strain type	<i>other</i>
	CP002467.1	<i>Terriglobus saanensis</i> SP1PR4	<i>t3pks</i>
	CP000360.1	<i>Candidatus Koribacter versatilis</i> Ellin345	<i>bacteriocin</i>
	CP003130.1	<i>Granulicella mallensis</i> MF5ACTX8	<i>otherks</i>
	-	<i>Acidobacterium</i> sp. S8	<i>nrps</i>
	-	<i>Terracidiphilus</i> sp. SS55	<i>t3pks</i>
	-	<i>Granulicella</i> sp S137	<i>bacteriocin-lanthipeptide</i>
	-	<i>Silvibacterium</i> sp. S15	<i>lassopeptide</i>
	-		<i>t1pks-nrps</i>
	-		<i>nrps</i>
	-	<i>Granulicella</i> sp. L46	<i>thiopeptide</i>
		<i>t3pks</i>	

### 3.1.3 Stammauswahl anhand von Resistenzgenen und Sequenzähnlichkeit

Tabelle 3.8: Stammauswahl des Phylums *Bacteroidetes* anhand von Resistenzgenen und Sequenzähnlichkeit, Teil 1

Phylum	NCBI-accession	Organismus	Gencluster
<i>Bacteroidetes</i>	NZ_KI1912105.1	<i>Alkaliflexus imshenetskyi</i> DSM 15055 strain Z-7010	<i>other</i>
	KQ474106.1	<i>Aquimarina longa</i> SW024 genomic scaffold scaffold12	<i>lanthipeptide</i>
	DF968183.1	<i>Bacteroidales bacterium</i> TBC1	<i>other</i>
	CP007619.1	<i>Bacteroides dorei</i> isolate HS1_L.1.B.010	<i>bacteriocin</i>
	CP008741.1	<i>Bacteroides dorei</i> isolate HS1_L.3.B.079	<i>bacteriocin</i>
	CP009057.1	<i>Bacteroides dorei</i> isolate HS2_L.2.B.045b	<i>bacteriocin</i>
	CR626927.1	<i>Bacteroides fragilis</i> NCTC 9343	<i>other</i>
	CP011073.1	<i>Bacteroides fragilis</i> strain BOB25	<i>other</i>
	AP006841.1	<i>Bacteroides fragilis</i> YCH46	<i>other</i>
	CP003345.1	<i>Bernardetia litoralis</i> DSM 6794 chromosome	<i>t3pks</i>
	CP001699.1	<i>Chitinophaga pinensis</i> DSM 2588	<i>lanthipeptide</i>
			<i>nrps</i>
			<i>lanthipeptide-t1pks</i> <i>-otherks-nrps</i>
	CP009928.1	<i>Chryseobacterium gallinarum</i> strain DSM 27622	<i>lanthipeptide</i>
			<i>lanthipeptide</i>
	AP014624.1	<i>Chryseobacterium</i> sp. StRB126 DNA	<i>nrps</i>
<i>other</i>			
<i>nrps</i>			
			<i>ladderane</i>

### 3.1.3 Stammauswahl anhand von Resistenzgenen und Sequenzähnlichkeit

Tabelle 3.9: Stammauswahl des Phylums *Bacteroidetes* anhand von Resistenzgenen und Sequenzähnlichkeit, Teil 2

Phylum	NCBI-accession	Organismus	Gencluster
<i>Bacteroidetes</i>	CP002528.1	<i>Dokdonia</i> sp. 4H-3-7-5 chromosome	<i>t3pks-aryl polyene</i>
	CP009301.1	<i>Dokdonia</i> sp. MED134	<i>t3pks</i>
	CP003346.1	<i>Echinicola vietnamensis</i> DSM 17526	<i>aryl polyene</i>
	CP011059.1	<i>Elizabethkingia miricola</i> strain BM10	<i>other</i>
	CP010798.1	<i>Flavobacterium psychrophilum</i> strain 4	<i>lanthipeptide</i>
	CP012586.1	<i>Flavobacterium psychrophilum</i> strain Z2	<i>t3pks</i>
	CP013909.1	<i>Hymenobacter sedentarius</i> strain DG5B chromosome	<i>terpene</i>
	CP006587.1	<i>Hymenobacter</i> sp. APR13	<i>t3pks</i>
	CP012623.1	<i>Hymenobacter</i> sp. DG25A	<i>t3pks</i>
	CP007145.1	<i>Hymenobacter swuensis</i> DY53	<i>t3pks</i>
	CP003178.1	<i>Niastella koreensis</i> GR20-10	<i>t3pks</i>
	CP003879.1	<i>Psychroflexus torquus</i> ATCC 700755	<i>aryl polyene</i>
	CP001807.1	<i>Rhodothermus marinus</i> DSM 4252	<i>terpene</i>
			<i>t3pks</i>
			<i>terpene</i>
			<i>lassopeptide</i>
			<i>terpene</i>
			<i>terpene</i>
			<i>t3pks</i>
			<i>microviridin</i>
		<i>t1pks-nrps</i>	
		<i>bacteriocin</i>	
		<i>bacteriocin</i>	
		<i>bacteriocin</i>	
		<i>other</i>	
		<i>lanthipeptide</i>	
		<i>Runella slithyiformis</i> DSM 19594	
		<i>Salinibacter ruber</i> DSM 13855	<i>terpene</i>
		<i>Salinibacter ruber</i> M8 chromosome	<i>terpene</i>
		<i>Solitalea canadensis</i> DSM 3403	<i>t3pks</i>
		<i>Sphingobacterium</i> sp. 21	<i>microviridin</i>
		<i>Spirosoma radiotolerans</i> strain DG5A	<i>t1pks-nrps</i>
		<i>Spirosoma radiotolerans</i> strain DG5A	<i>bacteriocin</i>
		<i>Tannerella forsythia</i> 92A2	<i>bacteriocin</i>
		<i>Tannerella forsythia</i> KS16 DNA	<i>bacteriocin</i>
		<i>Xylanibacter oryzae</i> DSM 17970 genomic scaffold	<i>other</i>
		<i>Zamogwangia profumda</i> SM-A87	<i>lanthipeptide</i>

### 3.1.3 Stammauswahl anhand von Resistenzgenen und Sequenzähnlichkeit

Tabelle 3.10: Stammauswahl des Phylums *Chloroflexi* und *PVC* anhand von Resistenzgenen und Sequenzähnlichkeit

Phylum	NCBI-accession	Stamm	Gencluster	
<i>Chloroflexi</i>	AP012029.1	<i>Anaerolinea thermophila</i> UNI-1	<i>other</i>	
	AP012337.1	<i>Caldilinea aerophila</i> DSM 14535	<i>terpene</i>	
	CP000804.1	<i>Roseiflexus castenholzii</i> DSM 13941	<i>otherks</i>	
	CP000686.1	<i>Roseiflexus</i> sp. RS-1	<i>otherks</i>	
<i>PVC</i>	CH672377.1	<i>Blastopirellula marina</i> DSM 3645 scf_1099463000750	<i>other</i>	
	CH672376.1	<i>Blastopirellula marina</i> DSM 3645 scf_1099463000751	<i>t1pks-nrps</i>	
	CP001032.1	<i>Opitutus terrae</i> PB90-1	<i>ladderane</i>	
	FR872580.1	<i>Parachlamydia acanthamoebae</i> UV-7	<i>t3pks</i>	
	AP012338.1	<i>Phycisphaera mikurensis</i> NBRC 102666	<i>t3pks</i>	
	CP011270.1	<i>Planctomyces</i> sp. SH-PL14	<i>terpene</i>	
	BX119912.1	<i>Rhodopirellula baltica</i> SH 1	<i>t3pks</i>	
	JH636434.1	<i>Schlesneria paludicola</i> DSM 18645 genomic scaffold	<i>otherks</i>	
	FR872582.1	<i>Simkania negevensis</i> Z main chromosome	<i>bacteriocin</i>	
	CP003364.1		<i>Singulisphaera acidiphila</i> DSM 18658	<i>other</i>
				<i>t1pks-nrps</i>
				<i>terpene</i>

### 3.1.4 Taxonomische Zuordnung von Bakterien mit Genclustern

In dem akquirierten Datensatz zur Abschätzung des Gencluster-Potentials von 361 Genomen, wurden 139 Genome aus dem Phylum *Bacteroidetes* mit Genclustern untersucht. Auf Basis der NCBI-*accession* wurde die entsprechende Taxonomie zugeordnet. Von 139 Stämmen konnten 138 auf Genusebene taxonomisch eingeordnet werden. Die Abbildung 3.6 zeigt das Ergebnis dieser Zuordnung, angefangen von der taxonomischen Ebene der Ordnung bis zum Genus. Zu erkennen ist, dass 50 % der Stämme mit Genclustern der taxonomischen Ordnung *Flavobacteriales* zugeordnet wurden. Jeweils 20 % Anteil der Stämme mit Genclustern wurden der taxonomischen Ordnung *Bacteroida* und *Cytophagales* zugeordnet. Der restliche Anteil von 10 % spaltet sich auf nach 5 % *Sphingobacteriaceae*, 2 % *Saprospirales*, 2 % *Salinibacter* und 1 % *Niabella*.

### 3.1.4 Taxonomische Zuordnung von Bakterien mit Genclustern

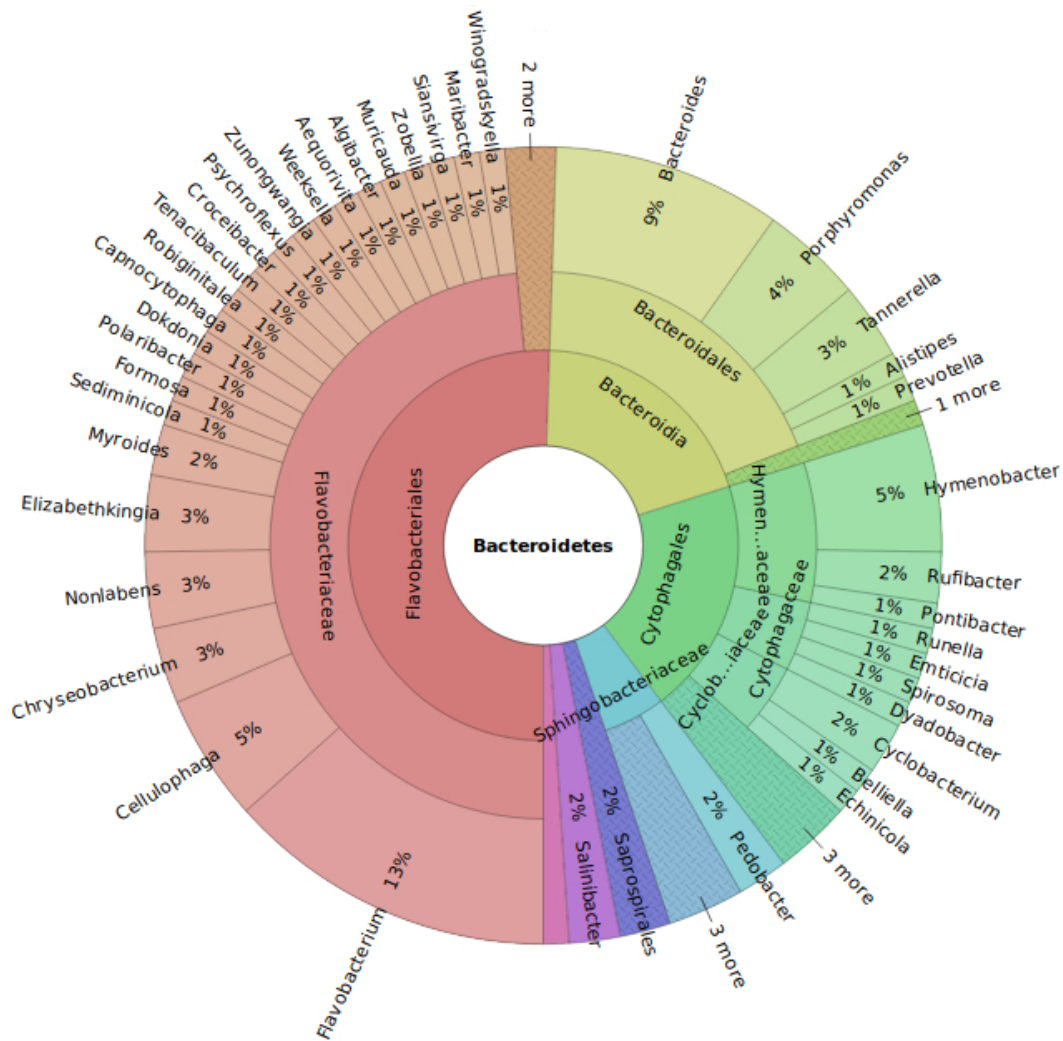


Abbildung 3.6: Taxonomische Zuordnung von Stämmen der *Bacteroidetes* die Biosynthese-Gencluster enthalten

Stämme der Ordnung *Flavobacteriales* sind größtenteils aerob und chemo-organotroph. Sie kommen ubiquitär in der Umwelt vor, wie zum Beispiel in Gewässern oder Sedimenten und können komplexe Biopolymere wie Cellulose, Chitin oder Pektin degradieren [210]. Eine weitere taxonomische Ordnung ist die der *Bacteroidia*, zu der mit 9 % der Genus *Bacteroidis* gehört. Dieser Genus umfasst vorwiegend Anaerobier, die zum Teil pathogene Bakterien wie *B. fragilis* beinhalten [211]. Die taxonomische Familie der *Sphingobacteriaceae* beinhaltet die Genera *Pedobacter*, *Sphingobacterium*, *Solitalea* und *Pseudopedobacter*. Stämme dieser taxonomischen Familie wurden aus Böden [212] oder Kompost isoliert [213]. Sekundärmetabolite mit antibakterieller Aktivität sind aus diesem Phylum nur von einzelnen Stämmen bekannt, die zum Teil nicht genomsequenziert sind. Bisher bekannte antibakterielle Metabolite sind beispielsweise das NRPS-

### 3.1.5 Charakterisierung von *trans*-AT PKS in Acidobakterien

synthetisierte Lysobactin/Katanosin B aus *Cytophaga* PBJ-5356, Formadicine aus *Flexibacter alginoliq uefaciens* sowie TAN-1057 A-D aus *Flexibacter* sp. und Lactivicin aus *Empedobacter lactamgenus* YK-258 [214]. Einige dieser Bakterien verfügen desweiteren über die Möglichkeit Biofilme auszubilden. Die Produktion von Biofilmen ist abhängig von Signalmolekülen zu denen in nicht inhibitorischen Konzentrationen auch Antibiotika zählen. Deshalb könnte die Eigenschaft der Biofilm-Produktion auch für die Akkumulation von BGCs mitverantwortlich sein, was zum Beispiel für den Genus *Flavobacterium* zuträfe [215]. Weitere Metabolite mit antibakterieller Eigenschaft sind die Elansolid-Metabolite aus dem Genus *Chitinophaga* [216]. Die ersten Metabolite Elansolid A1 und A2 aus *Chitinophaga sancti*, einstmalig dem Genus *Flexibacter* zugehörig [217], sind PKS-synthetisierte Macrolid-Antibiotika, von denen nur das Isomer A2 Aktivität gegen Gram-positive Bakterien zeigt [218]. In weiteren Studien zur Elansolidforschung konnten weitere stabile Derivate gefunden werden, die gegen den Methicillin-resistenten *Staphylococcus aureus* (MRSA) aktiv sind [219]. Ein weiteres Bakterium das zur Elansolidproduktion fähig ist, ist *Chitinophaga pinensis* DSM2588, welcher auch Vertreter dieses Datensatzes ist [220]. Dieser Stamm hat eine Genomgröße von 9,13 Megabasen und 19 Gencluster. Das durch einen *genome mining* Ansatz entdeckte *els* Gencluster, produziert das Elansolid D und ist ortholog zu dem *ela* Gencluster aus *Chitinophaga sancti* [221]. Eine weitere Besonderheit dieser beiden Gencluster bezieht sich auf die biochemische Eigenschaft der Polyketidsynthese, welche in ihrem modularen Aufbau eine Acyltransferase (AT) als *trans*-Domäne besitzen [222]. Das dazu korrespondierende Gencluster *els* aus *C. pinensis* ist in Tabelle 3.4 als *otherks-transatpks* bezeichnet. Das zweite *trans*-AT PKS Gencluster hat die Bezeichnung *nrps-transatpks-lanthipeptide*, von dem kein Naturstoff bekannt ist.

Polyketidsynthasen mit dieser Eigenschaft wurden zunächst nur in Gram-positiven Bakterien entdeckt. Durch weitere Genomanalysen relativiert sich die Anzahl gefundener Polyketidsynthasen hinsichtlich Gram-negativer Bakterien, was ebenfalls durch diese Arbeit bestätigt werden konnte. Das Kapitel 3.1.5 stellt die ersten drei identifizierten *trans*-AT Polyketidsynthasen aus Acidobakterien vor.

### 3.1.5 Charakterisierung von *trans*-AT PKS in Acidobakterien

In diesem Kapitel werden die Polyketidsynthasen der drei Gencluster *trans*-AT Polyketidsynthasen *t3pks-nrps-transatpks-t1pks* (*Terracidiphilus* sp. S55), *transatpks* (*Bryocella* sp. S190) und *t1pks-transatpks-nrps* (*Acidobacterium* sp. S8), charakterisiert. Diese *trans*-AT Polyketidsynthasen gehören zum Typ I und unterscheiden sich von *cis*-AT PKS Typ I durch diverse Mutationen, wodurch die AT-Domäne nicht mehr vollständig im Modul integriert ist [223]. Biochemisch ergeben sich durch diese Mutationen zur *trans*-AT Domäne ebenfalls Unterschiede, die stereochemisch bedingt andere Substrataffinitäten hervorrufen und im Vergleich zu *cis*-AT Domänen, die nach dem Co-Linearitätsprinzip katalysieren, eine Metabolitvorhersage erschweren [224] [225]. Ein Beispiel für ein *trans*-AT spezifisches Substrat ist Ethylmalonyl-CoA, welches in der Biosynthese von Kirromycin verwendet wird [226]. Polyketidsynthasen mit *trans*-AT wurden im Jahr 2010 durch die Publikation des *pksX* und *baeX* Genclusters in *Bacillus subtilis* und *Bacillus amylopha-*

### 3.1.5 Charakterisierung von *trans*-AT PKS in Acidobakterien

*ciens* bekannt [227]. Aktuell gibt es Hinweise darauf, dass diese einzelnen Domänen, auf genetischer Ebene im Verlauf der Evolution, durch horizontalen Gentransfer unabhängig von *cis*-AT Domänen weiterentwickelt worden sind, was den Unterschied in der Substrataffinität erklären könnte. Auf genetischer Ebene haben sich die *cis*-AT Domänen modulweise durch Genduplikation weiterentwickelt. Auf Grund der sich häufenden Funde von *trans*-AT PKS in Gram-negativen Bakterien lässt sich schlussfolgern, dass die evolutionären Veränderungen dieser Gene zunächst unter Gram-negativen Bakterien stattgefunden haben und nicht, wie vorher angenommen, in Gram-positiven Bakterien wie etwa Streptomyceten, die durch die Verwendung als Antibiotika-Produzenten mehr Aufmerksamkeit erlangt haben [214].

In Abbildung 3.7 sind die drei identifizierten Gencluster aus Acidobakterien dargestellt, welche das Gen für die Kodierung der *trans*-AT Domäne tragen. Zusätzlich werden deskriptiv die wesentlichen Merkmale erläutert. Das größte Gencluster ist das *t1pks-transatpks-nrps* aus *Acidobacterium* sp. S8 mit einer Länge von 78.969 Kilobasen und 36 Genen. Als Hybridcluster besitzt es außerdem Sequenzmotive für eine *cis*-AT Typ I Polyketidsynthase (*t1pks*) und für eine nicht-ribosomale Peptidsynthase. Das Gen *ppsD* wurde als Phtiocerol-Synthase annotiert und kodiert die *trans*-AT Domäne. Es hat eine Länge von 3.654 Basen. Die katalytische Sequenzabfolge des Enzyms wurde wie folgt vorhergesagt: KS-(*trans*-AT)-DH-ER-KR-ACP. Ein Proteinblast von Ppsd ergab mit einer Sequenzidentität von 33 % eine Typ I Polyketidsynthase aus *Hydrococcus rivularis* (NCBI-accession: WP\_073601769.1). Eine *knownclusterblast*-Analyse ergab eine Ähnlichkeit von 57 % mit Genen aus dem Gencluster für die Biosynthese von Myxothiazol [228], einem cytotoxischen Wirkstoff in der Krebsmedizin. Es wurde kein Resistenzgen identifiziert.

Der zweite Hybridcluster mit einem *trans*-AT ständigem Enzym, ist der *t3pks-nrps-transatpks-t1pks* Gencluster aus *Terracidiphilus* sp. S55. Der Gencluster hat eine Länge von 69.321 Kilobasen und 46 Genen. Das Gen *pks15* hat eine Länge von 2.265 Basen und wurde als Phenolphthiocerol-Synthase annotiert. Das Gen kodiert die katalytischen Domänen KS-(*trans*-AT)-ACP. Ein Proteinblast von Pks15 ergab mit einer Sequenzidentität von 39 % ein hypothetisches Protein aus *Calothrix* sp. PCC (NCBI-accession: WP\_019489582.1). Die *knownclusterblast*-Analyse zeigt eine Ähnlichkeit von 36 % mit Genen, die zum Myxalamid Gencluster zählen [229], dessen Produkt eine cytotoxische Wirkung zeigt. Es wurde kein Resistenzgen identifiziert.

Der dritte Gencluster aus *Bryocella* sp. S190 wurde als *transatpks* klassifiziert. Der Gencluster hat eine Länge von 44.268 Kilobasen und 22 Gene. Das Gen *pksN* ist 11.796 Basen lang und wurde als eine nicht näher gekennzeichnete Polyketidsynthase annotiert. Die katalytische Sequenzabfolge der Domänen lautet: KS-(*trans*-AT)-KR-MT-ACP-KS-(*trans*-AT)-ACP-KS-(*trans*-AT)-DH-KR. Der Proteinblast von Pksn ergab mit einer Sequenzidentität von 39 % das hypothetische Protein ADL29\_27455 aus *Streptomyces chatanoogensis* (NCBI-accession: KPC60798.1). Die *knownclusterblast*-Analyse ergab keine Treffer zu bekannten Genclustern. Es wurde kein Resistenzgen identifiziert.

### 3.1.5 Charakterisierung von *trans*-AT PKS in Acidobakterien

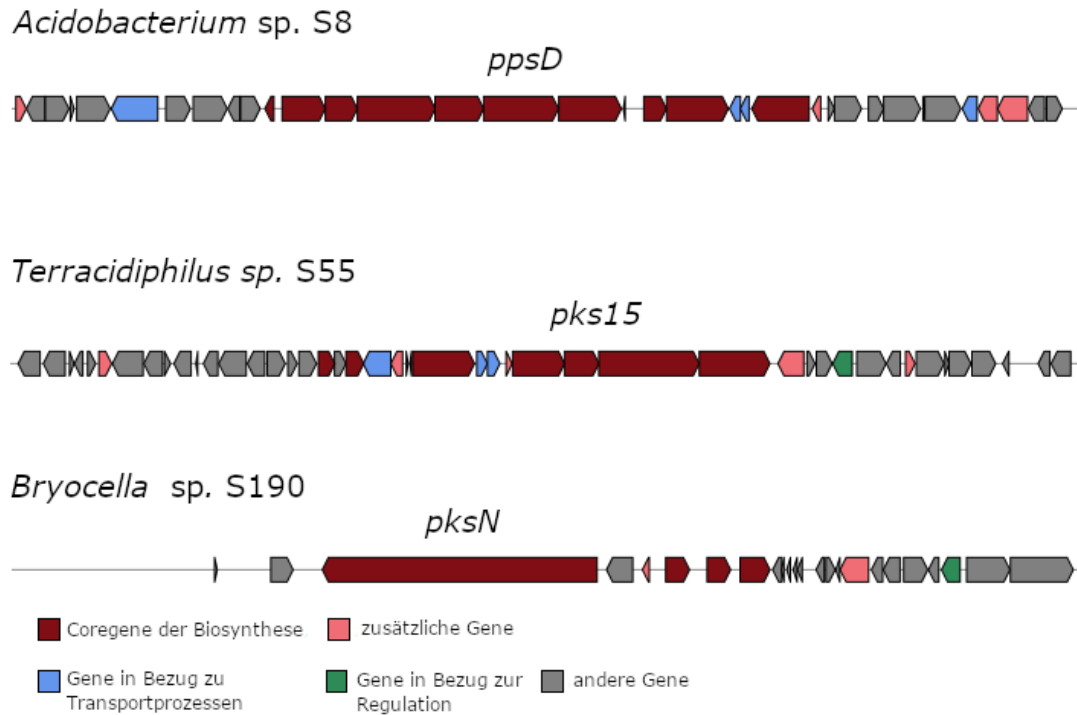


Abbildung 3.7: Gencluster mit *trans*-AT Polyketidsynthasen des Typs I aus drei Acidobakterien, *Acidobacterium* sp. S8, *Terracidiphilus* sp. S55 und *Bryocella* sp. S190. Die entsprechenden Gene sind hervorgehoben

In der weiterführenden Analyse wurden die entsprechenden Sequenzen der Ketosynthasen (KS) und *trans*-Acyltransferasen (AT) extrahiert und auf weitere Sequenzmotive durch ein *alignment* verglichen. Die Abbildung 3.8 stellt den Vergleich der Ketosynthasen dar, in der durch rote Rahmen die katalytischen Triaden gekennzeichnet sind. Ketosynthasen katalysieren die Kohlenstoff-Kohlenstoffbindung zwischen der wachsenden Polyketidkette und einer neuen Einheit mittels Claisenkondensation durch drei spezifische Aminosäuren, die in Form einer katalytischen Triade funktionieren. Die entsprechenden Sequenzmotive um diese Aminosäuren sind konserviert und konnten anhand der Angaben von Jenner (2016) [230] eindeutig zugeordnet werden. Daher ist anzunehmen, dass die katalytische Triade der fünf untersuchten Ketosynthasen aus den Aminosäuren Cystein-175-Histidin-310-Histidin-358 besteht.

### 3.1.5 Charakterisierung von *trans*-AT PKS in Acidobakterien

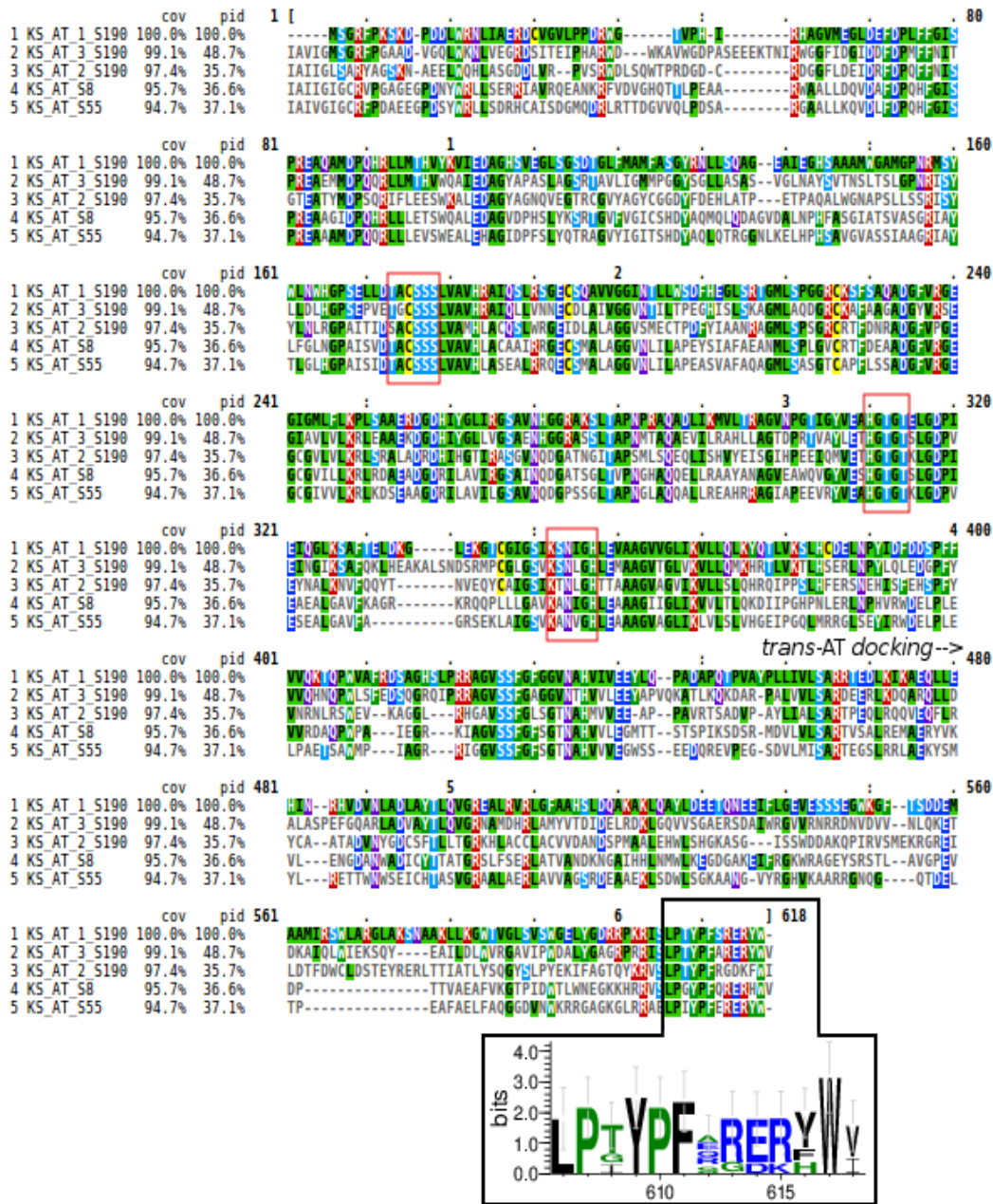


Abbildung 3.8: Sequenzvergleich der Ketosynthasen mit *trans*-AT docking Domänen aus *Acidobacterium* sp. S8, *Terracidiphilus* sp. S55 und *Bryocella* sp. S190. Herausgestellt ist jeweils der konservierte Bereich um die katalytische Triade (CHH) und der charakteristische C-Terminus

### 3.1.5 Charakterisierung von *trans*-AT PKS in Acidobakterien

Im *cis*-PKS-Typ bilden Ketosynthase und Acyltransferase ein Homodimer zwischen denen eine Linkersequenz ausgebildet ist. Zudem befinden sich am Homodimer sowohl am C- als auch am N-Terminus Linkersequenzen für die Kommunikation, wie zum Beispiel mit der ACP-Domäne [182]. Für den N-Terminus ist die Aminosäuresequenz IAIIG charakteristisch, während für den C-Terminus ein abschließendes Tryptophan typisch ist. Im PKS-Typ mit *trans*-AT ist der KS-AT-Linker durch eine AT-*docking* Domäne ersetzt (ab Position 458, Abbildung 3.8) und die Acyltransferase arbeitet als diskretes Enzym. Diese Domäne besteht aus etwa 100 Aminosäuren und besitzt zusätzlich zum *cis*-PKS-Typ die Sequenz LpxYPF vor dem abschließenden Tryptophan am C-Terminus [231]. Weitere Sequenz-Homologien sind im *trans*-AT PKS Typ nicht bekannt und können sich stark unterscheiden, was auf eine andauernde evolutionäre Entwicklung hindeutet [33]. Ein weiteres Merkmal für Polyketidsynthasen mit Kombination Ketosynthase und *trans*-AT Domäne ist ihre Funktion. Diese lässt sich anhand der katalytischen Triade der Ketosynthase determinieren. Neben der gewöhnlichen Funktion der Ketosynthasen zur Claisen-kondensation als Kettenverlängerung, welche durch die Triade Cystein-Histidin-Histidin vermittelt wird, sind auch Mutationen des Histidins zu anderen Aminosäuren bekannt, welche zum Funktionsverlust führen und als KS<sub>0</sub> bezeichnet werden [33]. Durch das *alignment* und des Vergleichs charakteristischer Sequenzbereiche aus Literaturangaben ist jedoch anzunehmen, dass es sich um valide katalytische Domänen handelt, welche die Claisen-kondensation durchführen.

Unter Betrachtung der durchgeführten Analyse hinsichtlich der speziellen Biochemie von *trans*-AT Polyketidsynthasen und geringer Sequenzähnlichkeit der Gene *ppsD*, *pks15* und *pksN*, als auch zu bekannten Genclustern, ist zusammenfassend das Potential zur Produktion eines neuen Metabolites aus diesen drei Acidobakterien gegeben und sollte weiterverfolgt werden. Deshalb wurde der Stamm *Terracidphilus* sp. S55 exemplarisch zur Erstellung einer Klonbibliothek verwendet (Kapitel 3.2).

### 3.1.6 Die Genomassemblierung von *Acidicapsa borealis*

Das Phylum *Acidobacteria* ist im Kontext der Antibiotika-Produktion wenig erforscht und beinhaltet in der bioinformatischen Analyse viele Stämme mit relevanten Genclustern. Dennoch gibt es in Relation zu anderen Phyla wenige Genomdaten, welche das Phylum weiter charakterisieren. Die Sequenzierung von *A. borealis* diente deshalb dazu, weitere Einblicke in den Metabolismus zu erhalten und diese für die Auswertung der Klonbibliothek zu nutzen. Zur Sequenzierung wurden zwei Technologien verwendet, die im Anschluss durch eine Hybridassemblierung, bestehend aus langen *reads* über 1.000 Basen (*Oxford Nanopore*) und kurzen, akkuraten *reads* (Illumina), für die vollständige Schließung des Genoms verwendet wurden.

Vor der Assemblierung wurden alle *reads* einem Filterschritt unterzogen, in dem *Nanopore reads* unter 1.000 Basen verworfen wurden. Als Resultat blieben 180.173 *reads* von ursprünglich 311.512 für die Assemblierung übrig. Die Abbildung 3.9 zeigt, dass der Großteil der Sequenzlängen ausgewogen um den Median von 2.392 Basen verteilt ist. Für die Hybridassemblierung wurden zusätzlich *reads* aus einem Illumina HiSeq-Lauf verwendet. Von ursprünglich 2x 27.210.738 *reads* wurden nach der Normalisierung 2x 1.832.463 *reads* eingesetzt. Die Normalisierung vor der Genomassemblierung bietet den Vorteil einer deutlich schnelleren Prozessierung und verringert die Gefahr einer Fehl-Assemblierung durch falsche Verzweigungen des de-Bruijn Graphen, verursacht durch *reads* mit geringer Sequenzabdeckung [160]. Das Genom wurde durch das Programm *Unicycler* zu einem circulärem Genom mit einer Größe von 6.103.018 Basen assembliert. Die Sequenzabdeckung, basierend auf reinen *Nanopore reads* nach Formel 2.1 und Verwendung des Medians als Sequenzlängen-Maß, betrug 296x. Die Genomabdeckung durch reine Illumina *reads* betrug 30x. Damit liegt die Sequenzabdeckung in dem Bereich anderer Hybridassemblierer, die in einem Durchgang ein vollständiges bakterielles Genom assemblieren [232].

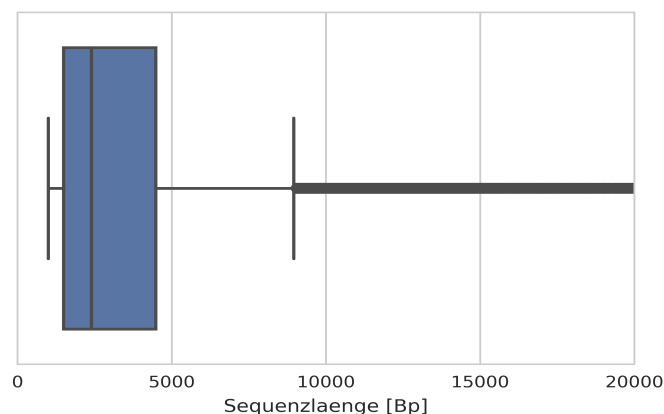


Abbildung 3.9: Sequenzlängen-Verteilung der gefilterten *reads* aus der *Oxford Nanopore* Sequenzierung für die Genomassemblierung von *A. borealis*

### 3.1.7 Die funktionale Annotation von *Acidicapsa borealis*

Für die Genvorhersage und funktionale Annotation des assemblierten Genoms wurde die Annotationsplattform *Genix* verwendet. Es wurde ein GC-Gehalt von 58,2 % berechnet und insgesamt 4.607 Gene vorhergesagt, von denen 4.560 kodierende Sequenzen (CDS) ausmachen. Von diesen CDS wurden 1.468 (32 %) als nicht charakterisiert bezeichnet und besitzen ausschließlich eine Uniprot-ID. Es wurden 2.263 kodierende Sequenzen in vorwärts-gerichteter Orientierung gefunden und 2.297 in rückwärts-gerichteter Orientierung. Desweiteren weist die Annotation 47 *tRNA*-Gene auf.

Um das Ergebnis der Annotation einzuordnen, wurde als Vergleich das Genom von *A. capsulatum* (NCBI-accession: CP001472) als erster genomsequenzierter Vertreter der Acidobakterien herangezogen. Aus diesem Genom sind 3.414 kodierende Sequenzen bekannt, von denen 1.010 (30 %) proteinogene Sequenzen als hypothetisch klassifiziert worden sind. Damit ist der Annotationserfolg von *A. borealis* als vergleichbar einzuschätzen.

#### Globale Untersuchung anhand von Genontologien

Im ersten Schritt der Genannotation wurden die Uniprot-IDs verwendet, um die korrespondierenden IDs für die Klassifizierung der Genontologien zu erhalten. Aus 4.560 Uniprot-IDs wurden 7.658 korrespondierende Genontologien extrahiert. Die hohe Anzahl an Genontologien gegenüber den Uniprot-IDs begründet sich darin, dass Proteine beziehungsweise Enzyme mehrfach an unterschiedlichen Prozessen beteiligt sein können. Diese Beobachtung wurde durch eine Normalisierung relativiert. Die Visualisierung auf der Abstraktionsebene der biologischen Prozesse erfolgte durch die Plattform REVIGO und ist in Abbildung 3.10 illustriert.

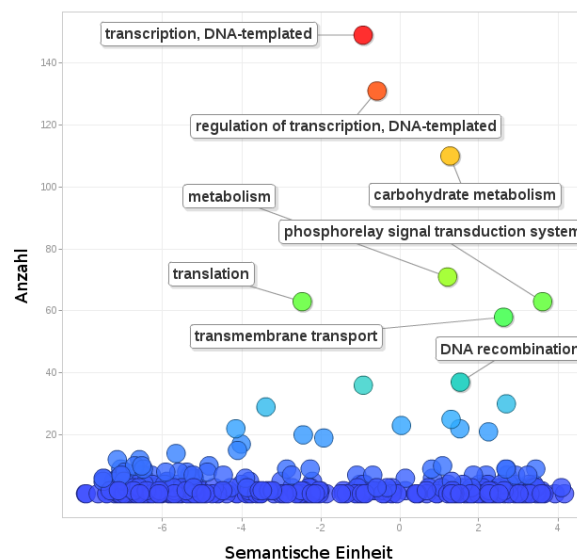


Abbildung 3.10: Quantifizierung biologischer Prozesse anhand von Genontologien des *A. borealis* Genoms

### 3.1.7 Die funktionale Annotation von *Acidicapsa borealis*

Die Abbildung zeigt als *Scatterplot* die Abundanzen der Genontologien in Abhängigkeit des semantischen Zusammenhangs [165]. Die Genontologien haben den Zweck, Proteine als Attribute in Kategorien einzuordnen, um ein systematisches Verständnis über konservierte Funktionen des Organismus zu erhalten [233] [234]. Die Abbildung 3.10 illustriert zusätzlich in Form des Farbverlaufs die Häufigkeit der Prozesse. Daraus ist zu entnehmen, dass der Carbohydrat-Metabolismus eine zentrale Rolle für dieses Bakterium einnimmt. Um den Zusammenhang der beteiligten Proteine und Enzyme besser zu verstehen, wurde deshalb dieser Stoffwechsel näher untersucht.

#### **Charakterisierung des Carbohydrat-Metabolismus**

Die Untersuchung der Genontologien auf der Ebene der biologischen Prozesse zeigte eine relativ hohe Abundanz von annotierten Proteinen, welche zu der Metabolisierung von Carbohydraten zugeordnet werden. Dieser Aspekt ist in zweifacher Hinsicht von Bedeutung. Zum einen können Proteine aus diesem Metabolismus Hinweise auf mögliche alternative Energiestoffquellen zu Glukose geben. Zum anderen ist die Glykosylierung in der Biosynthese von Naturstoffen ein wichtiger Faktor und entscheidet, wie am Beispiel von Erythromycin bekannt ist, über die antimikrobielle Wirkung des Moleküls.

Um den Carbohydrat-Metabolismus von *A. borealis* zu verstehen, wurden zwei Analysen durchgeführt, ein anhand von Uniprot-IDs durchgeführtes *mapping* auf den Stärke und Saccharose-Metabolismus, sowie eine HMM-basierte Untersuchung auf Proteine, hinsichtlich Carbohydrat-verstoffwechselnder Enzyme.

In der ersten Analyse des *mappings* konnten 21 Proteine dem Stärke und Saccharose-Metabolismus zugeordnet werden. Das Ergebnis ist als Karte in Abbildung 3.11 dargestellt und zeigt in grün hinterlegter Farbe die Enzyme, welche diesem Metabolismus zugeordnet worden sind. Daraus wird ersichtlich, dass eine Verstoffwechslung des komplexen Carbohydrats Cellulose über Cellodextrin und Cellubiose hin zu D-Glukose möglich ist. Desweiteren ist die Verstoffwechslung von Amylose über Glykogen, Maltose zu D-Glukose möglich. Das genetische Potential diese komplexen Moleküle zu verstoffwechseln, ist eine metabolische Besonderheit, die im Falle des Stärkeabbaus bereits experimentell nachgewiesen wurde [235]. Die Verstoffwechslung von Cellulose in kristalliner Form konnte in der Publikation nicht experimentell gezeigt werden. Jedoch kann der ausgeprägte Carbohydrat-Metabolismus als eine Adaption für moosreiche Waldhabitats angesehen werden [236]. Bakterien mit ähnlichen Eigenschaften wurden bisher aus dem Phylum *Firmicutes* beschrieben, darunter *Clostridium cellulolyticum* welches durch die Verwertung Cellulose auch von biotechnologischer Relevanz, wie etwa der Bioethanol-Herstellung ist [237].



### 3.1.7 Die funktionale Annotation von *Acidicapsa borealis*

In der zweiten Analyse wurden alle kodierenden Sequenzen des Genoms mittels einer HMM-Analyse gegen die *Carbohydrate-Active enZymes Database* (CaZy) verglichen und quantitativ in Form einer Tabelle dargestellt (3.11). Die Daten in der Tabelle zeigen 52 % aller gefundenen Einträge. Als allgemeine Beobachtung lässt sich feststellen, dass sehr häufig Enzyme mit hydrolytischer Funktion zu finden sind. Unter diesen Enzymfamilien stehen zwei Familien, welche Enzyme umfassen, die für die Metabolisierung von Cellulose benötigt werden. Darunter die Familie der Glykosidhydrolase Familie 9. Diese umfasst die Enzyme Cellodextrinase (EC 3.2.1.74) und Cellobiohydrolase (EC 3.2.1.91). Die zweite Familie ist die Glykosidhydrolase Familie 74, welche das Enzym Cellobiohydrolase (EC 3.2.1.150) beinhaltet.

Die Ergebnisse zeigen weitere Enzymfamilien für die Verstoffwechslung von Zuckermoleküle als Energiestoffquelle. Eine ist die Glykosidhydrolase Familie 3, welche Enzyme zur Spaltung von Xylan und Glucan beinhaltet. Die Glykosidhydrolase Familie 92 umfasst ausschließlich Enzyme für die Verstoffwechslung des Monosaccharids Mannose. Die Glykosidhydrolase Familie 72 beinhaltet Enzyme für die Verstoffwechslung von Galaktose. Die Glykosidhydrolase Familie 29 impliziert in diesem Zusammenhang zwei Enzyme für die Hydrolyse von Fruktose.

