Faculty 09 - Agricultural Sciences, Nutritional Sciences, and Environmental Management

Natural Product Discovery at the Intersection of Genomics and Synthetic Biology: Insights from Bacteroidota and Acidobacteriota

Dissertation

(Dr. rer. nat.)

Celine Mara Zumkeller

November 2025

1st Reviewer: Prof. Dr. Till Schäberle

Department of Natural Product Research

Institute for Insect Biotechnology

Justus-Liebig-University Giessen

2nd Reviewer: Prof. Dr. Sylvia Schnell

Department of General and Soil Microbiology

Institute for Applied Microbiology

Justus-Liebig-University Giessen

Für mein kleines Kunterbunt

Erklärung gemäß der Promotionsordnung des Fachbereichs 09 vom 07. Juli 2004 § 17 (2), geändert am 29. Mai 2019

Ich erkläre: Ich habe die vorgelegte Dissertation selbständig 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. 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.

_____ Gießen, den 06.10.2025

Celine Zumkeller

CONTENTS

LIST OF FIGURES

List of Tables

List of Supplementary Figures

List of Abbreviations and Acronyms

Abbreviation	Full term	Notes
ABS	Access and Benefit-Sharing	Regulatory framework
ACP	Acyl carrier protein	PKS domains carrying units
AI	Artificial intelligence	-
ANI	Average nucleotide identity	Genome similarity metric
antiSMASH	antibiotics & Secondary Metabolite Analysis Shell	BGC annotation tool
BGC	Biosynthetic gene cluster	encoding NPs
BiG-SCAPE	Biosynthetic Gene Similarity Clustering and Prospecting Engine	BGC networking
dDDH	digital DNA:DNA hybridisation	Genome similarity metric
DH	Dehydratase	PKS reductive domain
ER	Enoylreductase	PKS reductive domain
FHG	Fraunhofer-Gesellschaft	Strain collection context
GNPS	Global Natural Products Social molecular networking	MS/MS networking – similarity
GTDB-Tk	Genome Taxonomy Database Toolkit	Taxonomy assignment
HEL	Heterologous Expression Library	This thesis platform
HGT	Horizontal gene transfer	-
IAA	Indole-3-acetic acid	Auxin
iChip	Isolation Chip	In situ diffusion chamber
iP	N6-(Δ2-isopentenyl)adenine	Cytokinin
KR	Ketoreductase	PKS reductive domain
LC-MS/MS	Liquid chromatography/tandem mass spectrometry	-
MAG	Metagenome-assembled genome	-
MiBIG	Minimum Information about a Biosynthetic Gene Cluster	BGC reference database
MoClo	Modular Cloning	Golden Gate-like assembly

NRPS	Non-ribosomal peptide synthetase	-
NP	Natural product	-
OSMAC	One Strain-Many Compounds	-
PCP	Peptidyl carrier protein	NRPS domain
PGPT	Plant growth-promoting trait	-
PKS	Polyketide synthase	-
PPTase (Sfp)	Phosphopantetheinyl transferase	Modifying ACP/PCP domains
QC	Quality control	Reads/assemblies
QUAST	QUality ASsessment Tool	Assembly metrics
RiPP	Ribosomally synthesised and post-translationally modified peptide	-
RBS	Ribosome binding site	-
TAR	Transformation-associated recombination	Yeast-based BGC assembly
TE	Thioesterase	Chain release/cyclisation
Trp	Tryptophan	-
VM	Virtual machine	-

ABSTRACT

Microbial natural products remain a valuable source of bioactive compounds; however, discovery pipelines often encounter issues such as rediscovery, culture bias, and natural product (NP)-encoding but silent biosynthetic gene clusters (BGCs). This dissertation connects and applies an omics-guided, cultivation-dependent workflow for underexplored Bacteroidota and Acidobacteriota. It also establishes a heterologous expression toolkit for optimising the expression of silent biosynthetic gene clusters by applying synthetic biology principles.

The sections "Genome Mining of Bacteroidota" employ genome mining tools, including quality control, taxonomy, BGC prediction, and similarity networking, to profile and describe newly isolated marine Bacteroidota. This process expands genomic reference sequences and maps chemotaxonomic markers, such as flexirubin/flexixanthin BGCs, guiding strain prioritisation for natural product discovery and shaping study designs for functional investigations. Thus, the work establishes a foundation for systematic genomics documentation and reporting, enabling portfolio-style reporting of novel strains that are being included in the 'Fraunhofer Gesellschaft' (FHG) strain collection. A genus-wide analysis of the *Pedobacter* (phylum Bacteroidota) identified a branch genetically enriched in multimodular NRPS BGCs. Genetic validation connected one NRP cluster of interest to the novel cryopeptin lipopeptide family, positioning this clade as a genetically tractable hotspot for novel scaffolds.

The section "Profiling of Novel Acidobacteriota" investigates the biosynthetic potential of the Acidobacteriota. High-throughput selective cultivation yields novel *Acidobacteriaceae* representatives, which are profiled for their BGC composition in a comparative study of curated, public Acidobacteriota genomes. Metabolomics reveals the production of the phytohormones indole-3-acetic acid (IAA) and the cytokinin N6-(Δ2-isopentenyl)adenine (iP). Tryptophan supplementation drives a global metabolic shift, also increasing IAA production while generally decreasing iP. Absolute quantification of phytohormones places the amount in functionally relevant ranges, but crude extract testing does not increase barley seedling biomass under the tested conditions. These results refine the ecological and functional potential of Acidobacteriota and map the distributions of plant growth-promoting traits (PGPTs) across families.

The last section presents HEL, a modular, heterologous expression and library-style cloning platform that retrofits clones of BGCs with reusable and exchangeable expression cassettes (such as promoter libraries) and optional host-specific compatibility cassettes for broad chassis transfer. By doing so, HEL enables the tuning of expression by linking a prioritised BGC to standard, library-style phenotypes using modular cloning (MoClo), a Golden Gate-like cloning method.

Collectively, this dissertation provides a reproducible genomic workflow for underexplored phyla and the FHG strain collection (i); new genomic resources, genetic prioritisation, and tractability in *Pedobacter* (ii); integrated genomic–metabolomic insights into tryptophan-induced metabolic dynamics, phytohormone biology, and PGPT architecture in Acidobacteriaceae, and (iii) a flexible genetic toolkit (HEL) to induce activation or accelerate the expression of prioritised BGCs across different heterologous hosts.

1 INTRODUCTION

1.1 A Brief History of Natural Product Research

Microbial Natural Products (NPs) are specialised metabolites with diverse bioactivities (Seca and Pinto, 2019) produced by microorganisms (confined to bacteria and fungi in this thesis). The terms "secondary or specialised metabolite" and "natural product" are frequently used interchangeably to refer to specialised molecules synthesised beyond primary metabolism. In this thesis, the term "natural product" specifically denotes secondary metabolites. NPs have been shown to mediate a variety of functions, including morphogenic (Willey *et al.*, 2006), antimicrobial (Hibbing *et al.*, 2010), cytotoxic (Ueoka *et al.*, 2010), signalling, and other processes that shape the intricate interactome among bacteria (Traxler *et al.*, 2013) and with other organisms (Coolahan and Whalen, 2025; Sen *et al.*, 2009). The execution of these bioactivities relies on specific interactions between the NPs and their target, which has been shaped by evolution to act with high potency and specificity (Firn and Jones, 2003). For this reason, NPs are usually structurally complex (Feher and Schmidt, 2003) and not easy to mimic by synthetic chemistry approaches (Hong, 2014; Shenvi, 2024). Consequently, microbial NPs have a long-standing tradition and remain to this day an important source for discovering novel chemical scaffolds for human use (Newman and Cragg, 2020).

In the **20th century**, the discovery of antimicrobially active NPs (commonly called antibiotics) was a cornerstone in the history of public health (Hutchings *et al.*, 2019), revolutionising the treatment of infectious diseases. Thus, mortality by infectious diseases dropped significantly, and life expectancy increased due to the advent of effective antimicrobials, alongside sanitation and vaccination programs (CDC, 2025). This period, often called the "golden era" of antibiotics, relied heavily on a cultivation-first approach focused on a few productive microbial groups (Baltz, 2008)—primarily the Actinomycetota (Wohlleben *et al.*, 2016)(commonly known as Actinobacteria), as well as the Proteobacteria (Buijs *et al.*, 2019; Gross and Loper, 2009), and Myxococcota (Herrmann *et al.*, 2017) (commonly called Myxobacteria). It was also shaped by extensive bioprospecting campaigns (Schatz *et al.*, 1944; Schatz and Waksman, 1944; Waksman *et al.*, 1946, Lewis, 2020), often led by industry, to isolate previously uncultured microorganisms, characterise them, and incorporate them into established screening pipelines. These strategies generated many first-in-class NP scaffolds but ultimately laid the groundwork for future rediscovery of NP scaffolds, as discovery programs continued to explore the same lineages and reuse established screening protocols (Genilloud, 2019). **Beyond their use as anti-infectives** they have a long-standing tradition of aiding human needs, in various areas of human activity, most famously as anti-infectives, as therapeutics in general (Newman and Cragg, 2020), in agriculture as pesticides (Kirst, 2010; Mertz and Yao, 1990), in the animal health sector (McKELLAR and Benchaoui, 1996), and even for industrial (Mortensen *et al.*, 2017) and consumer (Nasser *et al.*, 2024) use. Their diverse applications are undergoing a renaissance as global pressures, including the antimicrobial-resistance (AMR) crisis (World Health Organization, 2025), neglected diseases (World Health Organization, 2020) (e.g., ivermectin for onchocerciasis (Lawrence *et al.*, 2015; World Health Organization, 2023);

amphotericin B for leishmaniasis (Sundar *et al.*, 2010)), and climate-driven environmental disruption (Intergovernmental Panel on Climate Change, 2023; Romanello *et al.*, 2024), demand sustainable solutions. However, discovery pipelines cannot keep up with this demand, as rediscovery of known chemotypes (Harvey *et al.*, 2015; Katz and Baltz, 2016) is common due to over-mined producer lineages and conventional screening setups. This is further amplified by a culture bias and the widespread presence of cryptic biosynthetic gene clusters (BGCs) (Medema and Fischbach, 2015a; Rutledge and Challis, 2015), encoding such natural products.

1.2 Microbial Groups of Interest for Novel Chemical Structures

The phylum **Bacteroidota** is an emerging NP producer (Brinkmann *et al.*, 2022a) that makes an excellent subject for a genome-guided, cultivation-dependent discovery project because of its generally straightforward growth in laboratory conditions (Beckmann *et al.*, 2017; Sack *et al.*, 2011). The chemical diversity within this phylum is considered promising and distinct from traditional NP-producing groups (Brinkmann *et al.*, 2022c, 2022a). However, only a relatively small number of compounds have been isolated and characterised so far (Brinkmann *et al.*, 2022d). Among these are complex compounds with known activities or ecological roles: for example, the elansolids (Steinmetz *et al.*, 2011) from *Chitinophaga* strains that are encoded by trans-acyltransferase polyketide synthase BGCs and show cytotoxic and antibacterial activity in experimental screenings. Another NP with an ecological function is the terpenoid thallusin (Matsuo *et al.*, 2005), produced by a *Flavobacterium*, which functions as an algal morphogen. The genus ***Pedobacter*** has been identified as a BGC hotspot within the Bacteroidota. And representatives are widely distributed across diverse environments. Though studied for industrially relevant enzymes (Zhu *et al.*, 2018a) and AMR genes (Ullmann *et al.*, 2020a; Viana *et al.*, 2018a) *Pedobacter* remains underexplored as a source of NPs. Aside from the cyclic lipodepsipeptides (iso)pedopeptins (Hirota-Takahata *et al.*, 2014; Kozuma *et al.*, 2014; Nord *et al.*, 2020a) — potent against MDR Gram-negatives — little is known despite genomic indications of a broader biosynthetic capacity.

The phylum **Acidobacteriota** is widespread across many environments (Huber *et al.*, 2022), including extreme habitats (Kishimoto *et al.*, 1991). First described in 1997 (Ludwig *et al.*, 1997) based on 16S rRNA sequences, its taxonomic classification is still evolving, currently divided into 15 class-level units (Dedysh and Yilmaz, 2018). However, these divisions are not fully aligned with the Genome Taxonomy database (Parks *et al.*, 2022), which will be used throughout this thesis for taxonomic classification. Early work relied on 16S surveys to map distribution and ecology (Jones *et al.*, 2009; Quaiser *et al.*, 2003) and later, metagenomic approaches provided initial insights into the potential functions (Crits-Christoph *et al.*, 2022; Gonçalves *et al.*, 2024; Reji and Zhang, 2022; Wong *et al.*, 2024). In addition to roles in sulphur (Hausmann *et al.*, 2018) and nitrogen cycling (Reji and Zhang, 2022), and H_2 consumption (Giguere *et al.*, 2021), members of this phylum possess a large number of carbohydrate-active enzymes (e.g. cellulases, chitinases, xylanases) (Eichorst *et al.*, 2018; Gonçalves *et al.*, 2024). At the time of the initial description, only three isolates (Coates *et al.*, 1999; Kishimoto *et al.*, 1991; Liesack *et al.*, 1994) were available, prompting efforts to cultivate and characterise new representatives. To date, there are 81 entries in LPSN (Leibniz Institute DSMZ) covering approximately 30 genera, which are matched by over 3175 genomes listed in the Genome Taxonomy Database (Parks *et al.*, 2022). As a result,

research — especially on hard-to-cultivate subdivisions — often depends on analysing metagenome-assembled genomes (Ruhl *et al.*, 2022).

After Parsley *et al.* (2011). first described structurally novel NRPS/PKS in soil metagenomes and traced them to Acidobacteriota lineages, NP research largely overlooked this phylum. Damsté *et al.* (2017) linked the first new metabolites to cultured Acidobacteria, specifically iso-diabolic acid (13,16-dimethyl octacosanedioic acid), a membrane-spanning lipid (Sinninghe Damsté *et al.*, 2011) and hopanoids, a bacterial sterol analogue (Damsté *et al.*, 2017; Sáenz *et al.*, 2015). The first, more complex NPs were then described 12 years later by (Leopold-Messer *et al.*, 2023), including Acidobactamides A-C, calyculin derivative 21-methoxy-calyculinamide, and six phorboxazole-like polyketides called phorbactazoles A–F, isolated from *Acanthopleuribacter pedis* KCTC 12899. These NPs were isolated and guided by predictions from trans-AT PKS gene clusters initially identified in sponge microbiomes. Reported activities were mainly cytotoxic; to our knowledge, no Acidobacteriota-derived natural product has shown antimicrobial activity.

Strain collections serve as strategic hubs for microbial biodiversity and biosynthetic potential (Boundy-Mills *et al.*, 2020). In the quest for new NPs, major industrial companies collected and stored numerous microbial isolates (Clardy *et al.*, 2006) from various habitats for cultivation and bioactivity testing. As rediscovery became common and synthetic chemistry libraries gained importance, industrial efforts towards the discovery of novel NP scaffolds gradually declined (Payne *et al.*, 2007). Today, most of these strain collections have transitioned to public institutions, foundations or service-oriented companies. This shift reduced corporate risk and opened new opportunities for these entities to revisit and utilise these resources using modern omics techniques (Steele *et al.*, 2019). Many organisations aim to transform their culture collections into genome-sequenced repositories to uncover the genetic blueprint of each strain and identify previously silent, undiscovered potential (Kalkreuter *et al.*, 2024). Achieving full omic integration requires infrastructure that links databases containing strain metadata (Parks *et al.*, 2022; Schober *et al.*, 2024) (taxonomy, origin, ABS) with computational workflows that connect omics (Nothias *et al.*, 2020; van Santen *et al.*, 2022; Zdouc *et al.*, 2025) and bioactivity data.

1.3 Biosynthetic Gene Clusters encoding NPs

NPs are encoded by BGCs, co-localised enzymes, regulators, and tailoring functions that together orchestrate the production and export of small-molecule NPs. While NPs originate from various biosynthetic routes, this chapter and most of the thesis focus on ribosomally produced and post-translationally modified peptides (RiPPs) and ribosome-independent, multimodular assembly line megaenzymes, because their hierarchical structure predetermines and shapes underlying methods, such as pattern-based detection of clusters and cloning challenges inherent in repetitive DNA fragments. One type of megaenzymes is the **Non-ribosomal peptide synthetases** (NRPS), producing peptide NPs (Miller and Gulick, 2016). Multimodular NRPSs consist of repetitive modules, with each one responsible for the assembly and integration of an amino acid. To do so, each module is commonly composed of an adenylation (A) domain, which selects and activates the amino acid, a peptidyl carrier (PCP) domain, moving the amino acid or the growing peptide chain from one condensation domain to the other, and a condensation (C) domain, which catalyses the peptide bond formation (reviewed in Bloudoff and Schmeing, 2017). This assembly line logic often extends peptides in a colinear order, frequently

terminating with thioesterase (TE) domains that also often catalyse cyclisation of the peptide (Little and Hertweck, 2022). Other domains, such as epimerisation or oxidation domains and adjacent tailoring enzymes (e.g., methyltransferases, halogenases and glycosyltransferases), further modify amino acids or peptides, thereby diversifying chemical scaffolds (T. Walsh, 2023).

The other type**, Polyketide synthase (PKS)** systems mirror this logic: type I modular PKSs carry repeated ketoacyl synthase (KS) (chain extension via Claisen condensation), acyltransferase (AT) (selection and loading of extender units), and acyl carrier protein (ACP) (carrier of extender units/growing chain) domains with optional reductive ketoreductase (KR), dehydratase (DH) and enolreductase (ER) domains (Khosla *et al.*, 2009). The absence, presence and functionality of these reductive domains determine the oxidation status of the final polyketide product (Keatinge-Clay, 2016). **Hybrid PKS–NRPS** clusters are commonly found and mediate chain transfer through the transfer of the growing NP by their carrier domains (ACP to PCP), enabling the seamless assembly of polyketide-peptide products (Bonhomme *et al.*, 2023). Beyond the core scaffold genes, BGCs typically include local regulation, precursor supply, and export/resistance genes; boundaries may be flanked by repeats or mobility elements, consistent with horizontal gene transfer (HGT) (Ziemert *et al.*, 2014). The modular and repetitive structure of NRPS and PKS systems allows detection of clusters within the genomes, while co-linearity also aids prediction of potential scaffold structure (Medema and Fischbach, 2015a); however, exceptions (skipping of modules, iterative use of modules) are common (Challis and Naismith, 2004), highlighting the need for experimental validation.

1.4 The Need for Novel Natural Products

This development is problematic because the antimicrobial resistance (AMR) crisis (World Health Organisation, 2025), the spread of infectious diseases due to globalisation (Frenk *et al.*, 2011), and climate-related pressures for agricultural systems (Mora *et al.*, 2022) continue to increase demand for new scaffolds. However, approvals in recent decades have primarily involved derivatives of known scaffolds, with only a few truly new classes featuring novel modes of action (Butler *et al.*, 2024, 2023; World Health Organization, 2024). The challenge for modern natural product research is to help close this innovation gap (Miethke *et al.*, 2021). Fortunately, the vast potential for chemical novelty exists (Cimermancic *et al.*, 2014; Kautsar *et al.*, 2021) within what is known as microbial dark matter, which remains largely undiscovered (Hug *et al.*, 2016; Lloyd *et al.*, 2018). However, accessing this microbial dark matter is difficult due to challenges in cultivating these organisms (Rinke *et al.*, 2013) or transferring BGCs of interest to suitable chassis (Zhang *et al.*, 2019). For this reason, current research focuses on underexplored taxa and bioresources (1.2) and on developing new tools to exploit them in a cultivation-dependent (1.4.1) and cultivation-independent (1.4.2) manner. Furthermore, sophisticated analytical and computational tools exist to connect and analyse complex data retrieved from both ends of the workflow (1.4.3).

1.4.1 Cultivation-dependent NP Discovery approaches

Since the great plate count anomaly revealed how few microbes form colonies compared to microscopic observations, microbiologists have tried to access this "**microbial dark matter**" (Rinke *et al.*, 2013). These efforts mainly focus on mimicking original isolation conditions that satisfy the environmental and nutritional needs of the microorganisms (Aoi *et al.*, 2009; Bollmann *et al.*, 2007; Kaeberlein *et al.*, 2002). Because nearby organisms can help (D'Onofrio *et al.*, 2010) but also hinder (Zengler *et al.*, 2002) growth, different methods try to separate single cells. The most well-known tool is the **iChip** (Nichols *et al.*, 2010): a small plate with wells designed to capture individual cells from the environment within an agarose matrix. The plate is then sealed with semi-permeable membranes and incubated in its natural environment, allowing cell growth by providing nutrients that diffuse through the membranes. After growth, the microcolonies are transferred to lab media for domestication and optional antimicrobial screening (Berdy *et al.*, 2017). Cultivation approaches using iChip recovered about 26–40% of cells and led to the discovery of teixobactin (Ling *et al.*, 2015) (Nauwynck *et al.*, 2025; Zengler *et al.*, 2002), an antibiotic active against Gram-positive bacteria.