Neben Enzyme für den Katabolismus wurden zwei Enzymfamilien für Transferreaktionen entdeckt. Diese sind die Glykosyltransferase Familien 2 und 4. Beide Familien beinhalten Enzyme mit diversen Aufgaben, wie zum Beispiel die Übertragung von Phosphat in Form von Guanidin-Diphosphat (GDP) auf Zuckermoleküle.

Weitere Enzymfamilien in der Tabelle umfassen die Carbohydrolase Familien 1 und 10. Zu diesen Familien zählen Enzyme, die unter anderem an dem Umbau von Fettsäuren beteiligt sind. Beispiele sind die Diacylglycerol O-Acyltransferase (EC 2.3.1.20) und die Sterolesterase (EC 3.1.1.13). Fettsäuren machen den Großteil der bakteriellen Zellmembran aus oder können als Signalmoleküle agieren.

Enzymfamilien, die im Zusammenhang mit der Glykosylierung von Peptiden oder Proteinen stehen, umfassen die Glycosyltransferase Familie 83 und die Carbohydrolase Familie 14.

Sowohl die KEGG-Analyse, als auch die Untersuchung durch CaZy bestätigt die Anwesenheit von Genen die für Enzyme kodieren, welche am Stärke- und Saccharose-Metabolismus beteiligt sind. Damit wird die experimentelle Beobachtung aus der Erstpublikation von *A. borealis* durch Kulichevskaya *et al.* (2012) [235] zum Stärkeabbau bioinformatisch erweitert.

### 3.1.7 Die funktionale Annotation von *Acidicapsa borealis*

Tabelle 3.11: Die 15 häufigsten Protein- oder Enzymfamilien des Carbohydrat-Stoffwechsels von *A. borealis*

Position	HMM	Beschreibung	Enzymklasse	Anteil [%]
1	CBM56	Carbohydrat-Bindemodul Familie 56	-	10
2	CE1	Carbohydratesterase Familie 1	Hydrolase EC 3	6
3	GT2	Glycosyltransferase Familie 2	Transferase EC 2	5
4	GH109	Glycosidhydrolase Familie 109	Hydrolase EC 3	5
5	CE10	Carbohydratesterase Familie 10	Hydrolase EC 3	5
6	GT4	Glycosyltransferase Familie 4	Transferase EC 2	4
7	GH3	Glycosidhydrolase Familie 3	Hydrolase EC 3	3
8	GT83	Glycosyltransferase Familie 83	Transferase EC 2	3
9	GH29	Glycosidhydrolase Familie 29	Hydrolase EC 3	3
10	GH92	Glycosidhydrolase Familie 92	Hydrolase EC 3	2
11	GH2	Glycosidhydrolase Familie 2	Hydrolase EC 3	2
12	CE14	Carbohydratesterase Familie 14	Hydrolase EC 3	2
13	GH27	Glycosidhydrolase Familie 27	Hydrolase EC 3	2
14	GH9	Glycosidehydrolase Familie 9	Hydrolase EC 3	1
15	GH74	Glycosidhydrolase Familie 74	Hydrolase EC 3	1

### Identifizierte Gencluster

Die Detektion von Biosynthese-Genclustern erfolgte mit *antiSMASH*. Es wurden zehn Gencluster detektiert (Tabelle 3.12). Um *A. borealis* mit dem Datensatz aus Kapitel 3.1.1 in Bezug zu setzen, wurde der Quotient aus Gencluster-Anzahl und Genomgröße berechnet. Dieser ergab einen Wert von 1,6 und liegt somit über dem Grenzwert der Kategorisierung von 0,88 aus dem Regressionsmodell der alternativen Bakterien. Daher wurde *A. borealis* ebenfalls in die Kategorie überdurchschnittlich eingestuft und entspricht dem Mehrheit von Acidobakterien als relevante Ressource für Antibiotika.

Im weiteren Verlauf der Analyse wurde eine funktionale Annotation von Resistenzgenen der Gencluster und des Genoms durchgeführt [238]. Neben der Zerstörung oder Modifikation von Antibiotika als Resistenz, werden auch *core*-Gene in Betracht gezogen, die an essentiellen Prozessen wie dem Aminosäuren- oder Energiemetabolismus sowie an der Transkription beteiligt sind und daher potentielle Ziele für Antibiotika darstellen. Diese Gene können im Genom mehrfach vorkommen, wenn sie beispielsweise der Angriffspunkt von Antibiotika im Organismus sind. Für diese Analyse wurde das Programm ARTS verwendet. Die Analyse durch das Programm basiert auf drei Kriterien, der Proximität von Markergenen (*core*- oder Resistenzgene) zum Gencluster, der Genduplikation und einer phylogenetischen Untersuchung, um festzustellen, ob *core*-Gene durch horizontalen Gentransfer in den Organismus eingebracht worden sind [103]. Zur Identifizierung dieser Markergene wurden zwei integrierte Datenbanken verwendet. Zum einen auf ein Actinobakterien-basiertes Referenzset, das aus bekannten Resistenzgenen besteht und der TIGRFAM Datenbank zur Identifizierung von *core*-Genen. Diese Analyse wurde mit dem assemblierten Genom durchgeführt. Es konnten drei Gene identifiziert werden, welche zu jeweils einen Gencluster bioinformatisch assoziiert sind (Tabelle 3.13).

Tabelle 3.12: Identifizierte BGC im assemblierten Genom von *A. borealis*

Position	BGC	Länge [Bb]	Genomlocus [Bb]
1	<i>terpene</i>	22.033	14.321 - 36.354
2	<i>t3pks</i>	41.083	35.523 - 76.606
3	<i>bacteriocin</i>	10.876	700.604 - 711.480
4	<i>terpene</i>	29.715	831.232 - 860.947
5	<i>nrps</i>	48.753	2.704.414 - 2.753.167
6	<i>lassopeptide</i>	21.987	2.794.335 - 2.816.322
7	<i>nrps-t1pks</i>	84.597	3.708.614 - 3.793.211
8	<i>other</i>	43.309	3.859.033 - 3.902.342
9	<i>nrps-t1pks</i>	66.854	5.246.806 - 5.313.660
10	<i>lassopeptide</i>	22.176	5.481.123 - 5.503.299

### 3.1.7 Die funktionale Annotation von *Acidicapsa borealis*

Tabelle 3.13: Identifizierte Markergene der Biosynthese-Gencluster in *Acidicapsa borealis*

Position	Gencluster	Kategorie	Beschreibung	Funktion
4	<i>terpene</i>	<i>core</i>	<i>uvra</i>	DNS Metabolismus
5	<i>nrps</i>	<i>core</i>	<i>efp</i>	Proteinsynthese
9	<i>nrps-t1pks</i>	Resistenz	ABC efflux	-

Das erste identifizierte Gen ist *uvra* aus dem Gencluster *terpene* (Position 4). Für dieses Gen wurden die Kriterien Gencluster-Proximität und Duplikation erfüllt, jedoch nicht das phylogenetische Kriterium. Das Gen *uvra* kodiert die A-Untereinheit die Bestandteil des bakteriellen SOS-Systems ist und Schäden an der DNS erkennt, die zum Beispiel durch UV-Bestrahlung entstanden sind [239]. Das zweite identifizierte Gen *efp*, kodiert den Elongations-Faktor P der in der Translation an Prolin-reichen Proteinen beteiligt ist [240] [241] [242]. Dieses Gen entspricht allen Kriterien, jedoch ist das Gen nicht direkt im Gencluster lokalisiert. Als drittes Gen, das der Kategorie Resistenz zugeordnet wird, wurde der ABC-Transporter als Efflux-Pumpe identifiziert. Diese Familie an Proteinen vermittelt einen ATP-getriebenen Transport durch die Zellmembran und kann je nach Substrat-, Transport- und Sequenzähnlichkeit, in weitere Untergruppen eingeteilt werden [243]. Eine weitere Spezifikation durch ARTS in eine spezifische Transporterfamilie ist nicht beschrieben. Literaturhinweise bezüglich ABC-Transporter vermittelte Resistenz gegenüber Streptogramin-Antibiotika [244], wie zum Beispiel das NRPS-PKS synthetisierte Virginiamycin S [245], implizieren aber in diesem Zusammenhang ein plausibles Resistenzgen, welches dem Gencluster-Typ entspricht.

### 3.1.8 Acidobakterien sind assoziiert mit *Coptotermis niger*

Die Isolation und Kultivierung neuer Bakterien unter Standard-Laborbedingungen ist stark limitiert (Kapitel 1.3). Daher bedarf es zusätzlicher Methoden um die Wahrscheinlichkeit für die Entdeckung neuer Antibiotika zu erhöhen. Durch die Anfertigung von Klonbibliotheken und der heterologen Genexpression von fremder bakterieller DNS, besteht die Möglichkeit, die genetische Information für die Biosynthese von neuen Metaboliten kultivierungs-unabhängig zu nutzen.

Neben der Anwendung molekularbiologischer Methoden als kultur-unabhängige Strategie, ist die Kenntnis und Erschließung von Habitaten mit hoher mikrobieller Diversität essentiell. In diesem Zusammenhang gelten Termitennest-assoziierte Bakterien als aussichtsreiche Quelle für neue Antibiotika. Termiten haben ein schwaches Immunsystem und werden symbiontisch durch Bakterien unterstützt, um pathogene Mikroorganismen abzuwehren. Deshalb wird eine hohe Anzahl an Antibiotika-Produzenten angenommen [246]. In diesem Experiment wurde dazu auf ein Termitennest des Genus *Coptotermes niger* zurückgegriffen und neben der Erstellung einer Klonbibliothek, das Termitennest-Habitat taxonomisch untersucht. Dieser Genus zählt als Anhänger der taxonomischen Familie *Rhinotermitidae* zu den niederen Termiten, ernährt sich primär von Holz und ist eng benachbart mit seinem Nahrungssubstrat [247]. Auf Grund der Komplexität des Termitennests, wurden die Bakterien durch eine Nykodenz-Dichtezentrifugation aus dem Nest gewonnen. Abbildung 3.12 zeigt die einzelnen Phasen nach der Nykodenz-Dichtezentrifugation. Für die anschließende Zellzahlbestimmung und DNS-Isolierung wurde die Zellschicht entnommen.

### 3.1.8 Acidobakterien sind assoziiert mit *Coptotermis niger*

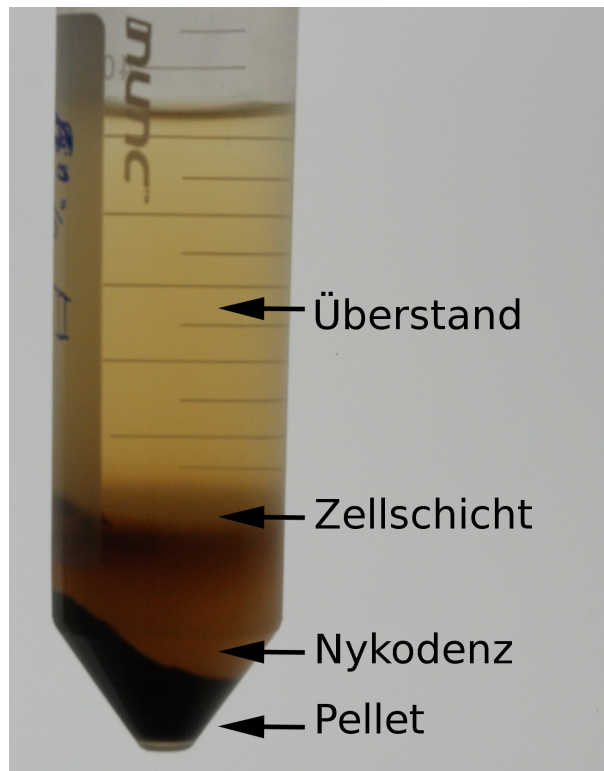


Abbildung 3.12: Gewinnung von Bakterien aus dem turaxiierten Termitennest nach Zentrifugation mit Nykodenzzusatz

Für die Bestimmung der Zellzahl, auf Basis der Fluoreszenz, wurde eine Verdünnung angefertigt und die isolierten Bakterien mit einem grün-fluoreszierenden SYTO-Farbstoff angefärbt. Zur Kalibrierung wurden *beads* als Referenz verwendet, die auf eine definierte Konzentration eingestellt wurden. Das fluoreszierende Signal wurde in einer Grafik dargestellt, welche die Intensität des Fluoreszenzsignals (FL1) in Abhängigkeit der Signalgröße als logarithmische Skalierung zeigt, erfasst und durch ein *gate* eingegrenzt (Abbildung 3.13). Durch das Verhältnis von Signalen der Zellen (7.204) und der *beads* als Referenz (2.264), wurde anhand der Formel 2.3 und des Verdünnungsfaktors von 80 eine Zellzahl von  $2,55 \times 10^8$  Zellen/mL berechnet, die aus der Nykodenz-Dichtezentrifugation gewonnen werden konnten.

### 3.1.8 Acidobakterien sind assoziiert mit *Coptotermis niger*

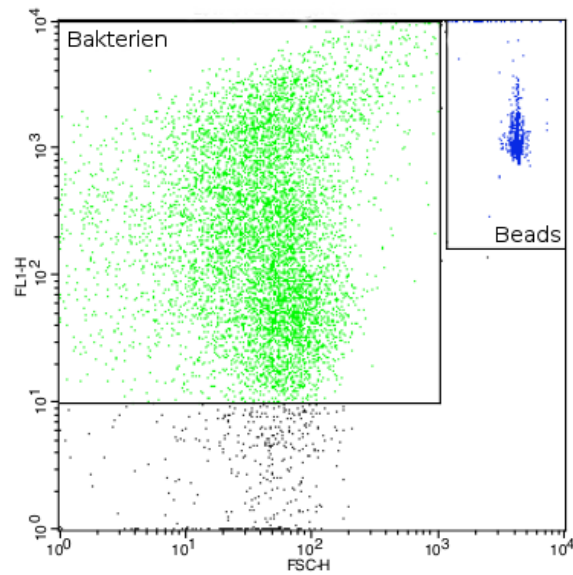


Abbildung 3.13: Zellzählung der mittels Nykodenz gewonnenen Zellen aus dem Termitennest. FL1: Fluoreszenzintensität, FSC: *forward scatter* beziehungsweise Signalgröße

Im weiteren Verlauf dieses Experiments wurde ein Aliquot für die Erstellung der Klonbibliothek (Kapitel 2.5) und ein weiteres Aliquot für die taxonomische Klassifizierung der gewonnenen Bakterien verwendet (Kapitel 2.4.4). Die taxonomische Klassifizierung der gewonnenen Bakterien basierte auf der Sequenzierung hypervariabler Regionen des 16S rRNA-Gens, die auch als *small-subunits* (SSU rRNA) bezeichnet werden. Diese sind unterschiedlich stark konserviert und gelten als Standard für die taxonomische Klassifizierung mittels Illuminasequenzierung [248] [249]. Auf Grund der Sequenzierung, die auf 300 – 400 Basenlänge limitiert ist und der hinreichenden, taxonomischen Auflösung [250], wurde die hypervariable Region V3 ausgewählt, durch *primer* amplifiziert und sequenziert.

Die bioinformatische Prozessierung der *reads* und taxonomische Klassifizierung erfolgte anhand der Erstellung von *operational taxonomic units* (OTUs). Dazu wurden alle *reads* durch einen paarweisen Sequenzvergleich untersucht und *reads* mit einer Sequenzähnlichkeit von >98 % zu einem OTU geclustert. Die Quantifizierung erfolgte durch ein konsekutives *mapping* der *reads* auf die OTUs. Das Ergebnis der taxonomischen Klassifizierung ist in Abbildung 3.14 dargestellt.

### 3.1.8 Acidobakterien sind assoziiert mit *Coptotermis niger*



Abbildung 3.14: Ergebnis der taxonomischen Klassifizierung auf Klassen-Ebene der gewonnenen Bakterien aus dem *C. niger* Termitennest

Das Ergebnis der taxonomischen Klassifizierung zeigt zwei dominante Phyla, die der Actinobakterien mit 37 % und die der Proteobakterien mit 47 %. Das Phylum *Proteobacteria* wird dominiert durch die taxonomische Klasse der *Alphaproteobacteria* (95 %) und einen geringen Anteil von *Gammaproteobacteria* (4 %). Mit einem marginalen Anteil von 0,8 % und 0,3 % sind *Deltaproteobacteria* und *Betaproteobacteria* vertreten. Abseits der beiden dominanten Phyla setzt sich der dritte Anteil an OTUs aus Vertretern alternativer Bakterien zusammen. 0,8 % der OTUs wurden dem Phylum *Chloroflexi* zugerechnet, 1 % dem Phylum *Bacteroidetes*. Mit einem Anteil von insgesamt 5,2 % ist

### 3.1.8 Acidobakterien sind assoziiert mit *Coptotermis niger*

das Phylum *PVC* vertreten, das sich in 3 % *Placntomycetes*, 2 % *Verrucomicrobia* und 0,2 % *Chlamydiae* aufteilt. Mit dem höchsten Anteil von 8 % ist das Phylum *Acidobacteria* unter den alternativen Bakterien klassifiziert worden. Das Phylum gliedert sich im weiteren taxonomischen Verlauf in die Familie der *Solibacteraceae* (54 %), *Acidobacteriaceae* (42 %), sowie mit jeweils 2 % in *Blastocatellaceae* und *Holophaga*.

Das Vorkommen an *Actinobacteria* in Termitennestern diverser Arten, konnte auch in anderen Studien belegt werden [174]. Bakterien, die in Termiten als Endosymbionten vorkommen können, gehören zu *Elusimicrobia*, *Bacteroidetes*, *Proteobacteria* sowie *Actinobacteria* und sind an der Verstoffwechslung von Lignocellulose, Aminosäuren, Stickstoff oder an der Cofaktor-Produktion beteiligt [246]. Die taxonomische Klassifizierung zeigt, dass das *C. niger* Termitennest alternative Bakterien beinhaltet, die besonders im Kontext neuer Gram-negativer Antibiotikaproduzenten eine bedeutende Ressource darstellen. Die Untersuchung bestätigt außerdem, dass Acidobakterien nicht nur im Habitat Boden, Sediment und der Rhizosphäre vertreten sind [251] [252] [253].

Als Ausblick für weiterführende Untersuchungen von Klonbibliotheken aus Metagenomen, ist eine 16S rRNA-Gen basierte Untersuchung eine zusätzliche Möglichkeit, die effektive Diversität von DNS innerhalb der Bibliothek abzuschätzen [254].

## 3.2 Erstellung einer Klonbibliothek mit DNS aus Acidobakterien

Für die Wahl der Bakterien zur Etablierung eines Protokolls für Klonbibliotheken, wurden die bioinformatischen Erkenntnisse aus Kapitel 3.7 berücksichtigt. Die Analyse zeigte, dass 80% an Acidobakterien der Kategorie überdurchschnittlich entsprechen, darunter auch der Stamm *Terracidiphilus* sp. S55, welcher außerdem ein Gencluster mit *trans*-AT Polyketidsynthase aufweist. Dieser Stamm wurde durch die Publikation von Garcia-Fraile *et al.* (2016) [255] zu *Terracidiphilus gabretensis* umbenannt. Jedoch sind keine Informationen bezüglich Sekundärmetabolite bekannt.

Als zweites Bakterium wurde der Stamm *Acidicapsa borealis* für die Erstellung einer Klonbibliothek verwendet. Dieser Stamm ist wie *Terracidiphilus* sp. S55 Gram-negativ und wurde im Rahmen der vollständigen Genomsequenzierung ausgewählt. Der Stamm wurde von Kulichevskaya *et al.* (2012) [235] publiziert und enthält ein rosafarbenes Pigment. Beide Stämme gehören der *sub division* 1 des Phylums an.

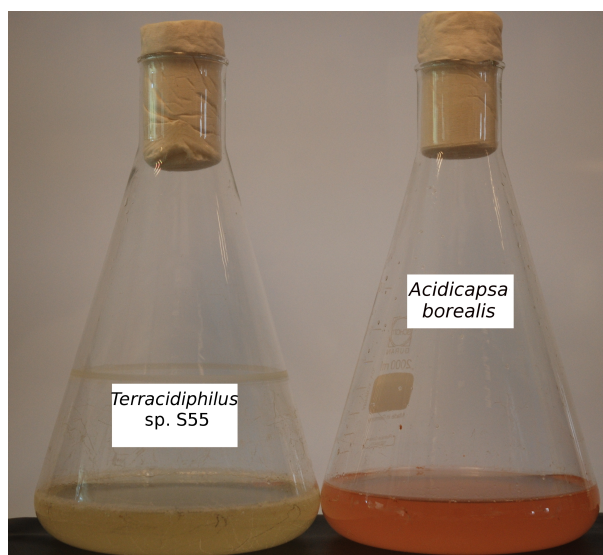


Abbildung 3.15: Kulturen der zwei Acidobakterien *Terracidiphilus* sp. S55 (links) und *A. borealis* (rechts) vor der Zellernte

### 3.2.1 Molekularbiologische Schritte der Erstellung von Klonbibliotheken

Die molekularbiologischen Schritte zur Etablierung der Klonbibliotheken umfassten die Isolation hochmolekularer DNS, Klonierung und Transformation (Kapitel 2.5). Diese Schritte wurden analog für jede Probe durchgeführt. Sowohl der Stamm *Terracidiphilus* sp. S55 als auch *A. borealis* wurden direkt nach der Kultivierung für die DNS-Isolation präpariert. Die Erstellung der Klonbibliothek aus Termitennest-assoziierten Bakterien erfolgte durch ein zuvor isoliertes Bakterienpellet mittels Nykodenz-Dichtezentrifugation. Für die erfolgreiche Expression ganzer Gencluster im Kilobasenbereich ist die Extraktion

### 3.2.1 Molekularbiologische Schritte der Erstellung von Klonbibliotheken

von intakter, hochmolekularer DNS die Voraussetzung [256]. Die Besonderheit hochmolekularer DNS besteht in der Empfindlichkeit gegenüber Scherkräften [257]. Eine oft verwendete Methodik ist daher die Immobilisierung des Zellpellets in Agarose, bei der die Zellen schonend lysiert werden [258] [259]. Dies erfolgt durch eine chemisch enzymatische Behandlung mit abwechselnden Waschschritten [260]. Die enzymatische Lyse ist der chemischen vorgeschaltet und besteht aus drei Schritten. Im ersten Inkubationsschritt erfolgte die Lyse der Zellwand durch Lysozym, das die Bindung zwischen N-Acetylmuraminsäure und N-Acetylglucosamin des Peptidoglykans hydrolysiert. Im zweiten Schritt wurde die Zellyse mittels Achromopeptidase gegenüber lysozymresistenten Bakterien, beziehungsweise Bakterien mit Gram-positiver Zellwandeigenschaft durchgeführt. Diese Behandlung war im Kontext gewonnener Bakterien aus dem Termitennest notwendig [261]. Im dritten Schritt wurden Proteine durch die Proteinase K verdaut und die DNS für die anschließende Restriktion von Zellproteinen befreit. Die Detergenzien *Sodium Dodecyl Sulfate* (SDS) und Sarkosyl verstärken die Wirkung des Enzyms und tragen simultan zur Disruption der Zellbestandteile bei.

Um die DNS klonierbar zu machen, erfolgte anschließend die partielle Restriktion. Da für diesen Schritt die *multiple cloning site* (MCS) von pBeloBAC11 (Abbildung 2.3) ausschlaggebend ist, wurde das Restriktionsenzym BamHI gewählt. Aus diesem Grund musste der partielle Verdau durch eine lange Äquilibrierung auf Eis und einer kurzen Reaktionszeit bei optimalen 37 °C erfolgen. Abbildung 3.16 zeigt den Effekt des Partialverdau nach Auftrennung durch die Pulsfeld-Gelelektrophorese (PFGE) [262].

### 3.2.1 Molekularbiologische Schritte der Erstellung von Klonbibliotheken

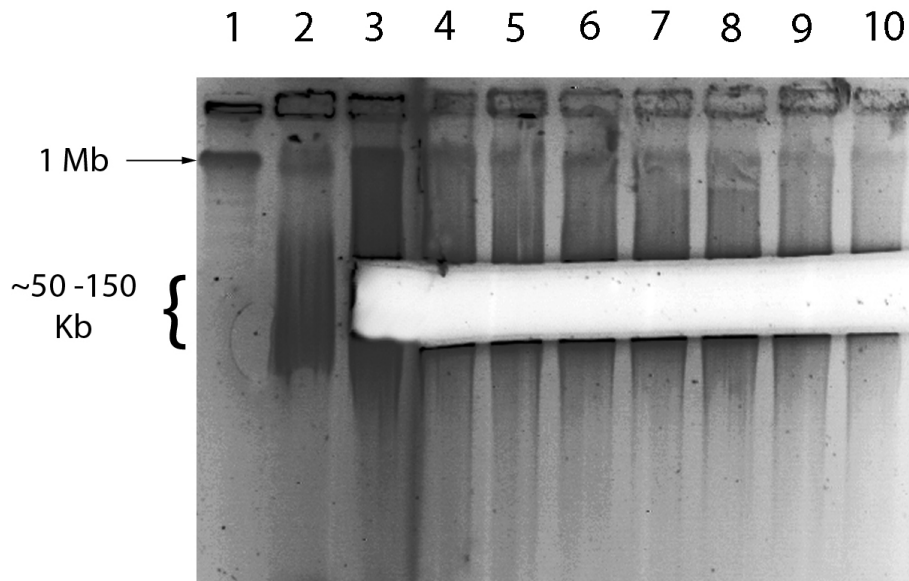


Abbildung 3.16: Beispiel der Auftrennung von partial verdauter DNS mittels PFGE. Tasche 1: Marker, 2: *A. borealis* DNS unverdaut, Tasche 3 - 10 *A. borealis* DNS partial verdaut und weiterverarbeitet für die Klonierung (Der Größenbereich zwischen 50 - 150 Kb wurde zuvor ausgeschnitten und anschließend gefärbt)

In Abbildung 3.16 ist die DNS nach Färbung mit Ethidiumbromid unter UV-Licht sichtbar. Der Vergleich der DNS in den Geltaschen 2 und 3 zeigt einen deutlich längeren Schmier, als in der Tasche 2, was durch BamHI verursacht wurde. Durch Abgleich der Markerbanden, konnte der Größenbereich von circa 50 - 150 Kilobasen abgeschätzt werden. Um die DNS nicht durch Mutation zu beschädigen, wurden zuvor die Geltaschen 1 - 3 angefärbt und als Schnittmuster für die Geltaschen 4 - 10 verwendet. Nach dem Ausschneiden wurden diese Geltaschen für eine Nachkontrolle ebenfalls gefärbt (Abbildung 3.16).

Die ausgeschnittenen Gelstücke wurden gewaschen, um sie von Borat zu befreien und dem Agaraseverdau zugesetzt. Anschließend konnte die isolierte DNS-Konzentration gemessen werden (Tabelle 3.14). Die erzielten Konzentrationen lagen zwischen 4 - 5 ng/ $\mu$ L. Auf Grund der geringen Menge wurde die Ligationsreaktion auf 100  $\mu$ L Gesamtvolumen erhöht, um das molare Verhältnis von DNS zu BAC-Vector zu erhöhen und die Reaktion zu optimieren. Eine genaue Berechnung konnte aufgrund der zufallsbedingten Insertlänge nicht bestimmt werden. Generell wird für eine BAC-Ligation ein molares Verhältnis, Insert zu BAC von 10 zu 1 angestrebt und die Ligationsreaktion auf mehrere Stunden ausgedehnt [263] [264] [265].

Um die Transformations-Effizienz zu optimieren, wurde der Ligationsansatz mittels Filterpapier und PEG 8000 aufkonzentriert. Dadurch zeigte sich nach der Transformation eine deutliche Steigerung der Kolonie-bildenden-Einheiten (KBE). Im weiteren Verlauf der

### 3.2.1 Molekularbiologische Schritte der Erstellung von Klonbibliotheken

Auswertung wurde für eine Quantifizierung der Klone durch eine Blau-Weiß-Selektion mittels 5-Brom-4-chlor-3-indoxyl- $\beta$ -D-galactopyranosid (X-Gal) durchgeführt, um Klone mit Insert von leeren Klonen zu unterscheiden (Abbildung 3.17).

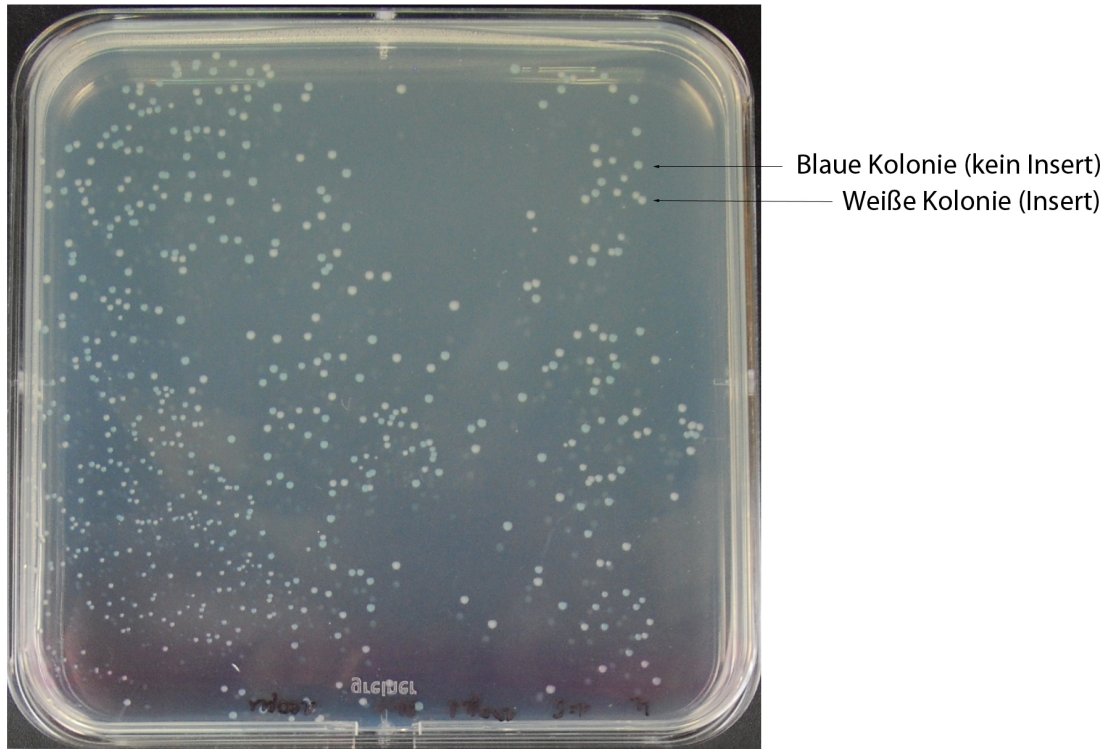


Abbildung 3.17: Blau-Weiß-Selektion anhand eines Ausstrichs von 100  $\mu$ L Transformations-Volumen (1 zu 5 verdünnt) auf eine Agarplatte (12 cm im Durchmesser), 24 Stunden nach der Transformation von klonierter DNS aus *A. borealis*

Blaue Kolonien resultieren durch die enzymatische Hydrolyse der beta-Galactosidase, welche durch Isopropyl- $\beta$ -D-thiogalactopyranosid (IPTG) induziert wird und X-Gal zu Galaktose und 4-Chloro-3-Brom-Indigo metabolisiert. Der synthetische Zucker IPTG bindet an den lacI-Repressor, wodurch der Laktose-Operator freigesetzt wird und die RNS-Polymerase II das lacZ-Gen transkribieren kann [266]. Die Expression erfolgt konstitutiv, da IPTG nicht metabolisiert werden kann. Die Restriktion der BamHI-Schnittstelle in dem lacZ-Gen, gefolgt von der DNS-Insertion, zerstört den Leserahmen des Gens worauf diese Reaktion nicht mehr stattfinden kann und die Färbung nicht erfolgt.

Auf diese Weise wurden insgesamt 960 Kolonien aus DNS von *A. borealis* und Zellen des *C. niger* Termitennests ausgewählt, konserviert und für die funktionale Untersuchung (Kapitel 3.3) verwendet. Die Ergebnisse aus insgesamt drei Ansätzen zur Erstellung von Klonbibliotheken sind in Tabelle 3.14 dargestellt. Die meisten Kolonien konnten mit der DNS aus *A. borealis* DNS angefertigt werden. Aus der Metagenomprobe des *C.*

### 3.2.1 Molekularbiologische Schritte der Erstellung von Klonbibliotheken

*niger* Termitennests wurden 80 Kolonien ausgewählt. Auf Grund der geringen Anzahl von neun Kolonien aus dem Ansatz von *Terracidiphilus* sp. S55, wurde auf eine weitere Untersuchung verzichtet. Die geringe Anzahl aus diesem Ansatz kann unterschiedliche Gründe haben. Erstens kann im Zuge der DNS-Isolation, etwa durch Kontamination mit DNase die DNS degradiert worden sein, weshalb weniger funktionsfähiges Insert für die Ligation zur Verfügung stand. Zweitens ist eine suboptimale Restriktion des pBelo-BAC11s möglich, bei der zu wenig Vector linearisiert oder durch Funktionsverlust der Phosphatase eine Re-Zirkularisation stattgefunden hat.

Tabelle 3.14: Zusammenfassung der Erstellung von Klonbibliotheken aus *Terracidiphilus* sp. S55, *A. borealis* und *C. niger* Termitennest

DNS-Quelle	<i>Terracidiphilus</i> sp. S55	<i>A. borealis</i>	Metagenom
Phylum	<i>Acidobacteria</i>	<i>Acidobacteria</i>	-
Material	400 mL Kultur	400 mL Kultur	291 mg <sup>1</sup>
Zellen/mL	3,94 x 10 <sup>4</sup>	1,52 x 10 <sup>7</sup>	2,55 x 10 <sup>8</sup>
Isolierte DNS (ng/µL)	4,1	4,2	5
DNS in Ligation (ng)	369	377	445
Totale KFU	9	1040	553
KFU/µg	25	363	1243
Mittel KBE/mL	183	2,08 x 10 <sup>4</sup>	1,1 x 10 <sup>4</sup>
STABW KBE/mL	238	7,19 x 10 <sup>3</sup>	6,69 x 10 <sup>3</sup>
Kolonien mit Insert	150	2,05 x 10 <sup>4</sup>	2,02 x 10 <sup>3</sup>
Kolonien ohne Insert	33	3,17 x 10 <sup>2</sup>	8,87 x 10 <sup>3</sup>
Kolonien mit Insert (%)	82	98	20
Generierte Platten	-	11	1
Untersuchte Klone	-	880	80
Klonbibliothek total	960		

<sup>1</sup>Zellpellet nach der Nykoden-Dichte-zentrifugation

### 3.3 Funktionale Untersuchung der Klonbibliotheken

#### 3.3.1 Restriktionsverdau zur Insertkontrolle

Nach der Quantifizierung der Klone erfolgte die Kultivierung im 1,4 mL Maßstab in 96-*deep well* Platten (Kapitel 2.6). Ausgehend von einer Kultivierung über 24 Stunden, wurde das Kulturvolumen für die funktionale Untersuchung eingesetzt.

Parallel dazu wurden exemplarisch Klone ausgewählt, um die Insertlänge der isolierten DNS durch einen vollständigen Restriktionsverdau zu überprüfen (Kapitel 2.7.1). Für diesen Versuch wurde das Restriktionsenzym NotI eingesetzt (Abbildung 3.18). Dieses Enzym benötigt eine Erkennungssequenz von acht Basen und schneidet deshalb relativ selten.

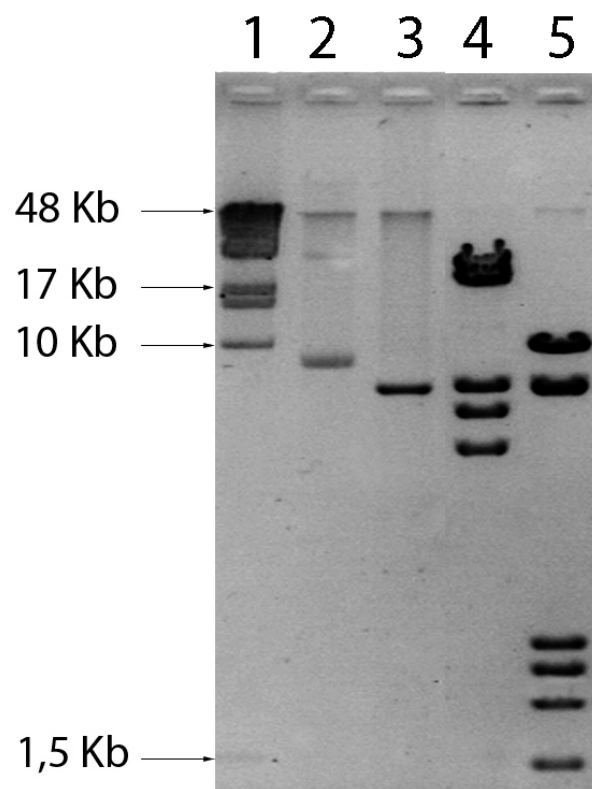


Abbildung 3.18: Restriktionsverdau zur Insertlängenkontrolle. Tasche 1: Marker, Tasche 2: pBeloBac11 zirkulär, Tasche 3: pBeloBac11 verdaut, Tasche 4: JF 3 A5, Tasche 5: Klon mit *Terracidiphilus* sp. S55 DNS

Das Insert aus JF 3 A5 zeigt drei Banden aus DNS von *A. borealis*, die zusammengerechnet etwa eine Länge von 40 Kilobasen ergeben. Der Klon mit DNS aus *Terracidiphilus* sp. S55 als DNS-Quelle zeigt fünf Banden mit einer Gesamtlänge von etwa 25 Kilobasen. Der verwendete Klon (JF 14 A4) aus dem Metagenom trug kein Insert und ist nicht gezeigt. Mit diesem Restriktionsverdau konnte exemplarisch bestätigt werden, dass die

### 3.3.1 Restriktionsverdau zur Insertkontrolle

Klonierung von bakterieller DNS aus Acidobakterien unter Verwendung der etablierten Methoden möglich ist. In Betracht der Größenordnung identifizierter Gencluster aus *A. borealis* (Tabelle 3.12), ist daher eine vollständige Abdeckung der Gencluster in 8 von 10 Beispielen möglich.

### 3.3.2 Detektion von NRPS-Genfragmenten durch PCR

Durch die Polymerase-Kettenreaktion zur Untersuchung der Klonbibliotheken können schnell und präzise Markergene von Biosynthese-Genclustern detektiert werden. Diese Methode wurde für Klone mit DNS aus *A. borealis*, sowie aus Bakterien mit Assoziation zu *C. niger* (Termitennest) im 96-well Format angewendet (Kapitel 2.7.2).

Die verwendeten *primer* binden auf der kodierten Adenylierungsdomäne der nicht-ribosomalen Peptidsynthase, die für den Substrateinbau während der Biosynthese verantwortlich ist und amplifizieren ein Fragment von 700 Basen [168]. In der Summe wurden 960 Klone durch diesen Ansatz untersucht. Bezüglich der Bibliothek aus *A. borealis* konnten vier Klone mit einer charakteristischen Bande gefunden werden (Abbildung 3.19 bis 3.21). Es wurden die Klone JF 7 A7, JF 7 E10, JF 8 B12 und JF 12 B9 mit einem spezifischen Bandenmuster identifiziert. Nach Abgleich der Gencluster-Analyse (Kapitel 3.1.7), kommen die Gencluster 5: *nrps*, 7: *nrps-t1pks* und 9: *nrps-t1pks* als Möglichkeit in Betracht, mindestens partiell kloniert worden zu sein. Die eindeutige Lokalisation der klonierten Bereiche ist im nachfolgenden Kapitel 3.3.3 mit Hilfe der Sequenzierung festgestellt worden.

Desweiteren wurden 80 Klone der Bibliothek aus dem Termitennest-Metagnom mittels PCR untersucht (Abbildung 5.2). Als Positivkontrolle wurde der Streptomyces FhG 336b mitgeführt (Abbildung 5.3). Die Untersuchung nach Markergenen ist ein komplementärer Ansatz zur direkten Untersuchung nach antibakterieller Aktivität (Kapitel 3.3.4). Die Markergen-basierte Untersuchung ermöglicht es zusätzlich, die genetische Information zu detektieren und zu konservieren, wenn auch diese keine antibiotische Wirkung verursacht. Dadurch eröffnet sich die Möglichkeit, unterschiedliche Wirtssysteme zu verwenden und die genetische Information gegebenenfalls zu modifizieren. Durch Verwendung weiterer Markergene zur Detektion auf Phenazine [267], Halogenasen [268], oder Cytochrome P450 Hydroxylasen [269] die im Kontext des Sekundärmetabolismus stehen, kann diese Methode erweitert werden.