More recently, droplet **microfluidics** technology has been used to create even smaller compartments (from microliters to femtoliters), encapsulating individual cells. The small size of these droplets offers the advantage of retrieving thousands of single cells for subsequent parallel cultivation under various conditions (Oberpaul *et al.*, 2022). This high-throughput technique has increased the diversity of microorganisms isolated, such as new representatives of Acidobacteriota (Dai *et al.*, 2025; Mahler *et al.*, 2021; Oberpaul *et al.*, 2022). After successful isolation, different cultivation strategies utilise inducer systems (Okada and Seyedsayamdost, 2017; Seyedsayamdost, 2014), co-cultivation (Onaka *et al.*, 2001; Schaenzer *et al.*, 2024; Traxler *et al.*, 2013), or media variations to maximise metabolic diversity output from isolate strains. One such approach is the One-strain-many-compounds (OSMAC) method (Bode *et al.*, 2002), which is cost-effective and widely used in NP research because it broadens chemical space and reduces rediscovery of known NPs.

1.4.2 Cultivation-independent NP Discovery approaches

When cultivation-dependent conditions do not activate a silent BGC, genetic activation strategies in the native host (homologous expression) and, if that is not feasible or unsuccessful, in a non-native host (heterologous expression) are the next step in the NP discovery process (Ochi, 2017; Rutledge and Challis, 2015). **Homologous expression** comes with the inherent advantage of the native physiological background of the host. Common measures include promoter replacement (M. M. Zhang *et al.*, 2017) and overexpression of BGC-associated regulators using CRISPR-Cas (Ameruoso *et al.*, 2022; Xie *et al.*, 2024) or transposon mutagenesis (Ahmed *et al.*, 2017; Xu *et al.*, 2017). These approaches, however, still depend on the genetic accessibility of the host, which might be limited, especially in underexplored groups.

Heterologous expression involves two main steps: first, capturing or cloning the BGC and transferring it into the heterologous host; second, optimising the BGC or host environments for successful expression, as the host's physiology may not be suited to the non-native BGC. Both steps present significant practical challenges. (i) Cloning or capturing large multimodular BGCs, with the average NRPS/PKS cluster size in bacteria of 38.7 kb (Wang *et al.*, 2014), but often up to 100kb (Schwecke *et al.*, 1995) and GC-rich, remains difficult. Thus, cloning usually involves multiple steps, either by assembling smaller fragments sequentially (e.g., using Gibson

assembly) or through more complex methods such as yeast homologous recombination (TAR) for assembly (Kim *et al.*, 2010; Yamanaka *et al.*, 2014) or CRISPR-Cas to capture the BGC from the producer's genome (CATCH for BGCs <50kb (Jiang *et al.*, 2015); CAPTURE for BGC >100 kb (Enghiad *et al.*, 2021) or ACTIMOT (Xie *et al.*, 2024)). (ii) Transferring a BGC involves carefully choosing a suitable chassis; For actinomycete BGCs, Streptomyces hosts (e.g., *S. coelicolor* M1146 (Gomez-Escribano and Bibb, 2011), *S. albus* J1074 (Wendt-Pienkowski *et al.*, 2005; Zaburannyi *et al.*, 2014), and more) are commonly selected. *Escherichia coli* BAP1 (Pfeifer *et al.*, 2001; Zhang *et al.*, 2008) is specifically engineered for NRPS/PKS expression by integrating the *sfp* gene (broad specificity PPTase, activating PCP domains). Other hosts, such as *B. subtilis* (Doekel *et al.*, 2002; Kumpfmüller *et al.*, 2016), *Pseudomonas putida* (Loeschcke and Thies, 2015), and *Myxococcus xanthus* (Julien and Shah, 2002; Stevens *et al.*, 2010; Wenzel *et al.*, 2005) can also be utilised. (iii) Often, selecting the right chassis alone is not sufficient for BGC activation (Smanski *et al.*, 2016). Researchers commonly modify the BGC by codon optimisation (Schmidt *et al.*, 2023), promoter or ribosomal binding site (RBS) exchange (Chiang *et al.*, 2009), or even substitute missing functional enzyme machinery such as the Sfp PPTase (see above) (Lambalot *et al.*, 1996; Pfeifer *et al.*, 2001), MbtH-like proteins (Herbst *et al.*, 2013; Mori *et al.*, 2018), P450 redox partners (Liu *et al.*, 2022; Rudolf *et al.*, 2017), tRNA pools (Gustafsson *et al.*, 2004; Kane, 1995), or efflux and self-resistance cassettes (Stein *et al.*, 2003; Xu *et al.*, 2020; Zhang *et al.*, 2010). If the BGCs originate from metagenomic DNA, the physiology of the original producer may be unknown, making design choices more challenging. In practice, successful activation typically requires iterative refactoring and, often, testing the construct across multiple hosts.

1.4.3 Analytical and Computational tools

Independent of their origin, recovered strain isolates/expression strains and their DNA, transcriptome, and metabolome require sophisticated analysis to prioritise discovery targets, avoid rediscovery, and verify expression outcomes. Across different molecule types (DNA → RNA → Metabolite), **omics** has followed similar developments in the data generation and analysis aspects. It evolved from low-throughput, expert-only assays (Sanger sequencing → early microarrays → GC-MS/NMR) to high-throughput (HT) platforms that trained technicians or specialised contract research organisations can perform. Analysis has developed in parallel: Specialised, command-line tools requiring advanced IT skills are now implemented as workbench pipelines (KBase (Arkin *et al.*, 2018), Galaxy (Goecks *et al.*, 2010)), also usable for interested individual researchers. Furthermore, large, public databases enable the comparison of datasets between research groups (when designs/formats align) or the dereplication of samples against curated reference databases. Table 1 presents selected workflows and databases (NP-focused) to guide wet-lab discovery and validation.

Table 1 Curated Natural-Product Pipelines and Repositories for Experiments

Molecule Type	Tools	Reference Databases
(Meta)Genomes	Data Acquisition: Illumina, Long Read Sequencing technologies: Oxford Nanopore, PacBIO Detect BGCs: antiSMASH, Cluster based on similarity: BiG-SCAPE (Navarro-Muñoz *et al.*, 2020a), BiG-SLICE (Kautsar *et al.*, 2021)	Genomes: NCBI, JGI BGCs: MIBiG, antiSMASH-Db
Aim: Prioritise novelty-rich strains and unknown clusters/clusters of interest		
(Meta)Transcriptomes	Data Acquisition: RNA-seq Analysis: BiG-MAP (Andreu *et al.*, 2021)	
Aim: Identify conditions for optimal BGC expression to guide cultivation efforts.		
Metabolomes	Data Acquisition: LC-MS/MS Analysis: GNPS	MassIVE (Choi *et al.*, 2020), GNPS (van Santen *et al.*, 2022), NP-Atlas (van Santen *et al.*, 2022)
Aim: Dereplication of known compounds, prevention of re-isolation		

To understand how these analytical pipelines emerged, it is useful to recall the evolution of computational genomics itself; With the availability of commercial Sanger sequencing services, researchers gained a first glimpse of how powerful such sequencing data could be. The biological blueprint of microorganisms does not only help to understand but also to predict fundamental biological processes (Sanger *et al.*, 1977). The increasing volume of sequencing data, along with its requirements and possibilities for analysis, sparked initiatives to standardise sequencing information and establish large, publicly available databases. This led to the foundation of large repositories such as NCBI (Smith, 2013), which soon became a central platform for sequence storage, analysis tools like BLAST (Altschul *et al.*, 1990) and curated genome reference standards such as RefSeq (Li *et al.*, 2021). Other initiatives followed (European Nucleotide Archive (O'Cathail *et al.*, 2025), Data bank of Japan (Ara *et al.*, 2024)) and complemented the database efforts, forming the International Nucleotide Sequence Database Collaboration (INSDC, (Karsch-Mizrachi *et al.*, 2025). Today, databases from JGI (Nordberg *et al.*, 2014) and the metagenome database MGnify (Mitchell *et al.*, 2020) further expand the availability of microbial genome datasets, while services such as Genome Workbench (Staff, 2020), Galaxy (Goecks *et al.*, 2010) and KBase (Arkin *et al.*, 2018) have made bioinformatic workflows more accessible to non-specialist users.

Due to their clustered nature and rule-based features, BGCs encoding NPs represent ideal targets for rule-based bioinformatic prediction tools. The antiSMASH tool (Medema *et al.*, 2011, antibiotics & Secondary Metabolite

Analysis Shell) was the first and remains the most widely used tool for BGC prediction. These earlier versions of antiSMASH, but also other applications such as CLUSEAN (Weber *et al.*, 2009), ClustScan (Starcevic *et al.*, 2008), np.searcher (Li *et al.*, 2009) and SMURF (Khaldi *et al.*, 2010) were limited in scope and most reliable with BGCs following the already derived rules, primarily within genomes of well-studied NP producers such as Actinomycetes (Medema and Fischbach, 2015b). Consequently, such algorithms failed to detect important but less canonical classes of SMs, such as the RiPPS (ribosomally-produced and post-translationally modified peptides). Today, antiSMASH detects over 101 different BGC types (Blin *et al.*, 2025), including RiPPS, using more advanced Hidden Markov Model–based algorithms.

Additional tools now implement structure predictions for produced SMs (PRISM)(Skinnider *et al.*, 2020), prediction of target interactions (Walker and Clardy, 2021). Similarity networking tools, such as BiG-SCAPE, selectively coupled with the public database MIBIG, estimate the novelty of a strain's biosynthetic potential (Navarro-Muñoz *et al.*, 2020a; Zdouc *et al.*, 2025). However, running these tools often requires specific computing environments, sufficient memory and labour-intensive steps to reformat generated output data for compatibility with other tools within the overall workflow. Nevertheless, establishing local workflow is critical when working with confidential genome data that does not allow uploading to web servers. This is particularly relevant for the Fraunhofer strain collection, where genome mining aims to discover and protect intellectual property from promising strains, making data security and reproducibility critical aspects in the work process.

Recent AI/ML tools now integrate these genomic outputs with metabolomic and transcriptomic data, creating a unified omics landscape. In **metabologenomics**, untargeted LC-MS/MS metabolomics is paired with genome data from the same strains to spot correlations between mass spectrometry (MS) features and BGCs occurrence/diversity. Rule-based genome mining tools map what a strain can potentially make, while the comparison with metabolomics data shows what is actually produced. Integration tools such as NP-Linker (Eldjárn *et al.*, 2021), Seq2PKS (Yan *et al.*, 2024) and NPOmix (Leão *et al.*, 2022) connect these layers to predict (i) which BGC makes which metabolite, (ii) identify encoded but silent chemistry, and (iii) guide targeted activation or design follow-up functional studies. Data can be stored in the "Paired Omics Data" platform (Schorn *et al.*, 2021).

Additionally, machine learning now fuels both sides: DeepBGC (Hannigan *et al.*, 2019), GECCO (Carroll *et al.*, 2021), or deepRiPP (Merwin *et al.*, 2020) are applied to find atypical or novel BGCs; Spec2Vec (Huber *et al.*, 2021a), MS2DeepScore (Huber *et al.*, 2021b), and MSNovelist (Stravs *et al.*, 2022) are used to improve MS/MS annotation. Together, these advances help exploit less-explored taxa, for which existing tools were not optimised.

1.5 Motivating Gaps

Despite rapid advances in omics techniques and increased data availability, several hurdles still hinder the efficient transformation of biosynthetic potential into molecules or functions in the Acidobacteriota as well as the Bacteroidota. (1) Challenges in cultivating Acidobacteriota: Although they are abundant in many habitats and play key roles in biogeochemical processes, there are few cultured representatives. When they are culturable, many strains exhibit slow growth or inconsistent growth patterns, characterised by low cell mass, which complicates the discovery of NPs through screening. This limits our understanding of metabolite production, including the potential of novel BGCs. (2) Compared to Acidobacteriota, the phylum Bacteroidota is better studied and already recognised as an emerging source of NPs. However, for some groups, BGC distribution and their links to specific molecules remain patchy, with only isolated examples described. (3) Access to NPs remains limited due to the inaccessibility of microbial dark matter and cryptic BGCs. Creating metagenomic libraries or transferring BGCs of interest into heterologous hosts is currently restricted to established vector-host systems, which offer only rough control over expression adjustments. When BGCs stay silent, a cycle of trial-and-error modifications often begins, making the process tedious and lengthy, and frequently failing to produce an NP.

1.6 Thesis Aim and Research Questions

To develop and validate a genome-guided discovery workflow that describes biosynthetic potential for the prioritisation of (A) strains, ultimately leading to molecules and phenotypes across two underexplored bacterial phyla, Bacteroidota and Acidobacteriota, and to establish and evaluate a synthetic biology toolkit for heterologous expression of prioritised BGCs (B), allowing expression modification and simple chassis transfer by retrofitting of BGC-encoding vectors.

1.6.1 Research Questions

⇒ From Genomes to Targets: How can standardised genome mining workflows prioritise BGCs and other traits of interest and trace them back to strains for experimental follow-up?

⇒ From Culture Isolates to NPs: How does Tryptophan supplementation change the global metabolome of Acidobacteriota and how does this translate to antimicrobially-associated traits?

⇒ From -omic traits to ecological function: Do the genomically and metabolically described PGPT support plant growth in plants?

⇒ From DNA of interest to NP: Can we use synthetic biology tools to simultaneously modulate expression of BGCs of interest while equipping them for broad chassis transfer? And what does this tell us about the functional requirements of the BGC?

2 METHODS AND RESULTS

Genome Mining of Bacteroidota

BACKGROUND

To this day, the massive and continuously growing volume of sequencing data, combined with increasingly advanced artificial intelligence (AI) tools, continues to fuel genome mining technologies, such as the discovery of NPs. These genome mining approaches remain highly relevant for unlocking the biosynthetic potential of microorganisms, as demonstrated throughout the thesis. Here we will focus on the phylum Bacteroidota, highlighting several results describing genomes from newly isolated marine Bacteroidota genera: *Sinomicrobium* (2.2), *Algoriphagus/Roseivirga* (2.3), *Galbibacter* (2.4) but also 21 *Pedobacter* strains of amphibian origin (2.5). These studies showcase the untapped biosynthetic potential of this phylum, especially in the genus *Pedobacter*. It also showcases how genome-guided approaches can be used to prioritise strains for natural product discovery. Building on these findings, the chapter 2.6 links BGCs to their respective NP. To run this analysis, a Linux-based computational infrastructure was established (Chapter 2.1). Here, bioinformatic tools were implemented to filter and select high-quality genome assemblies from large datasets, assign taxonomic affiliation automatically, detect BGCs, and perform a comparative analysis of the generated data. The overall workflow was also applied to establish a genome sequence handling workflow for strains from the Fraunhofer Biobank, to guide industrial small molecule discovery projects and promote the attractiveness of the strain collection for industrial partners.

2.1 Implementation of command line tools for computational genomics analysis

2.1.1 Aims

To apply the above-mentioned tools in a reproducible manner, I established a modular workflow assessing genome assemblies' quality and biosynthetic potential. The local bioinformatic workflow includes the following intermediate steps:

Environment setup: A Linux virtual machine (VM) was set up, and essential dependencies, including Biopython, Docker and Conda, were installed to ensure the functionality of downstream genome-processing tools:

1. CheckM2 - Used to evaluate genome completeness and contamination. Output: Percentage estimates for completeness and contamination of a given genome sequence.
2. QUAST – Applied to calculate classical assembly metrics such as N50, L50, GC content. Output: Table listing calculated metric of a given genome sequence
3. GTDB - Used for taxonomic affiliation based on whole genome marker genes. Output: Hierarchical taxonomic affiliations of a genome assembly.
4. OrthoANI – Performed pairwise genome alignments to identify redundant genome assemblies. Output: Average Nucleotide Identity (ANI) between to genome assemblies in percent.
5. antiSMASH - Predicted BGCs. Output: Predicted BGCs and an interactive HTML report.
6. BiG-SCAPE - Calculated similarities between given and publicly available BGCs to cluster them in Gene Cluster Families (GCFs). Output: Network files and HTML visualisations.
7. Corason – Used for comparative analysis of orthologous genes across related BGCs. Output: Gene cluster alignments (SVG) and evolutionary trees.

2.1.2 Results & Protocols

VM Access

The Linux VM can be accessed from a Windows system:

```
Windows→ Command Prompt

ssh celine@10.149.20.143
```

You will be asked to enter the password.

System Setup

The Linux VM runs on Ubuntu 22.04 and includes an external mount point for data storage. To mount the external storage:

```
sudo mount /dev/sdb ~/mount_point
```

Caution: The VM is not automatically backed up. Please ensure you manually snapshots.

File renaming

File names, including spaces, can break command line tools. To automatically remove spaces from a large set of genome files, use the following in Windows command prompt.

Open the Windows command prompt and navigate to your folder of interest by:

```
cd I:\Giessen\Projekte\Fraunhofer_NP\Sequencing
```

use

```
get-childitem *.gbk | foreach {rename-item $_ $_.name.replace(" ","_")}
```

→ *.gbk defines the format type that is selected

→ The first character within the quotation marks defines the character to be replaced (here space)

→ The second character within quotation marks defines the replacement character (here underscore)

Data Transfer

The current firewall and server settings allow a data transfer from the I:\ and Y:\ drivers only indirectly using scp. Example command:

```
scp -r DataSource DesiredLocation

#Example

scp -r I:\Giessen\Projekte\Fraunhofer_PhD\CheckM2Input celine@10.149.20.143:/home/celine/bin
```

Once transferred, large datasets should be moved to the mount point:

```
sudo mv CheckM2Input ~/mount_point
```

Conda Environments

Some genome analysis tools are preinstalled as Conda environments. To view available environments:

```
conda info –envs
```

Docker Tools

Some tools are run as docker containers; to list them and identify the installed version:

```
docker images
```

The following sections will introduce each bioinformatic tool, summarising the background context, required inputs, example commands and expected outputs. **Error! Reference source not found.** shows an overview of t he complete workflow - from quality control to optional visualisation tools (not discussed in this thesis).

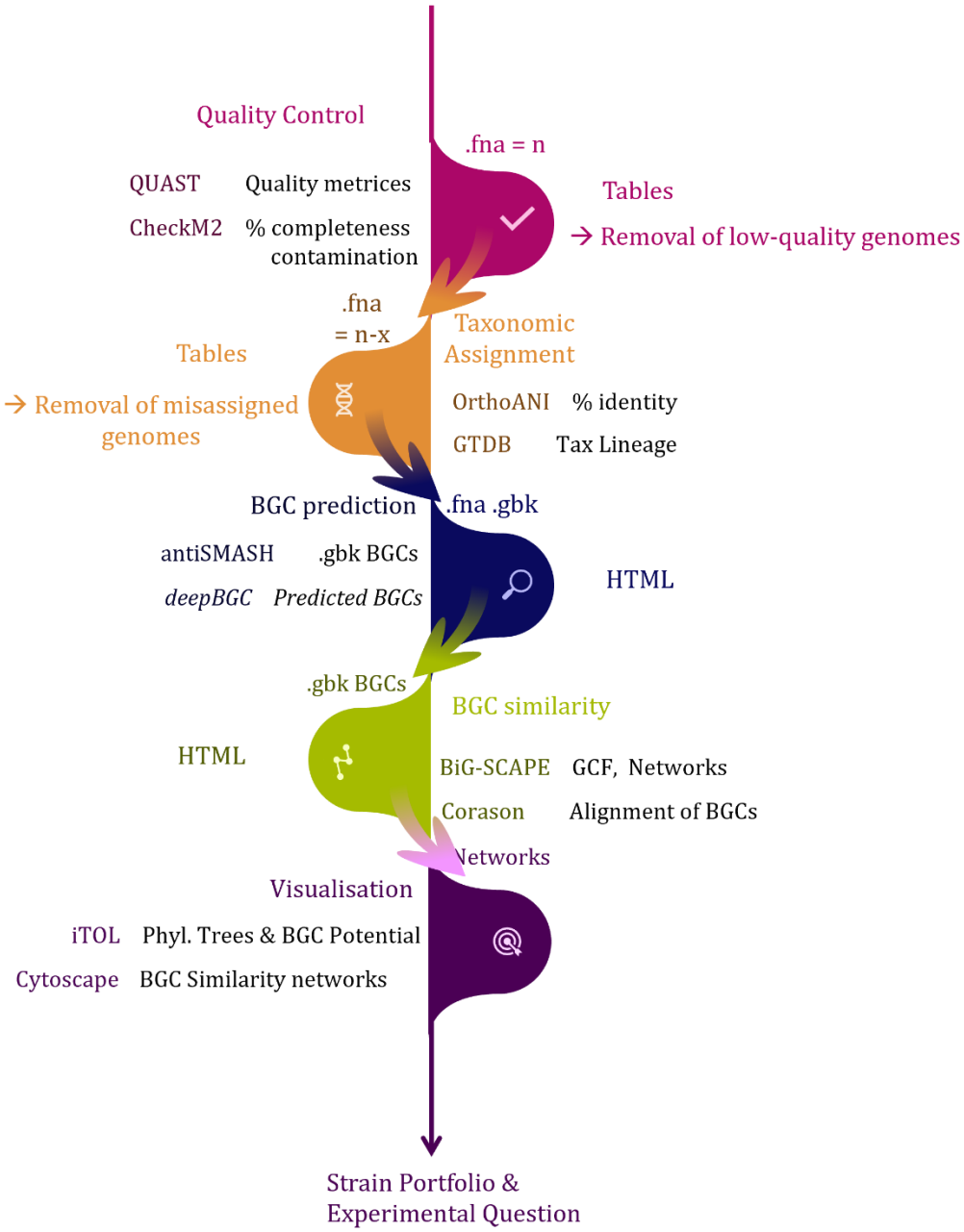

Figure 1: Overview of the bioinformatic workflow used in this study, illustrating the different analytical steps from top to bottom: Quality Control (pink), Taxonomic Assignment (orange), BGC prediction (blue), BGC analysis (green) and visualisation (purple). Key tools associated to each function are highlighted in the corresponding colours along with required input formats and generated output types.

2.1.2.1 CheckM2

CheckM2 (Chklovski *et al.*, 2023a) is the successor to the original CheckM tool (Parks *et al.*, 2015a) and was developed to improve genome quality estimation for lesser-known taxa and metagenome-assembled genomes (MAGs). While CheckM used the absence and presence of lineage-specific genetic markers to estimate completeness and contamination of genome assemblies, CheckM2 implements a machine-learning approach that also considers the number of coding sequences, the codon usage, the GC content, the k-mer frequencies and taxonomic estimates. By doing so, CheckM2 allows genome quality assessment across a wider taxonomic range.

→ Input: Assembled FASTA sequences (.fasta, .fa, .fna)

→ Usage: CheckM2 runs as a Conda environment. For activation:

```
cd ~/mount-point/software/checkm2 # directory of environment

conda activate checkm2

cd ~/mount-point

bin/checkm2 predict --threads No.ofThreads --input InputDir --output-directory OutputDir

#Example

bin/checkm2 predict --threads 30 --input ~/mount_point/CheckM2Input --output-directory
/home/celine/bin/CheckM2Output
```

→ Output: CheckM2 generates the following directories:

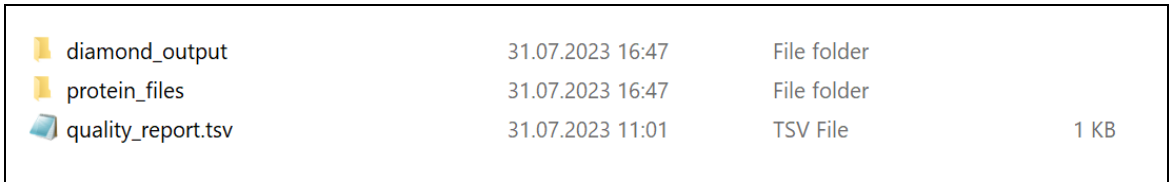

Figure 2: Screenshot of an output directory generated by CheckM2.

- Diamond_output: Summarises sequence similarities to reference sequences.
- Protein_files: Translated amino acid sequences (.faa files).
- Quality_report.tsv: Final summary presenting contamination and completeness values.

The quality_report is built up like this:

Table 2: Exemplary Output Table as generated by CheckM2

Name	Completeness	Contamination	Modell used	Translation_Table	Additional_ Notes
FH1421	100.0	0.25	Neural Network (Specific Model)	11	None
HAG010016	100.0	0.64	Neural Network (Specific Model)	11	None

→ Interpretation

CheckM2-evaluated genomes can be assigned according to standardised MIMAG (Bowers *et al.*, 2017), Minimum Information about a Metagenome-Assembled Genome) quality groups:

- High-quality genomes are defined by completeness of over 90% and contamination below 5%.
- Medium-quality genomes have a considerably lower completeness that is supposed to be over 50%, with a contamination value of less than 10%.
- Everything that falls below this standard is considered to be a low-quality draft. For our analyses, we only selected high-quality genomes to have a comprehensive and high likelihood of covering most of the genetic potential.

For further processing of genomes, we only selected high-quality genomes to ensure robust analysis of biosynthetic potential.

2.1.2.2 QUAST

QUAST (Quality Assessment Tool for Genome Assemblies) is a Python-based script used to compute quality metrics such as N50, genome size, number of contigs and GC content for user-submitted genomes. It also generates reports and graphs to easily compare genome assemblies across a dataset.

→ Input: Assembled genome sequences in FASTA format (.fasta, .fa, .fna), or FASTQ files for read mapping evaluations.

→ Usage: The tool and its dependencies are located at:

```
~/bin/quast-5.2.0
```

To run QUAST on a folder of genome sequences:

```
./quast.py ~/mount_point/genomes/QUASTInput/*.fasta -o ~/mount_point/QUAST_output
```

Help and available options/parameters are shown with:

```
./quast.py --help
```

→ Output: Quast generates the following output directory:

Figure 3: Screenshot of output directories generated by QUAST.

- Report files provide sequencing metrics in different formats for further processing.
- Basic_stats contains visualisation plots as PDF files.
- HTML files show interactive visualisations of contig structure and quality metrics.

Generated outputs can identify genome assemblies of poor quality. Such assemblies will be removed from further automated analysis.

2.1.2.3 GTDB-Tk

Once assemblies passed the quality checks with CheckM2 and QUAST, taxonomic classification was performed using GTDB-Tk (Genome Taxonomy Database Toolkit). GTDB-Tk allows consistent and accurate taxonomic classifications for user-submitted bacterial and archaeal genomes based on genome-wide comparisons rather than specific marker-based approaches. This enables robust classification for novel and under-characterised taxa.

The tool is particularly suited for large genome datasets because the generated outputs are compatible with downstream bioinformatic tools. In our setup, it allowed the issue of publicly available but misannotated genomes, which can skew comparative analyses if not properly curated. For this reason, not only internal, new assemblies but also reference genomes were reclassified and evaluated using GTDB-Tk.

→ Input: Assembled FASTA sequences (.fna).

All genomes must be placed collectively in:

`~/mount_point/GTDB_IO/GTDB-InputDirectory`

→ Usage: GTDB-Tk runs as a docker container.

The tool runs with:

```
docker run -v ~/mount_point/GTDB_IO:/data -v ~/mount_point/GTDB/release207_v2:/refdata ecogenomic/gtdbtk classify_wf --skip_ani_screen --genome_dir data/GTDB_Input --out_dir /data/GTDB_out --cpus 30
```

→ Output: GTDB-Tk generates the following output:

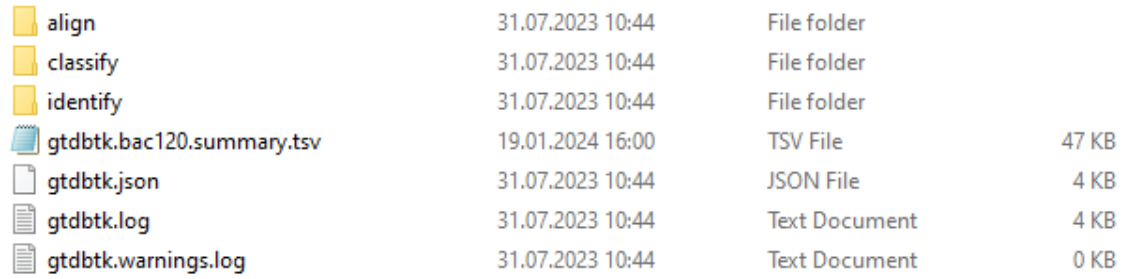

Figure 4: Screenshot of output directories generated by GTDB-Tk.

- Align directory: contains multiple sequence alignment with included reference genomes (gtdbtk.bac120.msa.fasta.gz) and without them (gtdbtk.bac120.user_msa.fasta.gz). These files can be used to calculate phylogenetic trees for visualisation in iTOL.
- Classify /Identify: contain intermediate results that were not used in our workflow.
- The tab-separated file gtdbtk.bac120.summary.tsv contains the final taxonomic assignments used for further processing.

→ Interpretation: The main classification result for further processing appears in the second column of the TSV file:

Table 3: Exemplary Output from GTDB-Tk

user_genome	**classification**
FH1421	d__Bacteria;p__Actinobacteriota;c__Actinomycetia;o__Streptomycetales;f__Streptomycetaceae;g__Streptomyces;s__
HAG010016	d__Bacteria;p__Actinobacteriota;c__Actinomycetia;o__Streptomycetales;f__Streptomycetaceae;g__Streptomyces;s__Streptomyces rimosus

Here, the determined taxonomic rank is summarised and separated by different prefixes:

- d__= domain.
- p__= phylum.
- c__ = class until s__ = species (empty if not assigned).

There are further 17 columns including detailed metadata like:

- Closest reference genome.
- Relative Evolutionary Divergence (RED) value indicating relative evolutionary distance to the nearest reference.

2.1.2.4 OrthoANI

During automatic processing, several public genomes appeared to be duplicates. To mitigate redundancy, we used OrthoANI to compare pairwise nucleotide similarity in our taxonomically classified genomes. Besides removing redundant genomes, this metric defines species boundaries, an information usually required to publish newly assembled genomes. OrthoANI (Lee *et al.*, 2016a) is a command-line tool implemented in a Java Runtime Environment to calculate the Average Nucleotide identity (ANI) between two genomes. The ANI is calculated by computationally splitting the genome into smaller fragments that are subsequently aligned. Alignments above a certain threshold are summarised, and the mean of these values represents the final ANI result. ANI values ≥95% commonly indicate same-species affiliation (Eichorst *et al.*, 2018).

→ Input: Assembled FASTA sequences (.fasta, .fa, .fna)

→ Usage: Desired input needs to be stored in celine@10.149.20.143:/home/celine/bin

It runs with:

```
java -jar OAU.jar -u /usr/bin/usearch11.0.667_i86linux32 -fd /home/celine/bin/InputOrthoANI -o /home/celine/bin/OrthoANI/tempOut/OrthoANI_Out
```

→ Output: It generates a single TXT file assigning numbers to the user-submitted genomes and provides OrthoANI values as matrix comparing all genomes against each other.

2.1.2.5 antiSMASH

After quality control and taxonomic filtering, antiSMASH was used to predict the biosynthetic potential present in the curated genome dataset. AntiSMASH is widely considered the gold standard for predicting, annotating and analysing bacterial and fungal BGCs from annotated user-provided genomes. It supports the identification of diverse cluster types and integrates other databases, such as MIBiG (Zdouc *et al.*, 2025), enabling comparative analysis of clusters, domain architecture and structural features. The generated output can serve as an input for tools like BiG-SCAPE, CORASON (Navarro-Muñoz *et al.*, 2020a) and NaPDOS (Ziemert *et al.*, 2012).

→ Input:

- Genome Assemblies as GenBank (.gbk, .gbff)
  or
- As FASTA (.fna. fasta .fa) with a corresponding annotation file (.gff) or if a gene prediction tool is specified by --genefinding-tool prodigal.

→ Usage:

antiSMASH runs as a docker container. Input genomes should be stored in ~/mount_point/genomes

If FASTA files are provided, it runs with:

```
run_antismash /home/celine/mount_point/genomes/Input.fna
/home/celine/mount_point/AntiSmashOutput_tmp_windows --genefinding-tool prodigal
```

Note: antiSMASH only accepts single genome files as input. To analyse multiple genomes, the commands should be chained using semicolons (;).

To view available options and help:

```
run_antismash . . –-help
```

Common analyses that are included in the web-based tool have to be specifically asked for and include:

Table 4: Exemplary options commonly used for antiSMASH

Flag	Analysis
--cassis	Motif-based prediction of NP gene cluster regions.
--cc-mibig	Comparison with the MIBiG dataset
--cb-general	Comparison of identified clusters with a database of antiSMASH-predicted clusters.
--cb-subclusters	Comparison of identified clusters with known subclusters responsible for synthesising precursors.

→ Output: antiSMASH generates the following output:

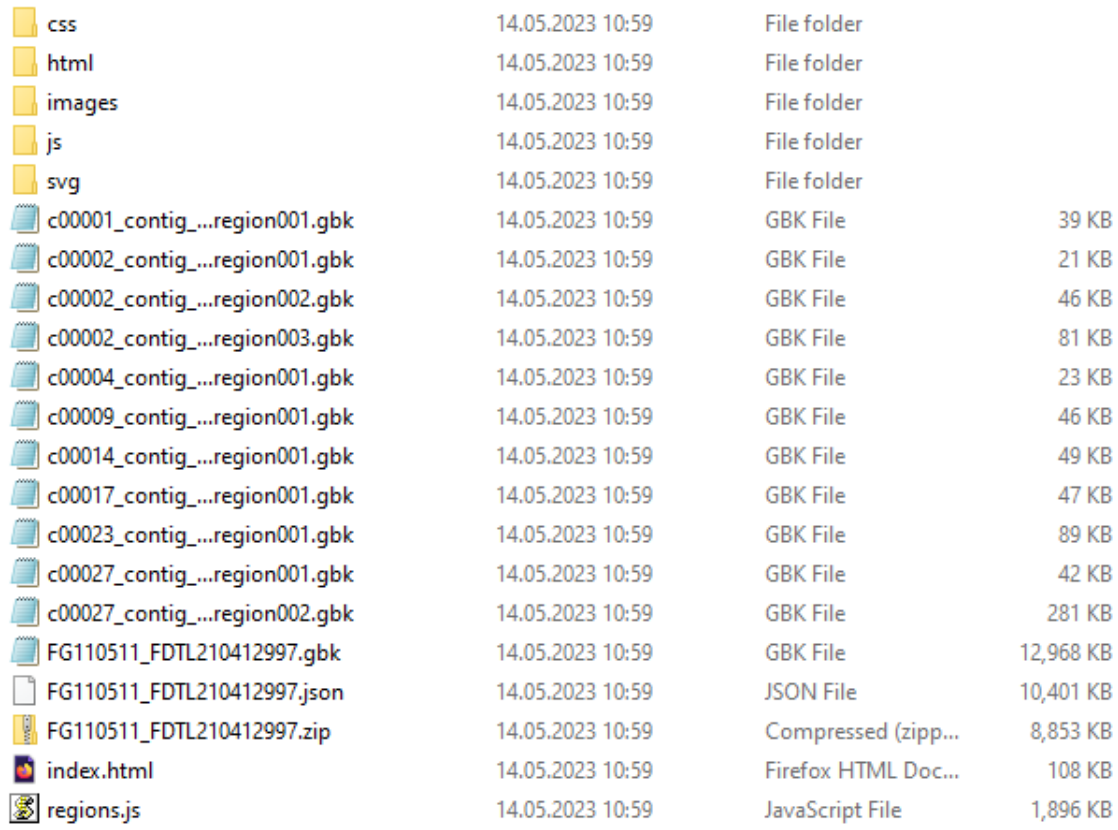

Figure 5: Screenshot of output directories generated by antiSMASH.

Key components to highlight include:

- Index.html: An interactive, graphical representation of the predicted BGCs and requested analyses.
- GenBank (.gbk): Extracted BGCs that can be used for further bioinformatic analyses.
- JSON files: Encoded cluster information for visualisation and downstream processing.

→ Python Script for a summary of BGC numbers and types.

To summarise and compare the predicted number and types of BGCs in the different Input genomes, a Python script was written (individualClusterCounter-CMZ.py) to automatically extract this information from user-defined antiSMASH output folders (SI 1:individualClusterCounter-CMZ.py).

To run it, use

```
sudo python3 individualClusterCounter-CMZ.py ~/DirectoryToantiSMASHOutputFolder OutputFileName
```

within the directory of the Python script. It generates an output table (space separated) listing each folder (strain) name and the number of clusters per BGC types RiPPs, NRPS, terpene, PKS, hybrid, and others.

Table 5: Exemplary Output from individualClusterCounter.py

Folder	**RiPP**	**NRPS**	**terpene**	**PKS**	**hybrid**	**other**
Input Folder A	5	1	3	1	1	0
Input Folder B	3	2	3	1	1	1
Input Folder C	2	1	1	1	0	0

If the script runs successfully, it also prints the number of folders analysed and those that contained the required input "index.html" SI 1:individualClusterCounter-CMZ.pySI 1:individualClusterCounter-CMZ.py).

2.1.2.6 BiG-SCAPE

For comparative analysis of clusters and for identification of already known BGCs, the annotated BGC outputs from antiSMASH were passed to BiG-SCAPE (Biosynthetic Gene Similarity Clustering and Prospecting Engine). BiG-SCAPE calculates and visualises the similarity of BGCs and groups them into gene cluster families (GCFs). The similarity is calculated based on domain similarity and organisation, gene content similarity (Jaccard Index) and sequence similarity between homologous genes. Upon request, it incorporates the MiBIG database to highlight similarity to known BGCs and potentially produced NPs.

→ Input:

- antiSMASH-detected BGCS in GenBank format (.gbk).
  Note: Pre-processing and renaming required for internal genomes.

GenBank file preprocessing is usually required when analysing internal genomes, because they often lack clear strain identifiers in their antiSMASH output file name. By default, antiSMASH assigns numbered contig names to-predicted BGCs (e.g. contig_001.region001.gbk). This leads to identical filenames in different strains. BiG-SCAPE only processes the first occurrence of this file name and skips the rest of the predicted BGCs. This is not clearly reported, resulting in a running, but incomplete analysis. To assign unique filenames to the input file, the strain identifier has to be added as prefix. The Python script RenameFilesWithDirectory.py (SI **2**: RenameFilesWithDirectory.py) can process .gbk files automatically:

```
sudo python3 RenameFileswithDirectory.py ~/mount_point/AntiSmashOutput_tmp_windows
```

This adds the directory name (usually the strain ID or GenBank number) to each .gbk file present. When run successfully, it lists the .gbk files that have been renamed.

→ Usage:

BiG-SCAPE was installed as a Docker container. The current version is: 1.1.2. To run it, use:

```
~/bin/run_bigscape InputDirectorycontainingantiSMASHoutput BigScapeOutputFolder
```

Various parameters can be changed. To view them, use:

```
run_bigscape . . –help
```

Commonly used parameters:

Table 6: Exemplary options for BiG-SCAPE

--mix	**An additional analysis across all BGC classes is calculated**
--cutoffs values --cutoffs 0.5 0.6 0.7	Set similarity thresholds for network generation. Thresholds from 0.0-1 can be selected. Usually, cutoffs between 0.5 and 0.7 were used.
--mibig	Includes reference clusters from MiBIG in analysis.

→ Output:

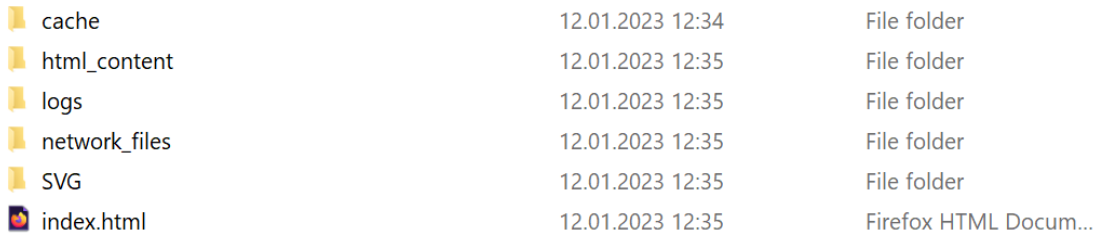

Figure 6: Screenshot of the BiG-SCAPE-generated output directory.

BiG-SCAPE generates five folders and an HTML file. The key components are:

- Network_files: contains network files and TSV annotation files for visualisation in Cytoscape.
- Index.html: HTML report visualising the similarity networks and GCFs interactively.

2.1.2.7 Corason

For selected clusters of interest, understanding their evolutionary relationship, and the presence and absence of tailoring genes can provide insights into the chemical diversity of the natural products and any corresponding structural variations. CORASON (CORe Analysis of Syntenic Orthologs to prioritize Natural products biosynthetic gene clusters) was co-published with BiGSCAPE to support detailed comparative analysis of gene clusters. It identifies core biosynthetic genes, aligns them across different BGCs and infers phylogenetic trees based on sequence similarity and gene synteny. The visualisation output colour codes homologous genes in the same colour, thereby highlighting more conserved core regions and variable accessory genes. This suggests structural diversification of the encoded NP in different strains.

→ Input: A Directory containing

- A “gbks” named folder containing antiSMASH-predicted BGCS.
- A core query gene as an amino acid FASTA file.

Note: The query gene must stem from one BGC in the gbks folder that will later be used as reference BGC.

→ Usage:

Corason is also run as a Docker container.

It runs with:

```
~/bin/run_corason QueryGene.fasta gbks gbks/ReferenceClusterOfQueryGene.gbk -g
```

- "QueryGene.fasta": defines the reference protein.
- "gbks": defines Corason's mode.
  Note: Besides the gbks mode, it also provides a "genome" that first scans genomes or larger contigs for BGCs and query genes without the requirement of analysed antiSMASH clusters. To use this, the user has to create a folder called "genome" containing sequence files such as GenBank files.
- "gbks/ReferenceClusterOfQueryGene.gbk": defines the BGC the query gene belongs to.

→ Output:

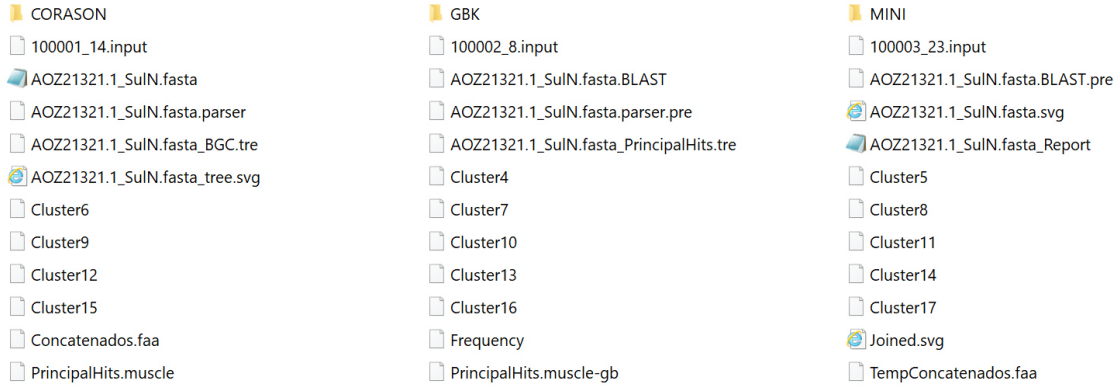

Figure 7: Screenshot of the Output directory generated by Corason

Relevant output files of Corason include:

- "joined.svg": shows the calculated phylogenetic tree and the distribution of identified genes in the different BGCs.
- "_Report" file: contains a summary of the run, including identified orthologs and parameters.