### 3.3.2 Detektion von NRPS-Genfragmenten durch PCR

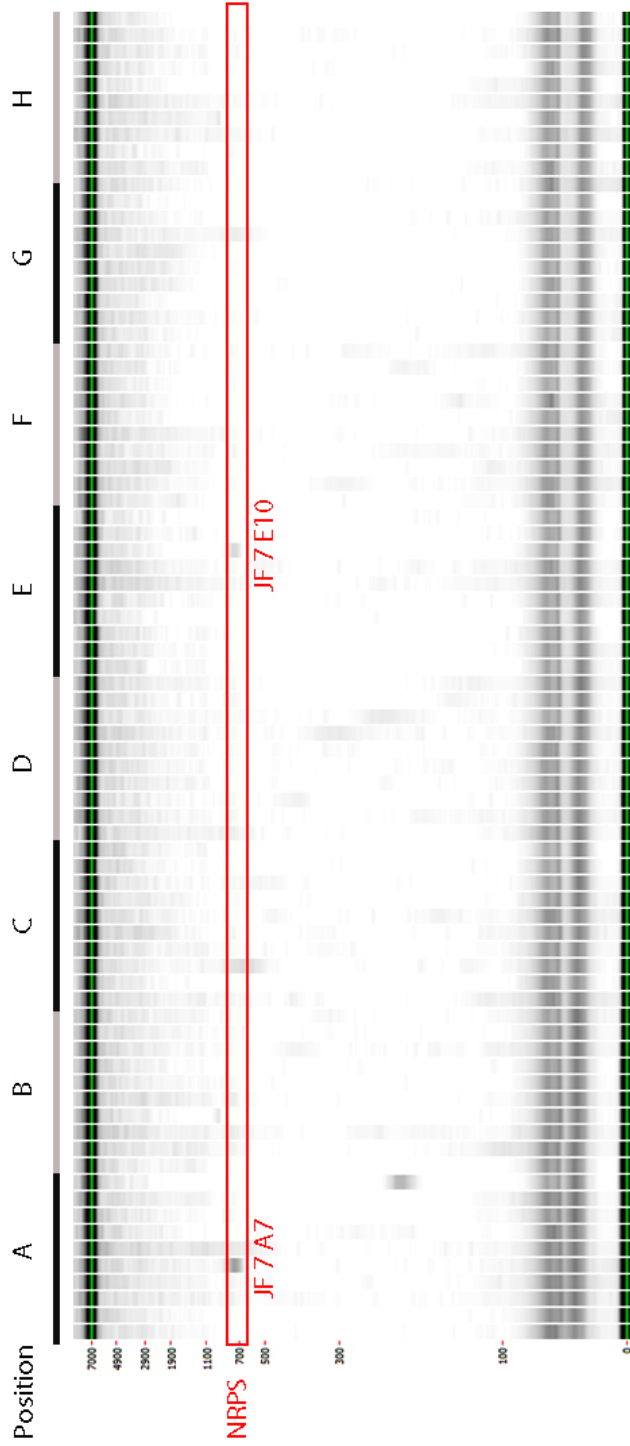


Abbildung 3.19: Platte JF 7: Markergen-spezifische Identifizierung der Adenyldomäne von nicht-ribosomalen Peptidsynthetasen aus der DNS von *A. borealis*. Es wurden zwei Signale identifiziert: JF 7 A7 und JF 7 E10. Auf Grund der Plattenkonfiguration beginnt die Nummerierung ab der Ziffer 3

### 3.3.2 Detektion von NRPS-Genfragmenten durch PCR

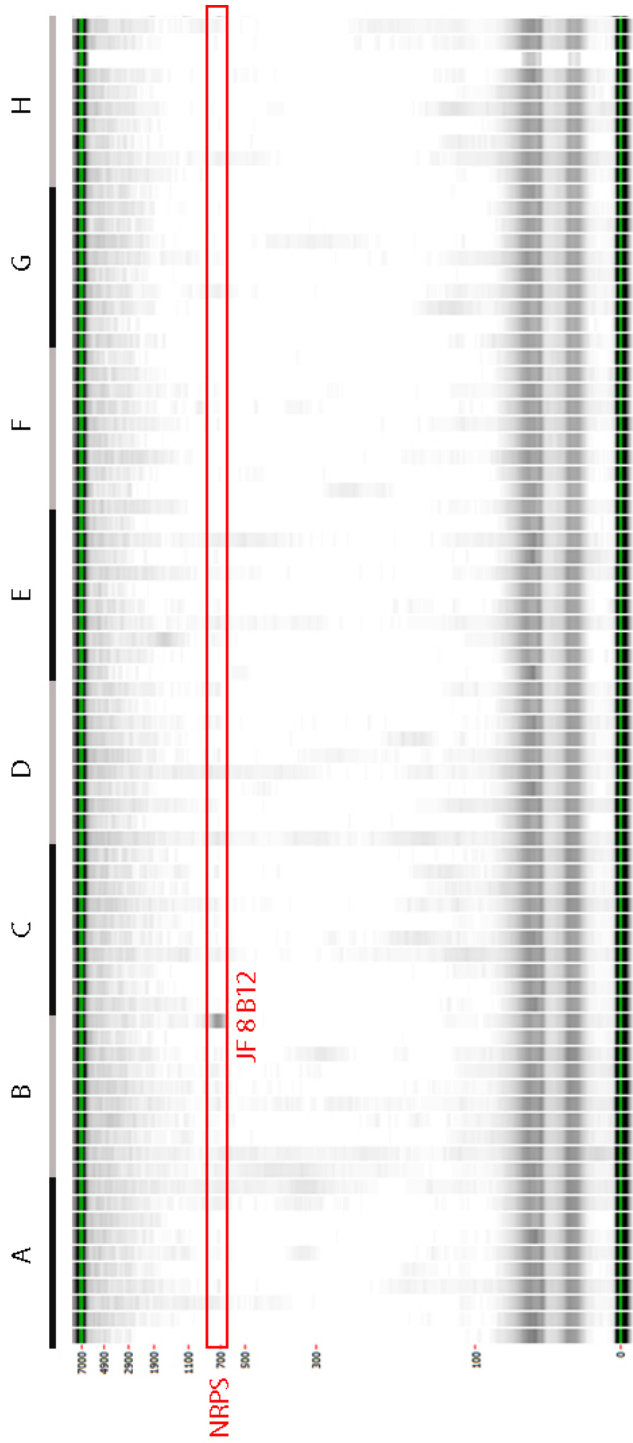


Abbildung 3.20: Platte JF 8: Markergen-spezifische Identifizierung der Adenyldomäne von nicht-ribosomalen Peptidsynthetasen aus der DNS von *A. borealis*. Es wurden ein Signal identifiziert: JF 8 B12. Auf Grund der Plattenkonfiguration beginnt die Nummerierung ab der Ziffer 3

### 3.3.2 Detektion von NRPS-Genfragmenten durch PCR

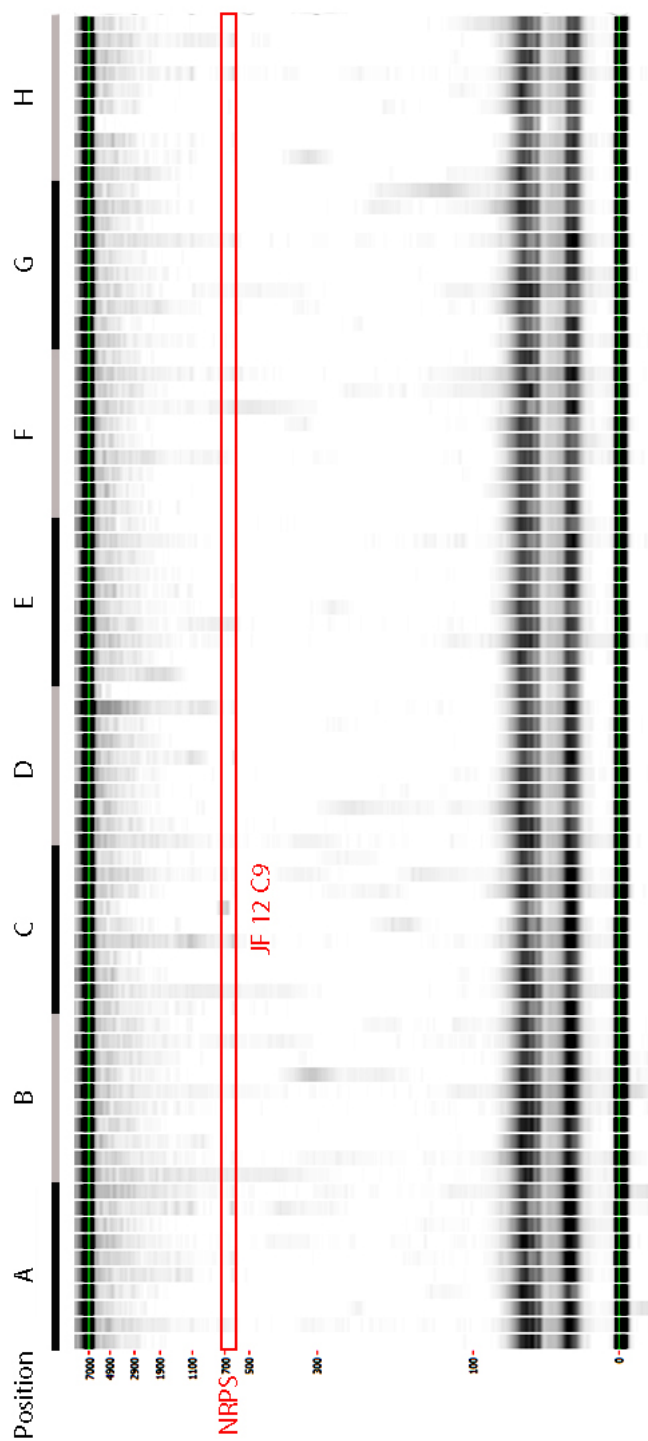


Abbildung 3.21: Markergen-spezifische Identifizierung der Adenyldomäne von nicht-ribosomalen Peptidsynthasen aus der DNS von *A. borealis* (Platte 12). Auf Grund der Plattenkonfiguration beginnt die Nummerierung ab der Ziffer 3

### 3.3.3 Bestimmung der Insertgröße von BACs mit NRPS-Signal

Aus der zuvor durchgeführten Untersuchung auf den konservierten Bereich der Adenyl-domäne, die Bestandteil von nicht-ribosomalen Peptidsynthasen ist, wurden vier eindeutige PCR-Signale den Klonen JF 7 A7, JF 8 B12, JF 7 E10 und JF 12 C9 zugeordnet. Um das Ergebnis zu bestätigen sowie den klonierten Genomloкус näher zu charakterisieren, wurden diese BAC-Inserts mittels Sangersequenzierung überprüft und die Sequenzdaten mit dem annotierten Genom von *A. borealis* verglichen (Kapitel 2.7.3). Unter Abgleich der Gencluster-Analyse aus Kapitel 3.3.2 konnte festgestellt werden, dass in diesem Genomabschnitt das NRPS-Gencluster fünf liegt. Dieses Gencluster ist in Abbildung 3.22 gelb hervorgehoben und nummeriert. Zusätzlich illustriert die Abbildung die Umfänge der BAC-Inserts auf dem Genomabschnitt. Es konnte festgestellt werden, dass 27 der 29 Gene dieses Genclusters durch vier BAC-Inserts konserviert worden sind. Die Abbildung zeigt, dass die Klone JF 7 A7 und JF 8 B12 das gleiche Insert tragen (BAC eins/drei) und den genomischen Bereich von 2.724.930 - 2.777.633 Basen abdecken. Die Insertsequenz hat eine Länge von 52.363 Basen.

BAC zwei zeigt den identifizierten Lokus des Klons JF 7 E10. Dieser deckt den Bereich von 2.715.373 - 2.732.894 Basen ab und hat eine Länge von 17.521 Basen.

Das Insert von BAC vier aus Klon JF 12 C9 deckt den Bereich von 2.731.193 - 2.777.633 Basen auf dem Genom ab und hat eine Länge von 46.440 Basen. Aus dem Sequenzvergleich zwischen *primer* und Genomsequenz wurde desweiteren festgestellt, dass alle vier Inserts in reverser Orientierung zur BAC-Sequenz eingebaut worden sind.

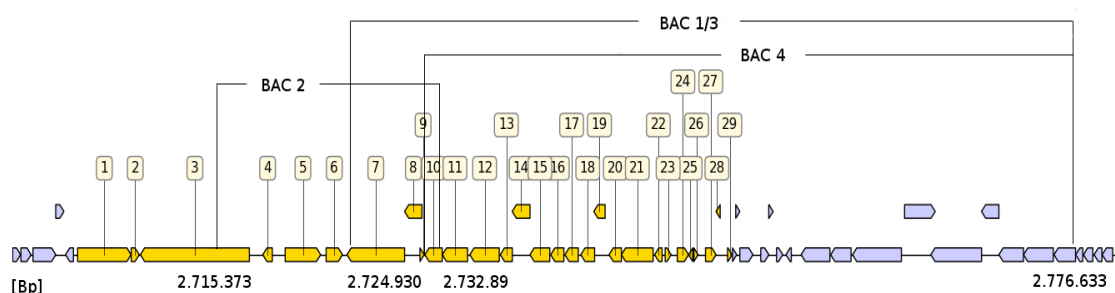


Abbildung 3.22: Lokalisation der klonierten Inserts ausgewählter BACs mit NRPS-Signalen aus der Klonbibliothek von *A. borealis*. Gelb: NRPS-Gencluster 5 aus *A. borealis*, Blau: weiterer genomischer Bereich

Um diesen konservierten Genomabschnitt näher zu charakterisieren wurde eine Annotation der Proteine durch Pfam durchgeführt. Die Ergebnisse sind in den Tabellen 3.15 und 3.16 dargestellt und entsprechen der Nummerierung in der Abbildung 3.22. Die Tabellen illustrieren, dass 20 der 29 Gene nicht eindeutig annotiert werden konnten. Dennoch lässt sich anhand der Annotation darauf schließen, dass es sich wahrscheinlich um ein Gencluster handelt, welches für Enzyme kodiert, die einen antibiotischen Metabolit produzieren könnten. Aus Abbildung 3.22 geht hervor, dass das Gen mit der Position vier auf dem Gencluster, vollständig durch den Klon JF 7 E10 (BAC zwei) abgedeckt worden ist.

### 3.3.3 Bestimmung der Insertgröße von BACs mit NRPS-Signal

Diesem Genprodukt sind zwei Pfamdomänen zugeordnet, eine ASH Domäne, welche im Zusammenhang mit der Flagellenfunktion steht und einer Domäne mit der Bezeichnung Glyco hydro 79C. Diese Domäne ist beschrieben für Glycosidhydrolasen (EC 3.2.1.) und katalysiert die Hydrolyse zwischen zwei Zuckermolekülen. Auf Position sechs des Genclusters wurde die Tyrocidin Synthase 3 als Genprodukt annotiert. Dieses Enzym ist an der Biosynthese des Peptidantibiotikums Tyrocidin beteiligt. Neben dem Gen *tycC* sind desweiteren *tycA* und *tycB* an der Biosynthese beteiligt, welche aber nicht durch die Annotation identifiziert werden konnten. Desweiteren wurden der Tyrocidin Synthase 3, fünf Pfameinträge zugeordnet, beginnt mit Condensation und AMP-*binding*. Im Falle von Condensation handelt es sich um ein Protein, welches eine Kondensationsreaktion zwischen Peptiden katalysiert. Die Familie AMP-*binding* gehört zu einer Gruppe an Enzymen, die Adenenin-Monophosphat (AMP) an ihr Substrat bindet. Analog dazu katalysiert diese Reaktion die Domäne AMP-*binding*\_C. Die katalytische Domäne PP-*binding* bindet 4'-Phosphopantethein als prosthetische Gruppe. Im Enzym fungiert sie als Arm für die Akquise von Substrat, wie zum Beispiel aktivierte Fett- oder Aminosäuren. Die Domäne Thioesterase ist die letzte Funktion dieses Genprodukts und ist typisch für die finale Katalyse einer Aminosäure an das Intermediat oder der Zyklisierung zu einem Endprodukt.

Auf den Positionen sieben und neun, wurde jeweils das Enzym Biotin Biosynthese- Cytochrom P450 annotiert. Biotin fungiert als Cofaktor für die Spaltung von Carbon-Carbon Verbindungen, wie beispielsweise Fettsäuren. Zusätzlich besitzen Proteine der Superfamilie P450 eine Eisenverbindung für Redoxreaktionen mit dem Cofaktor NAD/NADH. Auf Position zehn des Genclusters, das auf allen vier BAC-Inserts vorhanden ist, wurde das zweite NRPS-spezifische Enzym annotiert. Die Gramacidin Synthase Untereinheit B. Sie ist Bestandteil der Gramacidin-Biosynthese und katalysiert die Epimerisierung des antibiotischen Metabolits. Für dieses Genprodukt sind drei Pfam-Einträge zugeordnet, welche auf eine NRPS-basierte Biosynthese hindeuten.

### 3.3.3 Bestimmung der Insertgröße von BACs mit NRPS-Signal

Tabelle 3.15: Auflistung annotierter Proteine des NRPS-Genclusters fünf aus *A. borealis*, Teil 1

Position	Protein oder Enzym	Pfam-Familie
1	hypothetisches Protein	Beta_helix
		Big_3_5
2	hypothetisches Protein	OMP_b-brl
3	hypothetisches Protein	-
4	hypothetisches Protein	ASH
		Glyco_hydro_79C
5	hypothetisches Protein	-
6	Tyrocidinsynthase 3	Condensation
		AMP-binding
		AMP-binding_C
		PP-binding
		Thioesterase
7	Biotin Biosynthese Cytochrom P450	p450
8	hypothetisches Protein	-
9	Biotin Biosynthese Cytochrom P450	p450
10	Gramicidinsynthase Untereinheit B	AMP-binding
		AMP-binding_C
		PP-binding

Im weiteren Verlauf der Annotation wurden die Genprodukte auf Position 11 - 14 nicht weiter spezifiziert. Anhand der Pfam-Einträge Polysacc\_synt, Polysacc\_synt C auf der Position 12 ist es jedoch sehr wahrscheinlich, dass es sich hier um ein Membranprotein handelt, welches an der Biosynthese von Polysacchariden beteiligt ist. Ein weiteres hypothetisches Protein, welches ausschließlich durch einen Pfam-Eintrag charakterisiert werden konnte, ist auf Position 13 das Wzy\_C. Dieses Enzym ist zuständig für das O-Antigen, ein Lipopolysaccharid, welches an der äußeren Membran von Gram-negativen Bakterien zu finden ist.

Im folgenden Verlauf auf den Positionen 15 – 17 wurden weitere Proteine annotiert, welche dem Carbohydrat-Metabolismus zugeordnet werden. Im Falle von WfgD und WfaP handelt es sich um Glykosyltransferasen der Familie 2, welche aktivierte Zucker wie zum Beispiel UDP-Glukose auf andere Substrate übertragen. Im Falle des Genprodukts auf Position 17 handelt es sich ebenfalls um eine Glykosyltransferase, die der Familie 26 zugeordnet wird.

Ein weiteres Protein im Zusammenhang mit dem Transport von Polysacchariden ist auf Position 18 zu finden. Anhand des Pfam-Eintrags Poly\_export und SLBB lässt sich schlussfolgern, dass es sich hier um ein periplasmatisches Protein handelt, welches an der Biosynthese oder dem Transport beteiligt ist.

Position 19 beschreibt das Genprodukt Tyrosinkinase. Dieses Enzym ist an einer Vielzahl von zellulären Aktivitäten beteiligt, wie zum Beispiel der Translokation von Makromo-

### 3.3.3 Bestimmung der Insertgröße von BACs mit NRPS-Signal

lekülen.

Auf Position 20 wurde das Gen für den Transkriptions-Terminators NusG annotiert. An dieser Position liegt strangabwärts das nächste Protein im Zusammenhang der Transkriptions-Regulation. Das Protein DegU (Position 21) gehört zu einem Zweikomponentensystem, welches weitere Genkaskaden reguliert. Das Protein besitzt ein *helix-loop-helix* Motiv zur Interaktion mit DNS. In Gram-negativen Bakterien steht dieses Motiv zudem in Verbindung mit der Regulation von Exopolysacchariden.

Im letzten Teil der Annotation wurde die Position 25 durch den Pfam-Eintrag Metallophos charakterisiert. Dieser Eintrag beschreibt eine Phosphatase-Domäne welche als Exonuclease und Endonuclease fungieren kann. Die Annotation schließt mit der Position 28 und dem Pfam-Eintrag HTH\_17. Dieses Strukturmotiv bildet ähnlich wie GerE ein *helix-loop-helix* Motiv aus und ist an der Genexpression beteiligt.

### 3.3.3 Bestimmung der Insertgröße von BACs mit NRPS-Signal

Tabelle 3.16: Auflistung annotierter Proteine des NRPS-Genclusters fünf aus *A. borealis*, Teil 2

Position	Protein oder Enzym	Pfam-Familie
11	hypothetisches Protein	-
12	hypothetisches Protein	Polysacc_synt
		Polysacc_synt C
13	hypothetisches Protein	Wzy_C
14	hypothetisches Protein	-
15	UDP-Glc:alpha-D-GlcNAc-diphosphoundecaprenol beta-1 3-glucosyltransferase WfgD	Glycos_transf.2
16	UDP-Glc:alpha-D-GlcNAc-diphosphoundecaprenol beta-1 3-glucosyltransferase WfaP	Glycos_transf.2
17	N-acetylglucosaminyldiphosphoundecaprenol N-acetyl-beta-D-mannosaminyltransferase	Glyco_tran_WecB
18	hypothetisches Protein	SLBB
		Poly_export
19	Tyrosinkinase ptk	Wzz
		AAA_31
20	Transkriptions-Terminator NusG	NusG
21	hypothetisches Protein	-
22	Transkriptions-Regulator DegU	GerE
23	hypothetisches Protein	-
24	hypothetisches Protein	-
25	hypothetisches Protein	Metallophos
26	hypothetisches Protein	.
27	hypothetisches Protein	.
28	hypothetisches Protein	HTH_17
29	hypothetisches Protein	-

Die Ergebnisse dieser Annotation decken sich zusammenfassend mit der Gencluster-Vorhersage aus *antiSMASH*. Durch die Detektion der Tyrocidinsynthase 3 und der Gramacidinsynthase Untereinheit B, sind zwei Enzyme vorhanden, aus denen ein bekannter nicht-ribosomal synthetisierter Metabolit produziert werden kann [270] [271]. Dennoch konnte in der nachfolgenden Untersuchung auf antibakterielle Aktivität keine der vier sequenzierten Klone eine Aktivität nachgewiesen werden. Dies kann auf unterschiedliche Weise interpretiert werden.

Zum einen wurde durch keinen der Klone das Gencluster vollständig konserviert, so dass eventuelle Promotoren zur Aktivierung der Genexpression fehlen. Dies kann auch aufgrund des Einbaus der Sequenzen in reverser Orientierung der Fall gewesen sein. Grundsätzlich konnte in der Publikation von Sabree *et al.* (2006) [272] gezeigt werden,

### 3.3.3 Bestimmung der Insertgröße von BACs mit NRPS-Signal

dass *E. coli* zur genetischen Expression von DNS aus Acidobakterien fähig ist. Zum anderen könnte es sich um ein Gencluster handeln, von dem kein Metabolit produziert werden kann oder keine Aktivität ausgeht. Diese Interpretation liegt darin begründet, dass es sich um zwei Peptidsynthasen handelt, die an der Biosynthese von zwei unterschiedlichen Metaboliten beteiligt sind. Zudem wurden im Falle der Gramacidin-synthase Untereinheit B, keine weiteren Untereinheiten annotiert, weshalb die Funktion der Gramacidin Biosynthese in Frage gestellt ist. Aus der Publikation von Mootz *et al.* (1997) [271] geht hervor, dass die Tyrocidin-Biosynthese in Form eines Operons strukturiert ist, an dem strangabwärts zwei ABC-Transporter zur Vermittlung der Resistenz kodiert sind. Dieses Ergebnis konnte durch die Untersuchung auf Resistenzgene in Kapitel 3.1.7 nicht bestätigt werden.

### 3.3.4 Untersuchung von Klonbibliotheken hinsichtlich antibakterieller Aktivität

Die direkte Untersuchung nach antibakterieller Aktivität umfasste 960 Klone, welche aus genomischer DNS von *A. borealis* und gewonnenen Bakterien aus dem Termitennest erstellt wurden. Die nachfolgenden Kapitel stellen die Ergebnisse aus dem Kulturüberstands-basierten und der Extrakt-basierten Untersuchung zusammen.

#### Detektion von antibakterieller Aktivität durch Biolumineszenz

Die Biolumineszenz-basierte Untersuchung erfolgte durch die Anfertigung von zellfreien Überständen (Kapitel 2.7.4) im Anschluß an die Kultivierung. Als Testorganismus wurde die Aufnahme der Biolumineszenz erfolgte direkt nach dem Start ( $t_0$ ) und nach Stunde 6 ( $t_6$ ) im Duplikat.

Es zeigte sich, dass mindestens eine Komponente des Mediums einen Effekt auf den Testorganismus hatte, sodass die Inhibierung auf Basis der Biolumineszenz 60 % betrug. Aus diesem Grund wurde der Schwellenwert für die Aktivität im Duplikat auf >80 % gesetzt. Von 960 Klonen haben zwei Klone diese Voraussetzung erfüllt (Tabelle 3.17).

#### Detektion von antibakterieller Aktivität aus Kulturextrakten

Die Extrakterstellung der Klonbibliothek erfolgte mit den verbliebenen 1,3 mL Kulturvolumen im 96-well Format (Kapitel 2.7.4). Als Extraktionsmittel wurde 75 % (v/v) MeOH gewählt. Nach der Extraktion wurde zunächst im 384-well Format ein 2x konzentriertes Extrakt (im Verhältnis zum Volumen des Testorganismus) eingesetzt und eingedampft. Um die Inhibierung produzierter Metabolite zu charakterisieren wurden sowohl Gram-positive als auch Gram-negative Testorganismen eingesetzt, darunter zwei Mutanten der Stämme *E. coli* und *P. aeruginosa*. Im Falle dieser Mutanten wurden Schlüsselproteine des Transportsystems ausgeschaltet, um die Sensibilität gegenüber Antibiotika zu erhöhen [273] [274]. Der Stamm *E. coli*  $\Delta tolC$  besitzt eine Mutation des Gens *tolC*, das für einen Transporter der Zellmembran kodiert und Teil des Acridin-Resistenzkomplexes ist [275] [276]. Weil dieser Stamm nur eine Kopie dieses Gens besitzt, ist dieser sensitiver als das klinische Isolat *E. coli* ATCC 225922. Als weiterer Teststamm wurde *P. aeruginosa* PAO 750 eingesetzt. Pseudomonaden verfügen über hoch effiziente Effluxpumpen, die zu der Transportfamilie der *resistance-nodulation-division* (RND) gehören und eine Reihe von Antibiotika durch die Zellmembran aus der Zelle transportieren [277]. Im Falle des Typs *P. aeruginosa* PAO 750 wurde das Proteinsystem MexAB-OprM, MexCD-OprJ, MexEF-OprN und MexXY-OprM deletiert, wodurch der Stamm sensitiv gegenüber einer Reihe von Antibiotika wie Quinolone, Chloramphenicol, Trimethoprim, Imipenem und Tetracyclin ist [278] [279]. [280] [281]. Als Gram-positives Testbakterium wurde der Methicillin-sensitive *S. aureus* (MSSA) gewählt. Dieser Stamm besitzt, im Gegensatz zu MRSA, nicht das Gen *mecA*, welches für das Penicillin-Bindeprotein 2a kodiert und damit eine Resistenz gegenüber  $\beta$ -Lactam Anti-

### 3.3.4 Untersuchung von Klonbibliotheken hinsichtlich antibakterieller Aktivität

biotika vermittelt [282] [283].

Die Testergebnisse der Extrakt-basierten Untersuchung sind in Tabelle 3.17 dargestellt. Die Inhibierung auf Basis der optischen Dichte von Komponenten des reinen Mediums war im Vergleich zur Untersuchung durch Biolumineszenz deutlich niedriger. Im Vergleich zeigten die Extrakte aus den Klonen JF 8 A9, JF 9 H5, JF 9 C12, JF 11 B9, JF 13 B6, JF 13 G5 und JF 13 G11 eine Inhibierung auf die Testbakterien. Hierfür kann die unterschiedliche Methodik die Ursache sein. Um die Inhibierung dieser Klone auf die Testorganismen zu bestätigen, wurde im weiteren Verlauf für jeden als aktiv eingestuft Klon der Tabelle 3.17, eine 50 mL Kultur im Duplikat angefertigt und die Kulturen extrahiert (Kapitel 2.7.5). Die Extrakte wurden in den finalen Konzentrationen 1x, 2x und 10x getestet. Die Ergebnisse der Inhibierungen sind in Tabelle 3.18 dargestellt und zeigen, dass alle Extrakte als 10x Konzentrat aktiv gegen *P. aeruginosa* PAO 750 waren. Auch zeigten Extrakte, mit Ausnahme von JF 7 C4 und JF 8 A9 eine Inhibierung als 10x Konzentrat gegen beide *E. coli* Stämme. Eine Inhibierung gegen MSSA ATCC 25923 wurde nicht identifiziert.

Die Untersuchung auf antibakterielle Aktivität gegen Gram-positive als auch gegen Gram-negative Bakterien, ermöglicht es, das Wirkungsspektrum aktiver Metabolite einzugrenzen. Daher ist es möglich, das Testergebnisse wie Tabelle 3.17 darstellt, heterogen ausfallen. Es ist aber erkennbar, dass durch eine höhere Konzentrierung der Extrakte von 10x, die Heterogenität der antibakteriellen Aktivitäten verringert wurde, sodass eindeutiger zwischen Aktivität und keiner Aktivität diskriminiert werden kann. Insgesamt wurden in den Untersuchungen auf antibakterielle Aktivität 960 Klone getestet, von denen neun Klone bestätigte antibakterielle Aktivität zeigten. Dies entspricht einer Rate von einem Prozent.

Studien zu weiteren funktionalen Untersuchungen weisen ähnliche Statistiken auf, die je nach Fragestellung noch geringer ausfallen und eine Vielzahl an Klonen benötigen [284] [285]. Dies macht die Bedeutung der biologischen Komponenten, wie Ressource und Wirtssystem, offensichtlich, welche aufeinander abgestimmt werden müssen. Eine Möglichkeit die Wahrscheinlichkeit der Genexpression zu erhöhen und zu vergleichen, ist die Verwendung von *shuttle* Vektoren zwischen unterschiedlichen Wirtssystemen [286]. Ein weiterer Faktor ist die Methodik, effizient eine Vielzahl an Proben auf antibakterielle Aktivität zu untersuchen. Als Standard gelten Plattformen mit einer Kapazität von 10.000 Proben pro Woche (*high throughput screening*) [287]. Technische Innovationen wie Microfluidics-Systeme ermöglichen die Untersuchung von >50.000 Proben und werden als *ultra-high throughput* Prozesse bezeichnet [288].

### 3.3.4 Untersuchung von Klonbibliotheken hinsichtlich antibakterieller Aktivität

Tabelle 3.17: Ergebnisse der Inhibierung auf Basis der Biolumineszenz (primäre Untersuchung) und der Extrakt-basierten Inhibierung auf Basis der optischen Dichte (sekundäre Untersuchung). Die Ergebnisse sind als Duplikat dargestellt

Klon	Primäre Untersuchung		Sekundäre Untersuchung		
	<i>E. coli</i> DH5 $\alpha$ pFU166	<i>P. aeruginosa</i> PAO 750	<i>E. coli</i> ATCC 225922	<i>E. coli</i> $\Delta$ tolC	MSSA ATCC 25923
JF 7 C3	89 %	keine Aktivität	keine Aktivität	keine Aktivität	keine Aktivität
JF 7 C4	84 %	keine Aktivität	keine Aktivität	keine Aktivität	keine Aktivität
JF 8 A9	keine Aktivität	75 %	keine Aktivität	keine Aktivität	keine Aktivität
JF 9 H5	keine Aktivität	keine Aktivität	93 %	keine Aktivität	keine Aktivität
JF 9 C12	keine Aktivität	keine Aktivität	keine Aktivität	95 %	keine Aktivität
JF 11 B9	keine Aktivität	84 %	keine Aktivität	keine Aktivität	keine Aktivität
JF 13 B6	keine Aktivität	keine Aktivität	keine Aktivität	keine Aktivität	72 %
JF 13 G5	keine Aktivität	keine Aktivität	keine Aktivität	keine Aktivität	73 %
JF 13 G11	keine Aktivität	keine Aktivität	keine Aktivität	keine Aktivität	71 %
Müller Hinton II	61,1 %	15 %	-20 %	9,5 %	-13 %



### 3.3.5 Peptide: mögliche Verbindungen mit antibakterieller Aktivität

Ausgangspunkt für die chemische Analytik der Extrakte waren die Ergebnisse des *scale-up* Versuchs, welche in der Tabelle 3.18 dargestellt sind. Für die Analytik wurden diese Extrakte als 50x Konzentrat in 100 % MeOH gelöst und mittels UHPLC-MS/MS analysiert (Kapitel 2.7.5). Als Referenz wurde der Stamm *E. coli* MegaX DH10B T1 Electrocomp [pBeloBAC11] im Duplikat mitgeführt. Abbildung 3.23 zeigt die Chromatogramme als Liste. In rot ist jeweils die Probe JF 7 C3 markiert, als blau und schwarze Spur ist jeweils die Referenz zu erkennen. In Abbildung 3.24 sind die Signale herausgestellt erkennbar, jeweils einmal in blauer, sowie roter Farbe gezeigt. Sie stellen die Intensitäten der aufgetrennten Komponenten in Abhängigkeit der Retentionszeit dar. Aus der Abbildung geht hervor, dass jeweils zwei Signale detektiert worden sind, die nicht in den Referenz-Chromatogrammen zu erkennen sind. Ein Abgleich des Masse zu Ladungsverhältnis ( $m/z$ ) mit der Datenbank ergab, dass es sich um zwei Tripeptide Valylprolylleucin und Isoleucylprolylisoleucin handeln könnte (Tabelle 3.19). Um zu bestätigen, ob es sich bei Isoleucylprolylisoleucin, um das Tripeptid Diprotin A [289] mit den identischen Aminosäuren Isoleucin-Prolin-Isoleucin handelt, wurde ein MS/MS-Analyse durchgeführt und die Massenfragmente miteinander verglichen (Abbildung 3.19). Aus diesem Vergleich geht hervor, dass die Massenfragmente von Isoleucylprolylisoleucin nicht der des Diprotin A entsprechen. Das Masse zu Ladungsverhältnis ( $m/z$ ) von 245.1864 kann nicht erklärt werden, dennoch handelt es sich um die gleiche Stoffklasse.

Von Peptiden ist bekannt, dass aus diversen Verbindungen antimikrobielle Wirkung hervorgeht. Gencluster-Typen in diesem Zusammenhang, welche Aminosäuren verwenden um bioaktive Metabolite zu generieren, sind Ripp-Gencluster oder NRPS-Gencluster. Ein Beispiel eines bioaktiven Tripeptids ist das NRPS-synthetisierte Phosphinothricin [290]. Dennoch lässt sich anhand des negativen Signals aus der PCR-Untersuchung nach nicht-ribosomalen Peptidsynthasen, sowohl des Klons JF 7 C3 (Abbildung 3.19) als auch dessen BAC (Abbildung 5.3) das Fazit ziehen, dass es sich wahrscheinlich nicht um ein Gencluster diesen Typs handelt. Bezüglich Ripp-Gencluster wurden in dem Genom von *A. borealis* (Tabelle 3.12) die Typen *bacteriocin* und *lassopeptide* identifiziert. Die Klassifizierung des Produkts als Lassozeptid ist aufgrund fehlender Struktur nicht möglich und bedarf eines typischen Macrolactamrings, der für gewöhnlich aus acht oder neun Aminosäuren besteht [291]. Auch die Klassifizierung des Tripeptids als Bacteriocin, der Klasse 1 Lanthionin-haltiger Peptide ist ausgeschlossen. Aus Bacteriocine der Klasse 2 sind keine Tripeptide bekannt [292]. Als Fazit dieser Untersuchung sollten daher alle BACs der Klone mit identifizierter bakterieller Aktivität vollständig sequenziert werden.

3.3.5 Peptide: mögliche Verbindungen mit antibakterieller Aktivität

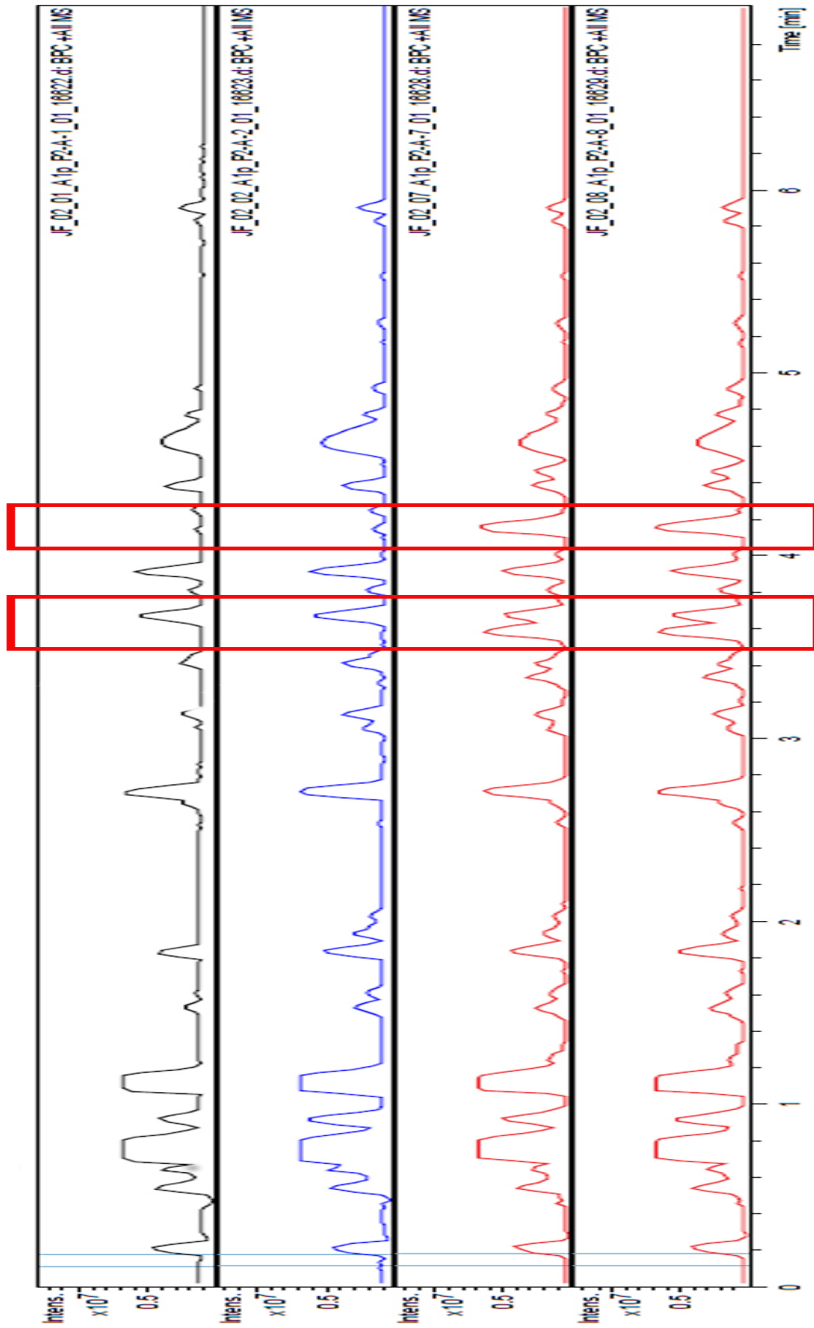


Abbildung 3.23: Chromatogramme als Liste Klons JF 7 C3 im Vergleich zur Referenz *E. coli* MegaX DH10B T1 [pBelob-AC11]

### 3.3.5 Peptide: mögliche Verbindungen mit antibakterieller Aktivität

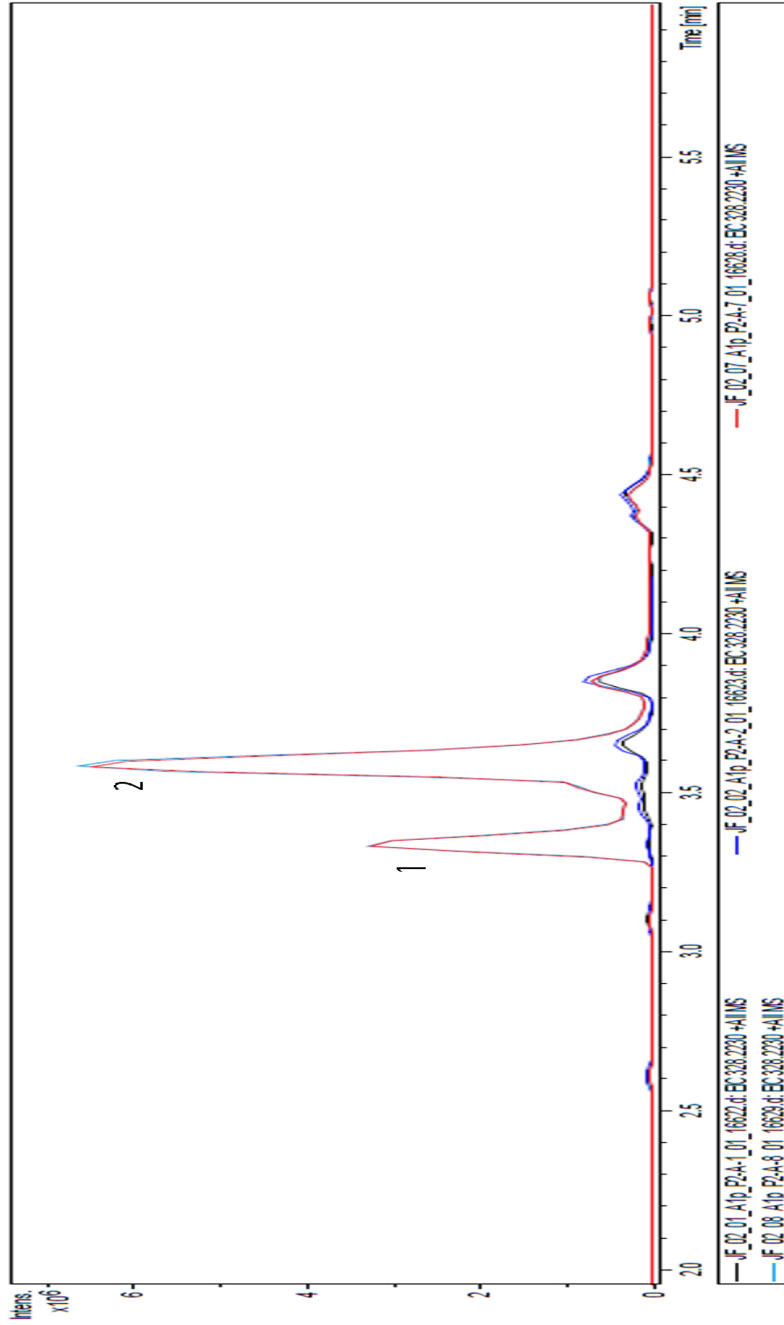


Abbildung 3.24: Signale Klons JF 7 C3 herausgestellt. Signal 1: Valylprolylleucin, Signal 2: Isoleucylprolyl isoleucin

### 3.3.5 Peptide: mögliche Verbindungen mit antibakterieller Aktivität

Tabelle 3.19: Identifizierte Massen des Klons JF 7 C3 im Duplikat

Probe	Retentionszeit [Min]	Masse [m/z]	Summenformel (neutral)	Vorhersage
Replikat 1, Signal 1	3,3	328,22	C <sub>16</sub> H <sub>29</sub> N <sub>3</sub> O <sub>4</sub>	Valylprolylleucin
Replikat 2, Signal 1	3,6	328,22	C <sub>16</sub> H <sub>29</sub> N <sub>3</sub> O <sub>4</sub>	Valylprolylleucin
Replikat 1, Signal 2	4,2	342,24	C <sub>17</sub> H <sub>31</sub> N <sub>3</sub> O <sub>4</sub>	Isoleucylprolylisoleucin
Replikat 2, Signal 2	4,5	342,24	C <sub>17</sub> H <sub>31</sub> N <sub>3</sub> O <sub>4</sub>	Isoleucylprolylisoleucin

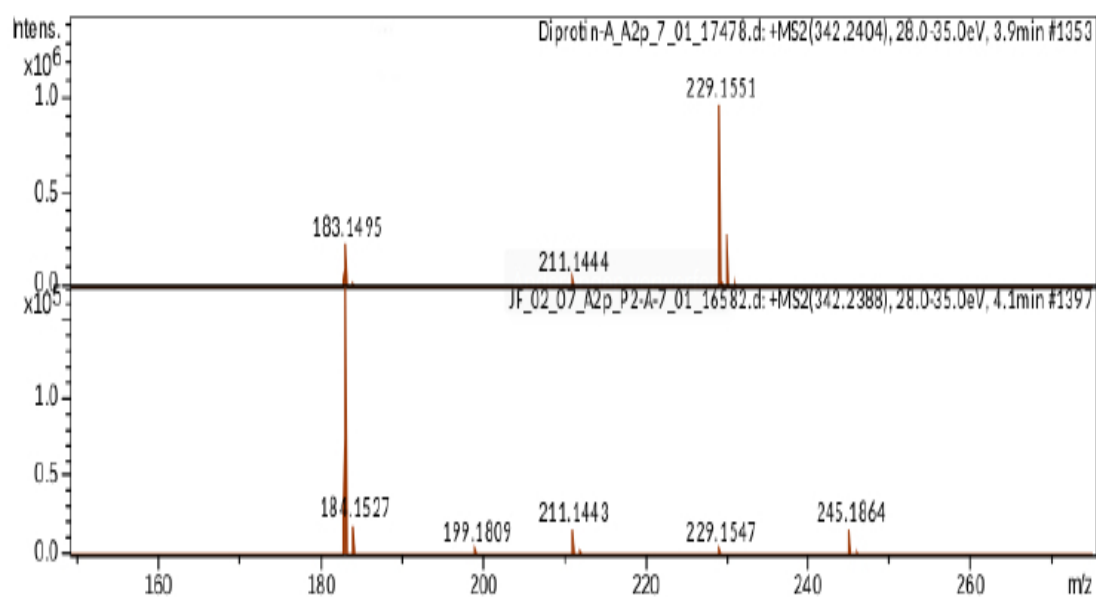


Abbildung 3.25: MS/MS-Spektrenvergleich von Diprotin A (oben) und Isoleucylprolylisoleucin (unten)

#### Analyse weiterer Extrakte mit antibakterieller Aktivität

Neben der Massenanalytik des Klons JF 7 C3, wurden weitere Extrakte untersucht die antibakterielle Aktivität zeigten. In der Tabelle 3.20 sind die Massen dargestellt welche im Vergleich zum Referenz-Chromatogramm identifiziert worden sind. Die Tabelle zeigt, dass der Klon JF 7 C4 die gleichen Massen wie der Klon JF 7 C3 generierte, die als Valylprolylleucin und Isoleucylprolylisoleucin identifiziert worden sind.