2.2 Draft Genome Sequence of *Sinomicrobium* sp. PAP.21, Isolated from a Coast Sample of Papua, Indonesia

Parts of this chapter have been published as an article in Microbiology Resource Announcements. The manuscript can be found at: DOI: 10.1128/mra.01268-22

AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequence of *Sinomicrobium* sp. Strain PAP.21, Isolated from a Coast Sample of Papua, Indonesia

Riyanti,[a] Celine M. Zumkeller,[b] Marius Spohn,[b,e] Sanja Mihajlovic,[b] Oliver Schwengers,[c,d] Alexander Goesmann,[c,d] Riviani Riviani,[a] Maria D. N. Meinita,[a] Till F. Schäberle,[b,d,e] Harwoko Harwoko[f]

[a]Faculty of Fisheries and Marine Science, Jenderal Soedirman University, Purwokerto, Indonesia
[b]Branch for Bioresources, Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Giessen, Germany
[c]Bioinformatics and Systems Biology, Justus Liebig University Giessen, Giessen, Germany
[d]German Center for Infection Research (DZIF), Partner Site Giessen-Marburg-Langen, Giessen, Germany
[e]Institute for Insect Biotechnology, Justus Liebig University Giessen, Giessen, Germany
[f]Department of Pharmacy, Faculty of Health Sciences, Jenderal Soedirman University, Purwokerto, Indonesia

Riyanti and Celine M. Zumkeller contributed equally to this work. Author order was determined both alphabetically and in order of increasing seniority.

ABSTRACT *Sinomicrobium* sp. strain PAP.21 (EXT111902) was isolated from the coast of Cenderawasih Bay National Park in West Papua, Indonesia. Its genome was assembled into 151 contigs with a total size of 5.439 Mbp, enabling the prediction of its specialized metabolite production capacity.

The phylum *Bacteroidetes* is a promising, although underexplored, bioresource for natural product discovery (1, 2). *Bacteroidetes* are among the most abundant bacteria within marine ecosystems (3), and their known bioactive natural products represent a remarkable diversity. This diversity is, e.g., exemplified by the lanthipeptide pinensin from *Chitinophaga pinensis* exhibiting antifungal activity (4). Expecting a positive correlation between the *Bacteroidetes* diversity and their produced chemical diversity, we aimed at accessing new strains.

In this study, *Sinomicrobium* sp. PAP.21 was isolated from marine sediment of the upper layer (5 to 10 cm) collected at a coast area of Cenderawasih Bay National Park, Papua, Indonesia (2°23′06.8″S 134°57′53.5″E). The strain was deposited in the Fraunhofer strain collection (5) under its identifier EXT111902.

The marine sediment sample was stored in a sterile plastic bag before being plated onto artificial seawater (1.5% agar, 0.01% KBr, 2.3% NaCl, 1.1% $MgCl_2 \cdot 6\ H_2O$, 0.1% $CaCl_2 \cdot 2H_2O$, 0.1% KCl, 0.004% $SrCl_2 \cdot 6\ H_2O$, 0.4% Na_2SO_4, 0.02% $NaHCO_3$, and 0.003% H_3BO_3 in H_2O), including *Escherichia coli* prey, and incubated at 30°C. Grown colonies were purified on marine agar (Carl Roth GmbH; product no. CP73.1 + 1.5% agar), and a sequencing sample was prepared by growing EXT111902 aerobically in marine broth (30°C for 24 h; Carl Roth GmbH; product no. CP73.1). The cell pellet was resuspended in ATL buffer (Qiagen) containing RNase A. BashingBead lysis tubes (Zymo Research) were used for cell disruptions. DNA was isolated using QIAmp 96 DNA QIAcube high-throughput (HT) kits with the addition of proteinase K (Qiagen). Libraries for short-read sequencing were prepared using the Illumina DNA prep tagmentation kit with 500 ng DNA input and 5 cycles of indexing PCR. Library quality was evaluated (Agilent 2100 bioanalyzer) and sequenced on an Illumina NovaSeq instrument using a NovaSeq 6000 SP v1 sequencing kit with a 2 × 150 bp read length and a depth of 4.0 to 5.0 Mio reads. For sequence processing and analysis, software tools were run with default settings unless otherwise stated. The sequencing was demultiplexed (Illumina bcl2fastq, v2.19.0.316), quality checked (Fastp [6] v0.20.1), and visualized (MultiQC [7] v1.7). A total of 11.98 million paired-end reads were quality filtered (Fastp [6] v0.20.1; additional parameter "–detect_adapter_for_pe –cut_by_quality5 –cut_by_quality3

Editor Frank J. Stewart, Montana State University

Address correspondence to Till F. Schäberle, till.schaeberle@ime.fraunhofer.de, or Harwoko Harwoko, harwoko.unsoed@gmail.com.

The authors declare no conflict of interest.

Received 6 December 2022
Accepted 22 February 2023
Published 21 March 2023

2.2.1 Background

The phylum Bacteroidetes is a promising, though underexplored bioresource for natural product discovery (Brinkmann *et al.*, 2022e, 2021). Bacteroidetes are among the most abundant bacteria within marine ecosystems (Fernández-Gómez *et al.*, 2013) and their known bioactive natural products represent a remarkable diversity. This is e.g. exemplified by the lanthipeptide pinensin from *Chitinophaga pinensis* exhibiting antifungal activity (Mohr *et al.*, 2015). Expecting a positive correlation between the Bacteroidetes diversity and their produced chemical diversity, we aimed at accessing new strains.

2.2.2 Methods and Results

2.2.2.1 Isolation and Sequencing

In this study, *Sinomicrobium* sp. PAP.21 was isolated from marine sediment of the upper layer (5-10 cm) collected at coast area of Cenderawasih Bay National Park, Papua, Indonesia (2°23'06.8"S 134°57'53.5"E). The strain was deposited in the Fraunhofer strain collection (Fox, 2014) under its identifier EXT111902.

A sequencing sample was prepared by growing EXT111902 aerobically in marine broth (30°C; 24 h; Carl Roth GmbH; Product No.: CP73.1). The cell pellet was resuspended in ATL buffer (Qiagen) containing RNAse A. ZR BashingBead Lysis Tubes (Zymo Research) were used for cell disruptions. DNA was isolated using QIAmp 96 DNA QIAcube HT kits with addition of proteinase K (Qiagen). Libraries for short-read sequencing were prepared using the Illumina DNA Prep Tagmentation kit with 500 ng DNA input and 5 cycles indexing PCR. Library quality was evaluated (Agilent 2100 Bioanalyzer) and sequenced at an Illumina NovaSeq using a NovaSeq 6000 SP v1 sequencing kit with 2x150 bp read length and a depth of 4.0–5.0 Mio reads. For sequence processing and analysis, software tools were run with default settings unless otherwise stated. The sequencing was demultiplexed (Illumina bcl2fastq, v2.19.0.316), quality-checked (Fastp (Chen *et al.*, 2018) v0.20.1) and visualized (MultiQC (Ewels *et al.*, 2016) v1.7). 11.98M paired-end reads were quality-filtered (Fastp (Chen *et al.*, 2018) v0.20.1; additional parameter: "--detect_adapter_for_pe --cut_by_quality5 --cut_by_quality3 --low_complexity_filter --length_required 21 --correction"), assembled (Unicycler (Wick *et al.*, 2017) v0.4.8) and annotated (Bakta (Schwengers *et al.*, 2021.) v1.5.1).

2.2.2.2 Computational Analysis

The genome consists of 5,438,544 bp (Coverage 320x; N50 141,686; L50 13) in 151 contigs and has a GC content of 44.4%. Using CheckM (v1.0.18) (Parks *et al.*, 2015b) the degree of genome completeness was determined at 99.34% with 3.9% contamination. The genome encodes 4,679 protein-coding genes, 47 tRNAs, 1 tmRNAs, 3 rRNAs and 6 ncRNAs. The taxonomical rank was established using the Type Strain Genome Server (Meier-Kolthoff and Göker, 2019a). This revealed *Sinomicrobium oceani* CGMCC 1.12145 (Xu *et al.*, n.d.) as closest related type strain. Digital DNA-DNA hybridization (dDDH) values exceed the species delineation threshold of 70% (76.9% (d0), 92.1% (d4) and 82.3% (d6)). An average nucleotide identity (ANI) (Lee *et al.*, 2016b) value of 98.89% supports affiliation of EXT111902 to the species *S. oceani*. antiSMASH v6.0 (Blin *et al.*, 2021a) was employed to predict the BGCs. BGCs annotation was achieved by their clustering with MIBiG (Kautsar *et al.*, 2020) reference clusters into gene cluster families (GCFs) using BiG-SCAPE (Navarro-Muñoz *et al.*, 2020b) setting a cutoff value of 0.6.

2.2.2.3 Biosynthetic Gene Cluster Prediction

EXT111902 carries one BGC clustering with BGC0001478, encoding the synthesis of the siderophore desferrioxamine E (Barona-Gómez *et al.*, 2004) and another BGC was annotated to BGC0000593, encoding microviridin J (Ziemert *et al.*, 2008a). The third database annotation refers to pinensin (BGC0001392) (Mohr *et al.*, 2015). Pinensin-like BGCs were previously detected in genomes of certain *Chitinophaga*, *Chryseobacterium, Elizabethkingia, Pedobacter*, and *Sinomicrobium* strains (Caetano *et al.*, 2020a). Variations in the amino acid sequence of the core peptides indicate a yet undiscovered structural diversity within this lantipeptide-type. Such alterations towards the known pinensins are also predicted for EXT111902 (Figure 8). Comparative alignment of the pinensin-type cluster in EXT111902 with that of *Chitinophaga pinensis* DSM2588 (BGC0001392) demonstrated the presence of homologous core biosynthetic genes (*pinA–C* and *pinT*), including the characteristic split dehydratase genes (*pinB1* and *pinB2*). Genes corresponding to *Cpin5321–5324*, proposed to mediate pinensin perception and import, were also identified, whereas homologs of *Cpin5330–5332*, implicated in compound export, were absent. The predicted core peptide encoded by the EXT111902 cluster consists of 24 amino acids, compared to the 22-amino-acid sequence of pinensin A.

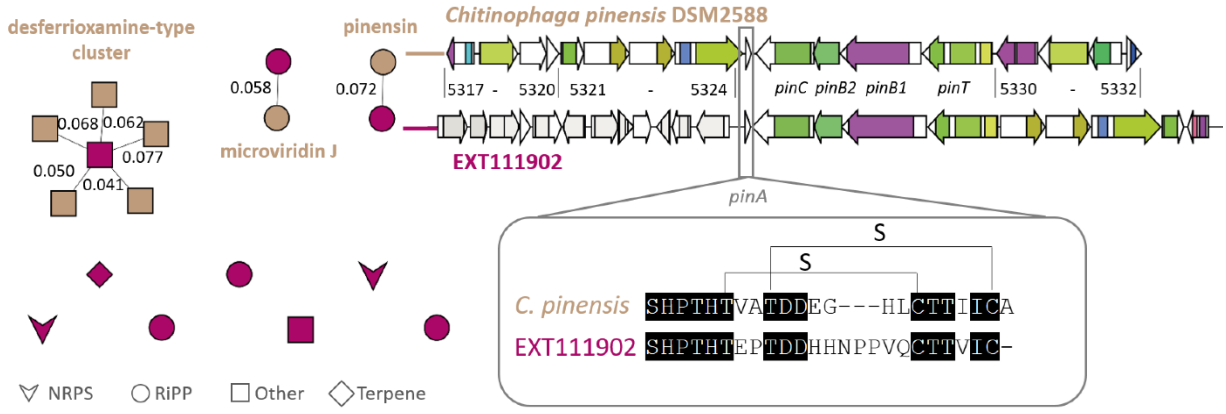

Figure 8: BiG-SCAPE network and comparative analysis of BGCs in *Sinomicrobium sp.* EXT111902. The genome encodes five RiPP-, two NRPS-, one terpene-, and two additional BGCs. Similarity networking with MIBiG references identified clusters related to desferrioxamine (BGC0001478), microviridin (BGC0000593), and pinensin (BGC0001392). Alignment of the pinensin-like cluster revealed conserved core genes (*pinA–C*, *pinT*) and absence of export-related genes (*Cpin5330–5332*), suggesting a modified pinensin variant.

This genome analysis established a high-quality, reproducibly processed assembly for *Sinomicrobium sp.* PAP.21 and expanded the genomic coverage of the genus within the Fraunhofer strain collection. The identification of multiple RiPP and siderophore gene clusters highlights the biosynthetic diversity of the species and provides a foundation for targeted metabolomic follow-up.

2.3 Draft Genome Sequences of *Algoriphagus sp*. PAP.12 and *Roseivirga sp*. PAP.19 Isolated from Marine Samples of Papua, Indonesia

Parts of this chapter have been published as an article in Microbiology Resource Announcements. The manuscript can be found at: DOI: 10.1128/mra.01264-22

AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequences of *Algoriphagus* sp. Strain PAP.12 and *Roseivirga* sp. Strain PAP.19, Isolated from Marine Samples from Papua, Indonesia

Riyanti,[a] Celine M. Zumkeller,[b] Marius Spohn,[b,f] Sanja Mihajlovic,[b] Oliver Schwengers,[c,d] Alexander Goesmann,[c,d] Nur A. Choironi,[e] Till F. Schäberle,[b,d,f] Harwoko Harwoko[e]

[a]Faculty of Fisheries and Marine Science, Jenderal Soedirman University, Purwokerto, Indonesia
[b]Branch for Bioresources, Fraunhofer Institute for Molecular Biology and Applied Ecology, Giessen, Germany
[c]Bioinformatics and Systems Biology, Justus Liebig University Giessen, Giessen, Germany
[d]German Center for Infection Research, Partner Site Giessen-Marburg-Langen, Giessen, Germany
[e]Department of Pharmacy, Faculty of Health Sciences, Jenderal Soedirman University, Purwokerto, Indonesia
[f]Institute for Insect Biotechnology, Justus Liebig University Giessen, Giessen, Germany

Riyanti and Celine M. Zumkeller contributed equally to this work. Author order was determined both alphabetically and in order of increasing seniority.

ABSTRACT *Algoriphagus* sp. strain PAP.12 (EXT111900) and *Roseivirga* sp. strain PAP.19 (EXT111901) were isolated from marine samples. Here, we report their draft genome sequences, 5.032 Mbp and 4.583 Mbp in size, respectively, and rate their specialized metabolite production capacity. Taxonomic ranks established by genome-based analysis indicate that *Algoriphagus* sp. strain PAP.12 represents a candidate new species.

The phylum *Bacteroidetes* is a proliferating but underexplored bioresource for natural product discovery (1, 2). The bacteria colonize diverse habitats and are among the most abundant groups of bacteria within marine ecosystems (3). In an effort to access new *Bacteroidetes* strains, strains PAP.12 and PAP.19 were isolated from Cenderawasih Bay National Park (Papua, Indonesia). The strains were deposited in the Fraunhofer strain collection (4) under the identifiers EXT111900 and EXT111901.

EXT111900 was retrieved from the upper layer (5 to 10 cm) of marine sediment (1°47′45.6″S, 134°25′12.0″E). EXT111901 was isolated from a sponge (1°48′03.6″S, 134° 06′03.6″E) by cutting using a diving knife while latex gloves were worn. The samples were put into sterile bags, brought to the surface to be kept at ambient seawater temperature, and processed within 2 hours after sampling, as follows. Sponge samples were cut into small pieces using a sterile scalpel, and all samples were rinsed with sterile seawater before being plated on artificial seawater (1.5% agar, 0.01% KBr, 2.3% NaCl, 1.1% $MgCl_2 \cdot 6H_2O$, 0.1% $CaCl_2 \cdot 2H_2O$, 0.1% KCl, 0.004% $SrCl_2 \cdot 6\ H_2O$, 0.4% Na_2SO_4, 0.02% $NaHCO_3$, and 0.003% H_3BO_3 in H_2O) with *Escherichia coli* prey and cycloheximide. Bacterial colonies were isolated, and sequencing samples were prepared by growing the strains aerobically for 24 h at 30°C in marine broth (product number CP73.1; Carl Roth GmbH). Cell pellets were resuspended in buffer ATL (Qiagen) containing RNase A. ZR BashingBead lysis tubes (Zymo Research) were used for cell disruption. DNA was isolated using QIAmp 96 DNA QIAcube high-throughput (HT) kits with the addition of proteinase K (Qiagen). Libraries for short-read sequencing were prepared using the Illumina DNA preparation tagmentation kit with 500 ng DNA as input and 5 cycles of indexing PCR. Library quality was evaluated (Agilent 2100 Bioanalyzer), and libraries were sequenced on an Illumina NovaSeq system using a NovaSeq 6000 SP v1 sequencing kit with 2 × 150 bp read length and a depth of 4.0 to 5.0 million reads per sample. For sequence processing and analysis, software tools were run with default settings unless otherwise stated. The sequencing was demultiplexed (Illumina

Editor Frank J. Stewart, Montana State University

Address correspondence to Till F. Schäberle, till.schaeberle@ime.fraunhofer.de, or Harwoko Harwoko, harwoko.unsoed@gmail.com.

The authors declare no conflict of interest.

Received 6 December 2022
Accepted 22 February 2023
Published 16 March 2023

2.3.1 Background

The phylum Bacteroidetes is a proliferating, though underexplored bioresource for natural product discovery (Brinkmann *et al.*, 2022b, 2022e). They colonize diverse habitats and are among the most abundant groups of bacteria within marine ecosystems (Fernández-Gómez *et al.*, 2013). Aimed at accessing new Bacteroidetes, strains PAP.12 and PAP.19 were isolated from Cenderawasih Bay National Park, Papua, Indonesia. The strains were deposited in the Fraunhofer strain collection (Fox, 2014) under their identifier EXT111900 and EXT111901.

2.3.2 Methods and Results

2.3.2.1 Isolation and Sequencing

Bacterial colonies were isolated and sequencing samples were prepared by growing the strains aerobically in marine broth (30°C; 24 h). Cell pellets were resuspended in ATL buffer (Qiagen) containing RNAse A. ZR BashingBead Lysis Tubes (Zymo Research) were used for cell disruption. DNA was isolated using QIAmp 96 DNA QIAcube HT kits with addition of proteinase K (Qiagen). Libraries for short-read sequencing were prepared using the Illumina DNA Prep Tagmentation kit with 500 ng DNA input and 5 cycles indexing PCR. Library quality was evaluated (Agilent 2100 Bioanalyzer) and sequenced at an Illumina NovaSeq using a NovaSeq 6000 SP v1 sequencing kit with 2x150 bp read length and a depth of 4.0–5.0 Mio reads per sample.

2.3.2.2 Computational Analysis

The sequencing was demultiplexed (Illumina bcl2fastq, v2.19.0.316), quality-checked (Fastq, v0.20.1) and visualized (MultiQC, v1.7). Paired-end reads were quality-filtered (Fastp (Chen *et al.*, 2018) v0.20.1), assembled (Unicycler (Wick *et al.*, 2017) v0.4.8), quality-checked (CheckM (Parks *et al.*, 2015b) v1.0.18) and genomes were annotated using Bakta (Schwengers *et al.*, n.d.) (v1.5.1). The genome of strain EXT111900 consists of 5,032,044 bp in 101 contigs (N_{50} 235,524) with a GC content of 39.4% revealing a completeness of 99.81% and contamination of 0.19%. It encodes 4,286 protein-coding genes, 41 tRNAs, 1 tmRNAs, 3 rRNAs and 3 ncRNAs. The genome of EXT111901 consists of 4,582,827 bp in six contigs (N_{50} 3,058,286) with a GC content of 42.1% revealing a completeness of 99.81% and no contamination. It encodes 3,944 protein-coding genes, 47 tRNAs, 1 tmRNAs, 6 rRNAs and 3 ncRNAs.

2.3.2.3 Taxonomic Classification

Taxonomical ranks were established using the Type Strain Genome Server (Meier-Kolthoff and Göker, 2019a), autoMLST (Alanjary *et al.*, 2019) and by determining the average nucleotide identity (ANI) (Lee *et al.*, 2016b). *Roseivirga pacifica* DY53 (Huo *et al.*, 2013) represents the closest related type strain of EXT111901. Digital DNA-DNA hybridization (dDDH) values and an ANI of 96.65%, all exceeding the species delineation threshold support affiliation of EXT111901 to the species of *R. pacifica*. EXT111900 closest related type strain is *Algoriphagus zhangzhouensis* DSM 25035 (Yang *et al.*, 2013). The dDDH values are 71.8% (d0), 35.8% (d4) and 62.9% (d6) and the ANI is 88.84%. Accordingly, EXT111900 possibly represents a candidate new *Algoriphagus* species.