### 3.3.5 Peptide: mögliche Verbindungen mit antibakterieller Aktivität

Tabelle 3.20: Massenspuren weiterer Klone mit antibakterieller Aktivität

Bezeichnung	Masse 1 [m/z]	Masse 2 [m/z]
JF 7 C4	328,223	342,2386
JF 9 H5	244,6471	358,1972
JF 13 C6	244,6472	358,1975
JF 13 G11	244,6470	358,1969
JF 13 G5	244,6472	358,1972
JF 11 B9	244,6472	-
JF 9 C12	358,1973	-
JF 8 A9	358,1973	-

Desweiteren sind die Massen der Klone JF 9 H5, JF 13 C6, JF 13 G11 und JF 13 G5 identisch. Die Klone JF 9 C12 und JF 8 A9 zeigten nur eine Masse als Produkt einer heterologen Expression. Auf Grund der gleichen Massen von 244,6471 und 358,1972 ist es wahrscheinlich, dass diese Klone die gleichen DNS-Bereiche als Inserts von *A. borealis* tragen. Eine Darstellung der Chromatogramme und Massen ist im Annex 5.7 dargestellt. Eine MS/MS-Analyse dieser Massen ist für eine weiterführende Untersuchung anzustreben.

## 4 Ausblick

Zur Etablierung der laborpraktischen Methoden, für die Erstellung rekombinanter Klone, wurden die Bioressourcen *A. borealis*, *Terracidiphilus* sp. S55 sowie Bakterien aus dem Termitennest von *C. niger* verwendet. Diese wurden zuvor bioinformatisch anhand ihrer Eigenschaften als potentielle Antibiotika-Produzenten eingestuft.

Aus der bioinformatischen Untersuchung gehen weitere 77 aufgelistete Gencluster aus Bakterien hervor, welche sich insbesondere durch die Eigenschaften qualifizieren, sehr wahrscheinlich Gencluster mit Resistenzgenen zu besitzen und keine Sequenzähnlichkeit zu bekannten Genclustern aufzuweisen. Diese Stämme stellen weitere Kandidaten für einen funktionalen Genomansatz dar.

In der Summe wurden 960 Klone durch die etablierte Methodik erstellt. Trotz der hohen Anzahl, fielen die Resultate für die Ansätze mit *Terracidiphilus* sp. S55 und des Metagenomansatzes gering aus. Deshalb sollte der Ansatz mit DNS aus *Terracidiphilus* sp. S55 wiederholt werden. Im Falle des Metagenomansatzes wurden 80 Klone konserviert, von denen aber nur einer exemplarisch untersucht worden ist und kein kloniertes Insert identifiziert wurde. Aufgrund der geringen Stichprobe kann deshalb nicht ausgeschlossen werden, dass dieser Ansatz gänzlich erfolglos war. Demnach wäre eine PCR-basierte Untersuchung im 96-*well* Format eine adäquate Herangehensweise, um den Metagenomansatz vollständig zu evaluieren.

Im Gegensatz dazu konnten aus dem Ansatz mit DNS aus *A. borealis*, 880 Klone generiert werden, von denen exemplarisch fünf Klone ausgewählt wurden und klonierte Inserts im gewünschten Größenbereich besitzen. Ferner wurden diese Klone funktional untersucht. Die erste Untersuchung erfolgte PCR-basiert und wurde im weiteren Verlauf durch eine Sequenzierung bestätigt. Weitere genannte Primer auf konservierte Bereiche von Genclustern könnten diesen Ansatz erweitern.

Der zweite funktionale Ansatz umfasste die Untersuchung auf antibakterielle Aktivität aus den neun Klone identifiziert worden sind. Eine vertiefte Untersuchung des Klons JF 7 C3 identifizierte zwei Tripeptide als sehr wahrscheinliche Auslöser dieser Aktivität. Dass diese Tripeptide Produkte eines Genclusters aus *A. borealis* sind, wurde diskutiert und bleibt unwahrscheinlich. Deshalb sollte an dieser Stelle eine Sequenzierung zur Identifizierung des klonierten genomischen Bereichs durchgeführt werden. Als abschließende Bewertung kann konstatiert werden, dass durch den Ansatz mit DNS aus *A. borealis* die methodischen Grundlagen erfolgreich realisiert worden sind, Klone zu generieren und funktional zu untersuchen. Im Falle von *E. coli* als Wirt erfolgte dies bereits unter der Prämisse, dass die DNS aus Acidobakterien kompatibel zu der mit *E. coli* ist. Diese Kompatibilität ist jedoch nur im geringen Maße phyla-übergreifend und wird zum Beispiel determiniert durch den GC-Gehalt sowie der *codon usage*. Daher sollten weitere Wirts-Vectorssysteme in Betracht gezogen werden, um das Ressourcen-Spektrum

### *Ausblick*

effektiver nutzen zu können. Auch hierzu eignet sich *E. coli* als Zwischenwirt für die Konservierung und Vervielfältigung der klonierten Inserts.

# 5 Annex

## 5.1 Bioinformatik

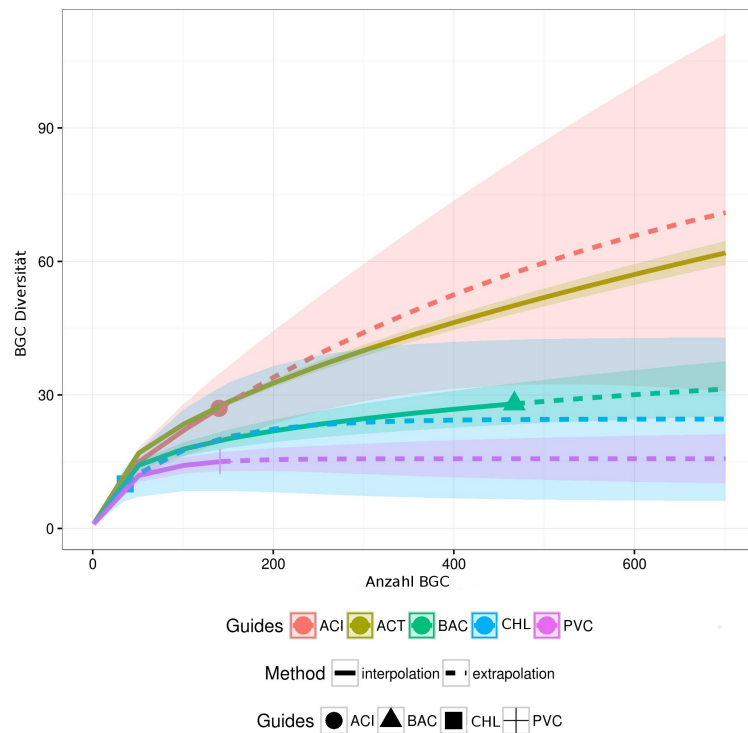


Abbildung 5.1: Hochrechnung des Gencluster-Potentials anhand des Datensatzes für jedes Phylum. ACI: *Acidobacteria*, ACT: *Actinobacteria*, BAC: *Bacteroidetes*, CHL: *Chloroflexi*, PVC: *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae*

Für die *rarefaction* (Abbildung 5.1) wurde die Gencluster-Anzahl und Diversität pro Phylum zur Berechnung verwendet. Der durchgängige Kurvenverlauf kennzeichnet die Interpolation, basierend auf realen Daten, der gestrichelte Verlauf der Funktion kennzeichnet die Extrapolation beziehungsweise Hochrechnung. Jeder Kurvenverlauf ist durch eine Farbcodierung begleitet, die den Vertrauensbereich darstellt. Analoge Berechnungen sind beispielsweise bei Doroghazi *et al.* 2014 [293] zu finden.

Tabelle 5.1: Klassifikation identifizierter Resistenzen in Gencluster, Teil 1

Resistenz-Mechanismus	Resistenz-Typ	PKS				NRPS				Ripp							
		<i>t3pks</i>	<i>t1pks</i>	<i>otherks</i>	<i>aryl polyene</i>	<i>ladderane</i>				<i>lassopeptide</i>	<i>lanthipeptide</i>	<i>bacteriocin</i>	<i>microviridin</i>	<i>thiopeptide</i>			
Molekül-Zerstörung	Klasse A																
	Klasse B																
	Subklasse B1	1															1
	Subklasse B2	1							1								
	Subklasse B3	1															
Transport	RND efflux	2							1	1	2						
	ABC efflux				1				1	3							
	Tet MFS efflux																
	MexH								1								
	MexE													1			
	Chlor Acetyltrans CAT	1															1
	VEB-PER																
Molekül-Modifikation	AAC6-I		1														
	AAC3																
	APH3	1															
	DNA_gyraseB	2															
	Pentapeptide 4																
sonstige	rpoB	1															1
	DNS topoisom IV	3		1										1			
	Proteasome																
	Carboxyl trans			4	1									1			
	Biotin lipoyl	5		4	2									1			1
	Gp dth C													1			
	OTCace																
	DNSP III, beta-Untereinheit	1															

Tabelle 5.2: Klassifikation identifizierter Resistenzen in Genclustern, Teil 2

Resistenz-Mechanismus	Resistenz-Typ	Hybride						bacteriocin-lanthipeptide
		t3pks-aryl polyene	t1pks-nrps	t3pks-terpene	t1pks-PuffA-others	t3pks-others	lanthipeptide-t1pks-others-nrps	
Molekül-Zerstörung	Klasse A		1					
	Klasse B							
	Subklasse B1							
	Subklasse B2							
Transport	Subklasse B3						1	
	RND efflux		1					
	ABC efflux		2					1
	Tet MFS efflux							
Molekül-Modifikation	MexH							
	MexE							
	Chlor							
	Acetyltrans CAT							
sonstige	VEB-PER						1	
	AAC6-I							
	AAC3							
	APH3							
sonstige	DNS gyraseB	1						
	Pentapeptide 4							
	rpoB							
	DNS topoisolV	1						
	Proteasome		1					
	Carboxyl trans							1
	Biotin lipoyl		1			1		
	Gp dh C							1
	OTCase		1					
	DNSP III beta-Untereinheit							

Tabelle 5.3: Klassifikation identifizierter Resistenzen in Genclustern, Teil 3

Resistenz-Mechanismus	Resistenz-Typ	sonstige		
		<i>ectoine</i>	<i>other</i>	<i>terpene</i>
Molekül-Zerstörung	Klasse A			
	Klasse B		1	
	Subklasse B1			
	Subklasse B2			
	Subklasse B3			
	RND efflux		3	3
Transport	ABC efflux	1	1	4
	Tet MFS efflux			1
	MexH			1
	MexE		1	
	Chlor Acetyltrans CAT		1	1
	VFB-PER			
Molekül-Modifikation	AAC6-I			
	AAC3		1	
	APH3			
	DNS gyraseB		1	3
	Pentapeptide_4			1
	rpoB			
sonstige	DNS topoisomIV			2
	Proteasome			2
	Carboxyl trans			1
	Biotinlipoyl		4	2
	Gp dh C		1	
	OTCace		1	2
DNBP III, beta-Untereinheit				1

Tabelle 5.4: Stämme mit Resistenzgenen und Genclusterrähnlichkeit, Teil 1

Phylum	NCBI-accession	Stamm	Gencluster	Ähnlichkeit zu Gencluster	Ähnlichkeit in %
<i>Acidobacteria</i>	NZ_KB906767.1	<i>Acidobacteriaceae</i> bacterium KBS 96	<i>other</i>	Lysobactin	5
	NZ_ATWD01000002.1	<i>Acidobacteriaceae</i> bacterium strain TAA 166	<i>terpene</i>	Malleobactin	11
	P001472.1	<i>Acidobacterium capsulatum</i> ATCC 51196	<i>t3pks-terpene</i>	Lysobactin	14
	-	<i>Terriglobus</i> sp. L56	<i>t3pks</i>	Nosiheptide	15
	KK211214.1	<i>Maribacter antarcticus</i> DSM 21422 genomic scaffold	<i>t1pks-PUFA-otherks</i>	Eicosapentaenoic	9
<i>Bacteroidetes</i>	CP011071.1	<i>Muricauda lutaomensis</i> strain CC-HSB-11	<i>terpene</i>	Carotenoid	28
	P001397.1	<i>Nonlabens dokdonensis</i> DSW-6	<i>terpene</i>	Carotenoid	28
	CP003156.1	<i>Owenweekia hongkongensis</i> DSM 17368	<i>aryl polyene-resorcinol</i>	Flexirubin	16
	CP001681.1	<i>Pedobacter heparinus</i> DSM 2366	<i>terpene</i>	Carotenoid	28
	H724148.1	<i>Polaribacter irgensii</i> 23-P scf_1099646005584 genomic Scaffold	<i>terpene</i>	Carotenoid	28
	CP004349.1	<i>Polaribacter</i> sp. MED152	<i>terpene</i>	Carotenoid	28
	CP009278.1	<i>Sphingobacterium</i> sp. ML3W	<i>terpene</i>	Carotenoid	28
	DF968182.1	<i>Bacteroidales</i> bacterium TBC1	<i>terpene</i>	Carotenoid	28
	CP001699.1	<i>Chitinophaga pinensis</i> DSM 2588	<i>aryl polyene-resorcinol</i>	Flexirubin	16

Tabelle 5.5: Stämme mit Resistenzgenen und Genclusterrähnlichkeit, Teil 2

Phylum	NCBI-accession	Stamm	Gencluster	Ähnlichkeit zu Gencluster	Ähnlichkeit in %
<i>Bacteroidetes</i>	CP002046.1	<i>Croceibacter atlanticus</i> HTCC2559	<i>aryl polyene</i>	Ravidomycin	5
	CP002961.1	<i>Emiticicia oligotrophica</i> DSM 17448	<i>terpene</i>	Carotenoid	28
	CP002542.1	<i>Fluviicola taffensis</i> DSM 16823	<i>aryl polyene-resorcinol</i>	Flexirubin	36
	CP002455.1	<i>Weeksella virosa</i> DSM 16922	<i>aryl polyene-resorcinol</i>	Flexirubin	61
	CP000875.1	<i>Herpetosiphon aurantiacus</i> DSM 785	<i>t1pks-nrps</i>	Pellasonen	75
	CP000875.1	<i>Herpetosiphon aurantiacus</i> DSM 785	<i>nrps</i>	Jagaricin	13
<i>Chloroflexi</i>	CP000875.1	<i>Herpetosiphon aurantiacus</i> DSM 785	<i>t1pks</i>	Lidamycin	8
	CP001032.1	<i>Opitutus terrae</i> PB90-1	<i>t1pks-nrps</i>	Arthrofactin	33
	CP001744.1	<i>Planctopirus limnophila</i> DSM 3776 chromosome	<i>t3pks-otherks</i>	Nosiheptide	11
<i>PVC</i>	CP002546.1	<i>Rubinisphaera brasiliensis</i> DSM 5305	<i>ectoine</i>	Ectoine	100

## 5.2 Klonbibliothek

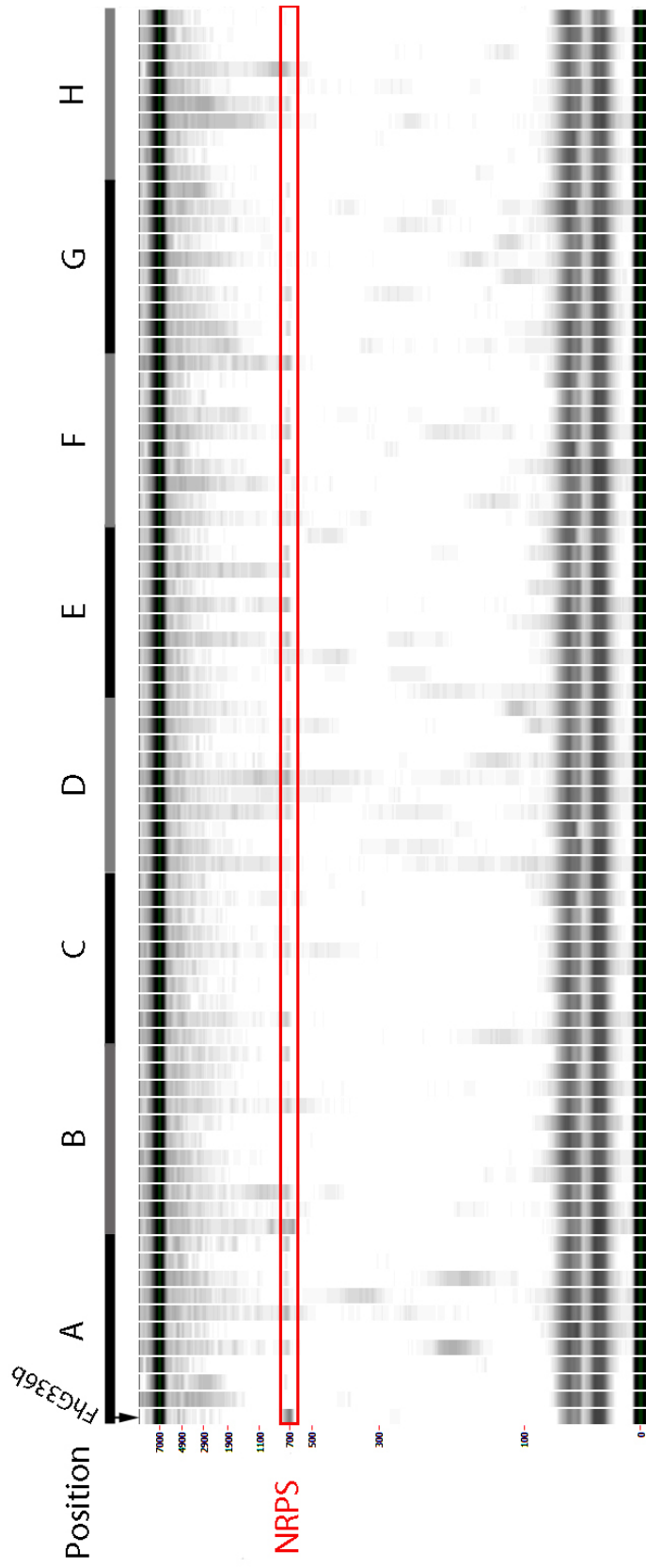


Abbildung 5.2: Markergen-spezifische Identifizierung der Adenyldomäne von nicht-ribosomalen Peptidsynthasen der Termitennest-Klonbibliothek

5.2 Klonbibliothek

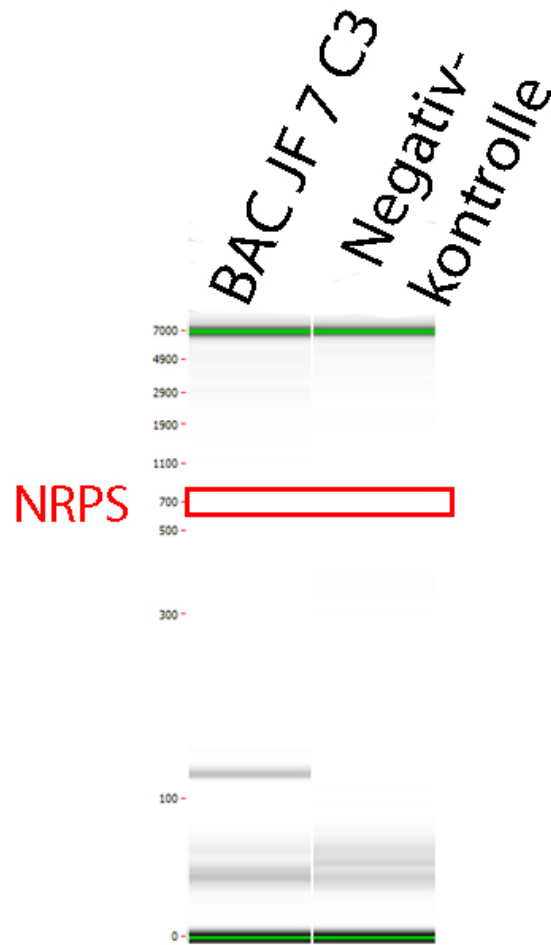


Abbildung 5.3: Markergen-spezifische Identifizierung von NRPS-Gencluster des isolierten BACs aus Klon JF 7 C3

### 5.3 Technische Daten der chemischen Analytik

HPLC	Agilent 1290 Infinity LC		
HR-QTOF-MS-MS	maXis IITM, Bruker Daltronics		
Säule	ACQuity UPLC BEH C 18 1,7 µm 21*100 mm		
Fluss	0,6 mL/min		
Injektionsvolumen	2 Liter		
Säulenofentemperatur	40 °C		
Eluenten	A: Wasser + 0,1% FA B: Acetonitril + 0,1% FA		
Gradient	Zeit [Min]	A [%]	B [%]
	0,0	95,0	5,0
	0,3	95,0	5,0
	18,0	4,75	95,25
	18,1	0,0	100
	22,5	0,0	100
	22,6	95,0	5,0
25	95,0	5,0	
Split	1 zu 10		
Kalibrant	Natriumformiat (positiv)		
Ionenquelle	Source Type	ESI	
	Set Dry Heater	250 °C	
	Set Dry Gas	7,5 L/min	
	Set Nebulizer	1,6 bar	
	Set Capillary	4500 V	
	Set End Plate Offset	-500 V	
Detektor	Set Target Mass Start	500 m/z	
	Set Target Mass End	2000 m/z	
MS-MS	Collision Energy	28,0 – 35,05 eV	
	Collision Gas	N <sub>2</sub>	

Abbildung 5.4: Technische Daten der verwendeten UHPLC-Methode

## 5.4 **Verwendete Chemikalien**

Tabelle 5.6: Verwendete Antibiotika

Chemikalie	Hersteller	Artikelnr. oder CAS
Gentamycin	Sigma-Aldrich	1405-41-0
Ampicillin	Sigma-Aldrich	69-53-4
Chloramphenicol	Sigma-Aldrich	56-75-7

## 5.4 Verwendete Chemikalien

Tabelle 5.7: Verwendete Chemikalien

Chemikalie	Hersteller	Artikelnr. oder CAS
Lösliche Stärke	Sigma-Aldrich	9005-84-9
Agar	<i>Fisher Scientific</i>	11793523
Hefeextrakt	Difco	LP0021T
Trypton	Sigma-Aldrich	93657-500G
Natriumchlorid	Sigma-Aldrich	7647-14-5
Natriumacetat	Sigma-Aldrich	127-09-3
Isopropanol	<i>Fisher Scientific</i>	67-63-0
Müller-Hinton II Medium	<i>Fisher Scientific</i>	10681675
Aceton	<i>Fisher Scientific</i>	67-64-1
Methanol	<i>Fisher Scientific</i>	10516279
Ethanol	<i>Fisher Scientific</i>	64-17-5
Ethylacetat	<i>Fisher Scientific</i>	141-78-6
Nykodenz	WAK-Chemie Medical GmbH	NC93227429
Tris-HCl	Sigma-Aldrich	1185-53-1
EDTA	Sigma-Aldrich	60-00-4
Ethidiumbromid	Sigma-Aldrich	1239-45-8
Spermidine	Sigma-Aldrich	124-20-9
SDS	Sigma-Aldrich	151-21-3
Sarkosyl	Sigma-Aldrich	137-16-6
PMSF	Sigma-Aldrich	329-98-6
Low-melting Agarose	Lonza	50100
Polyethylene glycol 8000	Sigma-Aldrich	25322-68-3
IPTG	<i>Fisher Scientific</i>	123-91-1
X-Gal	<i>Thermo Fisher</i>	R0941
Pepton	Sigma-Aldrich	91079-38-8
Casaminosäure	<i>Fisher Scientific</i>	11307719
Natriumpyruvat	Sigma-Aldrich	113-24-6
Di-Kaliumhydrogenphosphat	Sigma-Aldrich	7758-11-4
Magnesiumsulfat Heptahydrat	Sigma-Aldrich	10034-99-8
MES-Hydrat	Sigma-Aldrich	1266615-59-1
Essigsäure	<i>Fisher Scientific</i>	64-19-7
Trizma-Base	Sigma-Aldrich	77-86-1
Borsäure	Sigma-Aldrich	10043-35-3
Kalium-dihydrogenphosphat	Sigma-Aldrich	7778-77-0
Diprotin A	Enzo Life Science	90614-48-5
Ameisensäure	<i>Fisher Scientific</i>	64-18-6
Acetonitril	<i>Fisher Scientific</i>	75-05-8

## 5.5 Verwendete Geräte

Tabelle 5.8: Verwendete Geräte

Bezeichnung	Hersteller	Zweck
Victor2 1420 <i>Multilabel Reader</i>	Perkin Elmer	Primäre Untersuchung der Klonbibliotheken
Lumistar	BMG <i>Labtech</i>	Sekundäre Untersuchung der Klonbibliotheken
CHEF PFGE	Bio-Rad	Auftrennung hochmolekularer DNS
Savant SC210 P1	<i>Thermo Fisher</i>	Aufkonzentrieren von Extrakten
<i>Nanodrop</i>	<i>Thermo Fisher</i>	Quantifizierung von DNS
<i>Micropulser</i>	Bio-Rad	Elektroporation von <i>Escherichia coli</i>
<i>CyBi-Well vario</i>	CyBio AG	Verteilung von Extrakten
C1000 <i>Touch Thermal Cycler</i>	Bio-Rad	PCR
<i>TissueLyzer</i>	Qiagen	Extraktion von Proben
<i>LabChip GX Touch HT</i>	PerkinElmer	Analyse von PCR-Ergebnissen im 96-well/Format
FACScalibur	BD	Quantifizierung von Bakterien
RC 5 Zentrifuge	Sorvall	Zentrifugieren von Kulturen
5810R Zentrifuge	Eppendorf	Zentrifugieren im µL-mL Bereich
Alpha 2-4 LSCplus	Christ	Gefriertrocknung
Turrax T25	IKA	Turaxieren
MinION	<i>Oxford Nanopore</i>	Sequenzierung

## 5.6 Abkürzungsverzeichnis

Tabelle 5.9: Aminosäuren und Abkürzungen

einbuchstabiges Symbol	Aminosäure	dreibuchstabile Abkürzung
A	Alanin	Ala
R	Arginin	Arg
N	Asparagin	Asn
D	Asparaginsäure	Asp
B	Asparagin oder Asparaginsäure	Asx
C	Cystein	Cys
Q	Glutamin	Gln
E	Glutaminsäure	Glu
Z	Glutamin oder Glutaminsäure	Glx
G	Glycin	Gly
H	Histidin	His
I	Isoleucin	Ile
L	Leucin	Leu
K	Lysin	Lys
M	Methionin	Met
F	Phenylalanin	Phe
P	Prolin	Pro
S	Serin	Ser
T	Threonin	Thr
W	Tryptophan	Trp
Y	Tyrosin	Tyr
V	Valin	Val

## 5.6 Abkürzungsverzeichnis

Tabelle 5.10: Fachtermini, Teil 1

ABC	<i>ATP-binding cassette</i>
ACP	<i>Acyl-carrier-protein</i>
<i>antiSMASH</i>	<i>antibiotics &amp; secondary metabolite analysis shell</i>
ARTS	<i>Antibiotic Resistant Target Seeker</i>
AT	Acyltransferase
ATCC	<i>American Type Culture Collection</i>
BAC	<i>Bacterial artificial chromosome</i>
BGC	Biosynthese-Gencluster
<i>bla</i>	beta-Lactamase
C <sub>m</sub>	Chloramphenicol-Acetyltransferase
CBM	Carbohydrat-Bindemodul
CDS	<i>Coding Sequence</i>
CE	Carbohydratesterase
cm	centimeter
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
DMSO	Dimethylsulfoxid
EDTA	Ethylendiamin-tetraessigsäure
EMBL	<i>European Molecular Biology Laboratory</i>
<i>et al.</i>	<i>et alii</i> / und andere
ER	Enoylreduktase
ESI	Elektrospray-Ionisation
ESKAPE	<i>Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa, Enterobacter sp.</i>
EtAc	Ethylacetat
EtBr	Ethidiumbromid
ETOH	Ethanol
FACS	<i>Fluorescence assisted cell sorting</i>
FD	<i>fast digest</i>
FL1	Fluoreszenzintensität
FSC	<i>forward scatter</i>
GO	<i>Gene ontology</i>
GOI	<i>Gene of interest</i>
GH	Glykosylhydrolase
GT	Glykosyltransferase
IPTG	Isopropyl- $\beta$ -D-thiogalactopyranosid
KAS	Ketoacyl-Synthase
Kb	Kilobasen

## 5.6 Abkürzungsverzeichnis

Tabelle 5.11: Fachtermini, Teil 2

KBE	Kolonie-bildene-Einheiten
KEGG	<i>Kyoto Encyclopedia of Genes and Genomes</i>
KR	Ketoredukatase
KS	Ketosynthese
LB	Luria-Bertani Medium
LPS	Lipopolysaccharide
<i>luxCDABE</i>	Gene des Luciferase-Reportersystems
MCS	<i>Multiple Cloning Site</i>
Mb	Megabasen
MeOH	Methanol
MRSA	Methenicillin-resistenter <i>Staphylococcus aureus</i>
MS	Massenspektrometrie / Massenspektrometer
MSSA	Methenicillin-sensitiver <i>Staphylococcus aureus</i>
MT	Methyltransferase
NAD	<i>Nicotinamide adenine dinucleotide</i>
NCBI	<i>National Center for Biotechnology Information</i>
NRPS	Nicht-ribosomale Peptidsynthese
OD	optische Dichte
ORF	<i>Open Reading Frame</i>
OTU	<i>Operational Taxonomic Unit</i>
PCR	<i>Polymerase Chain Reaction</i>
PEG	Polyethylenglykol
PFGE	Pulsfeld Gelelektrophorese
PKS	Polyketidsynthese
PUFA	<i>Polyunsaturated fatty acids</i>
Ripp	<i>Ribosomally synthesized and post-translationally modified peptide</i>
rpm	<i>rounds per minute</i>
RND	<i>Resistance Nodulation Division</i>
RepE	Replicationsprotein
RT	Raumtemperatur
SDS	<i>Sodium Dodecyl Sulfate</i>
SOC	<i>Super Optimal Broth</i>
UHPLC	<i>Ultra-High Pressure Liquid Chromatography</i>
v/v	<i>volume per volume</i>
w/v	<i>weight per volume</i>
X-Gal	5-Brom-4-chlor-3-indoxyl- $\beta$ -D-galactopyranosid
x g	Vielfaches der Erdbeschleunigung
YAC	<i>Yeast Artificial Chromosome</i>