2.3.2.4 Biosynthetic Gene Cluster Prediction

Prediction of BGCs was performed using antiSMASH v6.0 (Blin *et al.*, 2021a). Their grouping into gene cluster families (GCFs) together with the MIBiG (Kautsar *et al.*, 2020) reference clusters using BiG-SCAPE (Navarro-Muñoz *et al.*, 2020b) allowed BGC-similarity determinations. EXT111900 carries one BGC clustering with BGC0000650, encoding the carotenoid flexixanthin (Tao *et al.*, 2006). Albeit less similar, the same match is detected from EXT111901 (Figure 9). A difference is the presence of a lycopene beta-cyclase in EXT111900 and its absence in EXT111901.

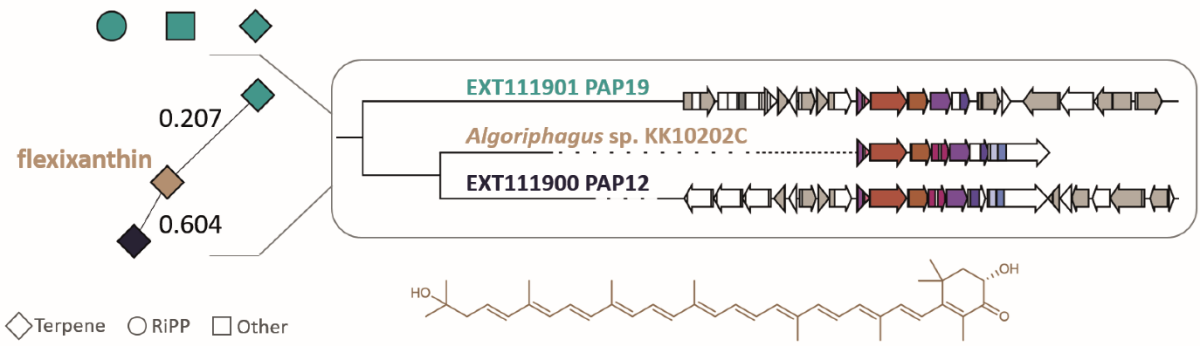

Figure 9: BiG-SCAPE network and comparative analysis of BGCs in *Roseivirga sp*. EXT111901 (turquoise) and *Algoriphagus* sp. EXT111900 (dark blue). Both genomes contain terpene, RiPP, and other BGC types. Terpene clusters from each strain group with the flexixanthin reference (MIBiG BGC0000650) from *Algoriphagus sp*. KK10202C, showing similarity indices of 0.207 and 0.604, respectively. A fusion-type lycopene β-cyclase gene (pink ORF) is present in EXT111900 but absent in EXT111901. The chemical structure of flexixanthin is shown at the bottom.

Both *Algoriphagus sp*. PAP.12 and *Roseivirga sp*. PAP.19 yielded high-quality genome assemblies that passed all workflow quality thresholds. Comparative genomics highlighted the taxonomic novelty of PAP.12 and confirmed the species identity of PAP.19. BGC analysis revealed conserved carotenoid biosynthetic potential, with subtle genomic variations suggesting species-specific metabolic traits. Together, these analyses extend the genomic framework of marine Bacteroidota within the Fraunhofer strain collection and demonstrate the reproducibility of the applied computational pipeline across multiple genera.

2.4 Genome Sequence of *Galbibacter sp*. PAP.153, Isolated from a Marine Sponge of Papua, Indonesia

Parts of this chapter have been published as an article in Microbiology Resource Announcements. The manuscript can be found at: DOI: 10.1128/mra.01297-23

AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

| Bacteriology | Announcement

Draft genome sequence of *Galbibacter* sp. PAP.153, isolated from a marine sponge in Papua, Indonesia

Riyanti ,[1] Celine M. Zumkeller,[2,3] Marius Spohn,[2,3] Sanja Mihajlovic,[2] Oliver Schwengers,[4,5] Alexander Goesmann,[4,5] Riviani Riviani,[1] Maria Dyah Nur Meinita,[1] Dewi Wisudyanti Budi Hastuti,[1] Isnaini Prihatiningsih,[1] Rose Dewi,[1] Till F. Schäberle[2,4,5]

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT *Galbibacter* sp. PAP.153 was isolated from a marine sponge. Here, we report its 4.12 Mbp draft genome sequence and rate its specialized metabolite production capacity with specific focus on the chemotaxonomic marker flexirubin.

KEYWORDS Bacteroidetes, *Galbibacter*, natural products, flexirubin, Indonesia

The phylum Bacteroidetes is a promising bioresource for discovering novel natural products (1–3) and is abundant in ocean ecosystems (4). The strain PAP.153 was isolated from Cenderawasih Bay National Park, Papua, Indonesia (2°23′6.8″S134°57′53.5″E) and deposited as EXT111903 in the Fraunhofer strain collection (5).

EXT111903 was isolated from sponge material retrieved with a diving knife while wearing latex gloves. The sample was placed in a sterile bag, carried to the surface, stored at ambient seawater temperature, and processed within 2 hours. The sponge material was washed with sterile seawater, sliced using a sterile scalpel, and plated on artificial seawater agar plates (1.5% agar, 0.01% KBr, 2.3% NaCl, 1.1% $MgCl_2{\cdot}6H_2O$, 0.1% $CaCl_2{\cdot}2H_2O$, 0.1% KCl, 0.004% $SrCl_2{\cdot}6\ H_2O$, 0.4% Na_2SO_4, 0.02% $NaHCO_3$, and 0.003% H_3BO_3 in H_2O), including *Escherichia coli* prey and cycloheximide. Distinct colonies were isolated, and sequencing samples were prepared by cultivating strains aerobically for 24 h at 30°C in marine broth (product number CP73.1; Carl Roth GmbH). EXT111903, distinguished by its yellow phenotype, was chosen because it was assigned to the Bacteroidetes phylum.

Cells were pelleted and resuspended in ATL buffer (Qiagen) containing RNase A. ZR BashingBead Lysis Tubes (Zymo Research) were used for cell disruption. DNA was isolated using QIAmp 96 DNA QIAcube HT kits with the addition of proteinase K (Qiagen). Libraries for short-read sequencing were prepared using the Illumina DNA Prep Tagmentation kit with 500 ng of DNA input and five cycles of indexing PCR. Library quality was evaluated (Agilent 2100 Bioanalyzer) and sequenced at an Illumina NovaSeq using a NovaSeq 6000 SP v1 sequencing kit with 2 × 150 bp read length and a depth of 4.0–5.0 Mio reads per sample. The sequencing was demultiplexed (Illumina bcl2fastq, v2.19.0.316), quality checked (Fastq, v0.20.1), and visualized (MultiQC, v1.7). Paired-end reads were quality-filtered [Fastp (6) v0.20.1], assembled [Unicycler (7) v0.4.8], annotated [PGAP (8)], and quality-checked [CheckM (9) v1.0.18]. The genome comprising 4.12 Mbp in 105 contigs (coverage: 351-fold and N_{50}: 145,281 bp) with a GC content of 38.6% revealed completeness of 100% and contamination of 0.47%. It encodes 3,664 protein-coding genes, 41 tRNAs, 1 tmRNA, 3 rRNAs, and 7 ncRNAs. The Type Strain Genome Server (10) identified *Galbibacter pacificus* CMA-7T (GCF_029603155.1) as the closest related type strain. High digital DNA-DNA hybridization values of 77.6% (d0), 83.3% (d4), and 81.5% (d6) and an ANI (11) of 98.02% surpassing the species delineation threshold confirm PAP153's association with the species of *G. pacificus*.

Editor Frank J. Stewart, Montana State University, Bozeman, Montana, USA

Address correspondence to Till F. Schäberle, till.f.schaeberle@agrar.uni-giessen.de, or Riyanti , riyanti1907@unsoed.ac.id.

Riyanti and Celine M. Zumkeller contributed equally to this article. Author order was determined both alphabetically and based on seniority.

The authors declare no conflict of interest.

Received 13 March 2024
Accepted 15 April 2024
Published 30 April 2024

2.4.1 Background

The phylum Bacteroidetes is a promising bioresource for discovering novel natural products (Brinkmann *et al.*, 2022b, 2022e; Silva *et al.*, 2022) and is abundant in ocean ecosystems (Fernández-Gómez *et al.*, 2013). The strain PAP.153 was isolated from Cenderawasih Bay National Park, Papua, Indonesia (2°23'6.8"S134°57'53.5" E) and deposited as EXT111903 in the Fraunhofer strain collection (Fox, 2014).

2.4.2 Methods and Results

2.4.2.1 Isolation and Sequencing

Distinct colonies were isolated, and sequencing samples were prepared by cultivating strains aerobically for 24 h at 30°C in marine broth (product number CP73.1; Carl Roth GmbH). EXT111903, distinguished by its yellow phenotype, was chosen because it was assigned to the Bacteroidetes phylum. Cells were pelleted and resuspended in ATL buffer (Qiagen) containing RNAse A. ZR BashingBead Lysis Tubes (Zymo Research) were used for cell disruption. DNA was isolated using QIAmp 96 DNA QIAcube HT kits with the addition of proteinase K (Qiagen). Libraries for short-read sequencing were prepared using the Illumina DNA Prep Tagmentation kit with 500 ng DNA input and 5 cycles indexing PCR. Library quality was evaluated (Agilent 2100 Bioanalyzer) and sequenced at an Illumina NovaSeq using a NovaSeq 6000 SP v1 sequencing kit with 2x150 bp read length and a depth of 4.0–5.0 Mio reads per sample.

2.4.2.2 Computational Analysis

The sequencing was demultiplexed (Illumina bcl2fastq, v2.19.0.316), quality checked (Fastq, v0.20.1), and visualized (MultiQC, v1.7). Paired-end reads were quality-filtered (Fastp (Chen *et al.*, 2018) v0.20.1), assembled (Unicycler (Wick *et al.*, 2017) v0.4.8), annotated (PGAP (Li *et al.*, 2021)), and quality-checked (CheckM (Parks *et al.*, 2015b) v1.0.18). The genome comprising 4.12 Mbp in 105 contigs (Coverage 351-fold, N_{50} 145,281 bps) with a GC content of 38.6%, revealed a completeness of 100% and contamination of 0.47%. It encodes 3664 protein-coding genes, 41 tRNAs, 1 tmRNA, 3rRNAs, and 7 ncRNAs. The Type Strain Genome Server (Meier-Kolthoff and Göker, 2019a) identified *Galbibacter pacificus* CMA-7T (GCF_029603155.1) as the closest related type strain. High Digital DNA-DNA hybridization (dDDH) values of 77.6% (d0), 83.3% (d4), and 81.5% (d6) and an ANI (Lee *et al.*, 2016b) of 98.02% surpassing the species delineation threshold confirm PAP153's association with the species of *G. pacificus*.

2.4.2.3 Biosynthetic Gene Cluster Prediction

antiSMASH v6.0 (Blin *et al.*, 2021a) was used to predict BGCs. *G. pacificus* PAP.153 contains 4 BGCs, two terpene-type, and two aryl polyene-type clusters. BiG-SCAPE analysis, including MiBIG reference BGCs (Kautsar *et al.*, 2020)(Cutoff: 0.6) grouped one BGC with the flexirubin BGC, a chemotaxonomic marker for Bacteroidetes (Schöner *et al.*, 2014) (Figure 10).

2.4.2.4 Colourimetric assay

To experimentally validate the genome-based prediction, a colourimetric assay for flexirubin-type pigments was performed. *Galbibacter sp*. PAP.153 was cultivated in nutrient broth (NB; 0.5% peptone, 0.3% malt extract) at 28 °C for 48 h under aerobic conditions. Cell pellets were harvested, and crude methanolic extracts were prepared by vortexing the biomass in methanol, followed by centrifugation.

An equal volume of 20% (w/v) KOH solution was added to the extract, resulting in a distinct orange-to-red color change characteristic of flexirubin-type pigments. Upon addition of 1 M HCl, the color reverted to pale yellow, confirming the reversible pH-dependent chromophore behaviour typical of flexirubin (Fautz and Reichenbach, 1980). Control assays using methanol alone or extracts from non-pigmented reference strains remained colorless throughout. This experimental validation supported the in silico BGC prediction and verified the active expression of the flexirubin pathway in *Galbibacter sp*. PAP.153.

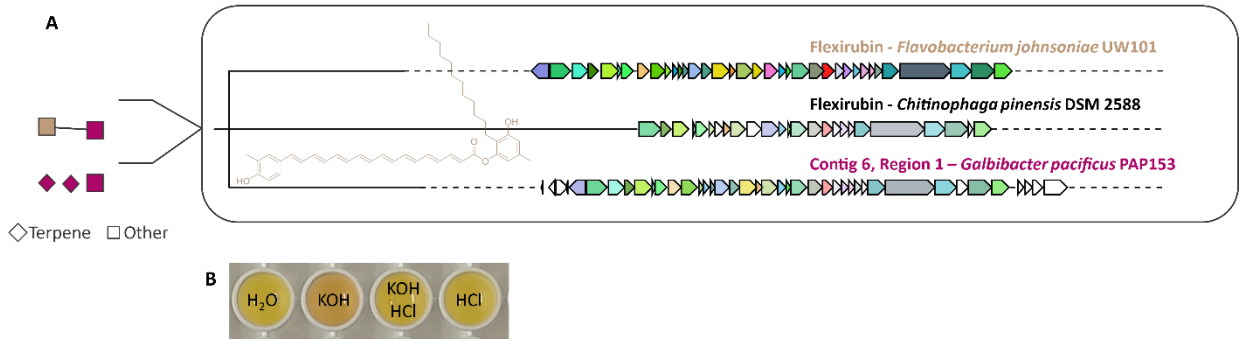

Figure 10: BiG-SCAPE network and validation of flexirubin BGC in *Galbibacter* sp. PAP.153.* **(A)** BiG-SCAPE network (left) showing one BGC (magenta) clustering with the flexirubin reference from Flavobacterium johnsoniae UW101 (similarity index = 0.430). Gene-level alignment with CORASON (right) confirmed a complete flexirubin locus. **(B)** Experimental validation via KOH-induced color change of methanolic extracts demonstrated flexirubin production; the orange coloration reverted to pale yellow upon acidification with HCl.

The genome of *Galbibacter sp*. PAP.153 represents a high-quality assembly with full biosynthetic annotation. Computational genome mining, supported by colourimetric validation, confirmed the presence and activity of the flexirubin biosynthetic pathway—a key chemotaxonomic marker within the Bacteroidota. These findings demonstrate the reliability of the applied analysis workflow and its value for linking genome mining predictions to experimentally verifiable phenotypes in underexplored taxa.

2.5 Genome Sequences of 21 *Pedobacter* strains isolated from amphibian specimens

Parts of this chapter have been published as an article in Microbiology Resource Announcements. The manuscript can be found at: DOI: 10.1128/mra.01185-23

AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

| Environmental Microbiology | Announcement

Draft genome sequences of 21 *Pedobacter* strains isolated from amphibian specimens

Celine M. Zumkeller,[1,2] Molly C. Bletz,[3] Andolalao Rakotoarison,[4] Joana Sabino-Pinto,[5] Silke Reiter,[1,2] Marius Spohn,[1,2] Oliver Schwengers,[6,7] Alexander Goesmann,[6,7] Miguel Vences,[8] Sanja Mihajlovic,[2] Till F. Schäberle[1,2,7]

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT The genomes of 21 *Pedobacter* strains isolated from the European salamander *Salamandra salamandra* and different Madagascan frog species were sequenced using Illumina sequencing. Here, we report their draft genome sequences (~4.7–7.2 Mbp in size) to allow comparative genomics and taxonomic assignment of these strains.

KEYWORDS *Pedobacter*, amphibians, Madagascar, Germany, beta-lactamases, natural products

The bacterial genus *Pedobacter* is widely distributed in many habitats and associated with macroorganisms, including amphibians (1, 2). *Pedobacter* has been identified as part of their cutaneous microbiome (3–5) and has been found to inhibit the growth of pathogenic fungi (3, 4). Certain strains are multidrug resistant and are not susceptible to beta-lactams, colistin, aminoglycosides, and ciprofloxacin (5). This phenotype is supported by a high and diverse number of antibiotic resistance genes detected in the genomes (6). Aimed at characterizing new *Pedobacter* strains, bacteria were isolated from skin swabs of salamanders and frogs from Germany and Madagascar.

For sampling and cultivation of strains with DE and EXT identifiers, see Bletz et al. (4). Briefly, amphibians, captured with clean nitrile gloves, were placed in sterile bags and rinsed with 50 mL of sterilized water before being swabbed. Swabs were stored in Tryptic Soy Yeast Extract + 20% glycerol and kept on ice before transferring to a −20°C freezer. FhG111542 and FhG11526 were isolated as follows: Salamander specimens were gently washed with sterile tap water, and bacteria were collected using sterile cotton swabs. Swabs were stored in sterile tap water at ambient temperature and processed on the same day. To release bacteria, swabs were vortexed, and the resulting bacterial suspensions were plated on agar plates (R2A: HiMedia Laboratories GmbH; Product No.: M1687 and 10% TSB: Thermo Fisher Scientific Inc.; Product No.: CM0129) and incubated at room temperature or 4°C for several days. Colonies were selected based on morphology and subcultured to obtain pure cultures. For sequencing, strains were grown aerobically in NB-medium (0.5% peptone, 0.3% malt extract, and 0.5% NaCl) at 18°C for 24–72 hours. Cell pellets were resuspended in ATL buffer (Qiagen) containing RNAse A. ZR BashingBead Lysis Tubes (Zymo Research) were used for cell disruption. DNA was isolated using QIAmp 96 DNA QIAcube HT Kits with the addition of proteinase K (Qiagen). Libraries for short-read sequencing were prepared using the Illumina DNA Prep Tagmentation Kit with 500 ng DNA input and five cycles indexing PCR. Library quality was evaluated (Agilent 2100 Bioanalyzer) and sequenced on an Illumina NovaSeq using a NovaSeq 6000 SP v1 Sequencing Kit with 2 × 150 bp read length and a depth of 4.0–5.0 million reads per sample. Unless otherwise stated, software tools were run with default settings for sequence processing and analysis. The sequence data were demultiplexed (Illumina bcl2fastq, v2.19.0.316), quality checked (Fastp, v0.20.1),

Editor Elinne Becket, California State University, San Marcos, California, USA

Address correspondence to Sanja Mihajlovic, Sanja.Mihajlovic@ime.fraunhofer.de, or Till F. Schäberle, till.schaeberle@ime.fraunhofer.de.

The authors declare no conflict of interest.

See the funding table on p. 4.

Received 8 December 2023
Accepted 6 February 2024
Published 27 February 2024

2.5.1 Background

The bacterial genus *Pedobacter* is widely distributed in many habitats and associated with macroorganisms, including amphibians (G. Huang *et al.*, 2022; Santibáñez *et al.*, 2022). *Pedobacter* has been identified as part of their cutaneous microbiome (3–5) and has been found to inhibit the growth of pathogenic fungi (Bletz *et al.*, 2017; Lauer *et al.*, 2007). Certain strains are multidrug-resistant and not susceptible to β-lactams, colistin, aminoglycosides, and ciprofloxacin(Ullmann *et al.*, 2020b). This phenotype is supported by a high and diverse number of antibiotic-resistance genes detected in the genomes (Viana *et al.*, 2018b). Aimed at characterizing new *Pedobacter* strains, bacteria were isolated from skin swabs of salamanders and frogs from Germany and Madagascar.

2.5.2 Methods and Results

2.5.2.1 Isolation and Sequencing

For sampling and cultivation of strains with DE and EXT identifiers, see Bletz *et al.* (Bletz *et al.*, 2017). For sequencing, strains were grown aerobically in NB-medium (0.5% Peptone, 0.3% Malt Extract, 0.5% NaCl) at 18°C for 24-72 hours. Cell pellets were resuspended in ATL buffer (Qiagen) containing RNAse A. ZR BashingBead Lysis Tubes (Zymo Research) were used for cell disruption. DNA was isolated using QIAmp 96 DNA QIAcube HT kits with the addition of proteinase K (Qiagen). Libraries for short-read sequencing were prepared using the Illumina DNA Prep Tagmentation kit with 500 ng DNA input and 5 cycles indexing PCR. Library quality was evaluated (Agilent 2100 Bioanalyzer) and sequenced on an Illumina NovaSeq using a NovaSeq 6000 SP v1 sequencing kit with 2x150 bp read length and a depth of 4.0–5.0 million reads per sample. Unless otherwise stated, software tools were run with default settings for sequence processing and analysis. The sequence data was demultiplexed (Illumina bcl2fastq, v2.19.0.316), quality-checked (Fastp, v0.20.1) and visualised (MultiQC, v1.7). Paired-end reads were quality-filtered (Fastp (Chen *et al.*, 2018) v0.20.1, additional 53 parameter: "--detect_adapter_for_pe --cut_by_quality5 --cut_by_quality3 --low_complexity_filter --54 length_required 21 --correction")), assembled (Unicycler (Wick *et al.*, 2017) v0.4.8) and quality-checked (CheckM2 (Chklovski *et al.*, 2023b) v1.0.18). Taxonomical ranks were established using the Type Strain Genome Server (Meier-Kolthoff and Göker, 2019a) and GTDB (Parks *et al.*, 2018).

2.5.2.2 Computational Analysis and Taxonomic Classification

Genome completeness exceeded 98% in all cases, with contamination levels below 2%. Assembly sizes ranged from 4.7 to 7.2 Mbp and GC contents from 35 to 44%, consistent with other *Pedobacter* reference genomes. Taxonomic placement was performed using the Type Strain Genome Server (TYGS) and Genome Taxonomy Database (GTDB) pipelines, confirming affiliation of all isolates to the genus *Pedobacter* with varying proximity to type strains such as *P. frigiditerrae*, *P. aquatilis*, and *P. heparinus* (Table 7).

Table 7: Overview of isolated *Pedobacter* strains and detected β-lactamase numbers and classes.

	Isolation				No. of detected genes/ β-lactamase class				
Isolate (ID)	Isolation Source	Sampling location (Lat., Long.)	Closest type strain (TYGS)	GenBank accession	A	B1 B2	B3	C	D
DE_ 0159	*Salamandra salamandra*	*Kottenforst, Germany*	*P. frigiditerrae* RP-1-13	JAVTSV000000000	0	0	1	3	0
DE_ 0302	*Salamandra salamandra*	*Harz, Germany*	*P. miscanthi* RS10	JAVTSU000000000	1	0	1	4	0
DE_ 0380	*Salamandra salamandra*	*Solling, Germany*	*P. gandavensis* LMG 31462 T	JAVTST000000000	1	0	0	4	1
DE_ 0385	*Salamandra salamandra*	*Solling, Germany*	*P. frigiditerrae* RP-1-13	JAVTSS000000000	0	0	1	3	0
DE_ 0392	*Salamandra salamandra*	*Solling, Germany*	*P. agri* DSM 19486	JAVTSR000000000	0	0	1	3	0
DE_ 0410	*Salamandra salamandra*	*Solling, Germany*	*P. nutrimenti* DSM 27372	JAVTSQ000000000	1	0	0	5	2
DE_ 0497	*Salamandra salamandra*	*Solling, Germany*	*P. frigoris* KACC 21154	JAVTSP000000000	0	0	0	2	0
DE_ 0550	*Salamandra salamandra*	*Harz, Germany*	*P. nototheniae* 36B243T	JAVTSO000000000	0	1	1	7	0
DE_ 0801	*Salamandra salamandra*	*Solling, Germany*	*P. psychrodurus* RP-3-21	JAVTSN000000000	1	0	1	7	0
DE_ 0989	*Salamandra salamandra*	*Eifel, Germany*	*P. antarcticus* DSM 11725	JAVTSM000000000	1	0	1	8	1
MADA_ 173	*Boophis williamsi*	*Ankaratra, Madagascar (-19.3463, 47.27705)*	*P. antarcticus* DSM 11725	JAVTSL000000000	1	0	1	7	1
MADA_ 278	*Mantidactylus aff . curtus 19*	*Ankaratra, Madagascar (-19.3463, 47.27705)*	*P. gandavensis* LMG 31462 T	JAVTSK000000000	2	0	0	4	0
MADA_ 852	*Aglyptodactylus madagascariensis*	*Andasibe, Madagascar (-18.9328, 48.41312)*	*P. aquatilis* CECT 7114	JAVTSJ000000000	1	0	1	3	0
MADA_ 1817	*Ptychadena mascareniensis*	*Andasibe, Madagascar (-18.9328, 48.41312)*	*P. agri* DSM 19486	JAVTSI000000000	0	0	1	3	0
MADA_ 1818	*Ptychadena mascareniensis*	*Andasibe, Madagascar (-18.9328, 48.41312)*	*P. agri* DSM 19486	JAVTSH000000000	0	0	1	3	0
MADA_ 2501	*Boophis goudoti*	*Ankaratra, Madagascar (-19.3463, 47.27705)*	*P. aquatilis* CECT 7114	JAVTSG000000000	0	1	1	3	0

MADA_2608	*Mantella aurantiaca*	*Breeding centre, Madagascar*	*P. nyackensis* DSM 19625	JAVTSF000000000	1	0	0	3	0
MADA_3128	*Spinomantis aglavei*	*Andasibe, Madagascar (-18.9328, 48.41312)*	*P. frigidisoli* RP-3-11	JAVTSE000000000	0	0	0	3	0
MADA_3506	*Boophis goudoti*	*Ankaratra, Madagascar (-19.3463, 47.27705)*	*P. aquatilis* CECT 7114	JAVTSD000000000	0	0	1	6	0
S10_4.1	*Salamandra*	*Schiffenberg Forest, Germany*	*P. heparinus* DSM 2366	JAVTSC000000000	0	1	0	3	0
S8_12.1	*Salamandra*	*Schiffenberg Forest, Germany*	*P. gandavensis* LMG 31462 T	JAVTSB000000000	1	0	0	2	0

2.5.2.3 Prediction of β-Lactamase Genes and Resistome Analysis

Since *Pedobacter* species are known to encode multiple β-lactamase genes, a dedicated resistome-profiling workflow was applied. open reading frames were screened with Resistance Gene Identifier using the Comprehensive Antibiotic Resistance Database (Alcock *et al.*, 2020) to predict β-lactamase genes and classify them into mechanistic families. All predicted β-lactamase sequences were extracted and analysed with the Enzyme Function Initiative–Enzyme Similarity Tool (EFI-EST) (Gerlt *et al.*, 2015). The network included proteins ranging from 200 to 440 amino acids and used an alignment score cutoff of 20. This computational network approach enabled the visual comparison of novel candidate β-lactamases with known reference enzymes and the classification of uncharacterised clusters based on sequence similarity, rather than annotation alone. The resulting SSN comprised 118 putative β-lactamases (average 5.6 ± 2.5 per genome) that clustered into 89 reference groups. In addition to analysing our 21 in-house *Pedobacter* strains, we also included publicly available genomes to assess the distribution of β-lactamase among the genus (Figure 11, left side). The majority of detected β-lactamases belonged to class C (609), followed by class B3 (72), class A (56), and class B1/B2 (32), consistent with previous findings (Viana *et al.*, 2018a). However, our expanded dataset additionally revealed 26 putative β-lactamases clustering with reference enzymes of the Ambler class D. The total number of predicted AMR genes detected by CARD and the curated β-lactamases included in the SSN were assigned to their taxonomic context within the genus based on the calculated phylogenetic tree (Figure 11, right side). The 26 predicted class D β-lactamases were distributed across 22 genomes, accumulating in distinct taxonomic branches. This includes our isolates (yellow shade), for example, *Pedobacter* sp. DEU0410, which carries eight predicted β-lactamases belonging to class C (5) and class D (2).

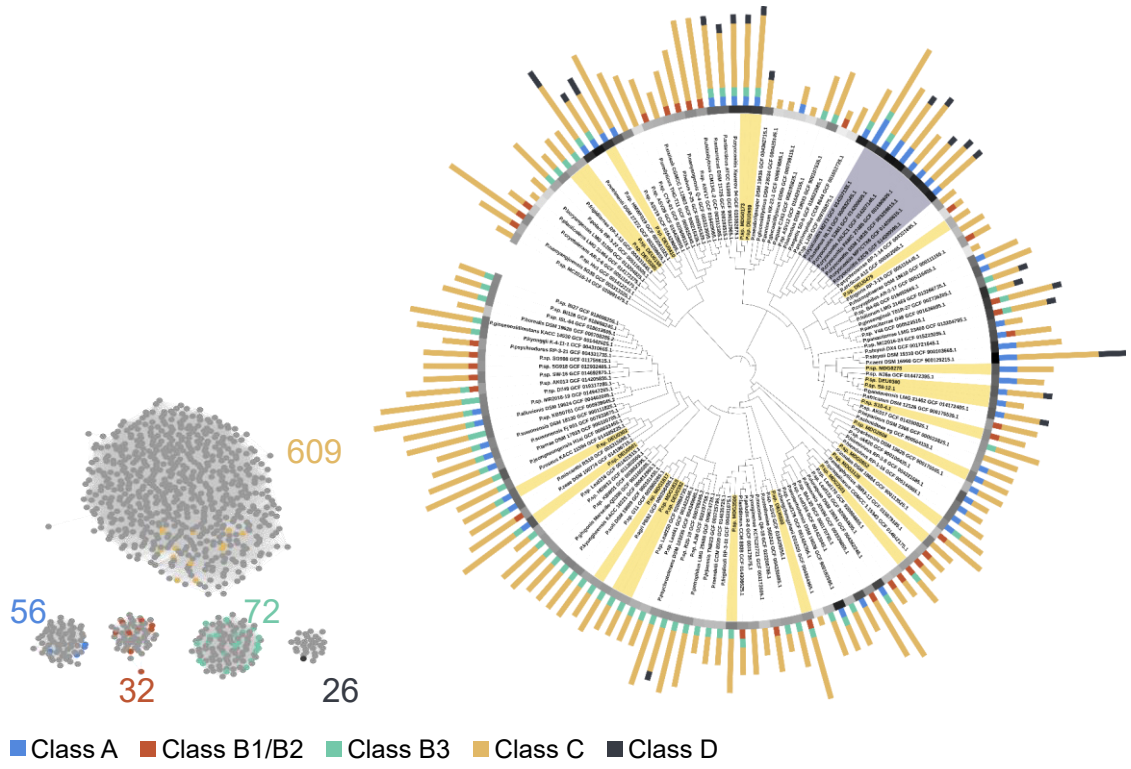

Figure 11: Distribution of β-lactamase enzymes across representatives of the *Pedobacter* genus (n = 143). **Left**: Sequence similarity network of *Pedobacter* β-lactamase enzymes (coloured by Ambler class) in relation to reference β-lactamases from the CARD database (grey), illustrating class assignment based on similarity calculated with EFI-EST. **Right**: Phylogenetic tree of *Pedobacter* representatives (n = 143) showing the density of antimicrobial resistance genes (first ring, black = high, grey = low), highlighting the 21 in-house Fraunhofer strains (yellow shading) and the *P. cryoconitis* clade (grey shading). The number of predicted β-lactamase enzymes per genome is displayed as bar charts, with colours corresponding to the predicted Ambler class.

The combination of CARD-based annotation and network-based classification demonstrated the diversity of resistance determinants within environmental *Pedobacter* isolates and within the genus *Pedobacter*, supporting the hypothesis that these organisms serve as reservoirs for novel β-lactamase variants.

This study established a high-throughput computational workflow for comparative genomics and resistome analysis across 21 *Pedobacter* isolates. The workflow integrated Illumina short-read assembly, automated annotation, and machine-learning-based genome quality assessment (CheckM2) with functional resistome mining and network visualisation. The data reveal a consistent presence of multiple β-lactamase genes per genome and identify unreported class D variants, expanding the known functional diversity of this enzyme family in the genus. These results highlight the value of integrating genome mining with sequence similarity networks for identifying and categorising resistance determinants in non-clinical bacterial taxa.

2.6 A genetically tractable branch of environmental *Pedobacter* from the phylum Bacteroidota represents a hotspot for natural product discovery

Parts of this chapter have been published as an article in Scientific Reports. The manuscript can be found at: DOI:10.1038/s41598-025-03955-z

www.nature.com/scientificreports

scientific reports

Check for updates

OPEN

A genetically tractable branch of environmental *Pedobacter* from the phylum Bacteroidota represents a hotspot for natural product discovery

Yang Liu[1,8], Luis Linares-Otoya[1,7,8], Christian Kersten[3], Michael Marner[1,2], Sanja Mihajlovic[2], Mohamed H. Abdeldayem[1], Sandra Semmler[2], Molly C. Bletz[4], Miguel Vences[5], Marius Spohn[1,2], Celine M. Zumkeller[1,2]✉ & Till F. Schäberle[1,2,6]✉

Scan for full publication

Emerging global challenges, such as antimicrobial resistance, have shifted the focus of natural product discovery from well-characterized microbial producers to underexplored taxonomic groups. Here, we computationally and experimentally characterize the biosynthetic potential of the genus *Pedobacter*, a Bacteroidota taxon known for antibiotic production that harbors numerous uncharacterized secondary metabolite (SM)-encoding biosynthetic gene clusters (BGCs). Through phylogenomic analysis of the genus *Pedobacter*, we identify a distinct clade enriched in lipopeptide-associated BGCs, most of which lack known chemical products. By developing de novo genetic tools and integrating metabolomics, we linked specific secondary metabolites to their corresponding BGCs. Using synthetic and analytical chemistry as proof of concept, we isolated and structurally characterized twelve linear lipopeptides (cryopeptins A–N), containing rare dehydrovalines from *Pedobacter cryoconitis* PAMC 27485. We demonstrate that all cryopeptins, despite their structural heterogeneity, are biosynthesized by a single multi-domain non-ribosomal peptide synthetase (NRPS) gene cluster. Mechanistically, we propose that this BGC drives chemical diversity through combinatorial fatty acid incorporation and iterative amino acid assembly, resulting in variable peptide chain lengths. This work highlights the biosynthetic versatility of *Pedobacter* and outlines methods for genetic manipulation in this genus to systematically access its cryptic natural product repertoire.

Keywords Bacteroidota, Lipopeptides, Natural products, NRPS, *Pedobacter*, Biosynthetic gene clusters (BGCs), Dehydrovaline

Secondary metabolites (SMs) are small molecules that often exhibit ecologically or pharmacologically relevant bioactivities[1,2]. These compounds have evolved in bacteria to possess complex structural features in order to achieve potent and specific molecular interactions with their targets[3–5]. Given the structural complexity of SMs, these molecules are hard to mimic by synthetic chemistry and, therefore, serve as a primary source for the discovery of novel lead structures[4,6].

Traditionally, actinobacteria, bacilli, and specific proteobacteria were major bacterial SM contributors. However, declining discovery rates prompt exploration of yet underexplored (bacterial) sources. Increasing

[1]Institute for Insect Biotechnology with Focus on Natural Product Research, Justus- Liebig-University Giessen, Ohlebergsweg 12, 35392 Giessen, Germany. [2]Natural Product Department, Fraunhofer-Institute for Molecular Biology and Applied Ecology (IME-BR), Ohlebergsweg 12, 35392 Giessen, Germany. [3]Institute of Pharmaceutical and Biomedical Sciences, Johannes Gutenberg University Mainz, 55128 Mainz, Germany. [4]Department of Environmental Conservation, University of Massachusetts Amherst, Amherst, MA, USA. [5]Zoological Institute, Technische Universität Braunschweig, Mendelssohnstr. 4, 38106 Braunschweig, Germany. [6]German Center for Infection Research (DZIF), Partner Site Giessen-Marburg- Langen, Ohlebergsweg 12, 35392 Giessen, Germany. [7]Present address: Department of Molecular Biology, Princeton University, Princeton, USA. [8]Yang Liu and Luis Linares-Otoya contributed equally to this work. ✉email: celine.zumkeller@ime.fraunhofer.de; Till.F.Schaeberle@agrar.uni-giessen.de

2.6.1 Background

Secondary metabolites (SMs) are small molecules that often exhibit ecologically or pharmacologically relevant bioactivities (Dinglasan *et al.*, 2025; Price-Whelan *et al.*, 2006). These compounds have evolved in bacteria to possess complex structural features in order to achieve potent and specific molecular interactions with their targets (Bushin *et al.*, 2020; Davies, 2013; S. Dickschat, 2010). Given the structural complexity of SMs, these molecules are hard to mimic by synthetic chemistry and, therefore, serve as a primary source for the discovery of novel lead structures (Davies, 2013; Stratton *et al.*, 2015).

Traditionally, actinobacteria, bacilli, and specific proteobacteria were major bacterial SM contributors. However, declining discovery rates prompt exploration of yet underexplored (bacterial) sources. Increasing capabilities to read, rate, and modify bacterial genomes are the keystone for a renaissance of SM discovery. Furthermore, the fact that most of the putative SM-encoding BGCs have not yet been assigned to any chemical product indicates a high biosynthetic potential within non-traditional SM producer taxa (Cimermancic *et al.*, 2014; Corre and L. Challis, 2009; Donadio *et al.*, 2007; Walsh and Fischbach, 2010). Among these recently recognized taxa are the Bacteroidota, in which the number of compounds isolated and characterized represent only a minor proportion of the predicted potential (Brinkmann *et al.*, 2022a). Moreover, its encoded chemical diversity is predicted to differ from traditional natural products (NPs) -producing taxa (S. Dickschat, 2010; Stratton *et al.*, 2015).

Our previous work analyzed representative genomes from the phylum Bacteroidota concerning their overall biosynthetic potential and identified the genus *Pedobacter* as a BGC hotspot (Brinkmann *et al.*, 2022a). Also, in 2021, Figueiredo *et al.* highlighted the *Pedobacter*, to be the genus with the highest BGC number within the family *Sphingobacteriaceae* (Figueiredo *et al.*, 2022). A recent pangenome study undermines this impression, suggesting a unique but uneven distribution of BGCs within the 41 *Pedobacter* genomes analyzed. *Pedobacter* species are widely distributed in many habitats, such as terrestrial (J. Huang *et al.*, 2022) and marine environments (He *et al.*, 2020), or associated with higher organisms (Corsaro *et al.*, 2017). Although *Pedobacter* strains have been studied for their potential to produce industrially relevant enzymes (Gu *et al.*, 2017; Zhu *et al.*, 2018b) and their relevance as antimicrobial resistance gene reservoirs (Ullmann *et al.*, 2020a; Viana *et al.*, 2018b), their potential for NPs discovery has been less explored. Members from this genus are known to produce antibiotics with potent activity against multidrug-resistant pathogens such as the cyclic lipodepsipeptides Pedopeptins and Isopedopeptins (from now on referred to (iso)pedopeptins (molecules) or pedopeptin-like (BGCs)). (Iso)pedopeptins are the sole bioactive specialized metabolite class identified from *Pedobacter* to this date (Hirota-Takahata *et al.*, 2014; Kozuma *et al.*, 2014; Nord *et al.*, 2020b).

Here, 143 *Pedobacter* genomes (including 21 project strains) were analyzed by genome mining, resulting in the identification of a *Pedobacter* branch enriched in yet uncharacterized non-ribosomal peptide synthetase (NRPS) gene clusters. To complement the traditional cultivation approaches, we developed a genetic toolkit for *Pedobacter* that enabled linking the natural products to their corresponding BGC in combination with metabolomics. This combined genetic-metabolomic approach led to the isolation of 12 lipopeptides that were fully characterized using nuclear magnetic resonance (NMR) experiments, Marfey´s analysis, and total synthesis.

2.6.2 Methods

2.6.2.1 Creation and Curation of the *Pedobacter* Assembly Dataset

149 *Pedobacter* assembly genomes matching our selection criteria (NCBI:txid84567, filter: RefSeq annotation) were obtained from NCBI in gbff file format on 28.11.2022. The dataset was complemented by 21 *Pedobacter* assemblies retrieved from amphibian specimens (Biosample accession 37505411-37505431) (n=171). To ensure data quality, genome assemblies were subject to quality control steps using CheckM2 (Chklovski *et al.*, 2023a), TYGS (Meier-Kolthoff and Göker, 2019b), OrthoANI (Lee *et al.*, 2016a), and Genome Taxonomy Data Base (GTDB-Tk) (Parks *et al.*, 2022). Firstly, we retained only assemblies meeting the minimum information about a metagenome -assembled genome (MIMAG) evaluation (Bowers *et al.*, 2017) criteria for high-quality genomes, demonstrating 90-100% completeness and 0-5% contamination according to CheckM2 analysis (n=168). Redundant genomes were identified and removed based on orthoANI scores greater than 99.9% and TYGS digital DNA-DNA hybridization (dDDH) values (d0, d4, d6 = 100%). For the remaining genomes, GTDB-Tk was used to confirm taxonomic affiliation. The final dataset consisted of 149 genome assemblies, all confirmed to belong to the genus *Pedobacter.* A whole-genome sequence-based phylogenetic tree was retrieved from TYGS and visualized using the interactive tree of life (iTOL v6(Letunic and Bork, 2024)) tool. Metadata such as assembly size and N50 were monitored using Geneious (v11.1.5).

2.6.2.2 Detection and Clustering of BGCs and Antibiotic Resistance Genes

Standalone antiSMASH v6.1.1[19] was used to detect BGCs in the *Pedobacter* genome dataset. Genome assemblies were provided as fasta files, genes were predicted using the integrated Prodigal Pipeline, and BGC detection was limited to clusters larger than 5 kb to minimize the number of incomplete BGCs (--minlength 5000 –genefinding-tool prodigal). Only contigs larger than five kbps were inspected, averaging 72 contigs (1-1481 contigs) per genome. We observed no correlation between the number of detected BGCs and the number of contigs per genome (R^2 = 0.0165; see Figure S1-1), warranting the robustness of the data. BGC numbers were correlated to the N50 values to detect broken NRPS clusters (Figure S1-2) and the genome assembly size (Figure S1-3). Assemblies showing a possible overestimation of BGCs (low N50 values but high numbers of BGCs, Figure S1-2), underwent a second round of BGC detection after contigs < 10 kb (“Filter Assembled Contigs by Length - v1.2.0" on kbase[42]). After manual curation of the respective antiSMASH outputs, all antiSMASH files were analyzed to extract multimodular NRPS BGCs. These BGCs (n=32) served as input for BiG-SCAPE to construct a similarity network. BiG-SCAPE was run with default parameters at a cutoff of 0.6, flagging –MIBiG, –mix to include MIBiG reference clusters and a similarity network independent of BGC-type. The Generated Network was visualized with Cytoscape (v3.8.2) using the yFiles organic Layout. . The cluster amount and types, specifically for multimodular NRPS and the assembly sizes were extracted and visualized as a heatmap in iTOL(Letunic and Bork, 2024).