## 5.7 Extraktanalytik

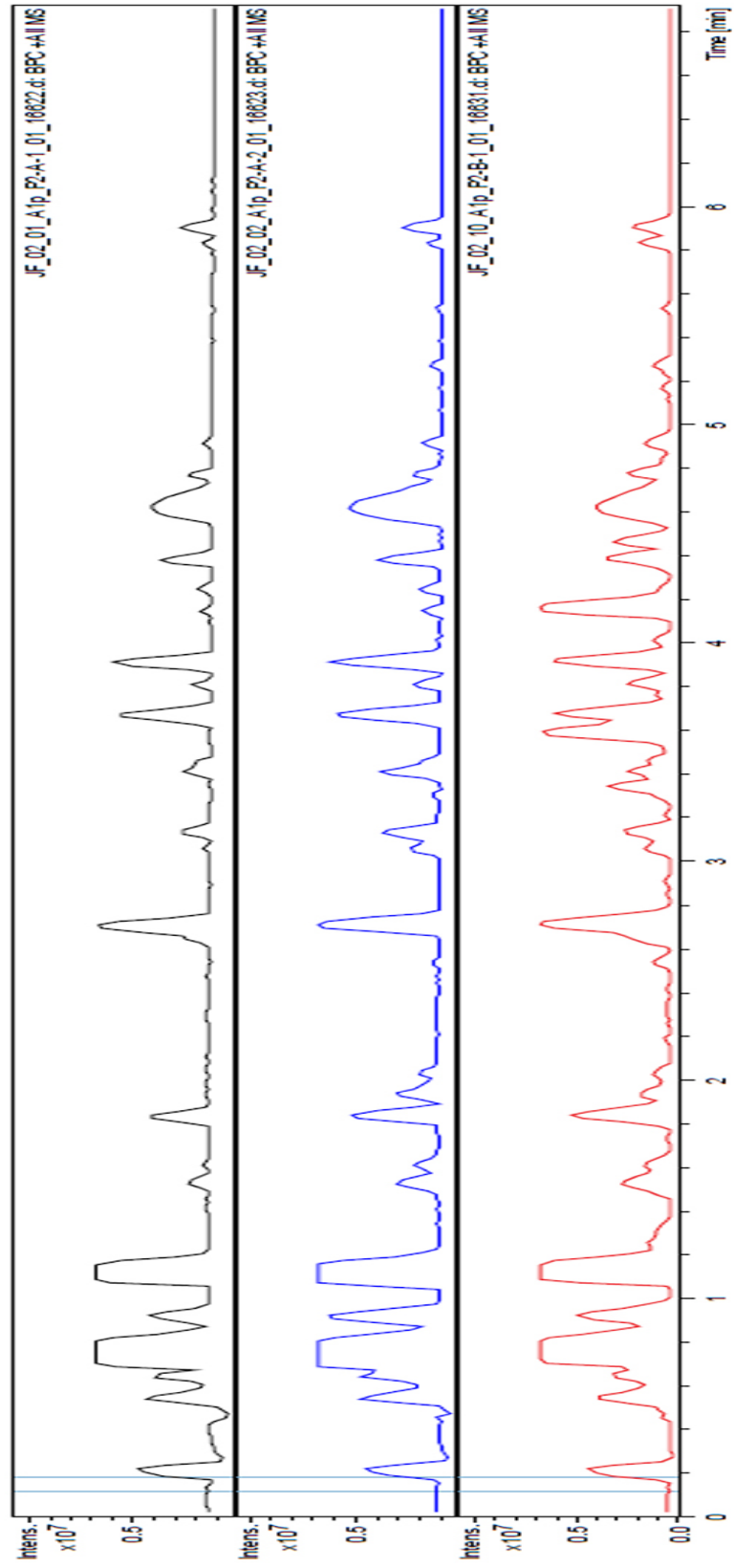


Abbildung 5.5: Chromatogramme als Liste des Klons JF 7 C4

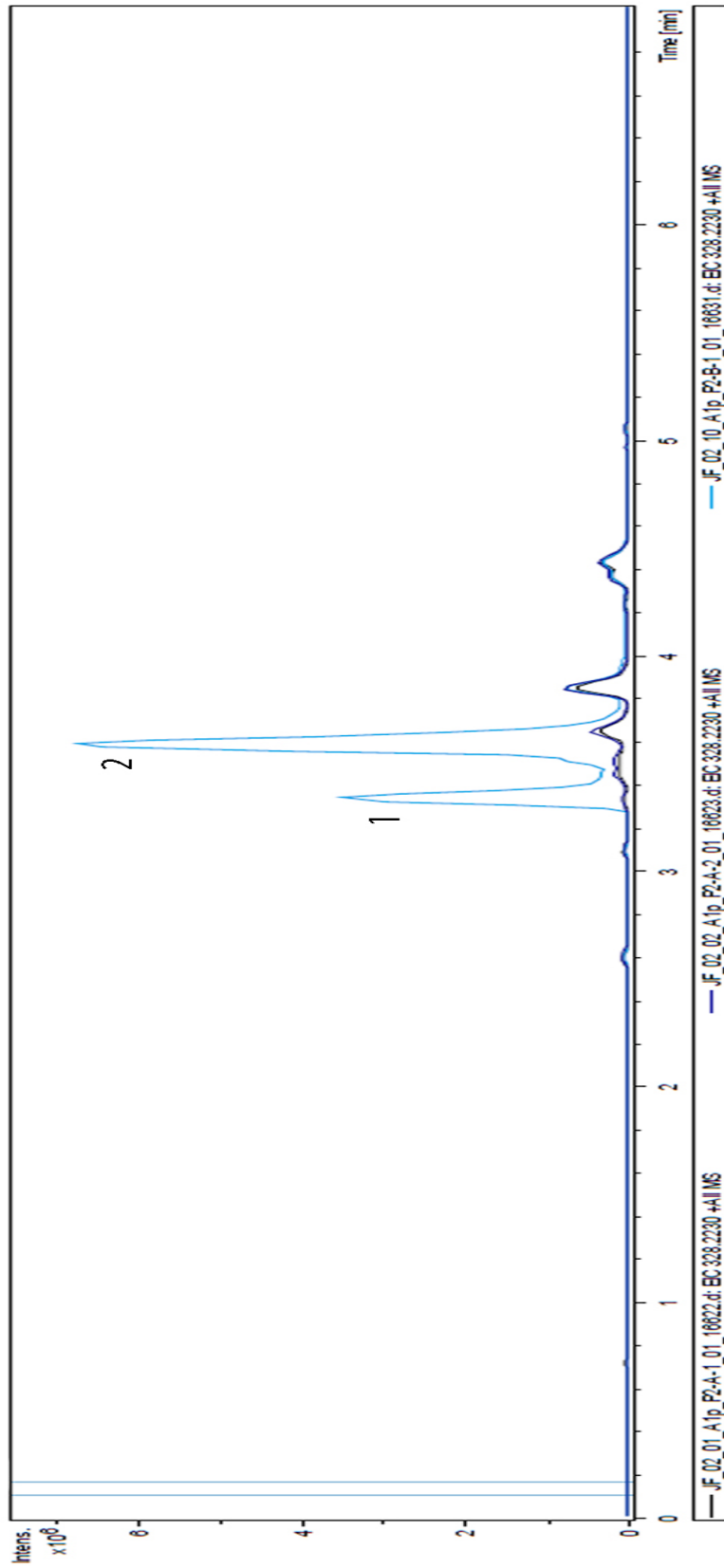


Abbildung 5.6: Signale des Klons JF 7 C4 hergestellt

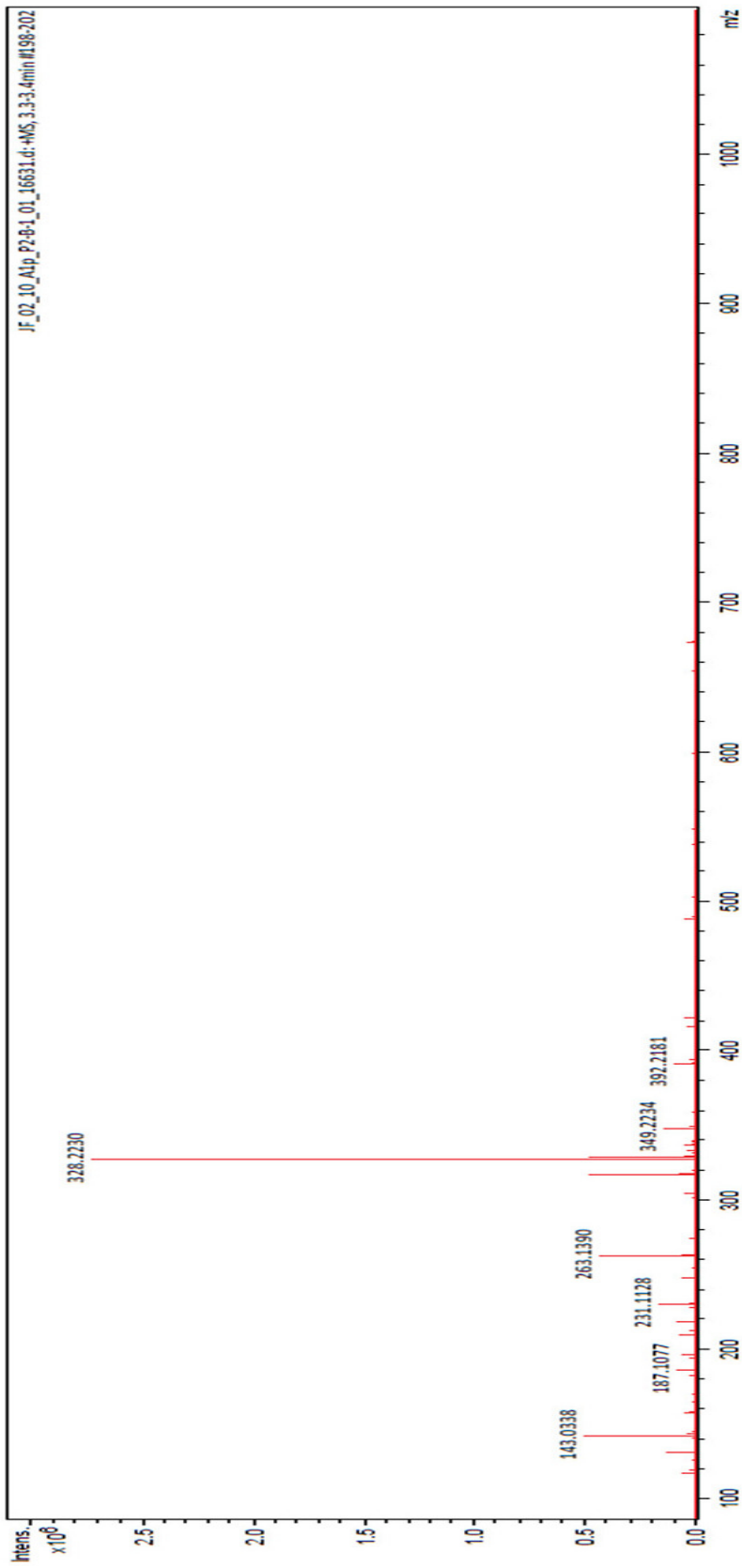


Abbildung 5.7: Massenspektrum 1 des Klons JF 7 C4

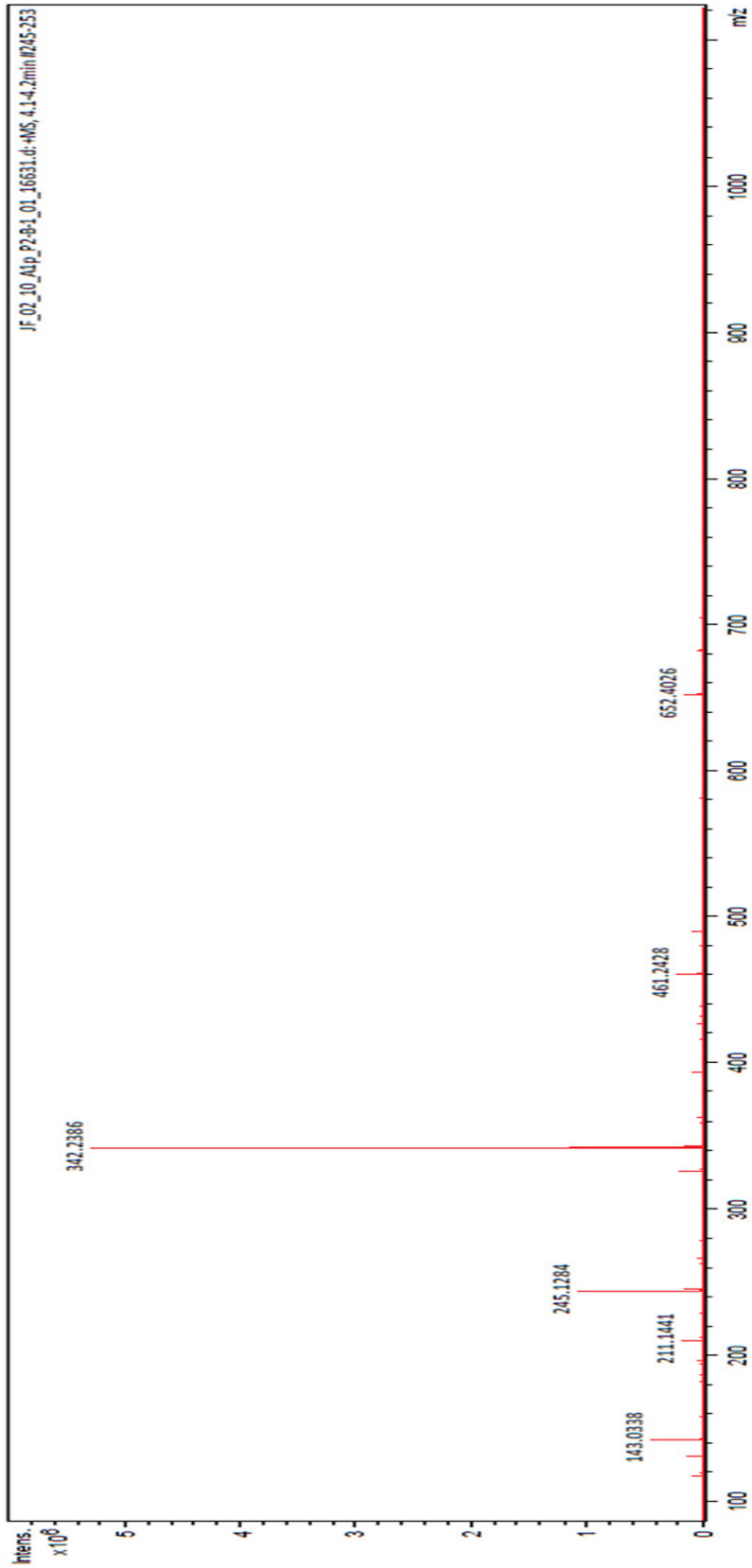


Abbildung 5.8: Massenspektrum 2 des Klons JF 7 C4

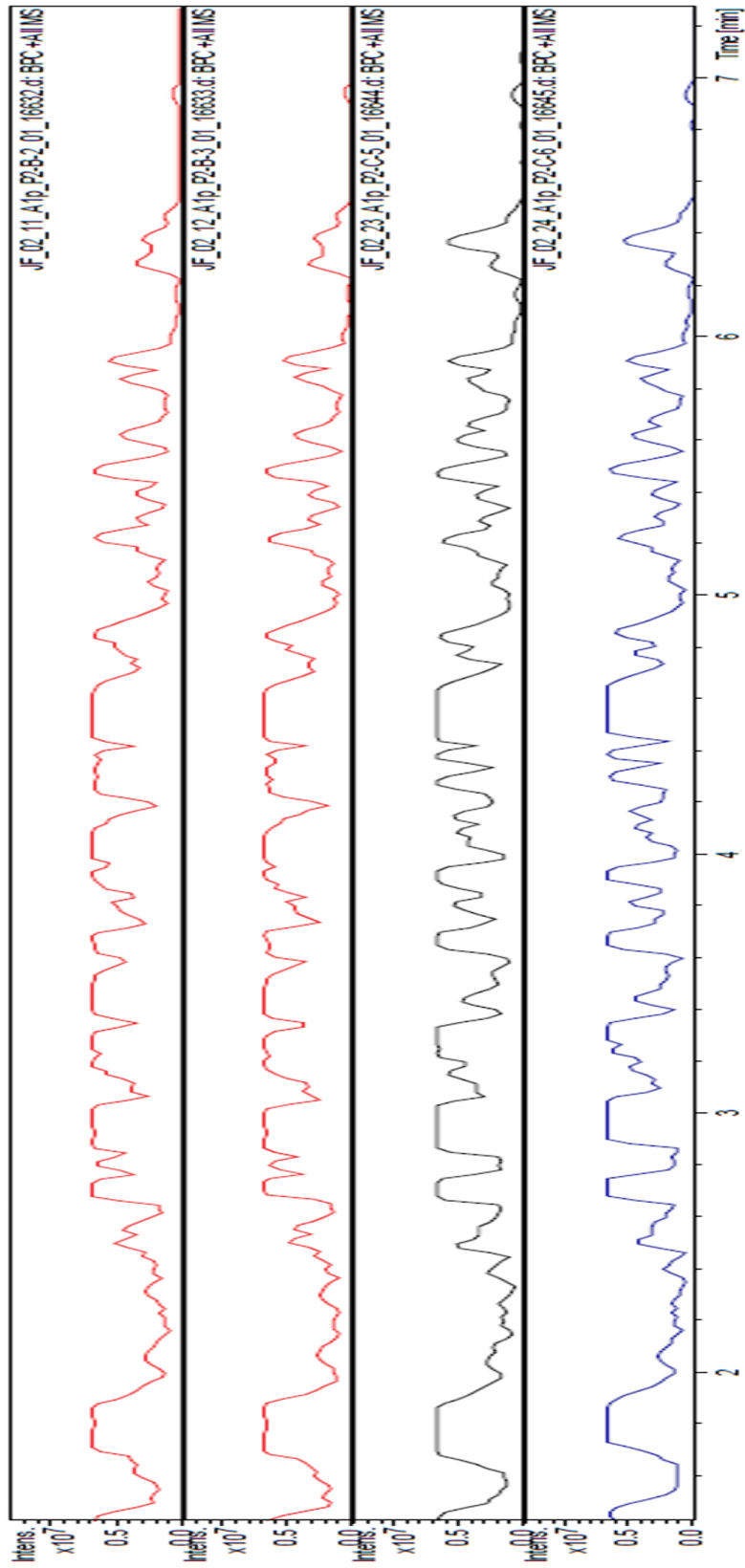


Abbildung 5.9: Chromatogramme als Liste des Klons JF 8 A9

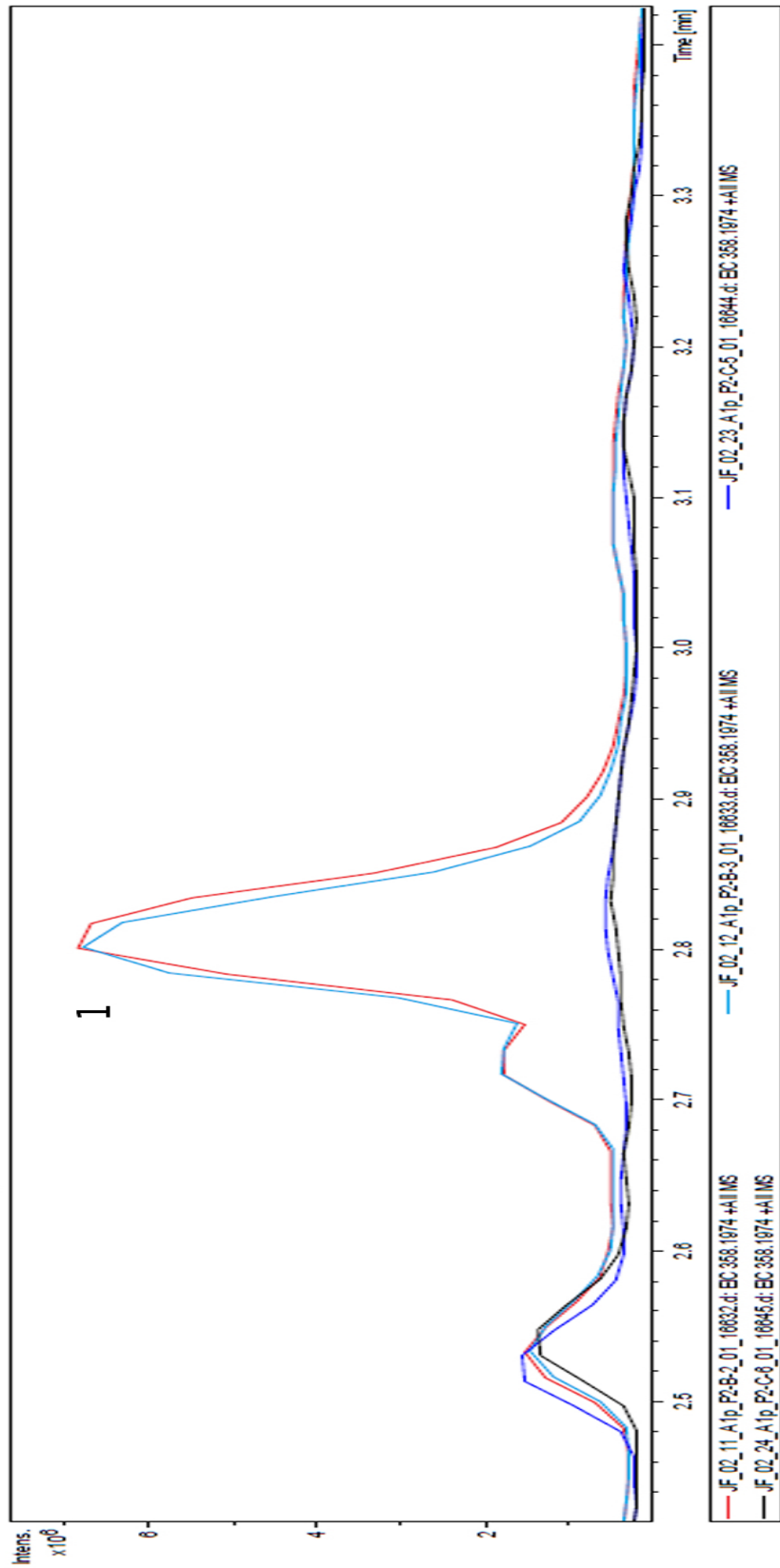


Abbildung 5.10: Signal 1 des Klons JF 8 A9 herausgestellt

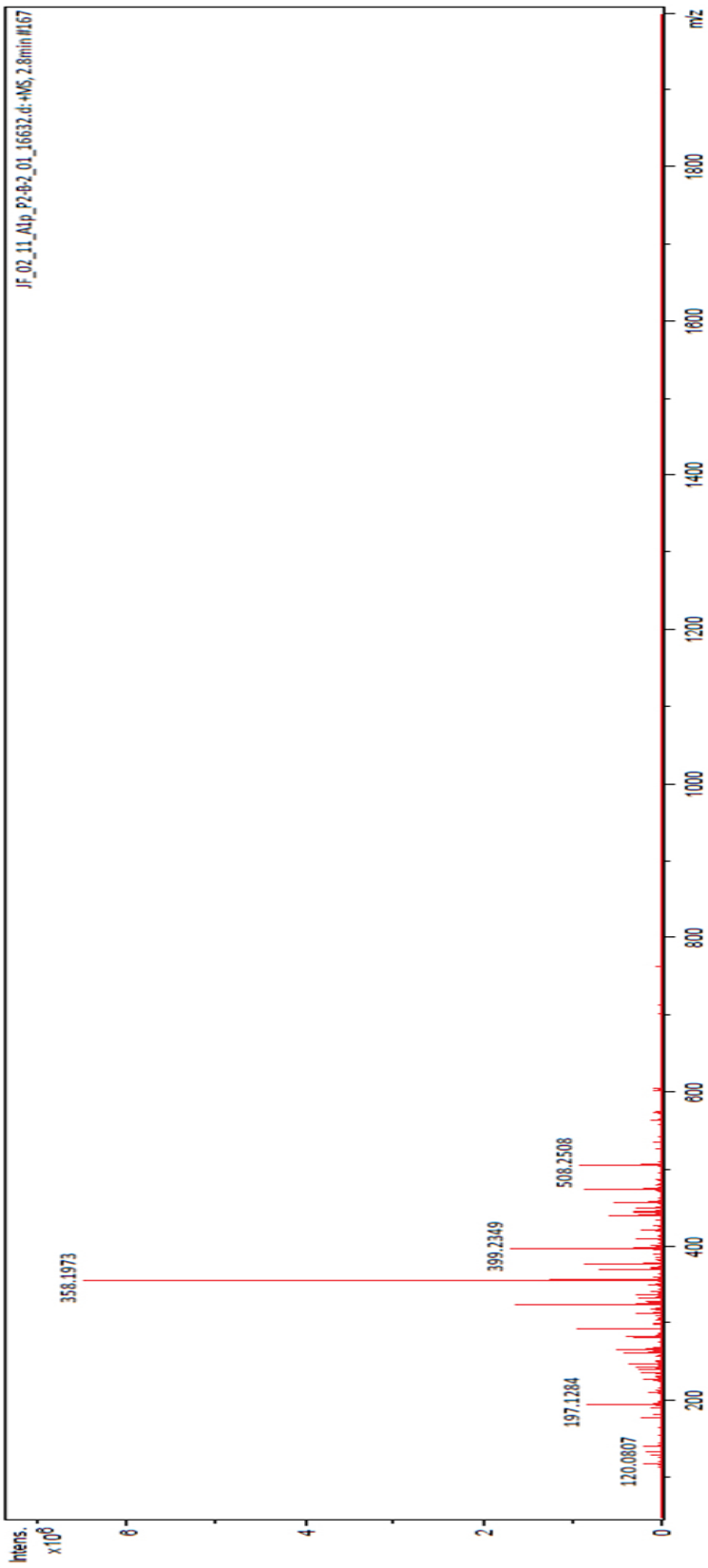


Abbildung 5.11: Massenspektrum 1 des Klons JF 8 A9

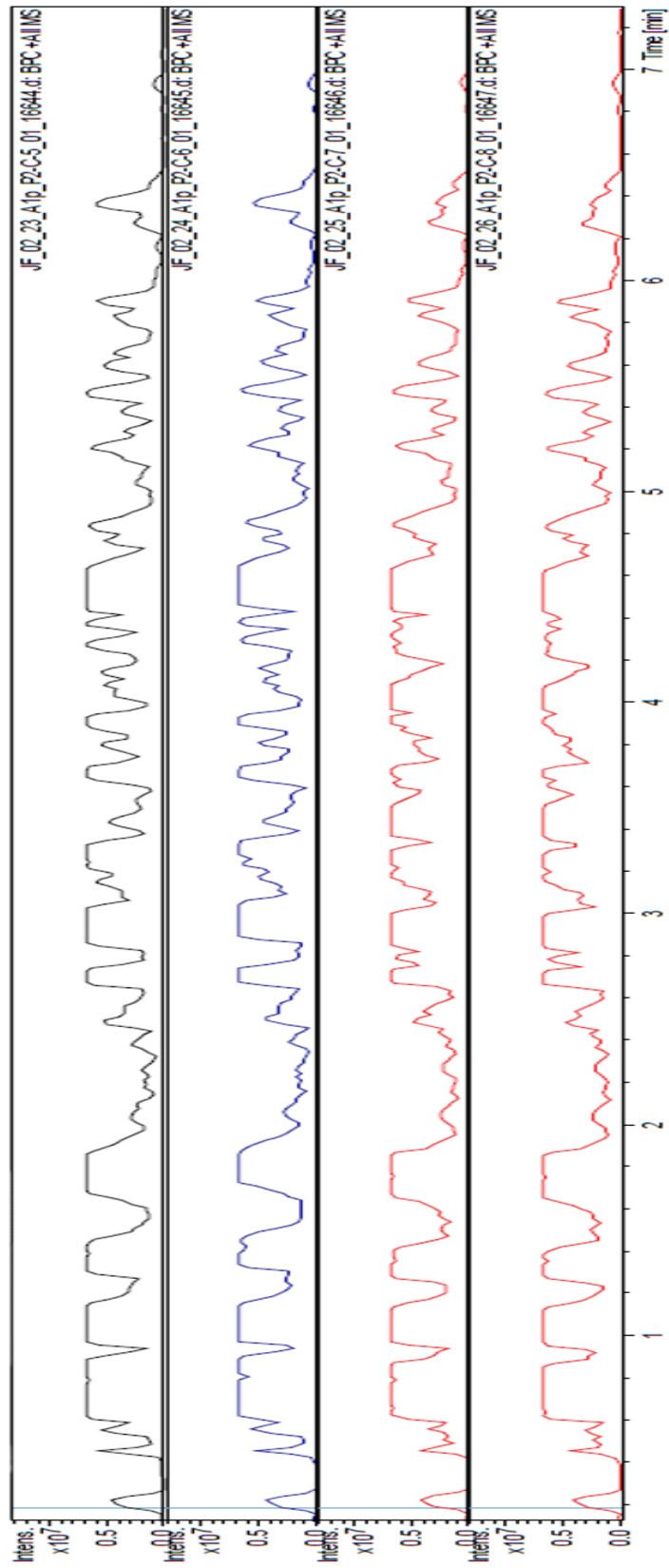


Abbildung 5.12: Chromatogramme als Liste des Klons JF 9 C12

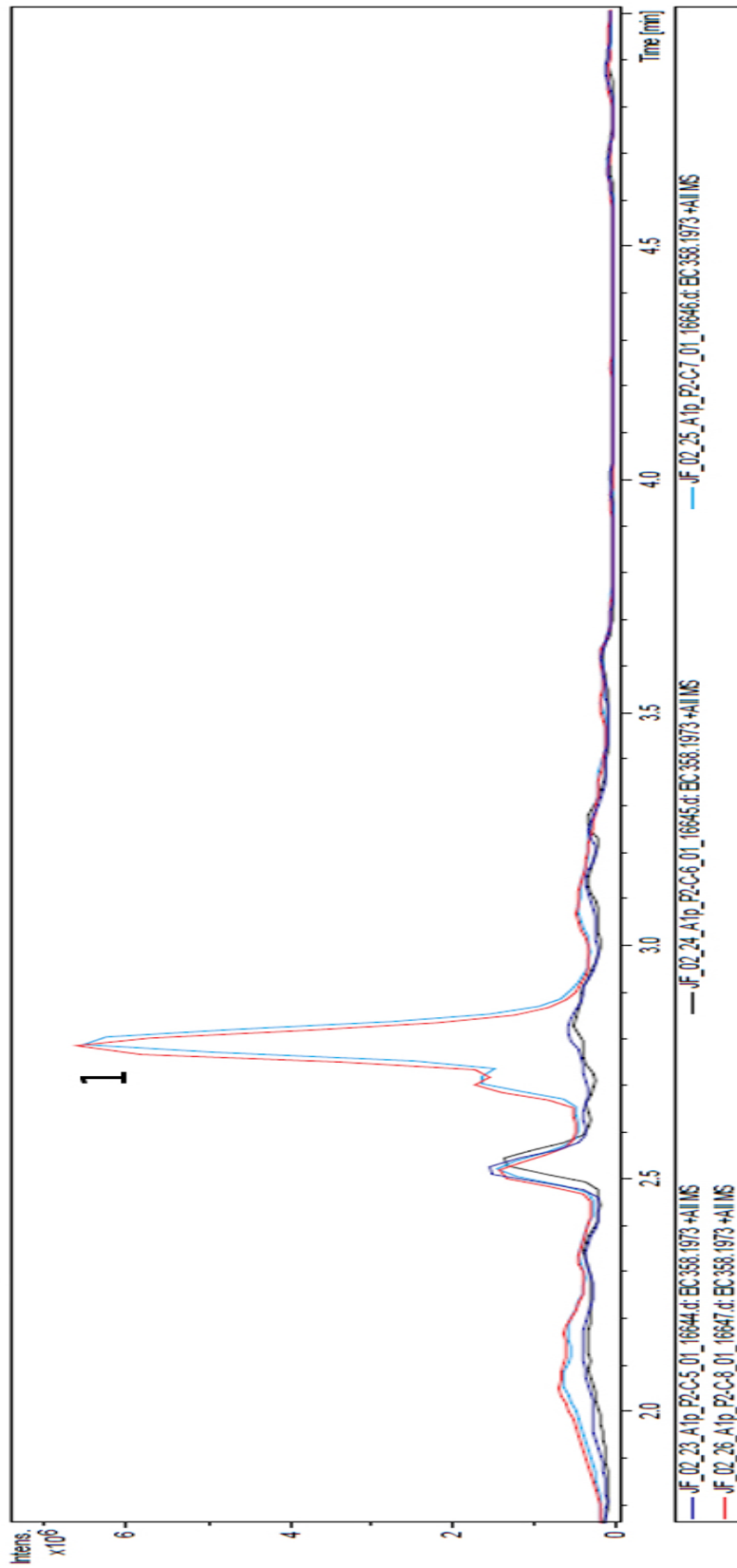


Abbildung 5.13: Signal 1 des Klons JF 9 C12 herausgestellt

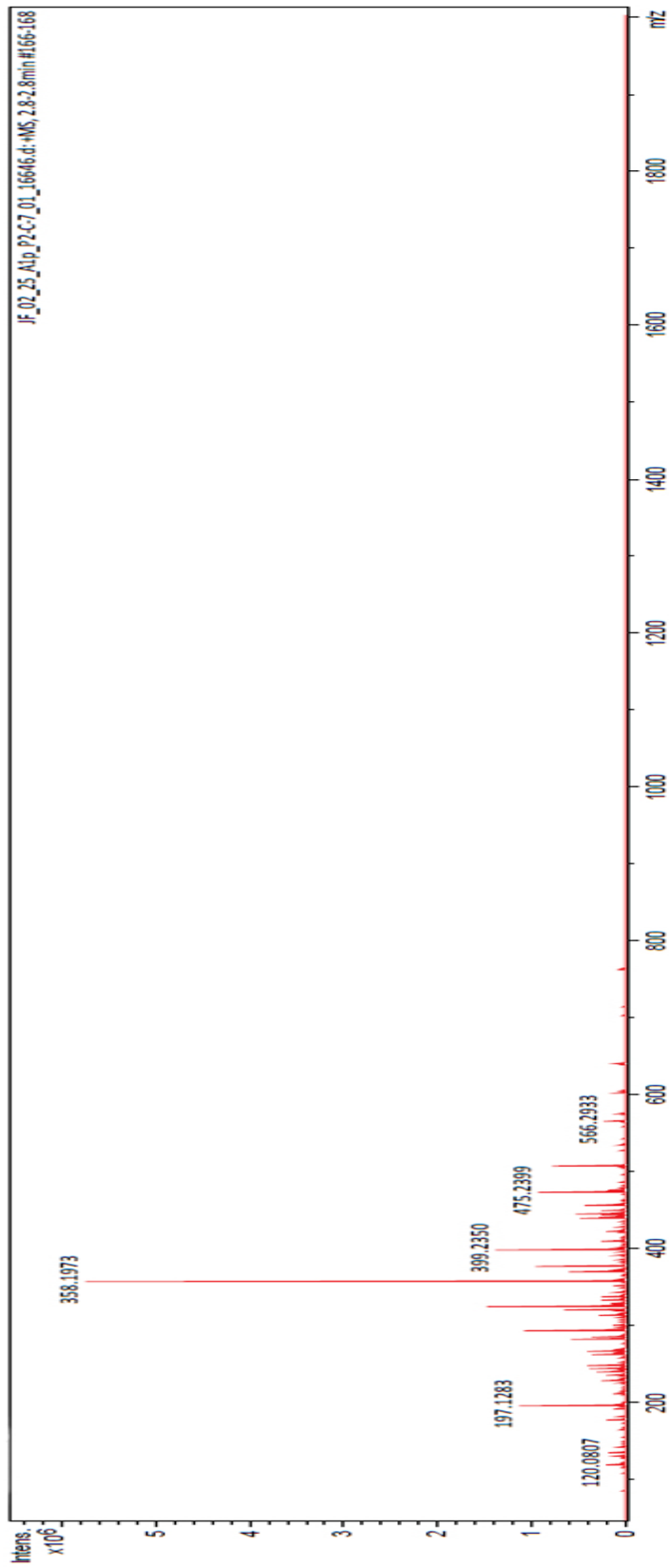


Abbildung 5.14: Massenspektrum 1 des Klons JF 9 C12

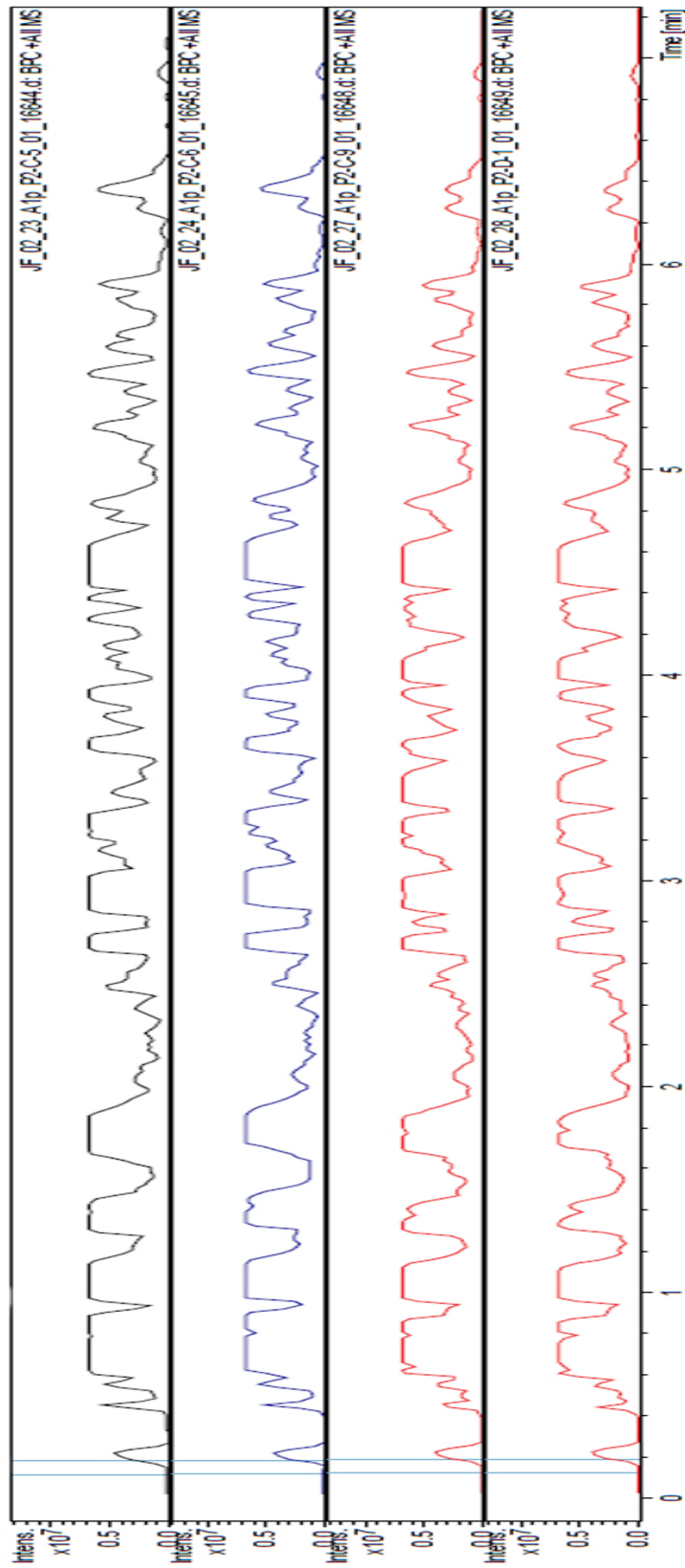


Abbildung 5.15: Chromatogramme als Liste des Klons JF 9 H5

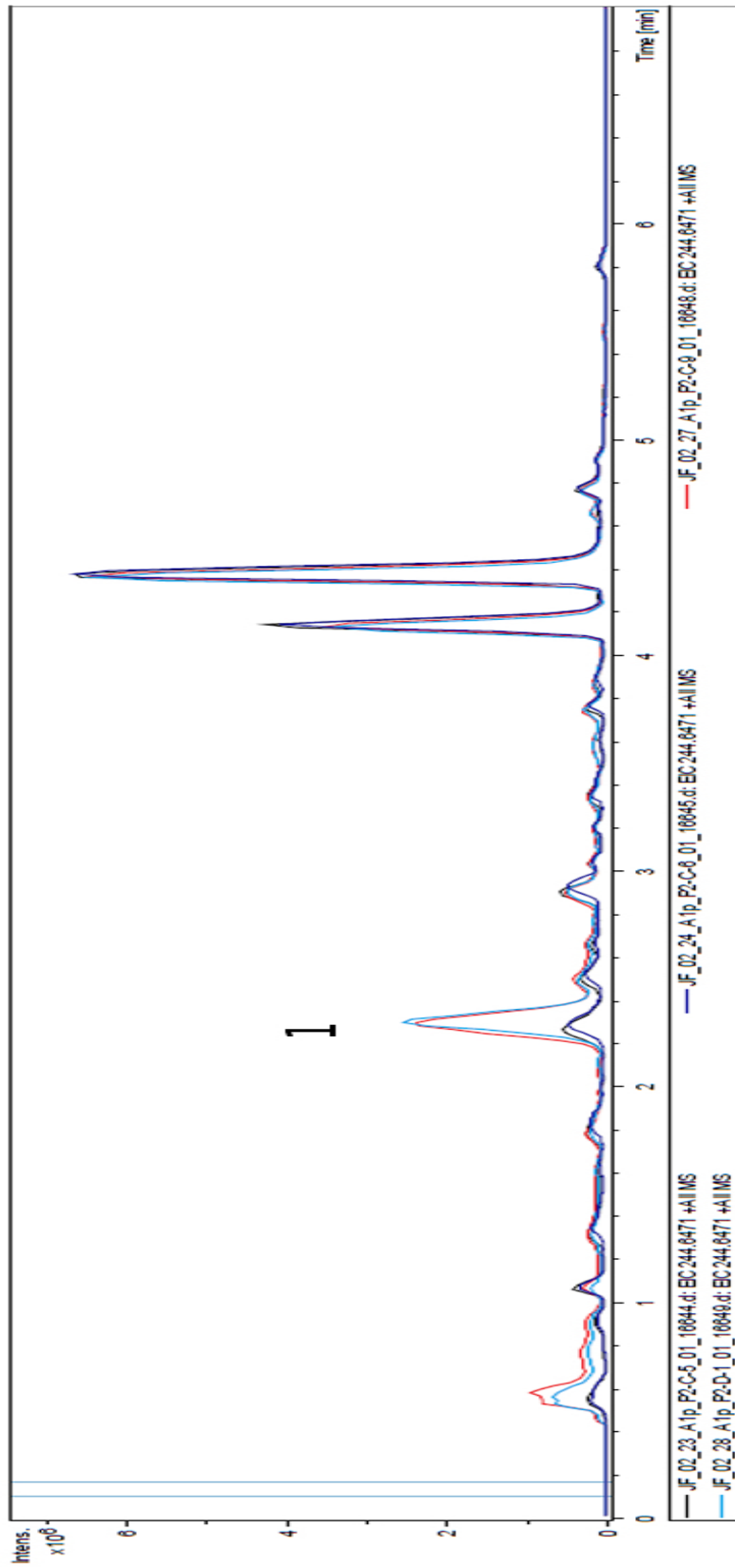


Abbildung 5.16: Signal 1 des Klons JF 9 H5 herausgestellt

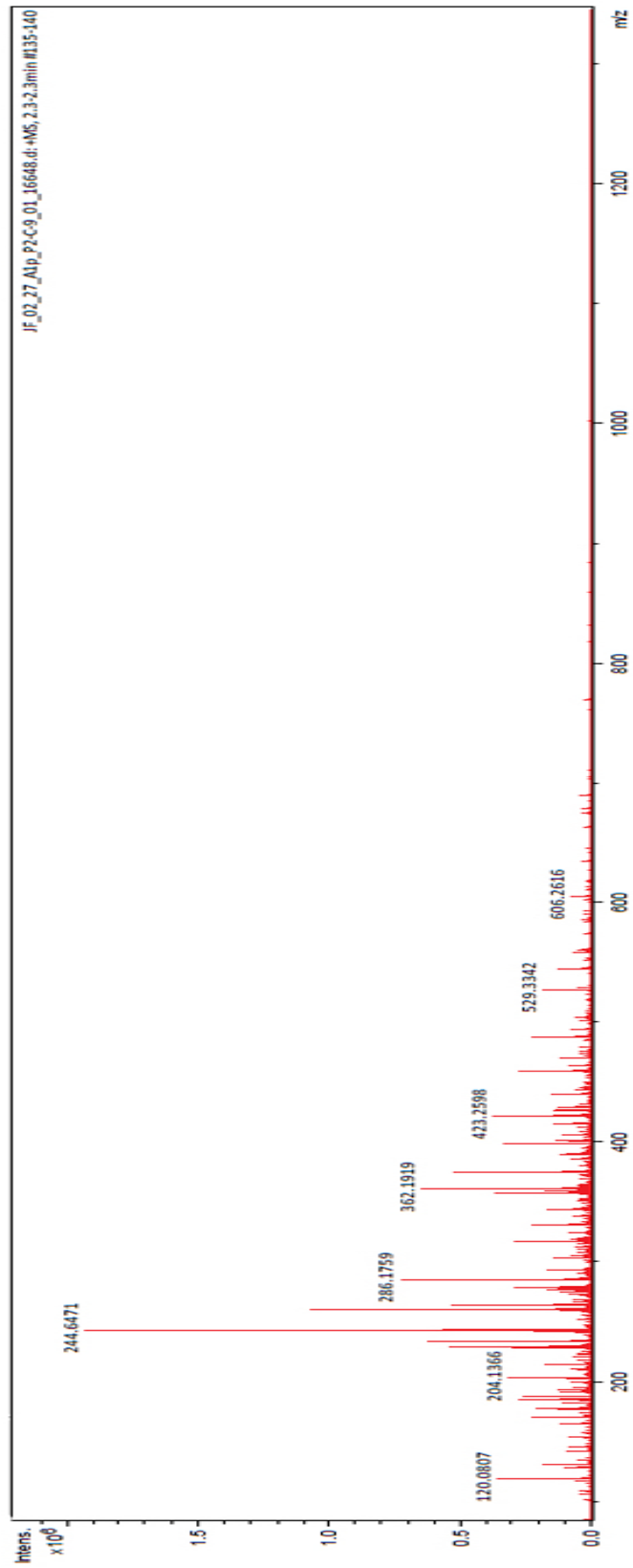


Abbildung 5.17: Massenspektrum 1 des Klons JF 9 H5

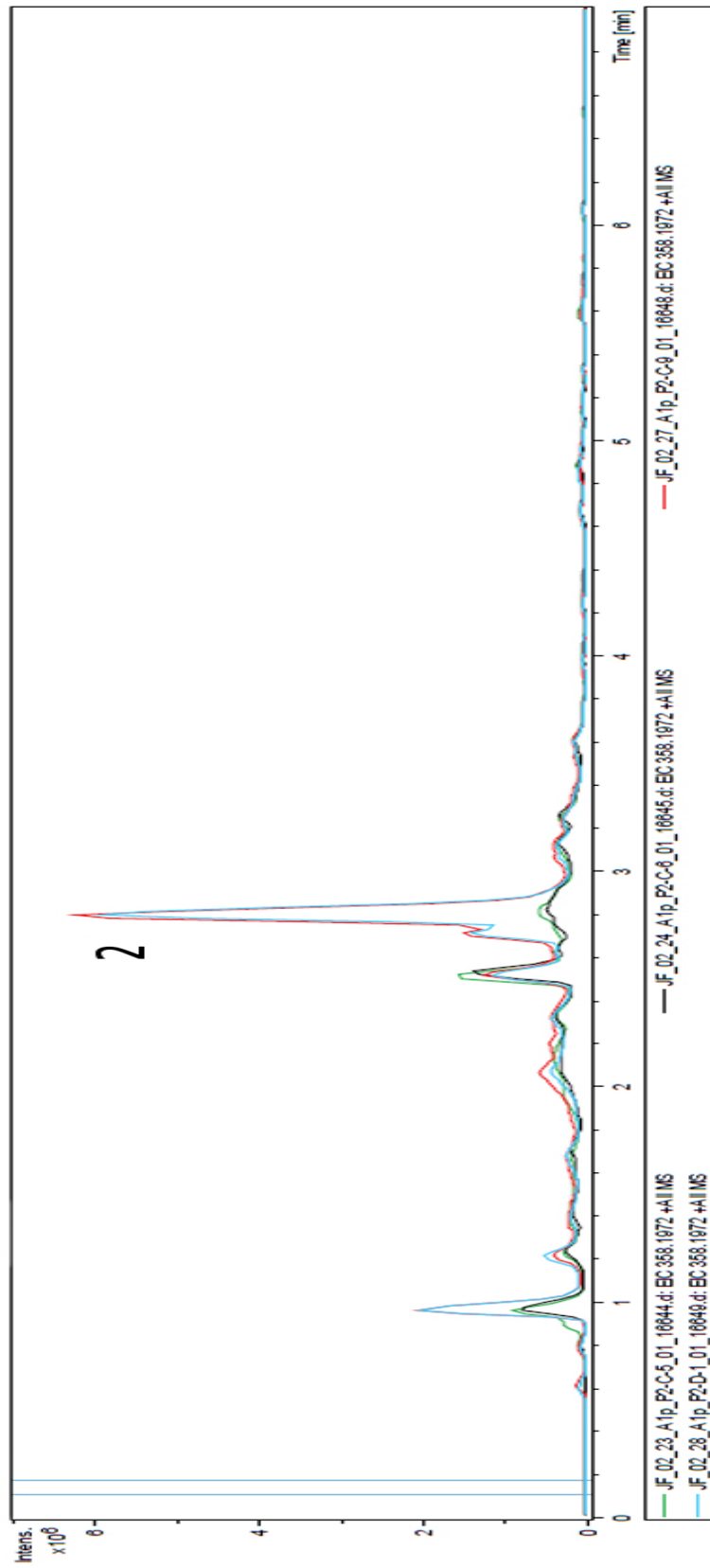


Abbildung 5.18: Massenspektrum 1 des Klons JF 9 H5

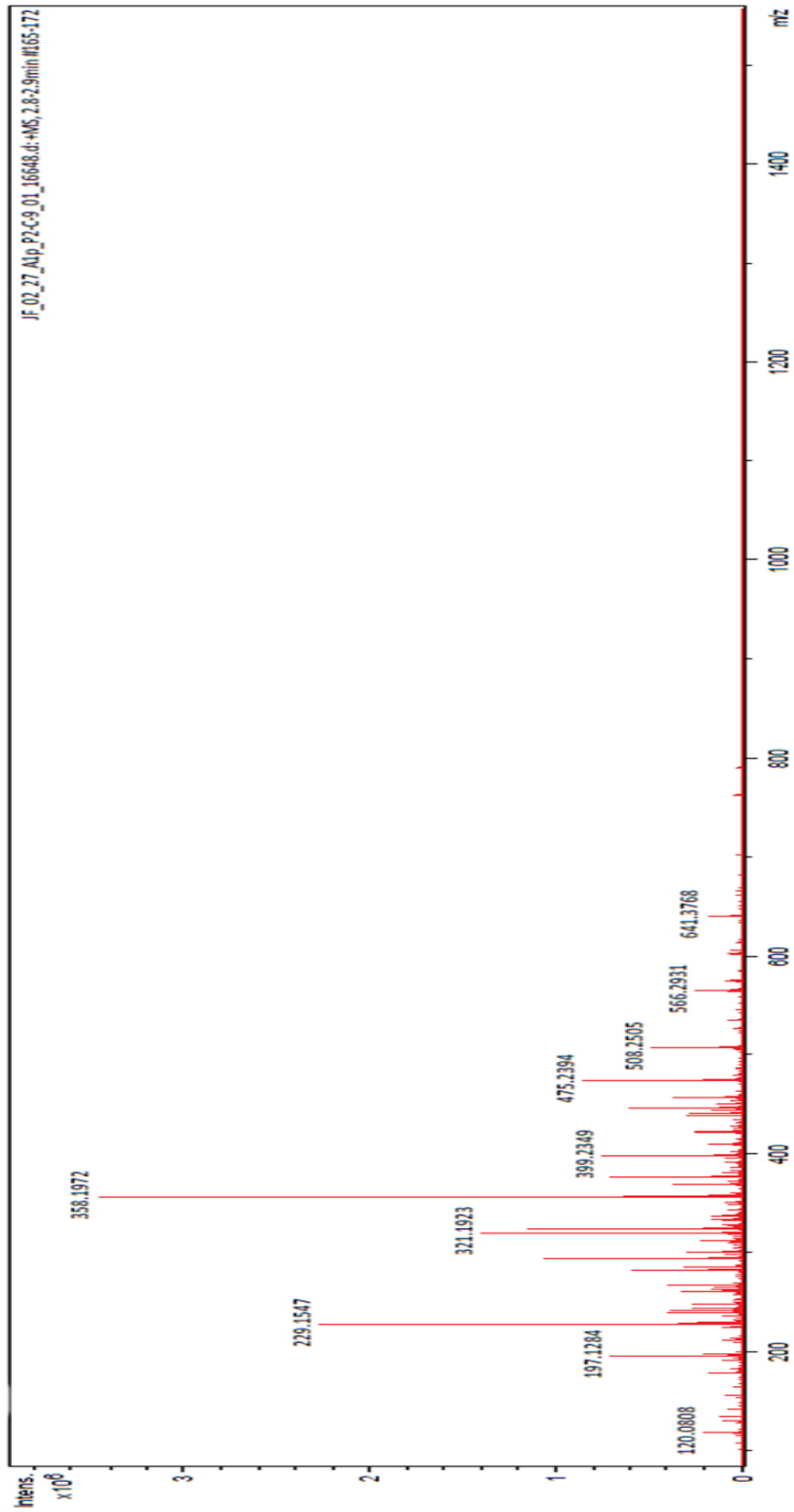


Abbildung 5.19: Massenspektrum 2 des Klons JF 9 H5

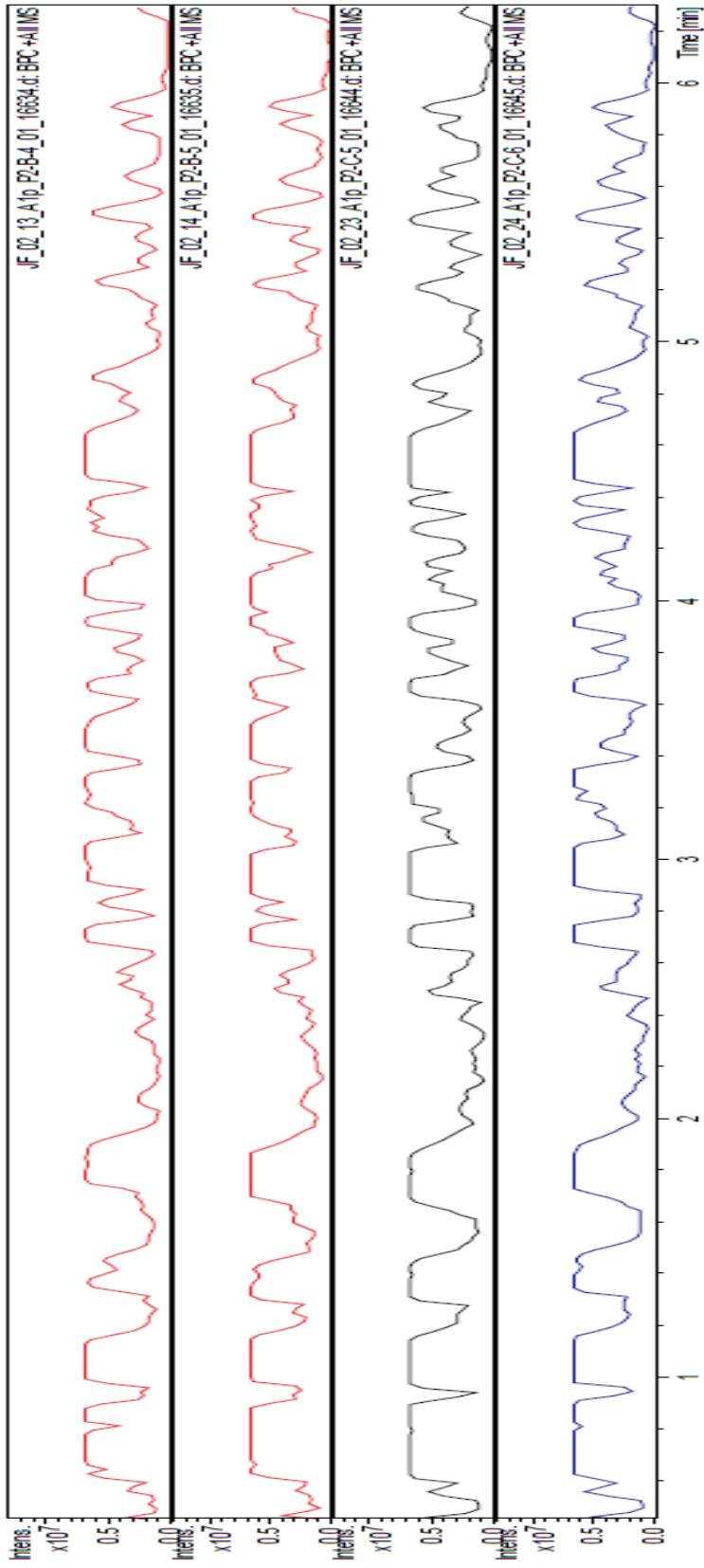


Abbildung 5.20: Chromatogramme als Liste des Klons JF 11 B9

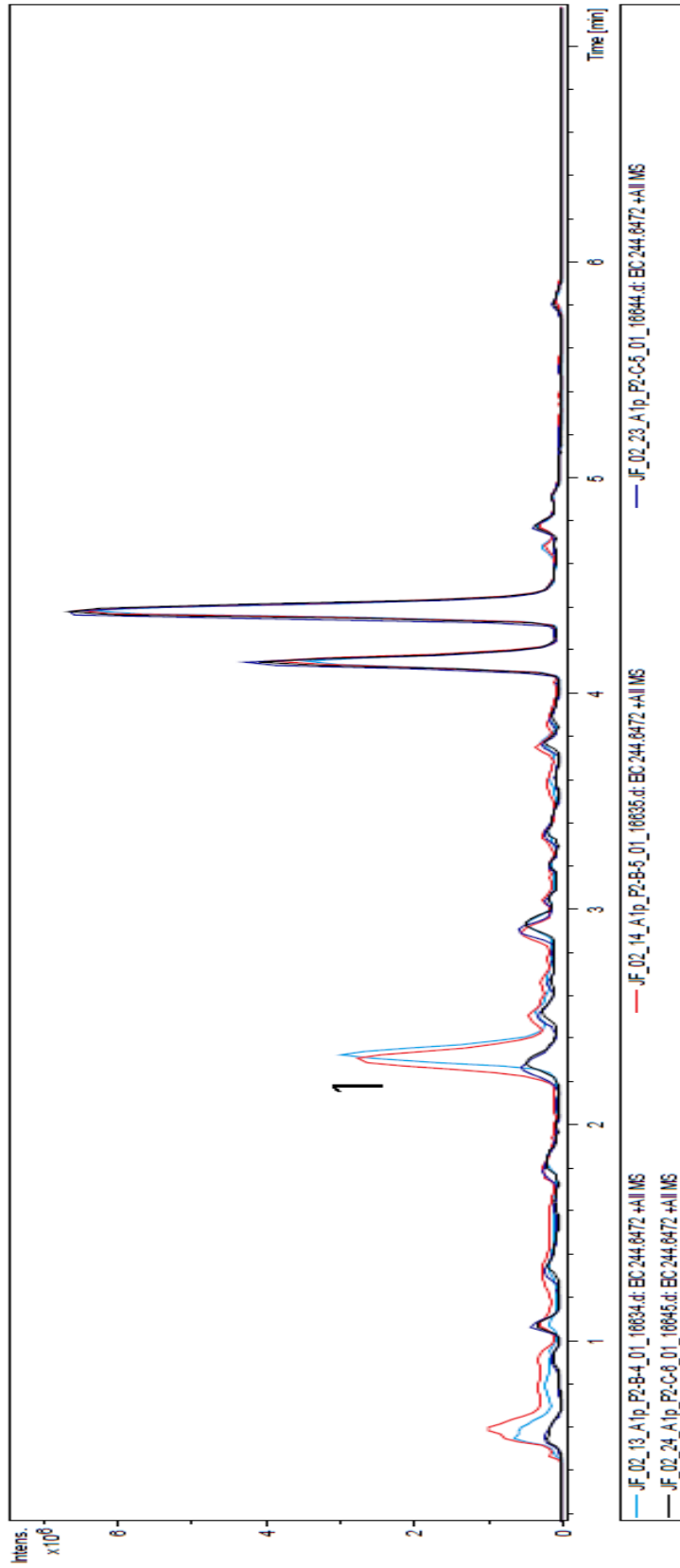


Abbildung 5.21: Signal 1 des Klons JF 11 B9 hergestellt

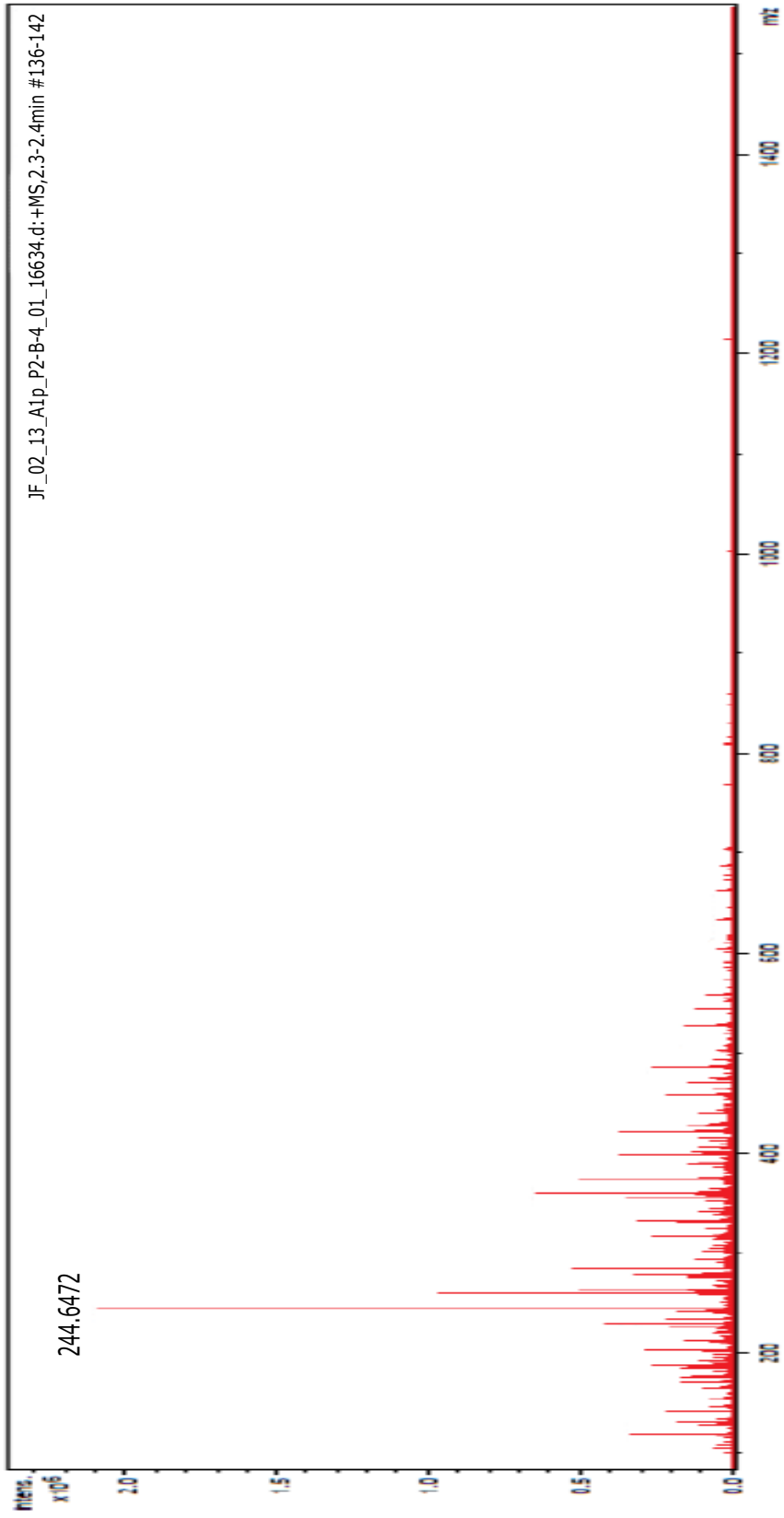


Abbildung 5.22: Massenspektrum 1 des Klons JF 11 B9

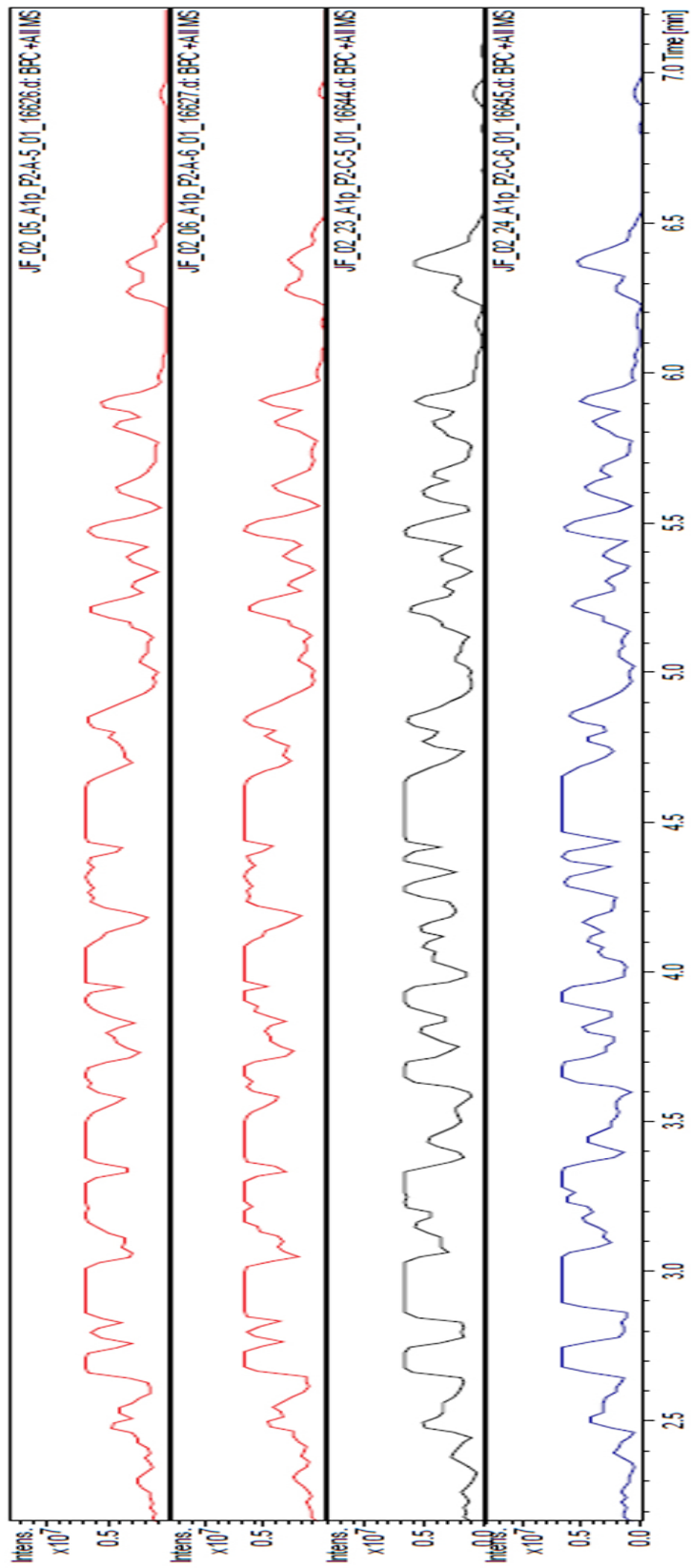


Abbildung 5.23: Chromatogramme als Liste des Klons JF\_13\_C6

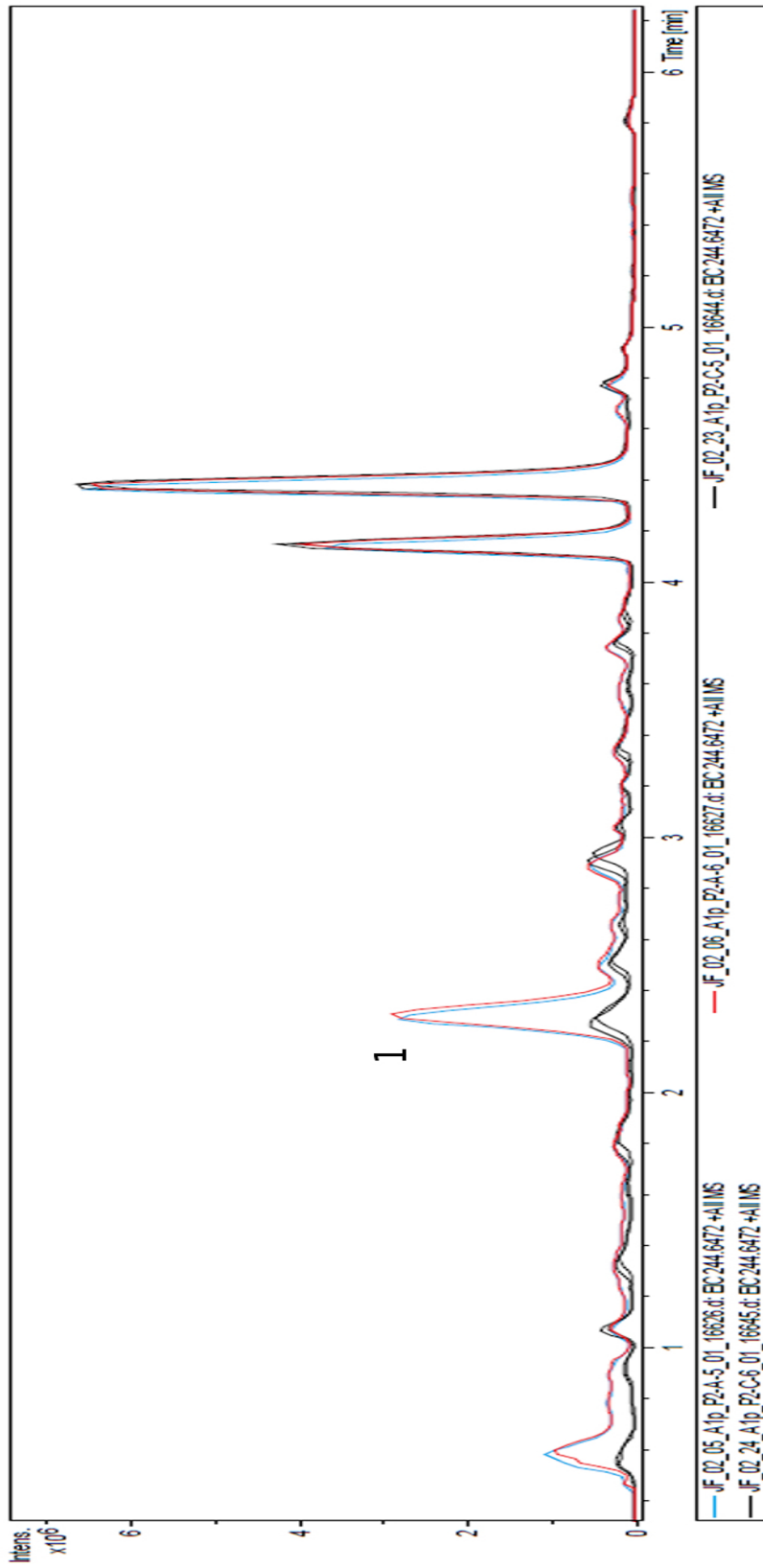


Abbildung 5.24: Signal 1 des Klons JF 13 C6 herausgestellt

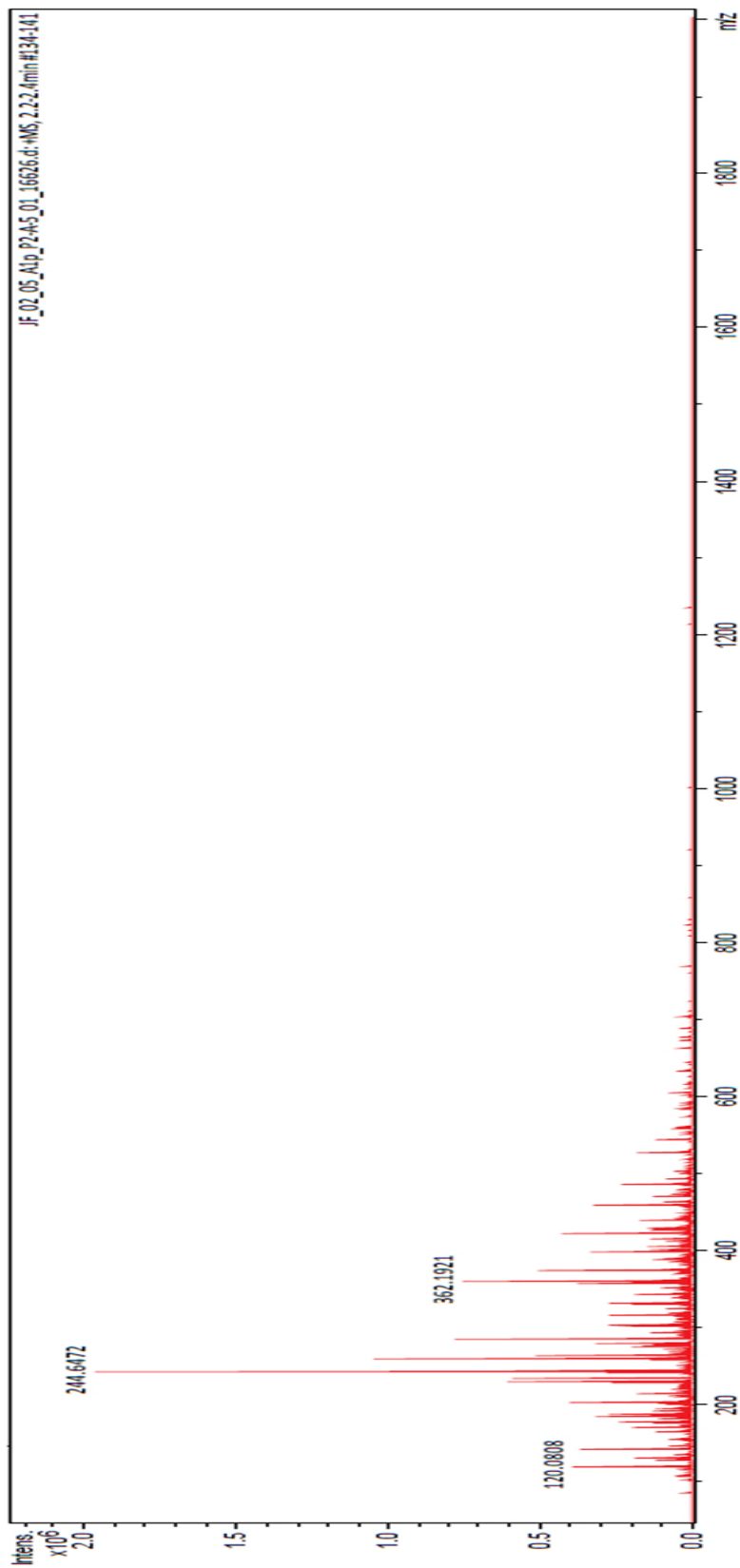


Abbildung 5.25: Massenspektrum 1 des Klons JF 13 C6

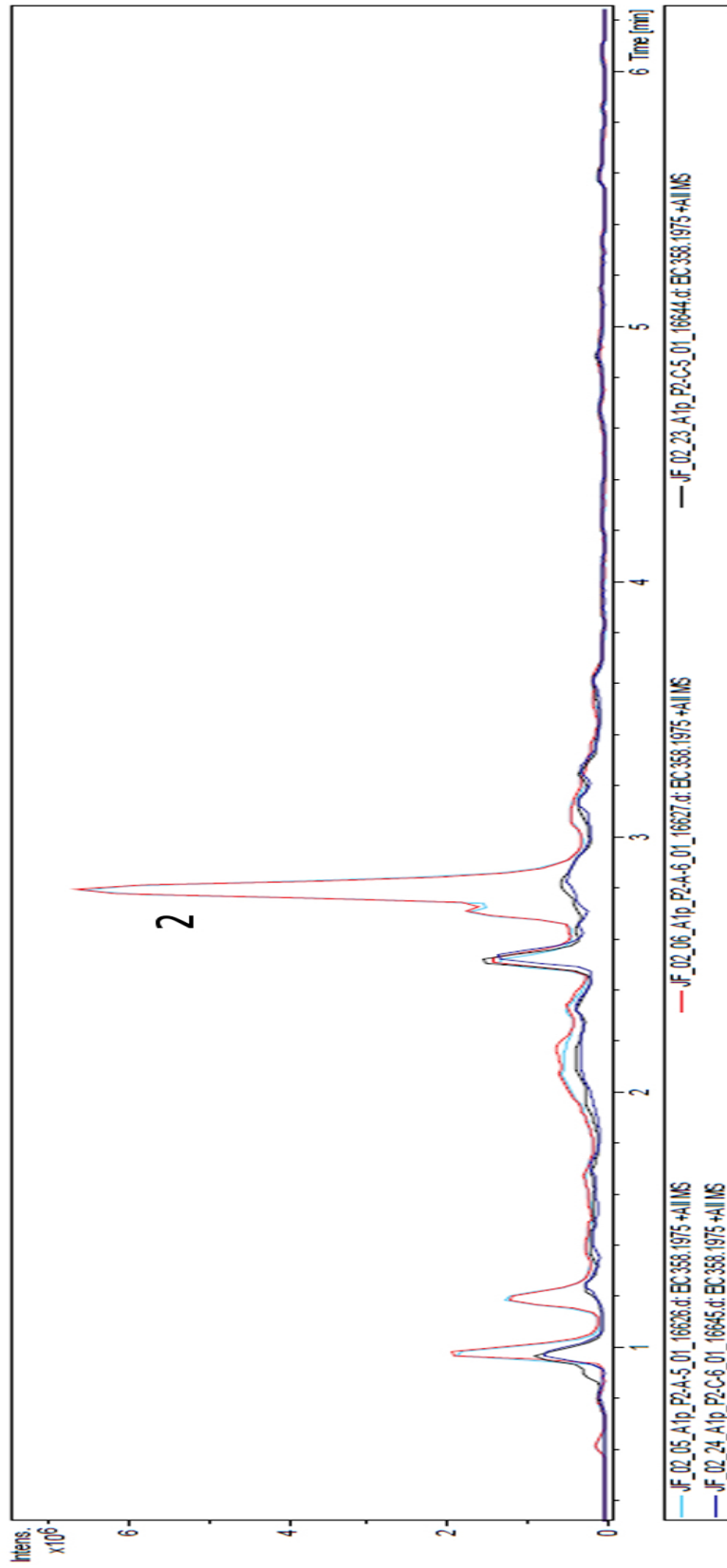


Abbildung 5.26: Signal 2 des Klons JF 13 C6 herausgestellt

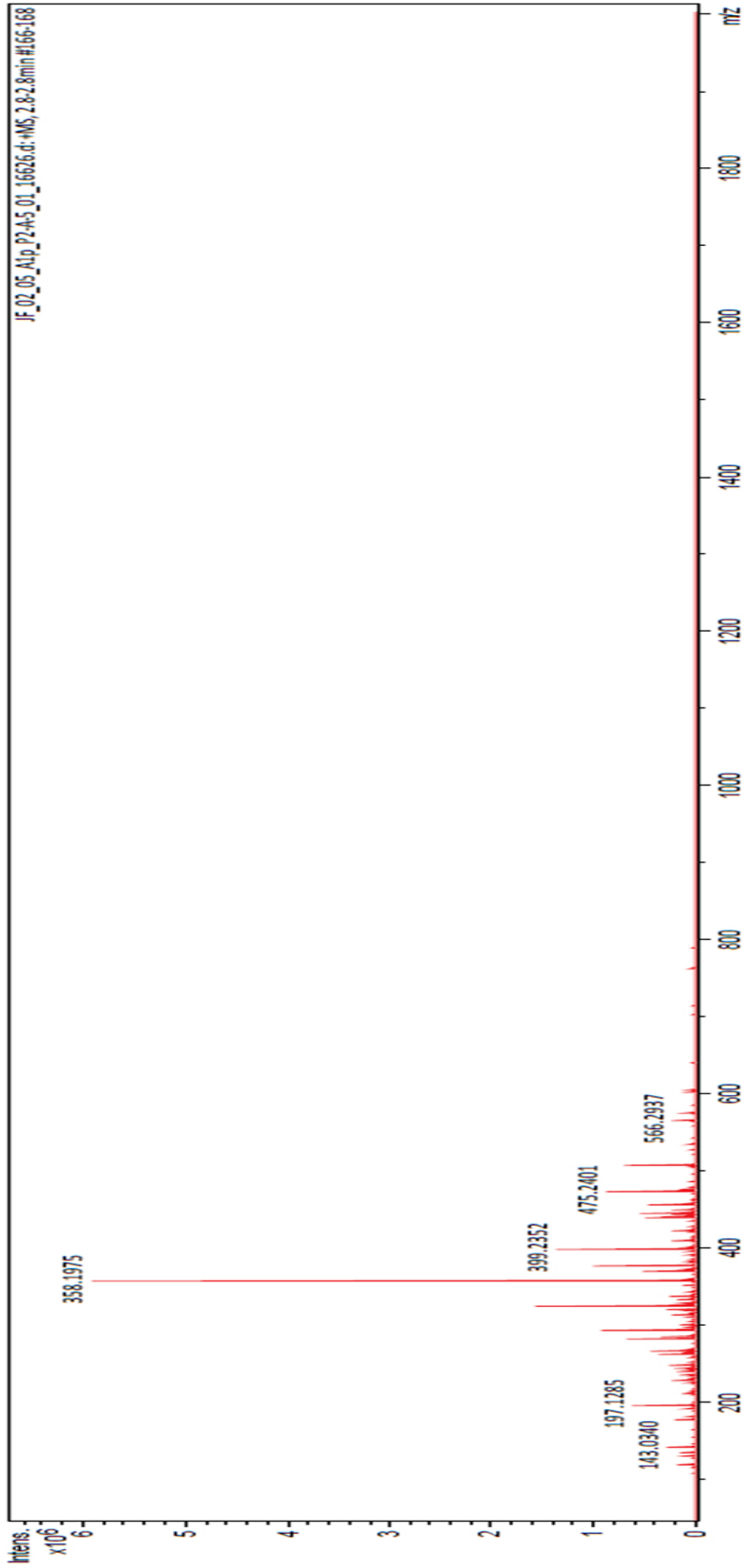


Abbildung 5.27: Massenspektrum 2 des Klons JF 13 C6

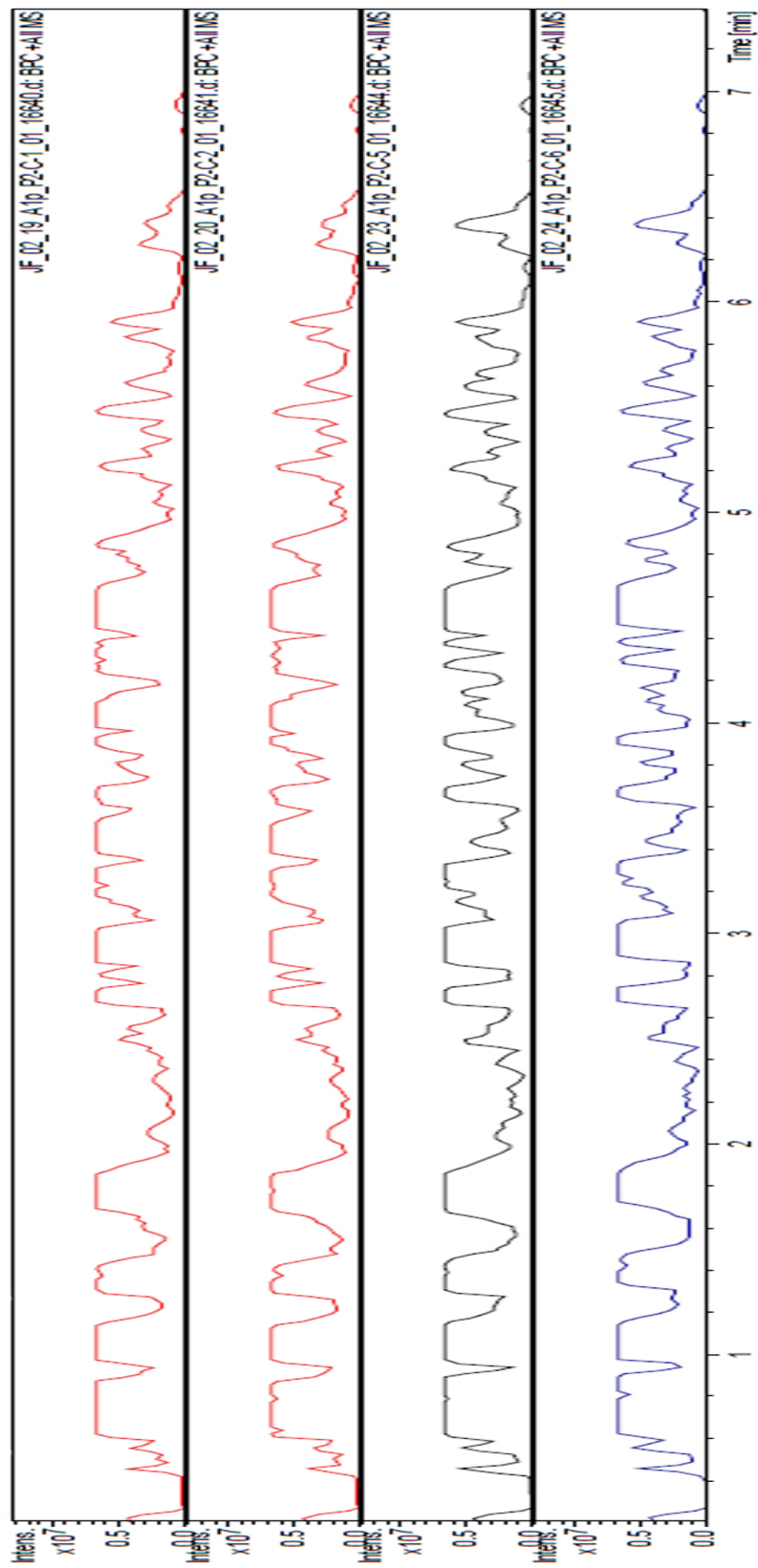


Abbildung 5.28: Chromatogramme als Liste des Klons JF 13 G11

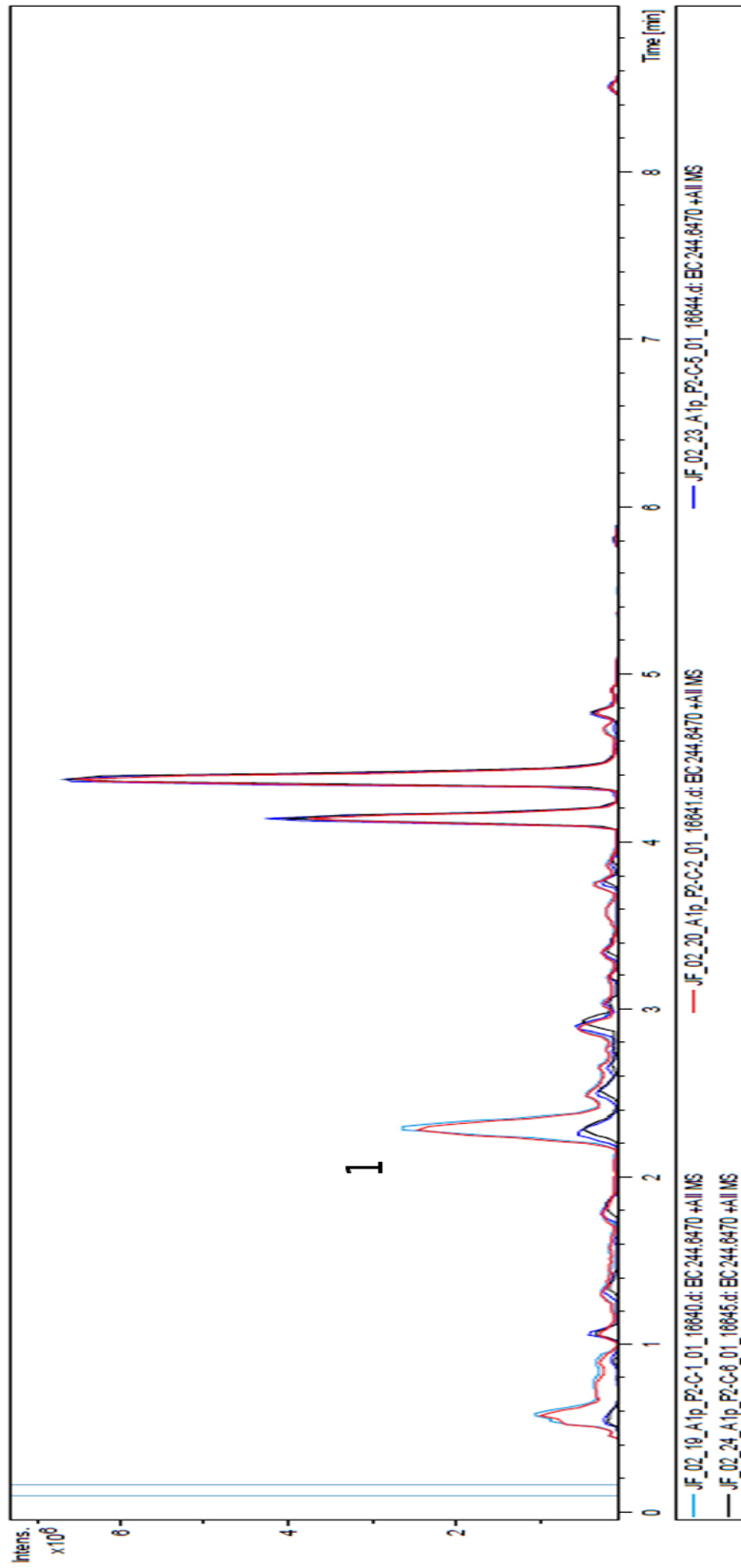


Abbildung 5.29: Signal 1 des Klons JF 13 G11 hergestellt

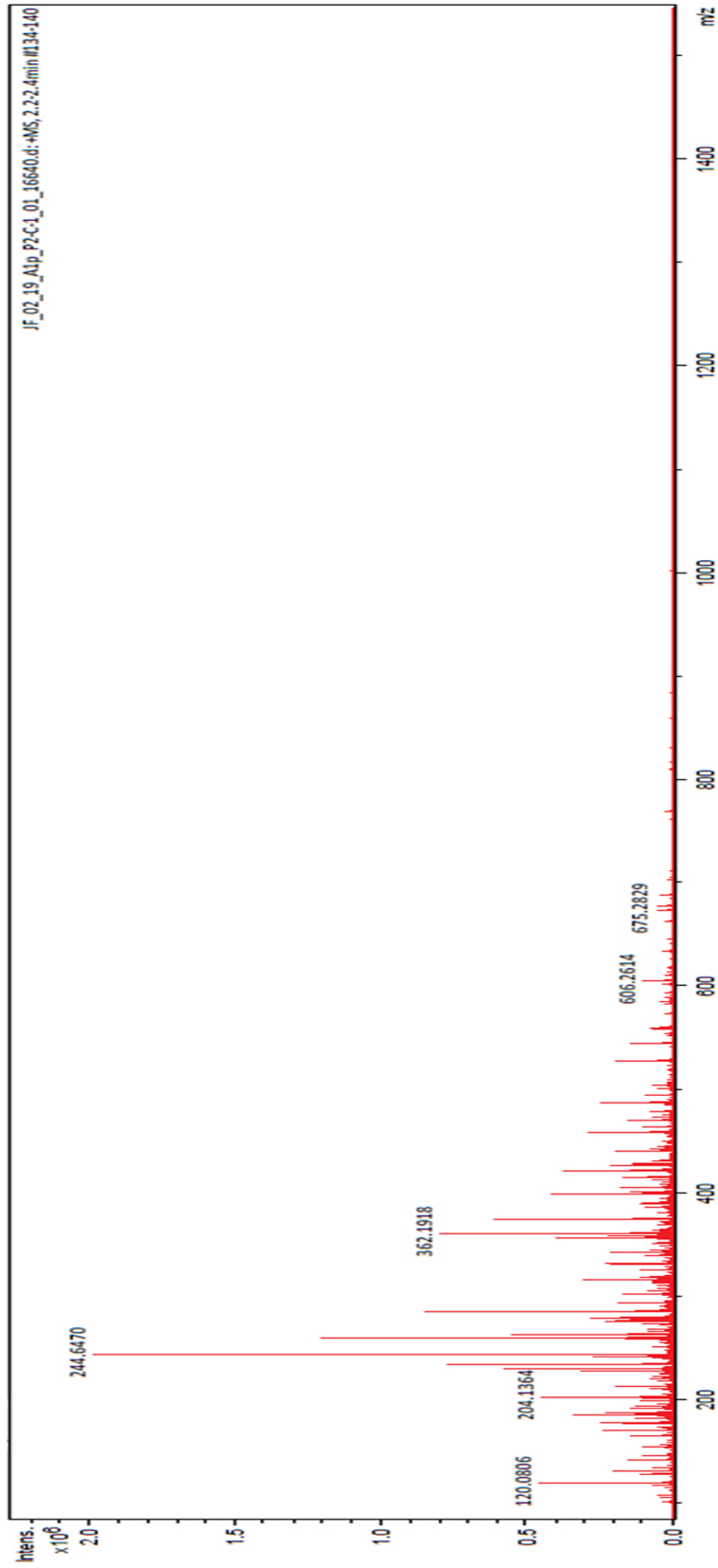


Abbildung 5.30: Massenspektrum 1 des Klons JF 13 G11

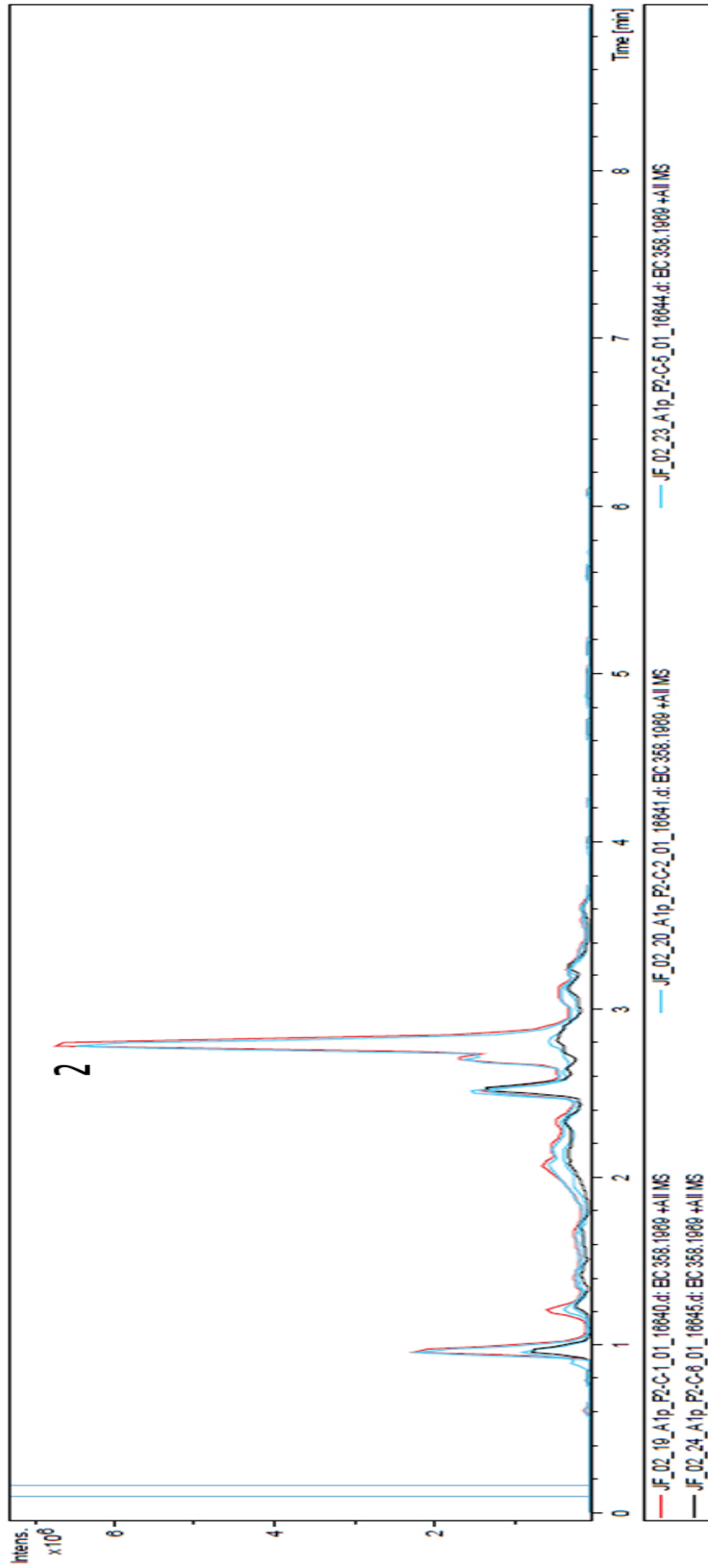


Abbildung 5.31: Signal 2 des Klons JF 13 G11 hergestellt

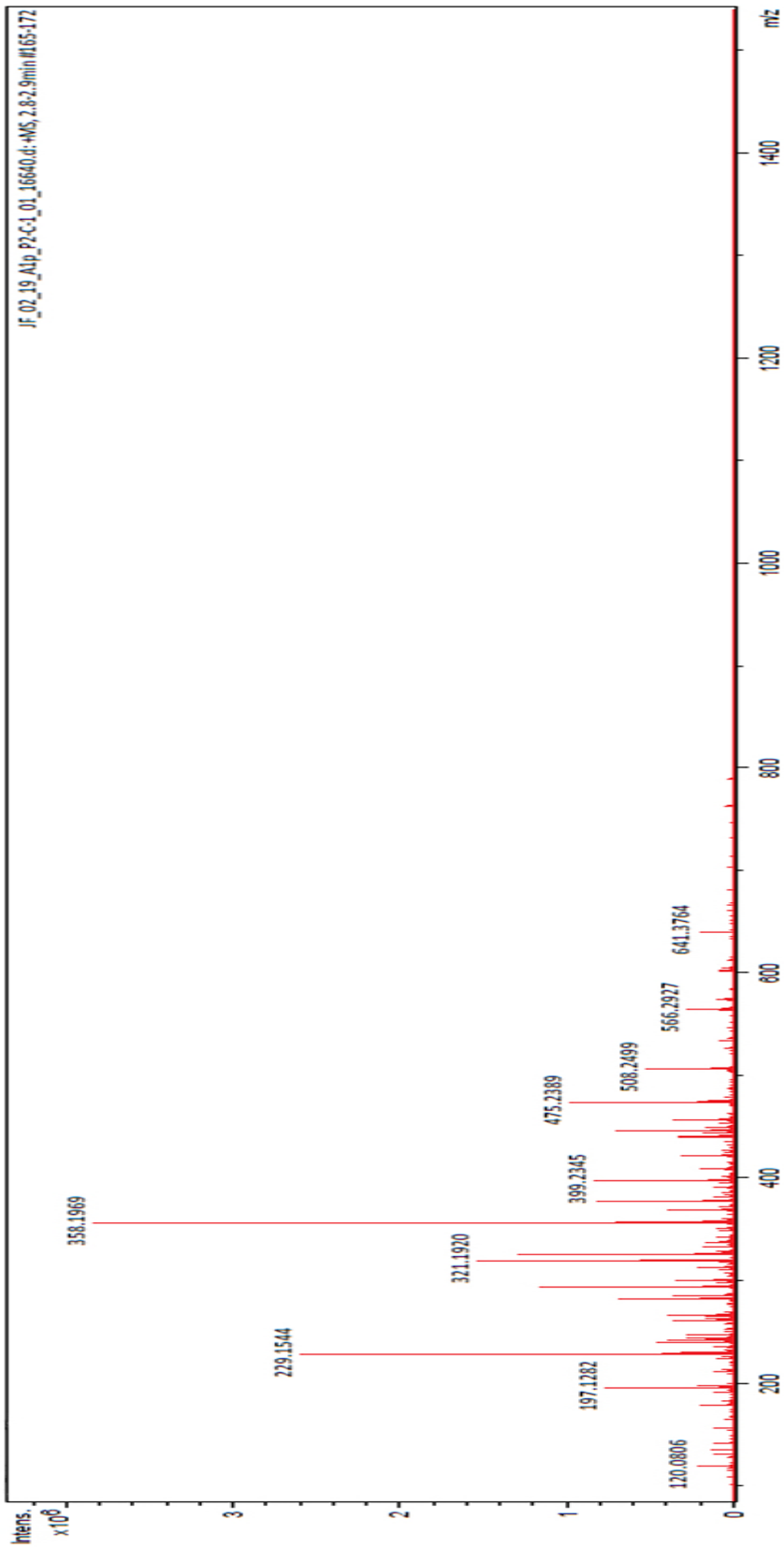


Abbildung 5.32: Massenspektrum 2 Klons JF 13 G11

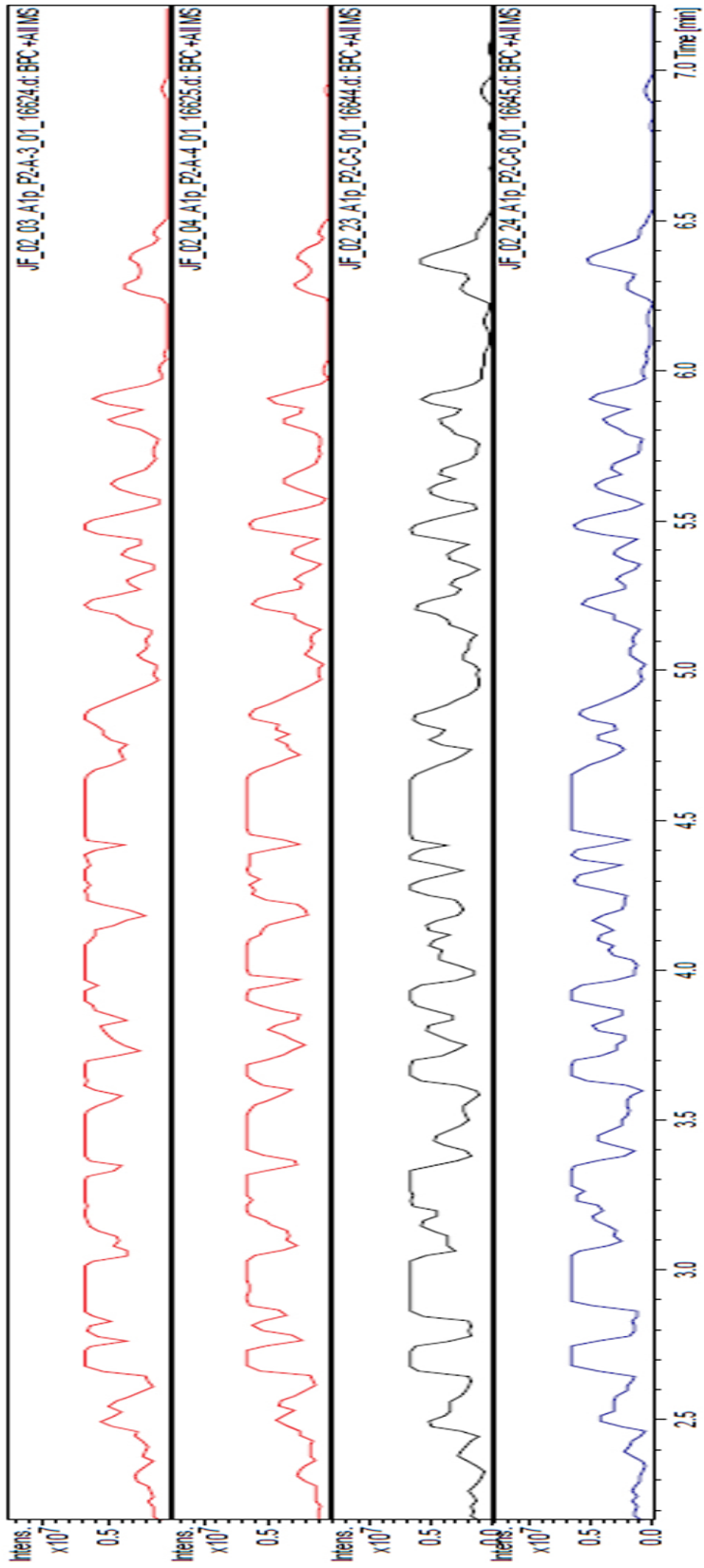


Abbildung 5.33: Chromatogramme als Liste Klons JF 13 G5

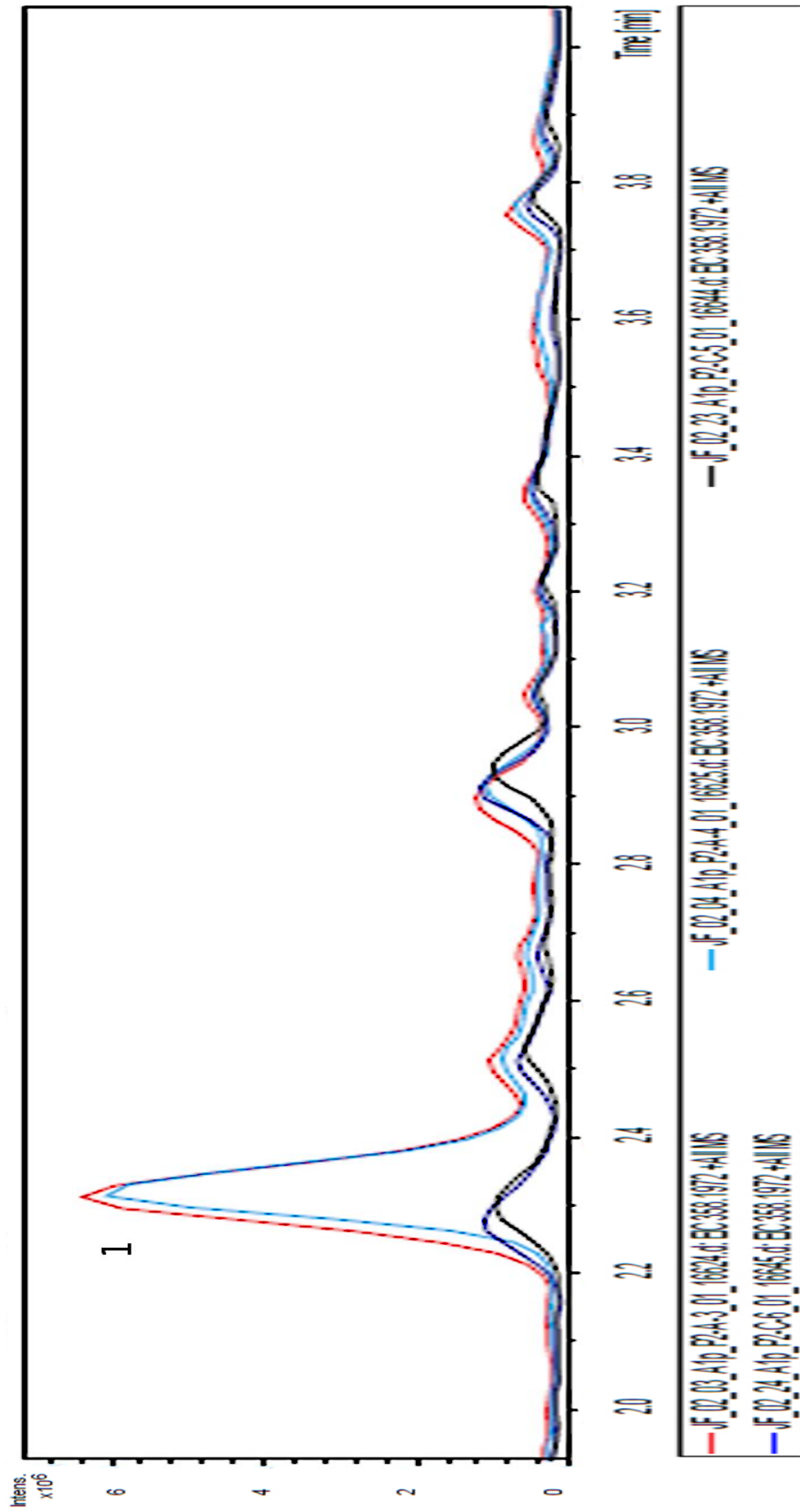


Abbildung 5.34: Signal 1 des Klons JF 13 G5 herausgestellt

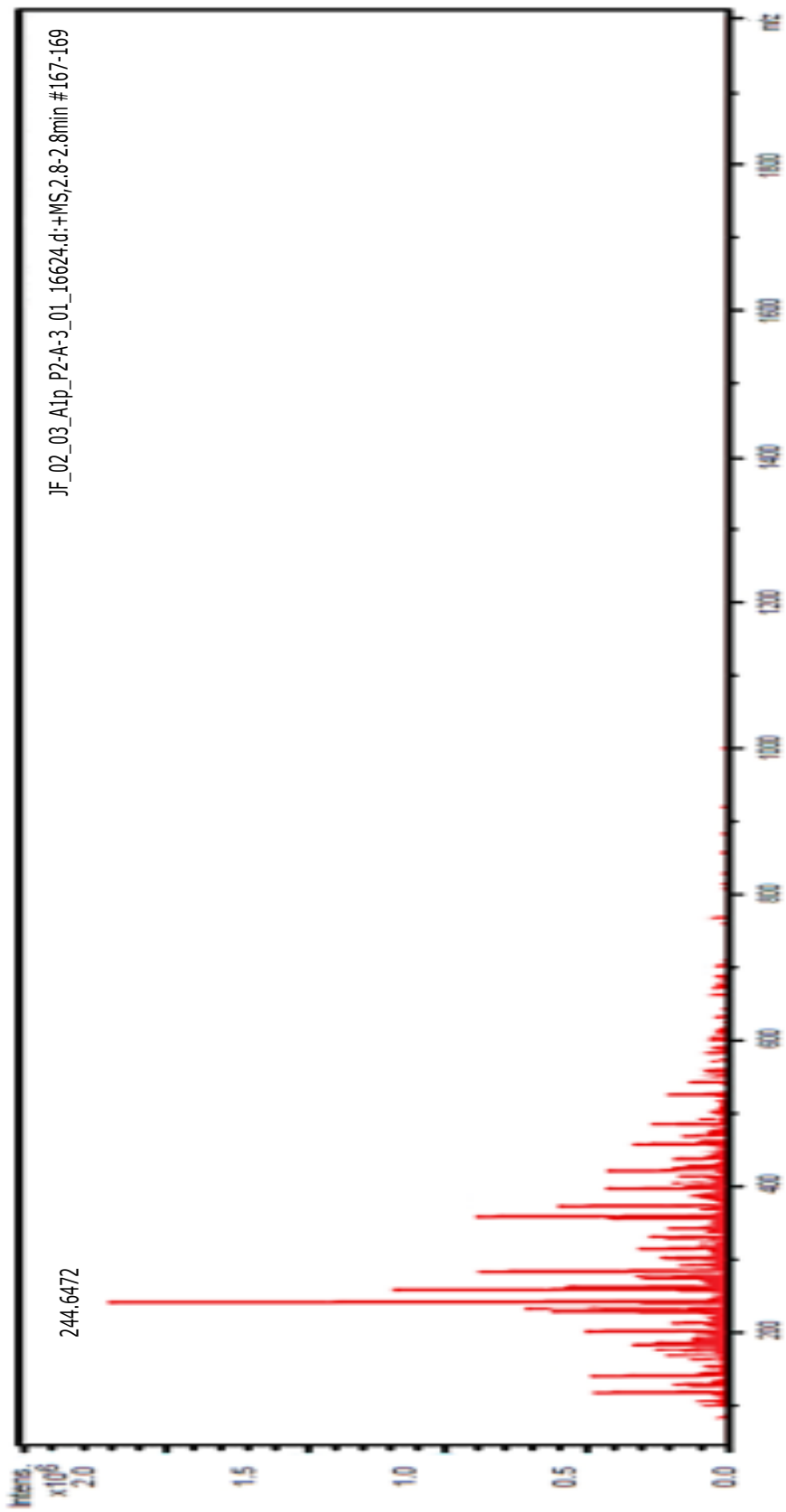


Abbildung 5.35: Massenspektrum 1 des Klons JF 13 G5

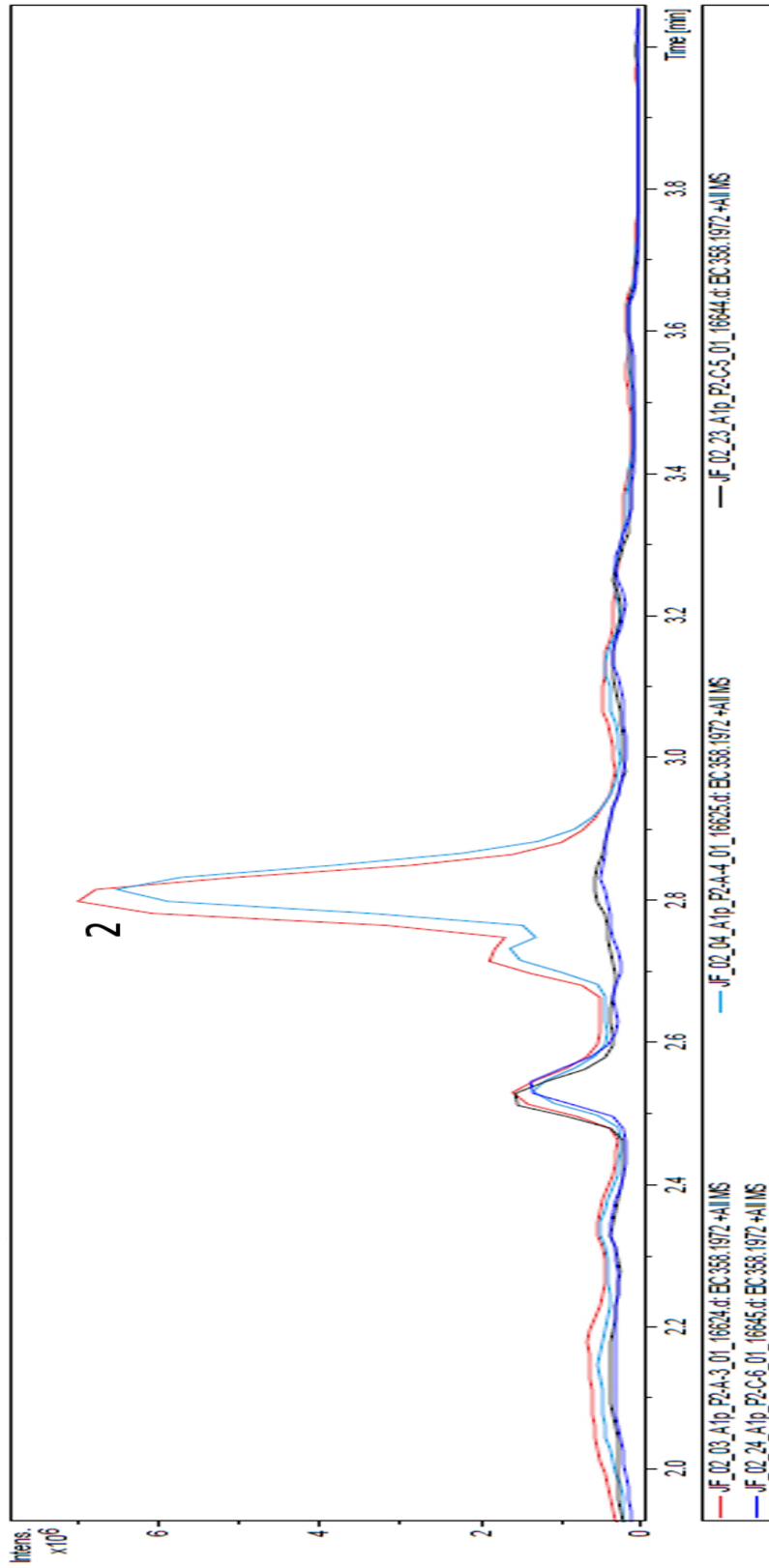


Abbildung 5.36: Signal 2 Klons JF 13 G5 hergestellt

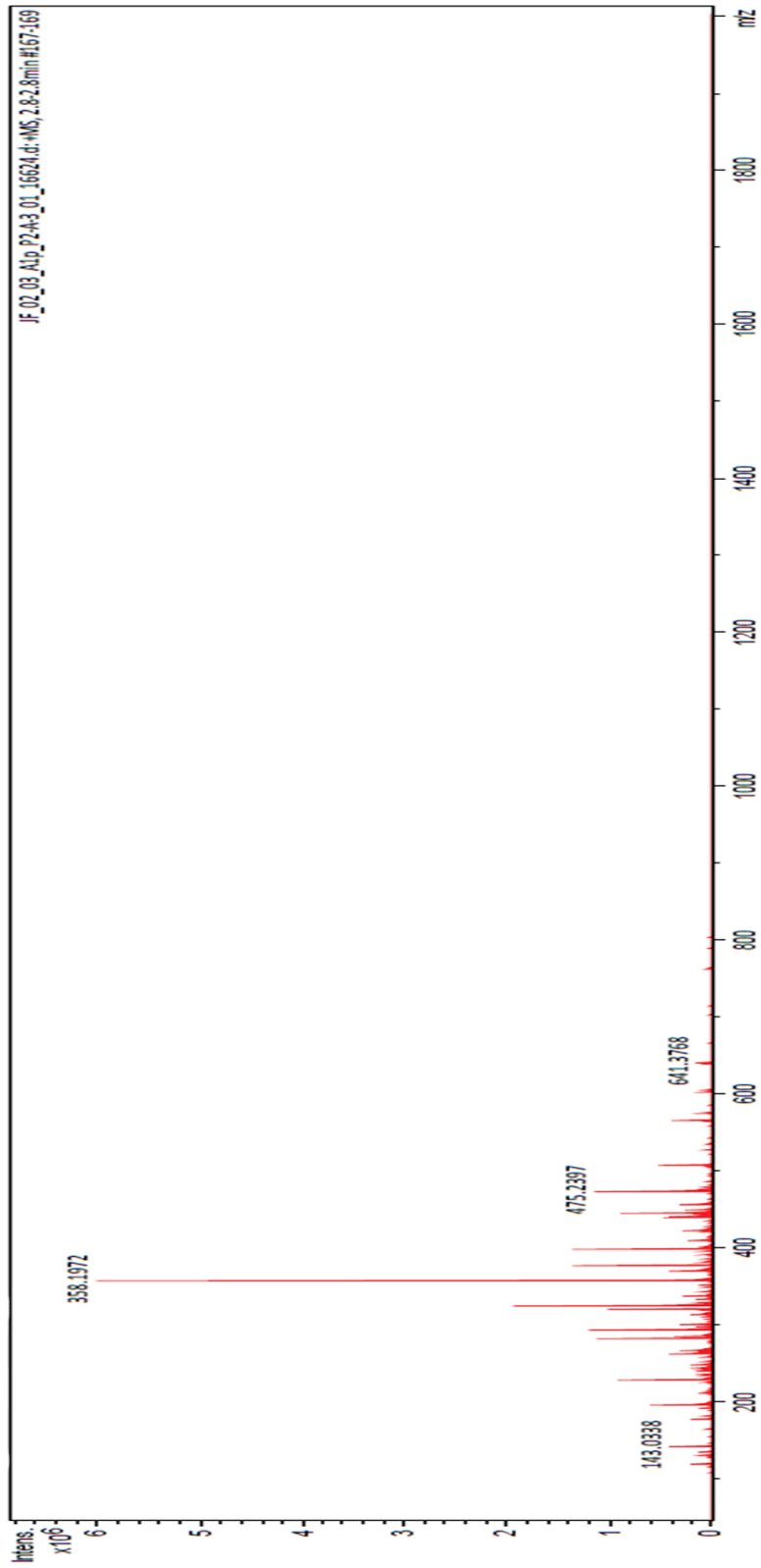


Abbildung 5.37: Massenspektrum 2 als Klons JF 13 G5

## Literaturverzeichnis

- [1] Antão EM, Wagner-Ahlf C. Antibiotikaresistenz. Bundesgesundheitsblatt - Gesundheitsforsch - Gesundheitsschutz. 2018;61(5):499–506. doi:10.1007/s00103-018-2726-y.
- [2] Aminov RI. A Brief History of the Antibiotic Era: Lessons Learned and Challenges for the Future. *Front Microbiol.* 2010;1:134. doi:10.3389/fmicb.2010.00134.
- [3] Pankey GA, Sabath LD. Clinical relevance of bacteriostatic versus bactericidal mechanisms of action in the treatment of Gram-positive bacterial infections. *Clin Infect Dis.* 2004;38(6):864–70. doi:10.1086/381972.
- [4] Sarkar FH, Li Y, Wang Z, Kong D. Cellular signaling perturbation by natural products. *Cell Signal.* 2009;21(11):1541–7. doi:10.1016/j.cellsig.2009.03.009.
- [5] de Lima Procópio RE, da Silva IR, Martins MK, de Azevedo JL, de Araújo JM, Procópio RE, et al. Antibiotics produced by *Streptomyces*. *Brazilian J Infect Dis.* 2012;16:466–471. doi:http://dx.doi.org/10.1016/j.bjid.2012.08.014.
- [6] Davies J. Where have All the Antibiotics Gone? *Can J Infect Dis Med Microbiol = J Can des Mal Infect la Microbiol medicale.* 2006;17(5):287–90. doi:10.1155/2006/707296.
- [7] Kumarasamy KK, Toleman MA, Walsh TR, Bagaria J, Butt F, Balakrishnan R, et al. Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. *Lancet Infect Dis.* 2010;10(9):597–602. doi:10.1016/S1473-3099(10)70143-2.
- [8] Garneau-Tsodikova S, Labby KJ. Mechanisms of Resistance to Aminoglycoside Antibiotics: Overview and Perspectives. *Medchemcomm.* 2016;7(1):11–27. doi:10.1039/C5MD00344J.
- [9] Zgurskaya HI, López CA, Gnanakaran S. Permeability Barrier of Gram-Negative Cell Envelopes and Approaches To Bypass It. *ACS Infect Dis.* 2015;1(11):512–522. doi:10.1021/acsinfecdis.5b00097.
- [10] Savage PB. Multidrug-resistant bacteria: overcoming antibiotic permeability barriers of gram-negative bacteria. *Ann Med.* 2001;33(3):167–71.
- [11] Fernandez L, Hancock REW. Adaptive and Mutational Resistance: Role of Porins and Efflux Pumps in Drug Resistance. *Clin Microbiol Rev.* 2012;25(4):661–681. doi:10.1128/CMR.00043-12.

- [12] Wieler LH, Broich K. Antibiotikaresistenzen: Große Herausforderungen erfordern komplexe Lösungen. *Bundesgesundheitsblatt - Gesundheitsforsch - Gesundheitsschutz*. 2018;61(5):495–498. doi:10.1007/s00103-018-2732-0.
- [13] Boucher HW, Talbot GH, Bradley JS, Edwards JE, Gilbert D, Rice LB, et al. Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. *Clin Infect Dis*. 2009;48:1–12. doi:10.1086/595011.
- [14] Cassini A, Högberg LD, Plachouras D, Quattrocchi A, Hoxha A, Simonsen GS, et al. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *Lancet Infect Dis*. 2018;0(0). doi:10.1016/S1473-3099(18)30605-4.
- [15] Kloß F, Gerbach S. Hürden und Aussichten neuer antimikrobieller Konzepte in Forschung und Entwicklung. *Bundesgesundheitsblatt - Gesundheitsforsch - Gesundheitsschutz*. 2018;61(5):595–605. doi:10.1007/s00103-018-2725-z.
- [16] Abu Sin M, Nahrgang S, Ziegelmann A, Clarici A, Matz S, Tenhagen BA, et al. Globale und nationale Strategien gegen Antibiotikaresistenzen. *Bundesgesundheitsblatt - Gesundheitsforsch - Gesundheitsschutz*. 2018;61(5):507–514. doi:10.1007/s00103-018-2722-2.
- [17] Simon Chorzelski, Benjamin Grosch, Heinrich Rentmeister, Simon Völler, Burkhard Landre, Julia Pfitzner, Claus Wechselmann, Reinhard Busse, Suzanne Edwards CH. *Breaking through the Wall Enhancing Research and Development of Antibiotics in Science and Industry*; 2015.
- [18] Högberg LD, Heddini A, Cars O. The global need for effective antibiotics: challenges and recent advances. *Trends Pharmacol Sci*. 2010;31(11):509–515. doi:10.1016/j.tips.2010.08.002.
- [19] Brady SF, Chao CJ, Clardy J. New Natural Product Families from an Environmental DNA (eDNA) Gene Cluster. *J Am Chem Soc*. 2002;124(34):9968–9969. doi:10.1021/ja0268985.
- [20] Mumm R, Dicke M. Variation in natural plant products and the attraction of bodyguards involved in indirect plant defenseThe present review is one in the special series of reviews on animal–plant interactions. *Can J Zool*. 2010;88(7):628–667. doi:10.1139/Z10-032.
- [21] Jwa NS, Agrawal GK, Tamogami S, Yonekura M, Han O, Iwahashi H, et al. Role of defense/stress-related marker genes, proteins and secondary metabolites in defining rice self-defense mechanisms. *Plant Physiol Biochem*. 2006;44(5-6):261–273. doi:10.1016/j.plaphy.2006.06.010.
- [22] Baltz RH. Antimicrobials from Actinomycetes: Back to the Future. *Microbe*. 2007;2 Number 3(3):125–131.

*Literaturverzeichnis*

- [23] Dingermann T, Hiller K, Schneider G. Schneider, Arzneidrogen. Spektrum, Akad. Verl; 2011. Available from: <http://www.springer.com/de/book/9783827427656>.
- [24] Cronan JE, Thomas J. Bacterial fatty acid synthesis and its relationships with polyketide synthetic pathways. *Methods Enzymol.* 2009;459:395–433. doi:10.1016/S0076-6879(09)04617-5.
- [25] Tohge T, Watanabe M, Hoefgen R, Fernie AR. Shikimate and Phenylalanine Biosynthesis in the Green Lineage. *Front Plant Sci.* 2013;4:62. doi:10.3389/fpls.2013.00062.
- [26] Rausch C, Weber T, Kohlbacher O, Wohlleben W, Huson DH. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res.* 2005;33(18):5799–808. doi:10.1093/nar/gki885.
- [27] Jenke-Kodama H, Sandmann A, Muller R, Dittmann E, Müller R, Dittmann E. Evolutionary implications of bacterial polyketide synthases. *Mol Biol Evol.* 2005;22(10):2027–2039. doi:10.1093/molbev/msi193.
- [28] Weissman KJ, Leadlay PF. Combinatorial biosynthesis of reduced polyketides. *Nat Rev Microbiol.* 2005;3(12):925–36. doi:10.1038/nrmicro1287.
- [29] Van Lanen SG, Shen B. Microbial genomics for the improvement of natural product discovery. *Curr Opin Microbiol.* 2006;9(3):252–60. doi:10.1016/j.mib.2006.04.002.
- [30] Lambalot RH, Gehring AM, Flugel RS, Zuber P, LaCelle M, Marahiel MA, et al. A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem Biol.* 1996;3(11):923–36.
- [31] Lim Y, Go M, Yew W. Exploiting the Biosynthetic Potential of Type III Polyketide Synthases. *Molecules.* 2016;21(6):806. doi:10.3390/molecules21060806.
- [32] Joshi AK, Zhang L, Rangan VS, Smith S. Cloning, expression, and characterization of a human 4'-phosphopantetheinyl transferase with broad substrate specificity. *J Biol Chem.* 2003;278(35):33142–9. doi:10.1074/jbc.M305459200.
- [33] Lohman JR, Ma M, Osipiuk J, Nocek B, Kim Y, Chang C, et al. Structural and evolutionary relationships of “AT-less” type I polyketide synthase ketosynthases. *Proc Natl Acad Sci.* 2015;112:12693–12698. doi:10.1073/pnas.1515460112.
- [34] Staunton J, Weissman KJ. Polyketide biosynthesis: a millennium review. *Nat Prod Rep.* 2001;18(4):380–416. doi:10.1039/a909079g.
- [35] Hertweck C. The biosynthetic logic of polyketide diversity. *Angew Chem Int Ed Engl.* 2009;48:4688–4716. doi:10.1002/anie.200806121.

- [36] Hertweck C, Luzhetskyy A, Rebets Y, Bechthold A. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat Prod Rep.* 2007;24(1):162–90. doi:10.1039/b507395m.
- [37] Hwang KS, Kim HU, Charusanti P, Palsson BØ, Lee SY. Systems biology and biotechnology of *Streptomyces* species for the production of secondary metabolites. *Biotechnol Adv.* 2014;32:255–268. doi:10.1016/j.biotechadv.2013.10.008.
- [38] Yu D, Xu F, Zeng J, Zhan J. Type III polyketide synthases in natural product biosynthesis. *IUBMB Life.* 2012;64(4):285–295. doi:10.1002/iub.1005.
- [39] Song L, Barona-gomez F, Corre C, Xiang L, Udway DW, Austin MB, et al. Type III Polyketide Synthase -Ketoacyl-ACP Starter Unit and Ethylmalonyl-CoA Extender Unit Selectivity Discovered by *Streptomyces coelicolor* Genome Mining. 2006; p. 14754–14755. doi:10.1074/jbc.M606500200.(4).
- [40] Nakano C, Ozawa H, Akanuma G, Funo N, Horinouchi S. Biosynthesis of aliphatic polyketides by type III polyketide synthase and methyltransferase in *Bacillus subtilis*. *J Bacteriol.* 2009;191(15):4916–23. doi:10.1128/JB.00407-09.
- [41] Li TL, Choroba OW, Hong H, Williams DH, Spencer JB. Biosynthesis of the vancomycin group of antibiotics: characterisation of a type III polyketide synthase in the pathway to (S)-3,5-dihydroxyphenylglycine. *Chem Commun (Camb).* 2001;(20):2156–7.
- [42] Wenzel SC, Bode HB, Kochems I, Müller R. A Type I/Type III Polyketide Synthase Hybrid Biosynthetic Pathway for the Structurally Unique Compound Kenodomycin. *ChemBioChem.* 2008;9(16):2711–2721. doi:10.1002/cbic.200800456.
- [43] Walsh CT, O’Brien RV, Khosla C. Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. *Angew Chem Int Ed Engl.* 2013;52(28):7098–124. doi:10.1002/anie.201208344.
- [44] Du L, Shen B. Identification and characterization of a type II peptidyl carrier protein from the bleomycin producer *Streptomyces verticillus* ATCC 15003. *Chem Biol.* 1999;6(8):507–517. doi:10.1016/S1074-5521(99)80083-0.
- [45] Donadio S, Monciardini P, Sosio M. Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. *Nat Prod Rep.* 2007;24:1073–1109. doi:10.1039/b514050c.
- [46] Walsh CT. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science.* 2004;303(5665):1805–10. doi:10.1126/science.1094318.
- [47] Bode HB, Müller R. The impact of bacterial genomics on natural product research. *Angew Chem Int Ed Engl.* 2005;44(42):6828–46. doi:10.1002/anie.200501080.

- [48] Masschelein J, Mattheus W, Gao LJ, Moons P, Van Houdt R, Uytterhoeven B, et al. A PKS/NRPS/FAS hybrid gene cluster from *Serratia plymuthica* RVH1 encoding the biosynthesis of three broad spectrum, zeamine-related antibiotics. *PLoS One*. 2013;8(1):e54143. doi:10.1371/journal.pone.0054143.
- [49] Fisch KM. Biosynthesis of natural products by microbial iterative hybrid PKS–NRPS. *RSC Adv*. 2013;3(40):18228. doi:10.1039/c3ra42661k.
- [50] Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep*. 2013;30(1):108–160. doi:10.1039/C2NP20085F.
- [51] Henriques ST, Melo MN, Castanho MARB. Cell-penetrating peptides and antimicrobial peptides: how different are they? *Biochem J*. 2006;399(1):1–7. doi:10.1042/BJ20061100.
- [52] Ben Lagha A, Haas B, Gottschalk M, Grenier D. Antimicrobial potential of bacteriocins in poultry and swine production. *Vet Res*. 2017;48(1):22. doi:10.1186/s13567-017-0425-6.
- [53] Kudryashova E, Seveau SM, Kudryashov DS. Targeting and inactivation of bacterial toxins by human defensins. *Biol Chem*. 2017;0(0). doi:10.1515/hsz-2017-0106.
- [54] Cheigh CI, Pyun YR. Nisin Biosynthesis and its Properties. *Biotechnol Lett*. 2005;27(21):1641–1648. doi:10.1007/s10529-005-2721-x.
- [55] Mohimani H, Kersten RD, Liu WT, Wang M, Purvine SO, Wu S, et al. Automated Genome Mining of Ribosomal Peptide Natural Products. *ACS Chem Biol*. 2014;9(7):1545–1551. doi:10.1021/cb500199h.
- [56] Ortega MA, van der Donk WA. New Insights into the Biosynthetic Logic of Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products. *Cell Chem Biol*. 2016;23(1):31–44. doi:10.1016/j.chembiol.2015.11.012.
- [57] Depoorter E, Bull MJ, Peeters C, Coenye T, Vandamme P, Mahenthiralingam E. Burkholderia: an update on taxonomy and biotechnological potential as antibiotic producers. *Appl Microbiol Biotechnol*. 2016;100(12):5215–5229. doi:10.1007/s00253-016-7520-x.
- [58] Mannanov RN, Sattarova RK. Antibiotics Produced by Bacillus Bacteria. *Chem Nat Compd*. 2001;37(2):117–123. doi:10.1023/A:1012314516354.
- [59] Zhang L, An R, Wang J, Sun N, Zhang S, Hu J, et al. Exploring novel bioactive compounds from marine microbes. *Curr Opin Microbiol*. 2005;8(3):276–281. doi:10.1016/j.mib.2005.04.008.

- [60] Graça AP, Calisto R, Lage OM. Planctomycetes as Novel Source of Bioactive Molecules. *Front Microbiol.* 2016;7:1241. doi:10.3389/fmicb.2016.01241.
- [61] Wagner-Döbler I, Beil W, Lang S, Meiners M, Laatsch H, Wagner-Dobler I, et al. Integrated approach to explore the potential of marine microorganisms for the production of bioactive metabolites. *Adv Biochem Eng Biotechnol.* 2002;74:207–238.
- [62] Kennedy J, Marchesi JR, Dobson ADW. Metagenomic approaches to exploit the biotechnological potential of the microbial consortia of marine sponges. *Appl Microbiol Biotechnol.* 2007;75(1):11–20. doi:10.1007/s00253-007-0875-2.
- [63] Hochmuth T, Piel J. Polyketide synthases of bacterial symbionts in sponges – Evolution-based applications in natural products research. *Phytochemistry.* 2009;70(15-16):1841–1849. doi:10.1016/J.PHYTOCHEM.2009.04.010.
- [64] Webster NS, Taylor MW. Marine sponges and their microbial symbionts: love and other relationships. *Environ Microbiol.* 2012;14(2):335–346. doi:10.1111/j.1462-2920.2011.02460.x.
- [65] Nichols D, Cahoon N, Trakhtenberg EM, Pham L, Mehta A, Belanger A, et al. Use of ichip for high-throughput in situ cultivation of “uncultivable” microbial species. *Appl Environ Microbiol.* 2010;76(8):2445–50. doi:10.1128/AEM.01754-09.