2.6.3 Results

2.6.3.1 A *Pedobacter* Branch is Enriched in Unique Multimodular NRPS Gene Clusters

To comprehensively rate the biosynthetic potential of the genus *Pedobacter*, we extracted 149 genomes from the National Center for Biotechnology Information (NCBI) reference genome database. This dataset was complemented with 21 project strain genomes (Zumkeller *et al.*, 2024) from amphibian specimens sampled either in Germany (ISO country-code DEU) or in Madagascar (ISO country-code MDG) (details under BioProject accession number PRJNA1019955). After data curation (Material and Methods), we constructed a whole-genome sequence-based phylogenetic tree, including 143 *Pedobacter* genome sequences (Figure 12). Some species affiliations published by NCBI were inconclusive when evaluated using different genome-based classification tools (Lee *et al.*, 2016a; Meier-Kolthoff and Göker, 2019b; Parks *et al.*, 2022) (Figure 12, grey squares). Supplementary Table 1 contains suggested species affiliations for these strains based on TYGS (Meier-Kolthoff and Göker, 2019a) and is published along with the manuscript (see page 50). We found the 21 *Pedobacter* strains from amphibian specimens (yellow shade) broadly distributed across the genus *Pedobacter*.

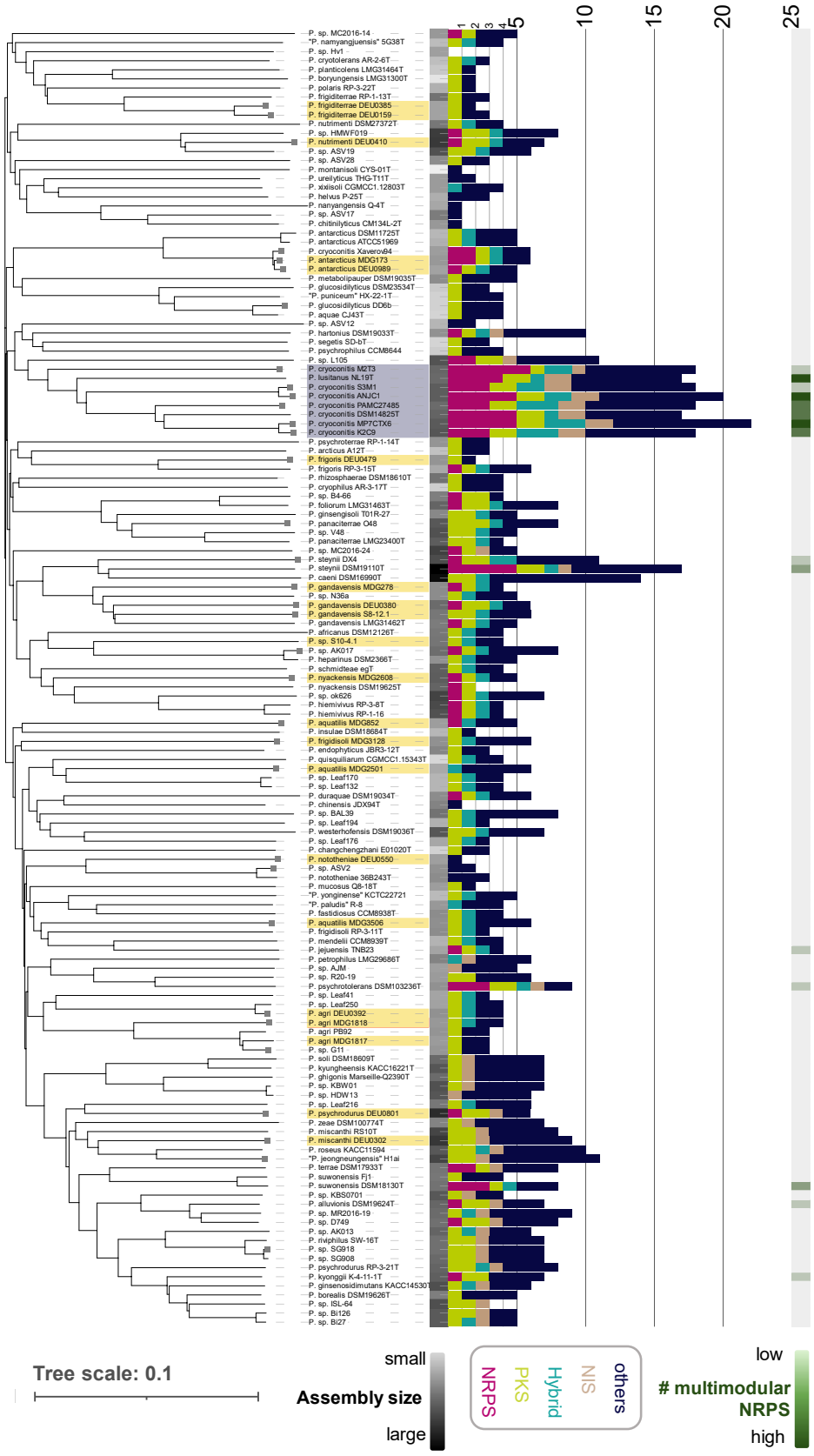
1
2
3
4
5
10
15
20
25
Tree scale: 0.1
small
Assembly size
large
NRPS
PKS
Hybrid
NIS
others
low
multimodular NRPS
high

Figure 12: Bioinformatics analysis of 143 *Pedobacter* genomes indicates a BGC hotspot of the *P. cryoconitis* branch. A consensus tree was calculated using the Type Strain Genome Server(Meier-Kolthoff and Göker, 2019b) based on the Genome BLAST Distance Phylogeny approach. Colors indicate project strains from amphibian specimens (yellow) and the *P. cryoconitis* branch (grey). The genome size (black heatmap), the number and types of biosynthetic gene clusters (bar charts), and the presence of multimodular NRPS gene clusters (green heatmap) are illustrated for each strain. Values were determined by antibiotics & Secondary Metabolite Analysis Shell (antiSMASH) v6.1.1 (Blin *et al.*, 2021b). The tree was annotated using the interactive tree of life (iTOL v6(Letunic and Bork, 2024)) tool. The species names are based on NCBI entries. Species affiliations of some strains remained inconclusive (marked with a square) if compared with different tools (OrthoANI (Lee *et al.*, 2016a), TYGS (Meier-Kolthoff and Göker, 2019b) and GTDB (Parks *et al.*, 2022)). The supplementary table also lists species affiliations based on TYGS for these strains.

Next, we assessed and rated the biosynthetic potential of the strains by BGC annotation analysis using antibiotics & Secondary Metabolite Analysis Shell (antiSMASH) 6.1 (Blin *et al.*, 2021b) (Supplementary Table 1). In summary, a *Pedobacter* genome featured 5.7 BGCs on average. The overall biosynthetic diversity is similar to the numbers reported by Covas *et al.* (This study vs. Covas *et al.*); While cluster types like NRPS (11% vs. 9%), Ribosomally synthesized and post-translationally modified peptides (RIPPs) (27% vs. 27%), Terpenes (22% vs. 22%) and Others (8% vs. 7%) are very similar, notable differences are observed in Polyketide synthases (PKS) (19% vs. 12%) and NRPS-PKS hybrids (12% vs. 17%). This discrepancy can be explained by the different size and composition of the genomic datasets. As indicated by Covas *et al.* (Covas *et al.*, 2023), there is no overall linear positive correlation between genome assembly size and the number of BGCs (R^2 = 0.29; Figure S1-3). This differentiates *Pedobacter* from other bacterial genera, such as, e.g., *Chitinophagia* (Brinkmann *et al.*, 2022a), *Nocardia* (Männle *et al.*, 2020), or *Amycolatopsis* (Adamek *et al.*, 2018), showing a positive correlation between the BGC number of each strain and the respective genome size. The accumulation of clusters in only a few genomes already hints towards an uneven distribution of BGCs within the genus. To identify strains with a high BGC load, the determined total BGC number was assigned to the strains and set to the taxonomic context of the genus based on the calculated phylogenetic tree (Figure 12). The eight strains exhibiting the highest BGC load (≥17 BGCs) cluster together in the phylogenetic tree (Figure 12, grey shade), including the two type strains, *P. cryoconitis* DSM 14825 (Margesin *et al.*, 2003) and *P. lusitanus* NL19^T(Covas *et al.*, 2017). This is in accordance with the analysis performed by Covas *et al.*, which also identified *P. cryoconitis* and *P. lusitanus* as hotspots of biosynthetic potential. Our analysis now extends this hotspot - the "*P. cryoconitis* branch" by six additional strains. As shown in Figure 12, all *P. cryoconitis* strains - except the type strains DSM 14825^T - exhibit taxonomic discrepancies. According to TYGS (Meier-Kolthoff and Göker, 2019b) analysis, these strains should be reassigned to *Pedobacter sp.* However, for consistency, we retained the NCBI naming throughout the manuscript. These eight strains show an average genome size of 6.13 ± 0.14 Mbps, encoding 18.25 ± 1.6 BGCs. All other *Pedobacter* strains (n=135) instead have an average genome size of 5.42 Mbps and a comparatively low BGC load of 5.1 ± 2.8 BGCs. Despite their close relationship, their BGC composition is still rather heterogeneous (cosine similarity score ranging from 0.5-1, Figure S1-7). Only *P. cryoconitis* K2C9 and *P. cryoconitis* MP7CTX6 share their complete BGC equipment. The *P. cryoconitis* branch accounts for 18% of all BGCs and a remarkably enriched amount of NRPS BGCs. Together, they carry 34 NRPS gene clusters (on average 4.25 ± 1.1 per strain), accounting for 43% of all NRPS gene clusters detected within the dataset.

The biosynthetic potential of *Pedobacter* in the production of terpenes and NRPS-independent siderophore (NIS) SMs was previously reported (Covas *et al.*, 2017). In this study, we focused on NRPS-derived SMs instead. NRPS biosynthetic machineries provide a tremendous structural complexity of peptides exhibiting diverse biological functions and a wide range of physio-chemical properties (Süssmuth and Mainz, 2017). Our in-depth analysis revealed the annotation of NRPS gene clusters within the genomes of 41 strains. Of these, 26 strains exclusively contained monomodular NRPS gene clusters while a rather exclusive amount of 15 strains carried 34 multimodular NRPS gene clusters. Most of these multimodular NRPS gene clusters are again encoded within the genomes of the *P. cryoconitis* branch (Figure 12, green heat map). Another seven strains carrying multimodular NRPS gene clusters are distributed over the genus. To determine the diversity of the 34 detected multimodular NRPS gene clusters, we used BiG-SCAPE (Navarro-Muñoz *et al.*, 2020c) (Biosynthetic Gene Similarity Clustering and Prospecting Engine) to analyze their sequential and compositional similarity and to assign them to gene cluster families (GCFs). At a cutoff of 0.6, 22 GCFs were generated. Three networks included GCFs from more than two strains, covering 16 BGCs in total (Figure 13, NRPS-1, -4, and -5). Corason (Navarro-Muñoz *et al.*, 2020c) alignment (Figure S1-6) of these GCFs showed that the GCF defining NRPS-1 actually contained three nodes with fused NRPSs (Figure 13A-chain symbol), while NRPS-1 is only partially covered in strain *P. cryoconitis* ANJC1 (Figure 13A-star). NRPS-1 GCF also comprises a duplet harbored by strains *P. cryoconitis* S3M1 and ANJC1 and another singleton in *P. cryoconitis* PAMC27485. NRPS-1 is most conserved and found in 7 of the 15 analyzed strains. Notably, 6 out of 8 strains of the *P. cryoconitis* branch encode this BGC, including *P. steynii* DSM19110^{T}, which does not belong to the same taxonomic group. NRPS-4 is a smaller yet unassigned BGC harbored by a third of the investigated strains, while NRPS-5 is present in four strains. Both, NRPS-4 and -5 are only found in members of the *P. cryoconitis* branch. Two GCFs were duplicates (NRPS-2 and -6), all encoded by *P. cryoconitis* ANJC1 and S3M1, respectively. The rest of the detected BGCs remained as singletons (n=17).

To assign the detected clusters to known BGCs, BigSCAPE (Navarro-Muñoz *et al.*, 2020c) was used to include and calculate the similarity of all reference BGCs available from the Minimum Information about a Biosynthetic Gene cluster (MIBiG (Terlouw *et al.*, 2023)) v 2.1 repository. Intriguingly, no deposited reference cluster was correlated to our data, suggesting a compositional uniqueness of the analyzed *Pedobacter* NRPS BGCs. Manual inclusion of the known Pedopeptin BGC, however, confirmed its presence in the described producer *P. lusitanus* NL19^T, and similar pedopeptin-like clusters in *P. cryoconitis* DSM 14825^T, *P. cryoconitis* K2C9 and *P cryoconitis* MP7CTX6 (Figure 13, NRPS-5, Figure S1-6C).Finally, we generated a cosine similarity score heatmap to visualize the similarity of the multimodular NRPS content of the strains ("https://software.broadinstitute.org/morpheus," n.d.) Figure 13B). The strains outside of the *P. cryoconitis* branch all encode unique multimodular NRPSs, resulting in their similarity score of zero for all strains. The strains of the *P. cryoconitis* branch, highly enriched in multimodular NRPSs, form a distinct group within the genus. Within this branch, two separate groups become obvious: One includes the five strains *P. cryoconitis* PAMC27485, *P. cryoconitis* DSM 14825^T, *P. cryoconitis* K2C9, *P. cryoconitis* MP7CTX6 and *P. lusitanus* NL19^T (similarity score range of 0.3 – 0.9). The strains PAMC27485 and DSM 14825^T share the lowest similarity because only NRPS-4 is shared between them. The group is formed by *P. cryoconitis* ANJC1 and *P. cryoconitis* S3M1, showing a high similarity score of 0.8. The wide range of similarities highlights the BGC heterogeneity within the *P. cryoconitis* branch.

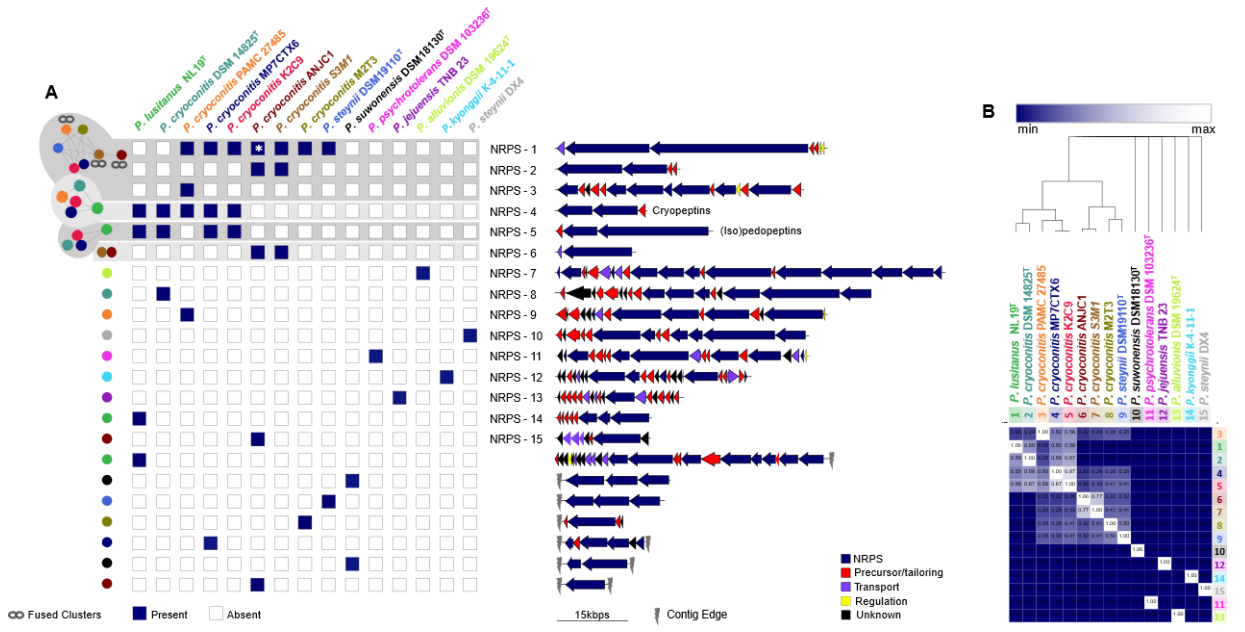

Figure 13: BGC Similarity of multimodular NRPS from the genus *Pedobacter* reveals a high diversity. A BiG-SCAPE (Navarro-Muñoz *et al.*, 2020c) analysis of 34 multimodular NRPS gene clusters from 15 strains was performed to analyze distribution and diversity (Strain-specific colored nodes and networks, left side). The presence of NRPS clusters was manually curated (blue and white boxes), thereby revealing the presence of fused clusters (chain sign, NRPS 1,2,3) and broken clusters (white star, NRPS-1). The architecture of multimodular NRPS clusters is highlighted on the left side (colored arrows), numbered from 1-15. Incomplete NRPS clusters at contig edges (grey flash) are not numbered. **B** Heat map illustrating the cosine similarity of the multimodular NRPS composition from blue (minimum similarity) to white (maximum similarity). The strains are clustered based on similarity, as highlighted by the tree ("https://software.broadinstitute.org/morpheus," n.d.). Strains 10-15 all encode unique multimodular NRPS clusters (similarity score: 0). Strains belonging to the *P. cryoconitis* branch and *P. steynii* DSM19110^T form a separate branch that share some of the multimodular NRPS between each other.

2.6.3.2 Summary - Genetic Validation of NRPS-4 and Functional Link to Cryopeptin Biosynthesis

Our *in-silico* analysis highlighted a *Pedobacter* branch possessing a great diversity of unique BGCs that differ from those in the MIBiG (Terlouw *et al.*, 2023) database (Figure S1-7A). This prompted a yet uncharted chemical space of SMs potentially produced by this *Pedobacter* branch. To link BGCs to their corresponding metabolites, we sought to disrupt the putative BGC and perform comparative metabolomics to identify the associated metabolites. To this end, we initially evaluated the metabolome of representative *Pedobacter* under diverse culture conditions, identified candidate NRPS-like metabolites via tandem mass spectrometry (MS/MS) analysis, and subsequently used this information to computationally identify expressed NRPS gene clusters that could be experimentally validated by genetic engineering. Therefore, NRPS-4 was selected as a representative and tractable target for genetic validation, as its predicted domain architecture and widespread occurrence within the *Pedobacter cryoconitis* branch strongly suggested a central role in the biosynthesis of the observed NRPS-like metabolites.

Through our combined bioinformatic and genetic analyses, we established a framework for linking biosynthetic gene clusters to their corresponding metabolites in *Pedobacter*. We identified a distinct *P. cryoconitis* branch enriched in multimodular NRPS clusters, representing a reservoir of previously uncharacterized biosynthetic diversity. By developing and optimizing transformation systems for *P. cryoconitis*, we enabled stable gene transfer and demonstrated targeted gene deletion using homologous recombination. Deletion of the *crpA* gene within the conserved NRPS-4 cluster abolished production of the associated lipopeptides, confirming its role in nonribosomal peptide biosynthesis. Large-scale cultivation and subsequent chemical characterization revealed a series of novel linear lipopeptides, the cryopeptins (A–N), thereby linking NRPS-4 to a distinct natural product family and establishing the *Pedobacter cryoconitis* lineage as a promising model for future natural product discovery. This illustrates how computational genomics can guide the discovery of novel compounds from less-explored taxonomic groups, such as *Pedobacter*. Here, we discovered the novel cryopeptins containing unusual dehydrovaline amino acids without relying on bioactivity-guided screening approaches. Instead, the development of genetic modification tools for *Pedobacter* enabled direct access to biosynthetic pathways predicted through genome mining. Moving forward, more sophisticated AI-guided computational approaches, such as metabologenomics can improve the targeted investigation of cryptic or silent BGCs. For example, NPOmix combines untargeted metabolomic data with annotated sequencing data to predict the biosynthetic origin of detected compounds. This integration not only enables targeted knockout designs but also provides important insights into the regulation of BGCs under various cultivation conditions, suggesting potential ecological functions of novel metabolites. These findings further highlight the growing power of computational tools to not only guide NP discovery of commercial interest but also to address fundamental microbial research questions, including microbial interactions and ecological functions.

This section outlines the setup of a computational workflow to analyse novel Bacteroidota strains obtained from ongoing bioprospecting efforts aimed at expanding the diversity of the FHG strain collection. It highlights the genetic characterisation of marine- and amphibian-derived strain isolates and concludes with the description of the cryopeptin NP and its BGC. Beyond our academic efforts investigating the Bacteroidota phylum – including the ongoing PhD project on Monobactam biosynthesis -this work lays the foundation for a broader bioinformatic and AI-driven research platform. The FHG strain collection, along with the already generated, mostly standardised data sets on genome sequencing, metabolomics and bioactivity datasets under different cultivation conditions, serves as a unique resource for these efforts. Expanding genome coverage will further support these endeavours and increase the strategic value and attractiveness of the FHG strain collection, both for industrial partners and as a tool to advance academic research.

Profiling of Novel Acidobacteriota

BACKGROUND

Acidobacteriota are prevalent and exhibit substantial phylogenetic diversity in various environments, including soils. However, limited research has been conducted due to their slow growth rates and challenges associated with cultivation. Metagenomic analyses and characterisation of isolated strains suggest that Acidobacteriota possess ecological adaptability and distinct metabolic capabilities, including the production of NPs and the presence of plant growth-promoting traits (PGPTs). This section presents two complementary results, covering a bioprospecting campaign (Chapter 2.7), yielding strains that were subsequently characterised (Chapter 2.8):

Three novel strains were successfully retrieved from termite nests and taxonomically assigned to the genus *Acidobacterium* and *Terracidiphilus*. This provided experimentally tractable strains from an ecologically rich microhabitat. These strains, along with one additional acidobacterial soil isolate, were profiled for their general metabolomic capacity using OSMAC and their specific metabolic response supplemented with Trp. The strains produced phytohormones iP and IAA, and supplementation with Trp induced a metabolic shift, resulting in the production of antifungal metabolites. Functional studies in plants and on plates did not show any plant growth-promoting traits. Complementarily, we employed an extensive genome mining approach to describe the genomic potential for PGP and BGC in acidobacterial strains.