- [66] Esteves AIS, Amer N, Nguyen M, Thomas T. Sample Processing Impacts the Viability and Cultivability of the Sponge Microbiome. *Front Microbiol.* 2016;7:499. doi:10.3389/fmicb.2016.00499.
- [67] van Vliet S. Bacterial Dormancy: How to Decide When to Wake Up. *Curr Biol.* 2015;25(17):R753–R755. doi:10.1016/j.cub.2015.07.039.
- [68] Lewis K, Epstein S, D’Onofrio A, Ling LL. Uncultured microorganisms as a source of secondary metabolites. *J Antibiot (Tokyo).* 2010;63(8):468–476. doi:10.1038/ja.2010.87.
- [69] Lage OM, Bondoso J. Bringing Planctomycetes into pure culture. *Front Microbiol.* 2012;3:405. doi:10.3389/fmicb.2012.00405.
- [70] Zhang J, Chiodini R, Badr A, Zhang G. The impact of next-generation sequencing on genomics. *J Genet Genomics.* 2011;38(3):95–109. doi:10.1016/j.jgg.2011.02.003.
- [71] Bode HB, Muller R, Müller R. The impact of bacterial genomics on natural product research. *Angew Chem Int Ed Engl.* 2005;44(42):6828–6846. doi:10.1002/anie.200501080.
- [72] Van Lanen SG, Shen B. Microbial genomics for the improvement of natural product discovery. *Curr Opin Microbiol.* 2006;9(3):252–260. doi:10.1016/j.mib.2006.04.002.

- [73] Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K, et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell*. 2014;158(2):412–421. doi:10.1016/j.cell.2014.06.034.
- [74] Jeske O, Jogler M, Petersen J, Sikorski J, Jogler C. From genome mining to phenotypic microarrays: Planctomycetes as source for novel bioactive molecules. *Antonie Van Leeuwenhoek*. 2013;104:551–567. doi:10.1007/s10482-013-0007-1.
- [75] Kielak AM, Barreto CC, Kowalchuk GA, van Veen JA, Kuramae EE. The Ecology of Acidobacteria: Moving beyond Genes and Genomes. *Front Microbiol*. 2016;7:744. doi:10.3389/fmicb.2016.00744.
- [76] O'Connor-Sánchez A, Rivera-Domínguez AJ, De los Santos-Briones C, López-Aguilar LK, Peña-Ramírez YJ, Prieto-Davo A. Acidobacteria appear to dominate the microbiome of two sympatric Caribbean Sponges and one Zoanthid. *Biol Res*. 2014;47(1):67. doi:10.1186/0717-6287-47-67.
- [77] Barns SM, Cain EC, Sommerville L, Kuske CR. Acidobacteria phylum sequences in uranium-contaminated subsurface sediments greatly expand the known diversity within the phylum. *Appl Environ Microbiol*. 2007;73(9):3113–6. doi:10.1128/AEM.02012-06.
- [78] Kishimoto N, Kosako Y, Tano T. *Acidobacterium capsulatum* gen. nov., sp. nov.: An acidophilic chemoorganotrophic bacterium containing menaquinone from acidic mineral environment. *Curr Microbiol*. 1991;22:1–7. doi:10.1007/BF02106205.
- [79] Ward NL, Challacombe JF, Janssen PH, Henrissat B, Coutinho PM, Wu M, et al. Three Genomes from the Phylum Acidobacteria Provide Insight into the Lifestyles of These Microorganisms in Soils. *Appl Environ Microbiol*. 2009;75(7):2046–2056. doi:10.1128/AEM.02294-08.
- [80] Parsley LC, Linneman J, Goode AM, Becklund K, George I, Goodman RM, et al. Polyketide synthase pathways identified from a metagenomic library are derived from soil Acidobacteria. *FEMS Microbiol Ecol*. 2011;78:176–187. doi:10.1111/j.1574-6941.2011.01122.x.
- [81] Crits-Christoph A, Diamond S, Butterfield CN, Thomas BC, Banfield JF. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature*. 2018;558(7710):440–444. doi:10.1038/s41586-018-0207-y.
- [82] Next generation sequencing technology: Advances and applications. *Biochim Biophys Acta - Mol Basis Dis*. 2014;1842(10):1932–1941. doi:10.1016/J.BBADIS.2014.06.015.
- [83] Marçais G, Yorke JA, Zimin A. QuorUM: An Error Corrector for Illumina Reads. *PLoS One*. 2015;10(6):e0130821. doi:10.1371/journal.pone.0130821.

## Literaturverzeichnis

- [84] Nagarajan N, Cook C, Di Bonaventura M, Ge H, Richards A, Bishop-Lilly KA, et al. Finishing genomes with limited resources: lessons from an ensemble of microbial genomes. *BMC Genomics*. 2010;11(1):242. doi:10.1186/1471-2164-11-242.
- [85] Munroe DJ, Harris TJR. Third-generation sequencing fireworks at Marco Island. *Nat Biotechnol*. 2010;28(5):426–428. doi:10.1038/nbt0510-426.
- [86] Deschamps S, Mudge J, Cameron C, Ramaraj T, Anand A, Fengler K, et al. Characterization, correction and de novo assembly of an Oxford Nanopore genomic dataset from *Agrobacterium tumefaciens*. *Sci Rep*. 2016;6:28625. doi:10.1038/srep28625.
- [87] Cherf GM, Lieberman KR, Rashid H, Lam CE, Karplus K, Akeson M. Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. *Nat Biotechnol*. 2012;30(4):344–8. doi:10.1038/nbt.2147.
- [88] Tyler AD, Mataseje L, Urfano CJ, Schmidt L, Antonation KS, Mulvey MR, et al. Evaluation of Oxford Nanopore’s MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. *Sci Rep*. 2018;8(1):10931. doi:10.1038/s41598-018-29334-5.
- [89] Kerstin Göpfrich KJ. Decoding DNA with a pocket-sized sequencer; 2018. Available from: <https://www.scienceinschool.org/content/decoding-dna-pocket-sized-sequencer>.
- [90] Wick RR, Judd LM, Gorrie CL, Holt KE, Keane J, Harris S. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Comput Biol*. 2017;13(6):e1005595. doi:10.1371/journal.pcbi.1005595.
- [91] Baichoo S, Ouzounis CA. Computational complexity of algorithms for sequence comparison, short-read assembly and genome alignment. *Biosystems*. 2017;156-157:72–85. doi:10.1016/j.biosystems.2017.03.003.
- [92] Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. 2010;11:119. doi:10.1186/1471-2105-11-119.
- [93] Salzberg SL, Delcher AL, Kasif S, White O. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res*. 1998;26(2):544–8.
- [94] Merkl R. *Bioinformatik: Grundlagen, Algorithmen, Anwendungen*. Wiley-Blackwell; 2015.
- [95] Genome annotation techniques: new approaches and challenges. *Drug Discov Today*. 2002;7(11):S70–S76. doi:10.1016/S1359-6446(02)02289-4.
- [96] Sasson O, Kaplan N, Linial M. Functional annotation prediction: all for one and one for all. *Protein Sci*. 2006;15(6):1557–62. doi:10.1110/ps.062185706.

### Literaturverzeichnis

- [97] Mulia S, Mishra D, Jena T. Profile HMM based Multiple Sequence Alignment for DNA Sequences. *Procedia Eng.* 2012;38:1783–1787. doi:10.1016/J.PROENG.2012.06.218.
- [98] Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. *Nucleic Acids Res.* 2014;42(Database issue):D222–30. doi:10.1093/nar/gkt1223.
- [99] Yoon BJ. Hidden Markov Models and their Applications in Biological Sequence Analysis. *Curr Genomics.* 2009;10(6):402–15. doi:10.2174/138920209789177575.
- [100] Eddy SR. Accelerated Profile HMM Searches. *PLoS Comput Biol.* 2011;7:e1002195. doi:10.1371/journal.pcbi.1002195.
- [101] Soding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics.* 2005;21(7):951–960. doi:10.1093/bioinformatics/bti125.
- [102] Blin K, Kim HU, Medema MH, Weber T. Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters. *Brief Bioinform.* 2017;33(7):933–41. doi:10.1093/bib/bbx146.
- [103] Alanjary M, Kronmiller B, Adamek M, Blin K, Weber T, Huson D, et al. The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery. *Nucleic Acids Res.* 2017;45(W1):W42–W48. doi:10.1093/nar/gkx360.
- [104] Boddy CN. Bioinformatics tools for genome mining of polyketide and non-ribosomal peptides. *J Ind Microbiol Biotechnol.* 2014;41(2):443–450. doi:10.1007/s10295-013-1368-1.
- [105] Li MHT, Ung PMU, Zajkowski J, Garneau-Tsodikova S, Sherman DH. Automated genome mining for natural products. *BMC Bioinformatics.* 2009;10:185. doi:10.1186/1471-2105-10-185.
- [106] Ziemert N, Podell S, Penn K, Badger JH, Allen E, Jensen PR. The Natural Product Domain Seeker NaPDoS: A Phylogeny Based Bioinformatic Tool to Classify Secondary Metabolite Gene Diversity. *PLoS One.* 2012;7(3):e34064. doi:10.1371/journal.pone.0034064.
- [107] Kim J, Yi GS. PKMiner: a database for exploring type II polyketide synthases. *BMC Microbiol.* 2012;12(1):169. doi:10.1186/1471-2180-12-169.
- [108] Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* 2011;39:W339–46. doi:10.1093/nar/gkr466.

- [109] The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production. *Synth Syst Biotechnol.* 2016;1(2):69–79. doi:10.1016/J.SYNBIO.2015.12.002.
- [110] Medema MH, Fischbach MA. Computational approaches to natural product discovery. *Nat Chem Biol.* 2015;11:639–648. doi:10.1038/nchembio.1884.
- [111] Medema MH, Takano E, Breitling R. Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol Biol Evol.* 2013;30(5):1218–23. doi:10.1093/molbev/mst025.
- [112] Overmann J. Principles of Enrichment, Isolation, Cultivation, and Preservation of Prokaryotes. In: Rosenberg E, DeLong E, Lory S, Stackebrandt E, Thompson F, editors. *The Prokaryotes*. Springer Berlin Heidelberg;. p. 149–207.
- [113] Watve MG, Tickoo R, Jog MM, Bhole BD. How many antibiotics are produced by the genus *Streptomyces*? *Arch Microbiol.* 2001;176(5):386–390. doi:10.1007/s002030100345.
- [114] Clardy J, Fischbach Ma, Walsh CT. New antibiotics from bacterial natural products. *Nat Biotechnol.* 2006;24(12):1541–50. doi:10.1038/nbt1266.
- [115] Zhang L, An R, Wang J, Sun N, Zhang S, Hu J, et al. Exploring novel bioactive compounds from marine microbes. *Curr Opin Microbiol.* 2005;8(3):276–81. doi:10.1016/j.mib.2005.04.008.
- [116] MacNeil Ia, Tiong CL, Minor C, August PR, Grossman TH, Loiacono Ka, et al. Expression and isolation of antimicrobial small molecules from soil DNA libraries. *J Mol Microbiol Biotechnol.* 2001;3(2):301–8.
- [117] Brady SF, Clardy J. Long-Chain N -Acyl Amino Acid Antibiotics Isolated from Heterologously Expressed Environmental DNA. 2000;(Table 1):12903–12904.
- [118] Riesenfeld CS, Goodman RM, Handelsman J. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environ Microbiol.* 2004;6(9):981–9. doi:10.1111/j.1462-2920.2004.00664.x.
- [119] Kennedy J, Marchesi JR, Dobson ADW. Metagenomic approaches to exploit the biotechnological potential of the microbial consortia of marine sponges. *Appl Microbiol Biotechnol.* 2007;75(1):11–20. doi:10.1007/s00253-007-0875-2.
- [120] Trindade M, van Zyl LJ, Navarro-Fernández J, Abd Elrazak A. Targeted metagenomics as a tool to tap into marine natural product diversity for the discovery and production of drug candidates. *Front Microbiol.* 2015;6:890. doi:10.3389/fmicb.2015.00890.
- [121] Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell.* 2014;157(6):1262–78. doi:10.1016/j.cell.2014.05.010.

- [122] Zhong C, Peng D, Ye W, Chai L, Qi J, Yu Z, et al. Determination of Plasmid Copy Number Reveals the Total Plasmid DNA Amount Is Greater than the Chromosomal DNA Amount in *Bacillus thuringiensis* YBT-1520. *PLoS One*. 2011;6(1):e16025. doi:10.1371/journal.pone.0016025.
- [123] Collins J, Hohn B. Cosmids: a type of plasmid gene-cloning vector that is packagable in vitro in bacteriophage lambda heads. *Proc Natl Acad Sci U S A*. 1978;75(9):4242–6.
- [124] Sabehi G, Béjà O. Preparation of BAC libraries from marine microbial populations. *Methods Enzym*. 2013;531:111–122. doi:10.1016/b978-0-12-407863-5.00006-x.
- [125] Kakirde KS, Wild J, Godiska R, Mead DA, Wiggins AG, Goodman RM, et al. Gram negative shuttle BAC vector for heterologous expression of metagenomic libraries. *Gene*. 2011;475:57–62. doi:10.1016/j.gene.2010.11.004.
- [126] Mancino V, Slepak T, Tachiiri Y. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in. 1992;89(September):8794–8797.
- [127] Ariyadasa R, Stein N. Advances in BAC-based physical mapping and map integration strategies in plants. *J Biomed Biotechnol*. 2012;2012:184854. doi:10.1155/2012/184854.
- [128] Marzluf GA, Reddy CA, Beveridge TJ, Schmidt TM, Snyder LR, Breznak JA, editors. *Methods for General and Molecular Microbiology*, Third Edition. American Society of Microbiology; 2007. Available from: <http://www.asmscience.org/content/book/10.1128/9781555817497>.
- [129] Mutka SC, Carney JR, Liu Y, Kennedy J. Heterologous production of epothilone C and D in *Escherichia coli*. *Biochemistry*. 2006;45(4):1321–1330. doi:10.1021/bi052075r.
- [130] Nishizaki T, Tsuge K, Itaya M, Doi N, Yanagawa H. Metabolic engineering of carotenoid biosynthesis in *Escherichia coli* by ordered gene assembly in *Bacillus subtilis*. *Appl Env Microbiol*. 2007;73:1355–1361. doi:10.1128/aem.02268-06.
- [131] Paget MSB, Helmann JDLBP. The sigma70 family of sigma factors. *Genome Biol*. 2003;4.
- [132] Wenzel SC, Müller R. Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. *Curr Opin Biotechnol*. 2005;16(6):594–606. doi:10.1016/j.copbio.2005.10.001.
- [133] Hatfield GW, Roth DA. Optimizing scaleup yield for protein production: Computationally Optimized DNA Assembly (CODA) and Translation Engineering. *Biotechnol Annu Rev*. 2007;13:27–42. doi:10.1016/s1387-2656(07)13002-7.

### Literaturverzeichnis

- [134] Muto A, Osawa S. The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc Natl Acad Sci U S A*. 1987;84(1):166–9.
- [135] Alibolandi M, Mirzahoseini H, Abad M, Movahed M. African journal of biotechnology.. vol. 9. Academic Journals; 2002. Available from: <https://www.ajol.info/index.php/ajb/article/view/78482>.
- [136] Mirzadeh K, Martínez V, Toddo S, Guntur S, Herrgård MJ, Elofsson A, et al. Enhanced Protein Production in *Escherichia coli* by Optimization of Cloning Scars at the Vector–Coding Sequence Junction. *ACS Synth Biol*. 2015;4(9):959–965. doi:10.1021/acssynbio.5b00033.
- [137] de Marco A. Molecular and Chemical Chaperones for Improving the Yields of Soluble Recombinant Proteins; 2011. p. 31–51. Available from: <http://link.springer.com/10.1007/978-1-61737-967-3>.
- [138] Gottesman S. PROTEASES AND THEIR TARGETS IN *ESCHERICHIA COLI*. *Annu Rev Genet*. 1996;30(1):465–506. doi:10.1146/annurev.genet.30.1.465.
- [139] Beld J, Sonnenschein EC, Vickery CR, Noel JP, Burkart MD. The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. *Nat Prod Rep*. 2014;31(1):61–108. doi:10.1039/c3np70054b.
- [140] Pfeifer Ba, Admiraal SJ, Gramajo H, Cane DE, Khosla C. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science*. 2001;291(5509):1790–2. doi:10.1126/science.1058092.
- [141] Durfee T, Nelson R, Baldwin S, Plunkett G, Burland V, Mau B, et al. The Complete Genome Sequence of *Escherichia coli* DH10B: Insights into the Biology of a Laboratory Workhorse. *J Bacteriol*. 2008;190:2597–2606. doi:10.1128/JB.01695-07.
- [142] Mashima J, Kodama Y, Kosuge T, Fujisawa T, Katayama T, Nagasaki H, et al. DNA data bank of Japan (DDBJ) progress report. *Nucleic Acids Res*. 2016;44(D1):D51–D57. doi:10.1093/nar/gkv1105.
- [143] Stoesser G, Baker W, van den Broek A, Camon E, Garcia-Pastor M, Kanz C, et al. The EMBL Nucleotide Sequence Database. *Nucleic Acids Res*. 2002;30(1):21–6.
- [144] Cochrane G, Karsch-Mizrachi I, Takagi T, International Nucleotide Sequence Database Collaboration IN. The International Nucleotide Sequence Database Collaboration. *Nucleic Acids Res*. 2016;44(D1):D48–50. doi:10.1093/nar/gkv1323.
- [145] Gibney G, Baxevanis AD. Searching NCBI Databases Using Entrez. In: *Curr. Protoc. Bioinforma.* vol. Chapter 1. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2011. p. Unit 1.3. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21633942><http://doi.wiley.com/10.1002/0471250953.bi0103s34>.

- [146] Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* 2015;43. doi:10.1093/nar/gkv437 LB - Weber2015.
- [147] Fuerst JA. The PVC superphylum: exceptions to the bacterial definition? *Antonie Van Leeuwenhoek.* 2013;104(4):451–466. doi:10.1007/s10482-013-9986-1.
- [148] McKinney W. Data Structures for Statistical Computing in Python. In: van der Walt S, Millman J, editors. *Proc. 9th Python Sci. Conf.*; 2010. p. 51–56.
- [149] Oliphant TE. Python for Scientific Computing. *Comput Sci Eng.* 2007;9(3):10–20. doi:10.1109/MCSE.2007.58.
- [150] Chao A, Chiu CH, Hsieh TC, Davis T, Nipperess DA, Faith DP. Rarefaction and extrapolation of phylogenetic diversity. *Methods Ecol Evol.* 2015;6(4):380–388. doi:10.1111/2041-210X.12247.
- [151] Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 2014;42:D206–D214. doi:10.1093/nar/gkt1226.
- [152] Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014;30(14):2068–2069. doi:10.1093/bioinformatics/btu153.
- [153] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32:1792–1797. doi:10.1093/nar/gkh340.
- [154] Brown NP, Leroy C, Sander C. MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics.* 1998;14(4):380–381. doi:10.1093/bioinformatics/14.4.380.
- [155] Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res.* 2004;14(6):1188–90. doi:10.1101/gr.849004.
- [156] Loman NJ, Quinlan AR. Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics.* 2014;30(23):3399–401. doi:10.1093/bioinformatics/btu555.
- [157] Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19(5):455–77. doi:10.1089/cmb.2012.0021.
- [158] Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS One.* 2014;9(11):e112963. doi:10.1371/journal.pone.0112963.
- [159] Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 2017;27(5):737–746. doi:10.1101/gr.214270.116.

- [160] Cha S, Bird DM. Optimizing k-mer size using a variant grid search to enhance de novo genome assembly. *Bioinformatics*. 2016;12(2):36–40. doi:10.6026/97320630012036.
- [161] Kremer FS, Eslabão MR, Dellagostin OA, Pinto LdS. Genix: a new online automated pipeline for bacterial genome annotation. *FEMS Microbiol Lett*. 2016;363(23):fnw263. doi:10.1093/femsle/fnw263.
- [162] Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, et al. UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View. In: *Methods Mol. Biol.*. vol. 1374; 2016. p. 23–54.
- [163] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27–30.
- [164] Nightingale A, Antunes R, Alpi E, Bursteinas B, Gonzales L, Liu W, et al. The Proteins API: accessing key integrated protein and genome information. *Nucleic Acids Res*. 2017;45(W1):W539–W544. doi:10.1093/nar/gkx237.
- [165] Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS One*. 2011;6(7):e21800. doi:10.1371/journal.pone.0021800.
- [166] Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*. 2012;40(W1):W445–W451. doi:10.1093/nar/gks479.
- [167] Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res*. 2013;41(12):e121–e121. doi:10.1093/nar/gkt263.
- [168] Ayuso-Sacido a, Genilloud O. New PCR primers for the screening of NRPS and PKS-I systems in actinomycetes: detection and distribution of these biosynthetic gene sequences in major taxonomic groups. *Microb Ecol*. 2005;49(1):10–24. doi:10.1007/s00248-004-0249-6.
- [169] Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41(Database issue):D590–6. doi:10.1093/nar/gks1219.
- [170] Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22(13):1658–1659. doi:10.1093/bioinformatics/btl158.
- [171] Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*. 2011;12:385. doi:10.1186/1471-2105-12-385.

- [172] Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J, et al. GenDB—an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res.* 2003;31(8):2187–2195.
- [173] Uliczka F, Pisano F, Kochut A, Opitz W, Herbst K, Stolz T, et al. Monitoring of Gene Expression in Bacteria during Infections Using an Adaptable Set of Bioluminescent, Fluorescent and Colorigenic Fusion Vectors. *PLoS One.* 2011;6(6):e20425. doi:10.1371/journal.pone.0020425.
- [174] Sujada N, Sungthong R, Lumyong S. Termite nests as an abundant source of cultivable actinobacteria for biotechnological purposes. *Microbes Env.* 2014;29(2):211–219. doi:DN/JST.JSTAGE/jsme2/ME13183 [pii].
- [175] Delacre M, Lakens D, Leys C. Why Psychologists Should by Default Use Welch’s t-test Instead of Student’s t-test (in press for the *International Review of Social Psychology*). 2017;doi:10.17605/OSF.IO/SBP6K.
- [176] van Wezel GP, McKenzie NL, Nodwell JR. Chapter 5 Applying the Genetics of Secondary Metabolism in Model Actinomycetes to the Discovery of New Antibiotics. *Methods Enzymol.* 2009;458:117–141. doi:10.1016/s0076-6879(09)04805-8.
- [177] Ni X, Jiang P, Xing L, Ou C, Yu H, Qi J, et al. Genome-wide mining of potential virulence-associated genes in *Riemerella anatipestifer* using random transposon mutagenesis. *Vet Microbiol.* 2016;189:52–58. doi:10.1016/j.vetmic.2016.04.014.
- [178] Dorn BR, Leung KPL, Progulsk-Fox A. Invasion of Human Oral Epithelial Cells by *Prevotella intermedia*. *Infect Immun.* 1998;66(12):6054–6057.
- [179] Elwell C, Mirrashidi K, Engel J. Chlamydia cell biology and pathogenesis. *Nat Rev Microbiol.* 2016;14(6):385–400. doi:10.1038/nrmicro.2016.30.
- [180] Löffler FE, Yan J, Ritalahti KM, Adrian L, Edwards EA, Konstantinidis KT, et al. *Dehalococcoides mccartyi* gen. nov., sp. nov., obligately organohalide-respiring anaerobic bacteria relevant to halogen cycling and bioremediation, belong to a novel bacterial class, *Dehalococcoidia* classis nov., order *Dehalococcoidales* ord. nov. and famil. *Int J Syst Evol Microbiol.* 2013;63:625–635. doi:doi:10.1099/ij.s.0.034926-0.
- [181] Donadio S, Staver MJ, McAlpine JB, Swanson SJ, Katz L. Modular organization of genes required for complex polyketide biosynthesis. *Science (80- )*. 1991;252(5006):675–679.
- [182] Chen Y, Kelly EE, Masluk RP, Nelson CL, Cantu DC, Reilly PJ. Structural classification and properties of ketoacyl synthases. *Protein Sci.* 2011;20(10):1659–67. doi:10.1002/pro.712.
- [183] Bretschneider T, Zocher G, Unger M, Scherlach K, Stehle T, Hertweck C. A ketosynthase homolog uses malonyl units to form esters in cervimycin biosynthesis. *Nat Chem Biol.* 2012;8(2):154–161. doi:10.1038/nchembio.746.

- [184] Campbell EL, Cohen MF, Meeks JC. A polyketide-synthase-like gene is involved in the synthesis of heterocyst glycolipids in *Nostoc punctiforme* strain ATCC 29133. *Arch Microbiol.* 1997;167(4):251–8.
- [185] Oves-Costales D, Kadi N, Challis GL. The long-overlooked enzymology of a non-ribosomal peptide synthetase-independent pathway for virulence-conferring siderophore biosynthesis. *Chem Commun.* 2009;(43):6530. doi:10.1039/b913092f.
- [186] Blin K, Kazempour D, Wohlleben W, Weber T. Improved Lanthipeptide Detection and Prediction for antiSMASH. *PLoS One.* 2014;9(2):e89420. doi:10.1371/journal.pone.0089420.
- [187] Milton DL, Hardman A, Camara M, Chhabra SR, Bycroft BW, Stewart GS, et al. Quorum sensing in *Vibrio anguillarum*: characterization of the *vanI/vanR* locus and identification of the autoinducer *N*-(3-oxodecanoyl)-*L*-homoserine lactone. *J Bacteriol.* 1997;179(9):3004–12. doi:10.1128/JB.179.9.3004-3012.1997.
- [188] Bibb MJ. Regulation of secondary metabolism in streptomycetes. *Curr Opin Microbiol.* 2005;8(2):208–215. doi:10.1016/j.mib.2005.02.016.
- [189] Pierson LS, Pierson EA, Pierson EA. Metabolism and function of phenazines in bacteria: impacts on the behavior of bacteria in the environment and biotechnological processes. *Appl Microbiol Biotechnol.* 2010;86(6):1659–70. doi:10.1007/s00253-010-2509-3.
- [190] Sadeghi A, Soltani BM, Nekouei MK, Jouzani GS, Mirzaei HH, Sadeghizadeh M. Diversity of the ectoines biosynthesis genes in the salt tolerant *Streptomyces* and evidence for inductive effect of ectoines on their accumulation. *Microbiol Res.* 2014;169(9-10):699–708. doi:10.1016/J.MICRES.2014.02.005.
- [191] Siegel AF, Siegel, F A. Rarefaction Curves. In: *Encycl. Stat. Sci.* Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2006. Available from: <http://doi.wiley.com/10.1002/0471667196.ess2195.pub2>.
- [192] Poulter CD. Biosynthesis of non-head-to-tail terpenes. Formation of 1'-1 and 1'-3 linkages. *Acc Chem Res.* 1990;23(3):70–77. doi:10.1021/ar00171a003.
- [193] Liu X, Gai Z, Tao F, Tang H, Xu P. Carotenoids Play a Positive Role in the Degradation of Heterocycles by *Sphingobium yanoikuyae*. *PLoS One.* 2012;7(6):e39522. doi:10.1371/journal.pone.0039522.
- [194] Frost LS, Leplae R, Summers AO, Toussaint A. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol.* 2005;3(9):722–732. doi:10.1038/nrmicro1235.
- [195] TOYOMASU T. Recent Advances Regarding Diterpene Cyclase Genes in Higher Plants and Fungi. *Biosci Biotechnol Biochem.* 2008;72(5):1168–1175. doi:10.1271/bbb.80044.

- [196] Janke R, Görner C, Hirte M, Brück T, Loll B. The first structure of a bacterial diterpene cyclase: CotB2. *Acta Crystallogr Sect D Biol Crystallogr*. 2014;70(6):1528–1537. doi:10.1107/S1399004714005513.
- [197] Dairi T, Hamano Y, Kuzuyama T, Itoh N, Furihata K, Seto H. Eubacterial Diterpene Cyclase Genes Essential for Production of the Isoprenoid Antibiotic Terpenecin. *J Bacteriol*. 2001;183(20):6085–6094. doi:10.1128/JB.183.20.6085-6094.2001.
- [198] Dairi T. Studies on Biosynthetic Genes and Enzymes of Isoprenoids Produced by Actinomycetes. *J Antibiot (Tokyo)*. 2005;58(4):227–243. doi:10.1038/ja.2005.27.
- [199] Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, et al. antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res*. 2013;41(W1):W204–W212. doi:10.1093/nar/gkt449.
- [200] Arakawa K. Manipulation of metabolic pathways controlled by signaling molecules, inducers of antibiotic production, for genome mining in *Streptomyces* spp. *Antonie Van Leeuwenhoek*. 2018;111(5):743–751. doi:10.1007/s10482-018-1052-6.
- [201] Okada BK, Seyedsayamdost MR, Shen A. Antibiotic dialogues: induction of silent biosynthetic gene clusters by exogenous small molecules. *FEMS Microbiol Rev*. 2017;41(1):19–33. doi:10.1093/femsre/fuw035.
- [202] Feng S, Xu C, Yang K, Wang H, Fan H, Liao M. Either fadD1 or fadD2, Which Encode acyl-CoA Synthetase, Is Essential for the Survival of *Haemophilus parasuis* SC096. *Front Cell Infect Microbiol*. 2017;7:72. doi:10.3389/fcimb.2017.00072.
- [203] Tran PN, Yen MR, Chiang CY, Lin HC, Chen PY. Detecting and prioritizing biosynthetic gene clusters for bioactive compounds in bacteria and fungi. *Appl Microbiol Biotechnol*. 2019;103(8):3277–3287. doi:10.1007/s00253-019-09708-z.
- [204] Wright GD. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nat Rev Microbiol*. 2007;5(3):175–186. doi:10.1038/nrmicro1614.
- [205] D’Costa VM, King CE, Kalan L, Morar M, Sung WWL, Schwarz C, et al. Antibiotic resistance is ancient. *Nature*. 2011;477(7365):457–461. doi:10.1038/nature10388.
- [206] Munita JM, Arias CA. Mechanisms of Antibiotic Resistance. *Microbiol Spectr*. 2016;4(2). doi:10.1128/microbiolspec.VMBF-0016-2015.
- [207] D’Costa VM, McGrann KM, Hughes DW, Wright GD. Sampling the antibiotic resistome. *Science*. 2006;311(5759):374–7. doi:10.1126/science.1120800.
- [208] Hughes VM, Datta N. Conjugative plasmids in bacteria of the ‘pre-antibiotic’ era. *Nature*. 1983;302(5910):725–726. doi:10.1038/302725a0.

- [209] Mazel D, Rowe-Magnus D. The Evolution of Antibiotic Resistance. In: *Evol. Microb. Pathog.* American Society of Microbiology; 2006. p. 221–241. Available from: <http://www.asmscience.org/content/book/10.1128/9781555815622.ch12>.
- [210] Kirchman DL. The ecology of Cytophaga-Flavobacteria in aquatic environments. *FEMS Microbiol Ecol.* 2002;39(2):91–100. doi:10.1111/j.1574-6941.2002.tb00910.x.
- [211] Wexler HM. Bacteroides: the good, the bad, and the nitty-gritty. *Clin Microbiol Rev.* 2007;20(4):593–621. doi:10.1128/CMR.00008-07.
- [212] Xiao T, He X, Cheng G, Kuang H, Ma X, Yusup K, et al. *Sphingobacterium hotanense* sp. nov., isolated from soil of a *Populus euphratica* forest, and emended descriptions of *Sphingobacterium daejeonense* and *Sphingobacterium shayense*. *Int J Syst Evol Microbiol.* 2013;63(Pt 3):815–820. doi:10.1099/ij.s.0.030155-0.
- [213] Lambiase A. The Family Sphingobacteriaceae. In: *The Prokaryotes.* Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. p. 907–914. Available from: [http://link.springer.com/10.1007/978-3-642-38954-2\\_136](http://link.springer.com/10.1007/978-3-642-38954-2_136).
- [214] Masschelein J, Jenner M, Challis GL. Antibiotics from Gram-negative bacteria: a comprehensive overview and selected biosynthetic highlights. *Nat Prod Rep.* 2017;34(7):712–783. doi:10.1039/C7NP00010C.
- [215] Townsley L, Shank EA. Natural-Product Antibiotics: Cues for Modulating Bacterial Biofilm Formation. *Trends Microbiol.* 2017;25(12):1016–1026. doi:10.1016/j.tim.2017.06.003.
- [216] Mohr KI, Volz C, Jansen R, Wray V, Hoffmann J, Bernecker S, et al. Pinensins: The First Antifungal Lantibiotics. *Angew Chemie Int Ed.* 2015;54(38):11254–11258. doi:10.1002/anie.201500927.
- [217] Kaempfer P, Young CC, Sridhar KR, Arun AB, Lai WA, Shen FT, et al. Transfer of [*Flexibacter*] *sancti*, [*Flexibacter*] *filiformis*, [*Flexibacter*] *japonensis* and [*Cytophaga*] *arvensicola* to the genus *Chitinophaga* and description of *Chitinophaga skermanii* sp. nov. *Int J Syst Evol Microbiol.* 2006;56(9):2223–2228. doi:10.1099/ij.s.0.64359-0.
- [218] Jansen R, Gerth K, Steinmetz H, Reinecke S, Kessler W, Kirschning A, et al. Elansolide28085A3, a Unique p-Quinone Methide Antibiotic from *Chitinophaga sancti*. *Chem - A Eur J.* 2011;17(28):7739–7744. doi:10.1002/chem.201100457.
- [219] Beckmann A, Hüttel S, Schmitt V, Müller R, Stadler M. Optimization of the biotechnological production of a novel class of anti-MRSA antibiotics from *Chitinophaga sancti*. *Microb Cell Fact.* 2017;16(1):143. doi:10.1186/s12934-017-0756-z.

### Literaturverzeichnis

- [220] Dehn R, Katsuyama Y, Weber A, Gerth K, Jansen R, Steinmetz H, et al. Molekulare Grundlage für die Biosynthese von Elansolid: Beweise für eine einzigartige, durch ein Chinonmethid initiierte intramolekulare Diels-Alder-Cycloaddition/Makrolactonisierung. *Angew Chemie*. 2011;123(17):3968–3973. doi:10.1002/ange.201006880.
- [221] Helfrich EJJ, Piel JJ, Monroe EA, Korobeynikov A, Sherman DH, Gerwick L, et al. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat Prod Rep*. 2016;33(2):231–316. doi:10.1039/C5NP00125K.
- [222] Teta R, Gurgui M, Helfrich EJJ, Künne S, Schneider A, Van Echten-Deckert G, et al. Genome Mining Reveals trans-AT Polyketide Synthase Directed Antibiotic Biosynthesis in the Bacterial Phylum Bacteroidetes. *ChemBioChem*. 2010;11(18):2506–2512. doi:10.1002/cbic.201000542.
- [223] Anand S, Prasad MVR, Yadav G, Kumar N, Shehara J, Ansari MZ, et al. SBSPKS: structure based sequence analysis of polyketide synthases. *Nucleic Acids Res*. 2010;38(Web Server issue):W487–96. doi:10.1093/nar/gkq340.
- [224] Jenner M, Frank S, Kampa A, Kohlhaas C, Pöplau P, Briggs GS, et al. Substrate Specificity in Ketosynthase Domains from trans-AT Polyketide Synthases. *Angew Chemie Int Ed*. 2013;52:1143–1147. doi:10.1002/anie.201207690.
- [225] Musiol E, Härtner T, Kulik A, Moldenhauer J, Piel J, Wohlleben W, et al. Supramolecular Templating in Kirromycin Biosynthesis: The Acyltransferase KirCII Loads Ethylmalonyl-CoA Extender onto a Specific ACP of the trans-AT PKS. *Chem Biol*. 2011;18(4):438–444. doi:10.1016/j.chembiol.2011.02.007.
- [226] Dunn BJ, Khosla C. Engineering the acyltransferase substrate specificity of assembly line polyketide synthases. *J R Soc Interface*. 2013;10(85):20130297. doi:10.1098/rsif.2013.0297.
- [227] Piel J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat Prod Rep*. 2010;27:996–1047. doi:10.1039/b816430b.
- [228] Silakowski B, Schairer HU, Ehret H, Kunze B, Weinig S, Nordsiek G, et al. New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1. *J Biol Chem*. 1999;274(52):37391–37399.
- [229] Bode HB, Meiser P, Klefisch T, Cortina NSdJ, Krug D, Göhring A, et al. Mutasynthesis-Derived Myxalamids and Origin of the Isobutyryl-CoA Starter Unit of Myxalamid B. *ChemBioChem*. 2007;8(17):2139–2144. doi:10.1002/cbic.200700401.
- [230] Jenner M. Using Mass Spectrometry for Biochemical Studies on Enzymatic Domains from Polyketide Synthases. Springer Theses. Cham: Springer International Publishing; 2016.

- [231] Gay DC, Gay G, Axelrod AJ, Jenner M, Kohlhaas C, Kampa A, et al. A close look at a ketosynthase from a trans-acyltransferase modular polyketide synthase. *Structure*. 2014;22(3):444–51. doi:10.1016/j.str.2013.12.016.
- [232] Sović I, Križanović K, Skala K, Šikić M. Evaluation of hybrid and non-hybrid methods for de novo assembly of nanopore reads. *Bioinformatics*. 2016;32(17):2582–2589. doi:10.1093/bioinformatics/btw237.
- [233] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25(1):25–9. doi:10.1038/75556.
- [234] Thomas PD. The Gene Ontology and the Meaning of Biological Function. In: *Methods Mol. Biol.* vol. 1446; 2017. p. 15–24. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27812932>[http://link.springer.com/10.1007/978-1-4939-3743-1\\_2](http://link.springer.com/10.1007/978-1-4939-3743-1_2).
- [235] Kulichevskaya IS, Kostina LA, Valaskova V, Rijpstra WIC, Sinninghe Damste JS, de Boer W, et al. *Acidicapsa borealis* gen. nov., sp. nov. and *Acidicapsa ligni* sp. nov., subdivision 1 Acidobacteria from Sphagnum peat and decaying wood. *Int J Syst Evol Microbiol*. 2012;62(Pt 7):1512–1520. doi:10.1099/ij.s.0.034819-0.
- [236] Beguin P, Aubert JP. The biological degradation of cellulose. *FEMS Microbiol Rev*. 1994;13(1):25–58. doi:10.1111/j.1574-6976.1994.tb00033.x.
- [237] Desvaux M, Guedon E, Petitdemange H. Cellulose catabolism by *Clostridium cellulolyticum* growing in batch culture on defined medium. *Appl Environ Microbiol*. 2000;66(6):2461–70.
- [238] Cruz-Morales P, Kopp JF, Martínez-Guerrero C, Yáñez-Guerra LA, Selem-Mojica N, Ramos-Aboites H, et al. Phylogenomic Analysis of Natural Products Biosynthetic Gene Clusters Allows Discovery of Arseno-Organic Metabolites in Model Streptomyces. *Genome Biol Evol*. 2016;8(6):1906–1916. doi:10.1093/gbe/evw125.
- [239] Agostini HJ, Carroll JD, Minton KW. Identification and characterization of *uvrA*, a DNA repair gene of *Deinococcus radiodurans*. *J Bacteriol*. 1996;178(23):6759–65.
- [240] Ude S, Lassak J, Starosta AL, Kraxenberger T, Wilson DN, Jung K. Translation Elongation Factor EF-P Alleviates Ribosome Stalling at Polyproline Stretches. *Science* (80- ). 2013;339(6115).
- [241] Lassak J, Wilson DN, Jung K. Stall no more at polyproline stretches with the translation elongation factors EF-P and IF-5A. *Mol Microbiol*. 2016;99(2):219–235. doi:10.1111/mmi.13233.
- [242] Bullwinkle TJ, Zou SB, Rajkovic A, Hersch SJ, Elgamal S, Robinson N, et al. (R)- $\beta$ -lysine-modified elongation factor P functions in translation elongation. *J Biol Chem*. 2013;288(6):4416–23. doi:10.1074/jbc.M112.438879.

*Literaturverzeichnis*

- [243] Delmar JA, Su CC, Yu EW. Bacterial multidrug efflux transporters. *Annu Rev Biophys.* 2014;43:93–117. doi:10.1146/annurev-biophys-051013-022855.
- [244] Li XZ, Elkins CA, Zgurskaya HI. Efflux-mediated antimicrobial resistance in bacteria : mechanisms, regulation and clinical implications;
- [245] Pulsawat N, Kitani S, Nihira T. Characterization of biosynthetic gene cluster for the production of virginiamycin M, a streptogramin type A antibiotic, in *Streptomyces virginiae*. *Gene.* 2007;393(1-2):31–42. doi:10.1016/j.gene.2006.12.035.
- [246] Peterson BF, Scharf ME. Lower Termite Associations with Microbes: Synergy, Protection, and Interplay. *Front Microbiol.* 2016;7:422. doi:10.3389/fmicb.2016.00422.
- [247] Watson JAL, Metcalf EC, Sewell JJ. Preliminary studies on the control of neotenic formation in *Mastotermes Darwiniensis* Froggatt (Isoptera). *Insectes Soc.* 1975;22(4):415–426. doi:10.1007/BF02224116.
- [248] Huber JA, Morrison HG, Huse SM, Neal PR, Sogin ML, Mark Welch DB. Effect of PCR amplicon size on assessments of clone library microbial diversity and community structure. *Environ Microbiol.* 2009;11(5):1292–1302. doi:10.1111/j.1462-2920.2008.01857.x.
- [249] Mizrahi-Man O, Davenport ER, Gilad Y. Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs. *PLoS One.* 2013;8(1):e53608. doi:10.1371/journal.pone.0053608.
- [250] Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 2013;41(1):e1–e1. doi:10.1093/nar/gks808.
- [251] Kielak AM, van Veen JA, Kowalchuk GA. Comparative analysis of acidobacterial genomic fragments from terrestrial and aquatic metagenomic libraries, with emphasis on acidobacteria subdivision 6. *Appl Environ Microbiol.* 2010;76(20):6769–77. doi:10.1128/AEM.00343-10.
- [252] Quaiser A, Ochsenreiter T, Lanz C, Schuster SC, Treusch AH, Eck J, et al. Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics. *Mol Microbiol.* 2003;50(2):563–575. doi:10.1046/j.1365-2958.2003.03707.x.
- [253] Challacombe JF, Eichorst SA, Hauser L, Land M, Xie G, Kuske CR. Biological Consequences of Ancient Gene Acquisition and Duplication in the Large Genome of *Candidatus Solibacter usitatus* Ellin6076. *PLoS One.* 2011;6(9):e24882. doi:10.1371/journal.pone.0024882.

- [254] Blackall LL, Burrell PC, Gwilliam H, Bradford D, Bond PL, Hugenholtz P. The use of 16S rDNA clone libraries to describe the microbial diversity of activated sludge communities. *Water Sci Technol.* 1998;37(4-5):451–454. doi:10.1016/S0273-1223(98)00144-9.
- [255] García-Fraile P, Benada O, Cajthaml T, Baldrian P, Lladó S. *Terracidiphilus gabretensis* gen. nov., sp. nov., an Abundant and Active Forest Soil Acidobacterium Important in Organic Matter Transformation. *Appl Environ Microbiol.* 2016;82(2):560–569. doi:10.1128/AEM.03353-15.
- [256] Béjà O, Suzuki MT, Koonin EV, Aravind L, Hadd A, Nguyen LP, et al. Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ Microbiol.* 2000;2:516–529. doi:10.1046/j.1462-2920.2000.00133.x.
- [257] Zhang M, Zhang Y, Scheuring CF, Wu CC, Dong JJ, Zhang HB. Preparation of megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics research. *Nat Protoc.* 2012;7:467–478.
- [258] Pollman MJ, Zuccarelli AJ. Rapid isolation of high-molecular-weight DNA from agarose gels. *Anal Biochem.* 1989;181:12–17. doi:10.1016/0003-2697(89)90386-2.
- [259] Schneegurt MA, Dore SY, Kulpa Jr CF. Direct extraction of DNA from soils for studies in microbial ecology. *Curr Issues Mol Biol.* 2003;5:1–8.
- [260] Reigstad LJ, Bartossek R, Schleper C. Chapter Thirteen - Preparation of High-Molecular Weight DNA and Metagenomic Libraries from Soils and Hot Springs. In: Martin GK, Lisa YS, editors. *Methods Enzymol.* vol. Volume 496. Academic Press; 2011. p. 319–344.
- [261] Salazar O, Asenjo JA. Enzymatic lysis of microbial cells. *Biotechnol Lett.* 2007;29(7):985–994. doi:10.1007/s10529-007-9345-2.
- [262] Herschleb J, Ananiev G, Schwartz DC. Pulsed-field gel electrophoresis. *Nat Protoc.* 2007;2:677–684.
- [263] Ma Z, Weining S, Sharp PJ, Liu C. Non-gridded library: a new approach for BAC (bacterial artificial chromosome) exploitation in hexaploid wheat (*Triticum aestivum*). *Nucleic Acids Res.* 2000;28(24):E106.
- [264] Cao W, Fu B, Wu K, Li N, Zhou Y, Gao Z, et al. Construction and characterization of three wheat bacterial artificial chromosome libraries. *Int J Mol Sci.* 2014;15(12):21896–912. doi:10.3390/ijms151221896.
- [265] Martinez A, Kolvek SJ, Yip CLT, Hopke J, Brown KA, MacNeil IA, et al. Genetically Modified Bacterial Strains and Novel Bacterial Artificial Chromosome

- Shuttle Vectors for Constructing Environmental Libraries and Detecting Heterologous Natural Products in Multiple Expression Hosts. *Appl Environ Microbiol.* 2004;70:2452–2463. doi:10.1128/aem.70.4.2452-2463.2004.
- [266] Kercher MA, Lu P, Lewis M. Lac repressor-operator complex. *Curr Opin Struct Biol.* 1997;7(1):76–85.
- [267] Schneemann I, Wiese J, Kunz AL, Imhoff JF. Genetic Approach for the Fast Discovery of Phenazine Producing Bacteria. *Mar Drugs.* 2011;9(12):772–789. doi:10.3390/md9050772.
- [268] Hornung A, Bertazzo M, Dziarnowski A, Schneider K, Welzel K, Wohlert SEE, et al. A genomic screening approach to the structure-guided identification of drug candidates from natural sources. *Chembiochem.* 2007;8(7):757–766. doi:10.1002/cbic.200600375.
- [269] Lee MY, Myeong JS, Park HJ, Han K, Kim ES. Isolation and partial characterization of a cryptic polyene gene cluster in *Pseudonocardia autotrophica*. *J Ind Microbiol Biotechnol.* 2006;33(2):84–87. doi:10.1007/s10295-005-0018-7.
- [270] Kessler N, Schuhmann H, Morneweg S, Linne U, Marahiel MA. The linear pentadecapeptide gramicidin is assembled by four multimodular nonribosomal peptide synthetases that comprise 16 modules with 56 catalytic domains. *J Biol Chem.* 2004;279(9):7413–9. doi:10.1074/jbc.M309658200.
- [271] Mootz HD, Marahiel MA. The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J Bacteriol.* 1997;179(21):6843–50.
- [272] Sabree ZL, Bergendahl V, Liles MR, Burgess RR, Goodman RM, Handelsman J. Identification and characterization of the gene encoding the *Acidobacterium capsulatum* major sigma factor. *Gene.* 2006;376:144–151. doi:10.1016/j.gene.2006.02.033.
- [273] Delcour AH. Outer membrane permeability and antibiotic resistance. *Biochim Biophys Acta.* 2009;1794(5):808–16. doi:10.1016/j.bbapap.2008.11.005.
- [274] Zgurskaya HI, López CA, Gnanakaran S. Permeability Barrier of Gram-Negative Cell Envelopes and Approaches To Bypass It. *ACS Infect Dis.* 2015;1(11):512–522. doi:10.1021/acsinfecdis.5b00097.
- [275] Du D, Wang Z, James NR, Voss JE, Klimont E, Ohene-Agyei T, et al. Structure of the AcrAB–TolC multidrug efflux pump. *Nature.* 2014;509(7501):512–515. doi:10.1038/nature13205.
- [276] Zgurskaya HI, Krishnamoorthy G, Ntrel A, Lu S. Mechanism and Function of the Outer Membrane Channel TolC in Multidrug Resistance and Physiology of Enterobacteria. *Front Microbiol.* 2011;2:189. doi:10.3389/fmicb.2011.00189.

- [277] Nikaido H, Takatsuka Y. Mechanisms of RND multidrug efflux pumps. *Biochim Biophys Acta - Proteins Proteomics*. 2009;1794(5):769–781. doi:10.1016/j.bbapap.2008.10.004.
- [278] Lyu Y, Yang X, Goswami S, Gorityala BK, Idowu T, Domalaon R, et al. Amphiphilic Tobramycin–Lysine Conjugates Sensitize Multidrug Resistant Gram-Negative Bacteria to Rifampicin and Minocycline. *J Med Chem*. 2017;60(9):3684–3702. doi:10.1021/acs.jmedchem.6b01742.
- [279] Rice L. Federal Funding for the Study of Antimicrobial Resistance in Nosocomial Pathogens: No ESKAPE. *J Infect Dis*. 2008;197(8):1079–1081. doi:10.1086/533452.
- [280] Chuanchuen R, Murata T, Gotoh N, Schweizer HP. Substrate-dependent utilization of OprM or OpmH by the *Pseudomonas aeruginosa* MexJK efflux pump. *Antimicrob Agents Chemother*. 2005;49(5):2133–6. doi:10.1128/AAC.49.5.2133-2136.2005.
- [281] Fyfe C, Sutcliffe JA, Grossman TH. Development and characterization of a *Pseudomonas aeruginosa* in vitro coupled transcription-translation assay system for evaluation of translation inhibitors. *J Microbiol Methods*. 2012;90(3):256–61. doi:10.1016/j.mimet.2012.05.018.
- [282] de Souza CSM, Fortaleza CMCB, Witzel CL, Silveira M, Bonesso MF, Marques SA, et al. Toxigenic profile of methicillin-sensitive and resistant *Staphylococcus aureus* isolated from special groups. *Ann Clin Microbiol Antimicrob*. 2016;15:9. doi:10.1186/s12941-016-0125-5.
- [283] Elhassan MM, Ozbak HA, Hemeg HA, Elmekki MA, Ahmed LM. Absence of the *mecA* Gene in Methicillin Resistant *Staphylococcus aureus* Isolated from Different Clinical Specimens in Shendi City, Sudan. *Biomed Res Int*. 2015;2015:895860. doi:10.1155/2015/895860.
- [284] Li JWH, Vederas JC. Drug Discovery and Natural Products: End of an Era or an Endless Frontier? *Science* (80- ). 2009;325(5937).
- [285] Uchiyama T, Miyazaki K. Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr Opin Biotechnol*. 2009;20(6):616–622. doi:10.1016/j.copbio.2009.09.010.
- [286] Troeschel SC, Thies S, Link O, Real CI, Knops K, Wilhelm S, et al. Novel broad host range shuttle vectors for expression in *Escherichia coli*, *Bacillus subtilis* and *Pseudomonas putida*. *J Biotechnol*. 2012;161:71–79. doi:10.1016/j.jbiotec.2012.02.020.
- [287] Mishra KP, Ganju L, Sairam M, Banerjee PK, Sawhney RC. A review of high throughput technology for the screening of natural products. *Biomed Pharmacother*. 2008;62(2):94–98. doi:10.1016/j.biopha.2007.06.012.

*Literaturverzeichnis*

- [288] Lan F, Demaree B, Ahmed N, Abate AR. Single-cell genome sequencing at ultra-high-throughput with microfluidic droplet barcoding. *Nat Biotechnol.* 2017;35(7):640–646. doi:10.1038/nbt.3880.
- [289] Rónai AZ, Timár J, Makó E, Erdö F, Gyarmati Z, Tóth G, et al. Diprotin A, an inhibitor of dipeptidyl aminopeptidase IV(EC 3.4.14.5) produces naloxone-reversible analgesia in rats. *Life Sci.* 1999;64(2):145–52.
- [290] Schinko E, Schad K, Eys S, Keller U, Wohlleben W. Phosphinothricin-tripeptide biosynthesis: An original version of bacterial secondary metabolism? *Phytochemistry.* 2009;70(15-16):1787–1800. doi:10.1016/j.phytochem.2009.09.002.
- [291] Hegemann JD, Zimmermann M, Zhu S, Steuber H, Harms K, Xie X, et al. Xanthomoninse28085I-III: A New Class of Lasso Peptides with a Seven-Residue Macrolactam Ring. *Angew Chemie Int Ed.* 2014;53(8):2230–2234. doi:10.1002/anie.201309267.
- [292] Oppegård C, Rogne P, Emanuelsen L, Kristiansen PE, Fimland G, Nissen-Meyer J. The Two-Peptide Class II Bacteriocins: Structure, Production, and Mode of Action. *J Mol Microbiol Biotechnol.* 2007;13(4):210–219. doi:10.1159/000104750.
- [293] Doroghazi JR, Albright JC, Goering AW, Ju KS, Haines RR, Tchaluikov KA, et al. A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat Chem Biol.* 2014;10:963–968.

## Danksagung

Nach mehr als drei Jahren harter Arbeit, möchte einige Kollegen hervorheben, mit denen ich intensiv und konstruktiv zusammengearbeitet habe. Ein besonderes Dankeschön gilt Dr. Benedikt Leis für die Unterstützung in der Sequenzierung von *A. borealis* und BAC-Inserts an der auch Celine Zumkeller beteiligt war. Ich danke Michael Marnier für die Unterstützung während der Extrakterstellung, sowie Stefanie Krahn und Mareike Lang für das Biolumineszenz-Assay. Bei Rolf Hirsch und Kirsten Bommersheim bedanke ich mich für die Beratung während der Testung der Extrakte. Desweiteren möchte ich mich bei Markus Oberpaul für die sehr gute Zusammenarbeit bei der Gewinnung von Bakterien aus Termitennestern bedanken. Dr. Marius Spohn und Jennifer Kuhn danke ich für die Hilfe während der PCR-basierten Untersuchung der Klonbibliotheken. Zu guter Letzt möchte ich Dr. Luigi Toti für die allgemeine sehr gute Zusammenarbeit erwähnen, sowie Dr. Gerrit Jürjens für seine unterstützende Fachkompetenz.