2.7 High-throughput cultivation for the selective isolation of Acidobacteria from termite nests as a potential new source of natural products

Parts of this chapter have been published as an article in Frontiers in Microbiology. The manuscript can be found at: DOI: 10.3389/fmicb.2020.597628.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 06 November 2020
doi: 10.3389/fmicb.2020.597628

Scan for full publication

High-Throughput Cultivation for the Selective Isolation of Acidobacteria From Termite Nests

Markus Oberpaul[1], Celine M. Zumkeller[1], Tanja Culver[1], Marius Spohn[1], Sanja Mihajlovic[1], Benedikt Leis[1], Stefanie P. Glaeser[2], Rudy Plarre[3], Dino P. McMahon[3,4], Peter Hammann[5†], Till F. Schäberle[1,6], Jens Glaeser[1*†] and Andreas Vilcinskas[1,6]**

[1] Branch for Bioresources, Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Giessen, Germany, [2] Institute of Applied Microbiology, Justus Liebig University Giessen, Giessen, Germany, [3] Bundesanstalt für Materialforschung und -prüfung, Berlin, Germany, [4] Institute of Biology, Free University of Berlin, Berlin, Germany, [5] Sanofi-Aventis Deutschland GmbH, R&D Integrated Drug Discovery, Hoechst Industrial Park, Frankfurt am Main, Germany, [6] Institute for Insect Biotechnology, Justus Liebig University Giessen, Giessen, Germany

Microbial communities in the immediate environment of socialized invertebrates can help to suppress pathogens, in part by synthesizing bioactive natural products. Here we characterized the core microbiomes of three termite species (genus *Coptotermes*) and their nest material to gain more insight into the diversity of termite-associated bacteria. Sampling a healthy termite colony over time implicated a consolidated and highly stable microbiome, pointing toward the fact that beneficial bacterial phyla play a major role in termite fitness. In contrast, there was a significant shift in the composition of the core microbiome in one nest during a fungal infection, affecting the abundance of well-characterized *Streptomyces* species (phylum Actinobacteria) as well as less-studied bacterial phyla such as Acidobacteria. High-throughput cultivation in microplates was implemented to isolate and identify these less-studied bacterial phylogenetic group. Amplicon sequencing confirmed that our method maintained the bacterial diversity of the environmental samples, enabling the isolation of novel Acidobacteriaceae and expanding the list of cultivated species to include two strains that may define new species within the genera *Terracidiphilus* and *Acidobacterium*.

Keywords: termites, *Coptotermes*, core microbiome, natural products discovery, Acidobacteria, underexplored phyla, social insects, termite-associated microbes

OPEN ACCESS

Edited by:
George Tsiamis, University of Patras, Greece

Reviewed by:
Tomislav Cernava, Graz University of Technology, Austria
Javier Pascual, Darwin Bioprospecting Excellence, Spain

***Correspondence:**
Till F. Schäberle
Till.F.Schaeberle@agrar.uni-giessen.de
Jens Glaeser
Jens.glaeser@evotec.com
Andreas Vilcinskas
Andreas.Vilcinskas@agrar.uni-giessen.de

†Present address:
Peter Hammann, Evotec International GmbH, Göttingen, Germany
Jens Glaeser, Evotec International GmbH, Göttingen, Germany

Specialty section:
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

Received: *21 August 2020*
Accepted: *19 October 2020*
Published: *06 November 2020*

Citation:
Oberpaul M, Zumkeller CM, Culver T, Spohn M, Mihajlovic S, Leis B, Glaeser SP, Plarre R, McMahon DP, Hammann P, Schäberle TF, Glaeser J and Vilcinskas A (2020) High-Throughput Cultivation for the Selective Isolation of Acidobacteria From Termite Nests. Front. Microbiol. 11:597628. doi: 10.3389/fmicb.2020.597628

INTRODUCTION

Subterranean termites play a key role in the decomposition of plant biomass (Scheffrahn et al., 2015; Kuwahara et al., 2017). Their ingestion and degradation of wood can change the composition of soils and remodel entire landscapes (Bonachela et al., 2015). Furthermore, some termites cause damage valued at more than US$ 22 billion p.a. by attacking wooden structures (Chouvenc et al., 2011; Chouvenc et al., 2016). The genus *Coptotermes* contains the largest number of economically destructive termite species (Scheffrahn et al., 2015), including *C. formosanus* (Su, 2003) and *C. gestroi* (Jenkins et al., 2007; Vargo and Husseneder, 2009), both of which are native to Asia but have spread to other areas as invasive pests. Furthermore, *C. testaceus* is the dominant termite species infesting living trees in the central Amazonian rain forests (Apolinário and Martius, 2004).

Termites are eusocial insects with worker, soldier and reproductive castes. Their social lifestyle includes allogrooming and trophallaxis, which requires frequent direct contact among individuals

2.7.1 Isolation Background

Samples of *Coptotermes* termite nest material (surface, wood, and carton compartments) were collected from long-term laboratory colonies maintained at the Federal Institute for Materials Research and Testing (Berlin, Germany). These systems provided stable, microbially diverse environments that allowed the targeted enrichment of underexplored bacterial phyla, particularly Acidobacteriota. Nest material from *C. testaceus* was chosen for detailed cultivation and sequencing analyses due to its consistently high Acidobacteriota abundance.

2.7.2 Phylogenetic Reconstruction

To resolve the phylogenetic placement of the Acidobacteriota isolates obtained through high-throughput cultivation, nearly full-length 16S rRNA gene sequences were amplified using primer pair E8F (5′-GAG TTT GAT CCT GGC TCA G-3′) and 1492R (5′-ACG GYT ACC TTG TTA CGA CTT-3′) (Lane, 1991). Sequences were compared against the NCBI 16S rRNA reference database using BLASTn, and the 40 closest matches were retrieved for phylogenetic analysis. Multiple sequence alignment was carried out in ClustalW using default parameters, and phylogenetic trees were constructed in MEGA v7.0.26 employing the maximum-likelihood method and the Tamura–Nei substitution model with 1,000 bootstrap replicates to assess branch support. The resulting tree was graphically edited and annotated using iTOL v4.4.2**.**

The analysis placed the isolates FhG110202**,** FhG110206, and FhG110214 within Acidobacteriota subgroup 1, closely related to the genera *Acidobacterium* and *Terracidiphilus*. Pairwise sequence identities of 97.1–98.9 % indicated that at least two of the isolates may represent novel species-level taxa within these genera (Figure **14**).

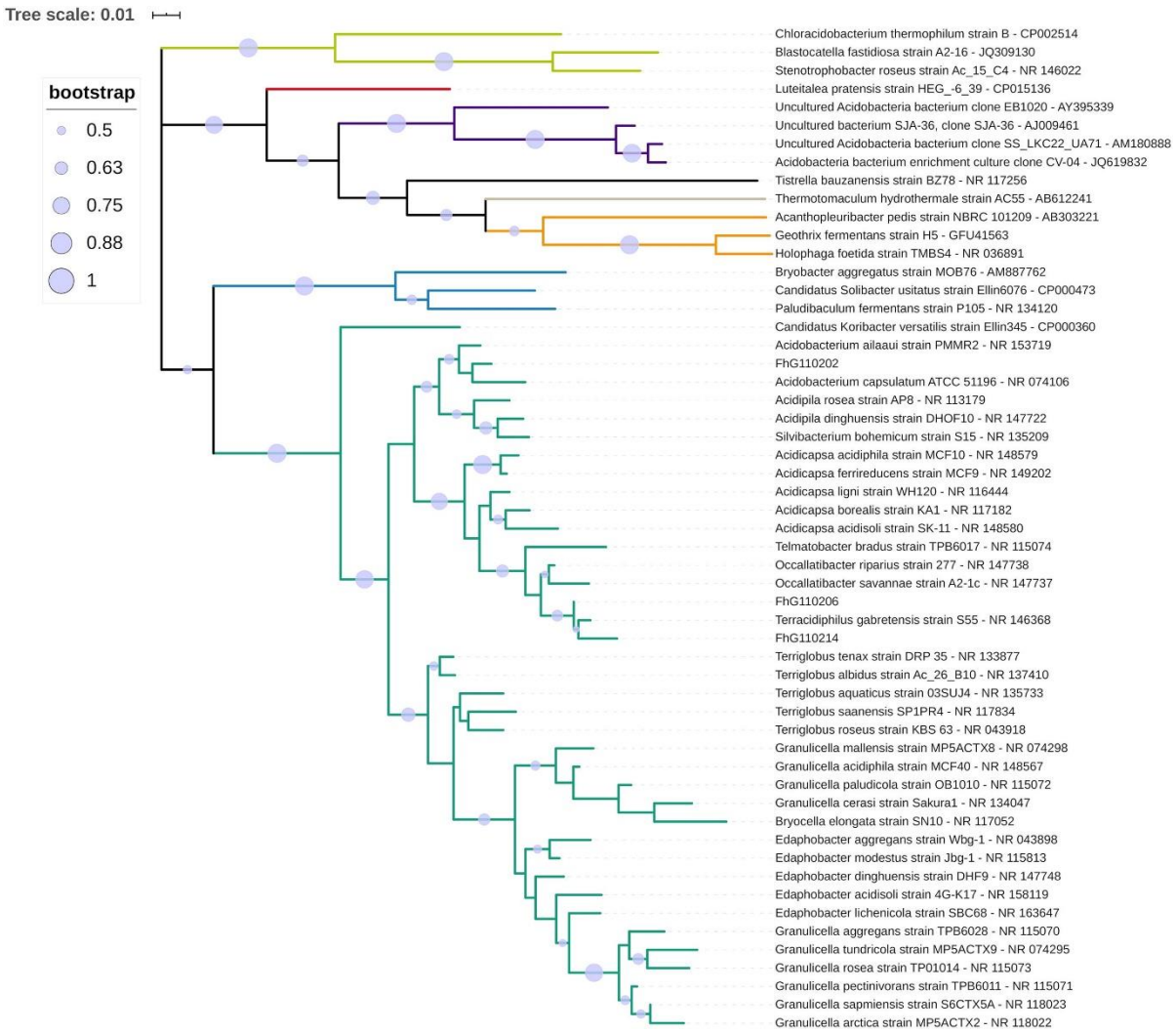

Figure 14: Phylogenetic placement of Acidobacteriota isolates FHG110202, FHG110206, and FHG110214. The maximum-likelihood tree was constructed from a ClustalW alignment of 16S rRNA gene sequences spanning positions 113–1,357 (based on *E. coli* numbering; Brosius *et al.*, 1978). The dataset included the three isolates, the 40 most similar reference sequences from NCBI, and representatives of Acidobacteriota subgroups 3, 4, 6, 7, 8, and 10. Tree reconstruction was performed in MEGA v7.0.26 using the Tamura–Nei model with 1,000 bootstrap replicates. Bootstrap support values >50% are indicated by filled circles, and subgroup affiliations are color-coded. Newly isolated strains are marked with green arrows. Branch lengths represent the number of substitutions per site

This phylogenetic framework confirmed the taxonomic affiliation of the cultivated strains and extended the diversity of available Acidobacteriota representatives for future genome sequencing and functional analysis (see Result 2.8).

2.8 Tryptophan-Driven Metabolomic Shift in Acidobacteriaceae Reveals Phytohormones and Antifungal Metabolites

Parts of this chapter have been published as a preprint article on bioRxiv and can be found at: DOI:10.1101/2025.10.02.680145. It was submitted for publication in November 2025.


bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

Scan for

full publication

HOME| SUBI

New Results

Follow this preprint

Tryptophan-Driven Metabolomic Shift in *Acidobacteriaceae* Reveals Phytohormones and Antifungal Metabolites

Celine M. Zumkeller, Christoph Hartwig, Walter Lanzalonga, Maria A. Patras, Yang Liu, Michael Marner, Sanja Mihajilovic, Marius S. Spohn, Till F. Schäberle

doi: https://doi.org/10.1101/2025.10.02.680145

This article is a preprint and has not been certified by peer review [**what does this mean?**].

Abstract | Full Text | Info/History | Metrics | Preview PDF

Abstract

Acidobacteriota is one of the most abundant phyla in soils and has recently attracted attention for its potential role in promoting phytosanitary benefits. The metabolomic capabilities of this phylum remain poorly characterised, with few experimentally confirmed metabolites described. To address these gaps, we combined metabolomic profiling with comparative genomic analysis to explore the functional potential of novel strains within the *Acidobacteriaceae* family. Genome mining across the phylum for plant-growth-promoting traits revealed the presence and taxon-specific enrichment of genes related to phytohormone production, as well as other genes associated with plant-beneficial microorganisms. When tryptophan was added to the cultivation medium, it triggered a strong metabolic response in the strains, resulting in measurable changes in the production levels of phytohormones such as indole-3-acetic acid and indole-3-pyruvate. Building on these findings, we examined the metabolic reprogramming caused by tryptophan supplementation, which suppressed the growth of phytopathogenic fungi, leading to the identification of malassezindoles and pityriacitrins as active agents. This was confirmed by isolating and elucidating the structure of pityriacitrin B and one of its methyl esters through NMR studies. Overall, these findings shed light on the previously unexplored metabolic potential of the Acidobacteriota phylum, emphasising its ecological importance for phytosanitary applications.

Competing Interest Statement

The authors have declared no competing interest.

2.8.1 Background

Ubiquitous across various ecosystems, including soil (Eichorst *et al.*, 2007a; Jones *et al.*, 2009; Koch *et al.*, 2008), marine sediments (Flieder *et al.*, 2021), wetlands (Belova *et al.*, 2018; Mehta-Kolte and Bond, 2012; Pankratov and Dedysh, 2010), as well as extreme environments such as permafrost (Sipes *et al.*, 2024) and acidic mine drainages (Kishimoto *et al.*, 1991), Acidobacteriota represent one of the most widespread bacterial groups (Ludwig *et al.*, 1997). The phylum is phylogenetically diverse (Quaiser *et al.*, 2003), classified into 26 subdivisions (Dedysh and Yilmaz, 2018) based on 16S rDNA sequencing. Despite their ubiquity, the Acidobacteriota have historically been underexplored due to the limited availability of cultured representatives, which has restricted experimental characterisation and genomic insights (Ward *et al.*, 2009). Today, the availability of high-quality metagenome-assembled genomes, combined with the growing number of cultured representatives and isolate genomes, has significantly broadened our understanding of the metabolic capabilities and ecological roles of this phylum (Dyksma and Pester, 2023; Lee *et al.*, 2008; Ward *et al.*, 2009).

Initially described as slow-growing oligotrophs (Foesel *et al.*, 2013) With limited metabolic dynamics, adapted to nutrient-depleted environments, modern insights now reveal their versatility and critical involvement in biogeochemical processes, including sulphur (Hausmann *et al.*, 2018) and nitrogen (Reji and Zhang, 2022) cycling as well as H_2 consumption (Giguere *et al.*, 2021). Furthermore, they can degrade plant-derived polysaccharides using their rich repertoire of carbohydrate-active enzymes, such as cellulases, chitinases, and xylanases (Gonçalves *et al.*, 2024). Finally, early studies dating back to 2006 proposed a substantial potential for secondary metabolite production by biosynthetic gene clusters, identifying novel Polyketide synthase gene clusters of acidobacterial origin from soil (Crits-Christoph *et al.*, 2018a; Parsley *et al.*, 2011). Subsequent genomic (McReynolds *et al.*, 2025) and metagenomic studies confirmed this finding, identifying BGC-enriched lineages such as the *Acanthopleuribacteraceae* (Leopold-Messer *et al.*, 2023) and the genus *Candidatus Angelobacter* (Crits-Christoph *et al.*, 2018a). Yet, translating bioinformatic findings into experimental validation remains difficult and still requires the availability of culturable representatives. Notably, Leopold-Messer *et al.* (Leopold-Messer *et al.*, 2023)successfully retrieved marine sponge symbiont strains from the *Acanthopleuribacteraceae* family, marking the first description of non-ribosomal peptide (NRP)-polyketide (PK)-derived secondary metabolites from this phylum (Leopold-Messer *et al.*, 2023).

To overcome these gaps, various cultivation approaches have been developed to selectively favour acidobacterial growth and retrieve single-isolate strains from complex environmental samples (Campanharo *et al.*, 2016; Eichorst *et al.*, 2007a; Oberpaul *et al.*, 2020; Sait *et al.*, 2002). These approaches have significantly increased the number of isolated strains; however, they still only represent a fraction of the predicted phylum diversity. Most of the strain isolates are retrieved from soil (McReynolds *et al.*, 2025), where Acidobacteriota have consistently shown high abundances, ranging from 20% to 40% of the detected 16S rRNA(Felske *et al.*, 2000; Kuske *et al.*, 1997) sequences. Cultivation strategies often utilise plant polymers for Acidobacteriota enrichment (Campanharo *et al.*, 2016; Eichorst *et al.*, 2011; Oberpaul *et al.*, 2020), providing insight into their predicted function in enhancing soil health by degrading undigestible plant polymers and improving nutrient content in the soil. Consequently, several studies investigate the potential plant-growth-promoting (PGP) effects of the phylum. In 2016, the first evidence of a direct plant-interaction was described by Kielak *et al.*

(Kielak *et al.*, 2016b). They found that some acidobacterial strains produced IAA and also demonstrated a PGP effect on the model plant *Arabidopsis thaliana* (Kielak *et al.*, 2016b). Moreover, Acidobacteriota are recognised for synthesising hopanoids (Damsté *et al.*, 2017), bacterial analogues of sterols, known to enhance stress tolerance in soil bacteria as an integral component of the bacterial membrane. Their unique exopolysaccharide production (Kalam *et al.*, 2020; Kielak *et al.*, 2017) further suggests a potential ability to endure dry conditions and possible roles in sustaining plant health, e.g., protecting against desiccation. Complementary metagenomic and meta-transcriptomic studies have revealed a significant enrichment of genes related to PGP. Gonçalves *et al.* (2024) profiled 758 soil-associated metagenome-assembled genomes (MAGs) and identified plant-growth-promoting-traits (PGPTs) involved in phytohormone production, nitrogen fixation, phosphorus solubilisation, siderophore production, and exopolysaccharide biosynthesis (Gonçalves *et al.*, 2024). These traits were explicitly enriched in certain families of the Acidobacteriota, namely *Acidobacteriaceae*, *Pyrinomonadaceae*, and *Koribacteriaceae*. Meta-transcriptomic evidence shows that these traits are actively transcribed in their respective environmental context (Gonçalves *et al.*, 2024). Furthermore, transcriptomic research has demonstrated the metabolic dynamics and responsiveness to external stimuli of cultured Acidobacteriota strains. For instance, *Acidobacterium capsulatum* responds to oxygen deprivation by switching to a respiro-fermentative state, resulting in the production of acetate and ethanol —a strategy that allows for survival in habitats with fluctuating O_2 levels, such as soil (Trojan *et al.*, 2024). The oligotrophic lifestyle typical of the Acidobacteriota has been investigated in a *Granulicella* species, which responds to high-carbon availability by upregulating stress-related genes, indicating sensitivity to nutritional changes (Costa *et al.*, 2020). While many ecological functions are inferred from meta(transcriptomics) data, experimentally generated metabolomic data remain limited, primarily focusing on the characterisation of membrane lipids and their production in response to different cultivation parameters such as temperature, pH, and O2 (Halamka *et al.*, 2024; Sinninghe Damsté *et al.*, 2011).

To address these gaps, we integrated metabolomic profiling with comparative genomic analysis to investigate the functional potential of novel strains of the *Acidobacteriaceae* family. Phylum-wide genome mining for PGPTs revealed the presence and taxon-specific enrichment of genes associated with phytohormone production, as well as additional genes related to PGP effects. The addition of tryptophan (Trp) to the cultivation medium elicited a strong metabolic response in the strains, resulting in quantitatively determined shifts in the production levels of the phytohormones indole-3-acetic acid (IAA) and indole-3-pyruvate (iP). Prompted by these findings, we evaluated the metabolic reprogramming induced by Trp supplementation, revealing suppressive effects on the growth of phytopathogenic fungi, and subsequently identified pityriacitrins and malassezindoles as the causative agents. Collectively, these findings highlight the previously underexplored metabolic space of the Acidobacteriota phylum, underscoring its ecological significance for phytosanitary purposes.

2.8.2 Methods

2.8.2.1 Microbial Cultivation of Acidobacterial Strains

Acidobacterial strains (Table 8) were routinely cultivated in R2A Medium (R2A Broth, HIMEDIA, Maharashtra, India) buffered with MES-hydrate(2-(N-Morpholino)ethanesulfonic acid hydrate, CAS: 1266615-59-1) (1.95

g/l) at pH 6.0. Cultures were incubated at 28°C with 75 rpm. To sustain healthy and continuous growth, 10% of the 7-day-old preculture (30 ml in 100 ml Erlenmeyer flask) was used to inoculate the main cultures (100 ml in 300 ml Erlenmeyer flasks). For quality control, we transferred 100 µl after each inoculation or sampling step to an R2A-MES-buffered agar plate to observe colony morphologies and identify the growth of contaminant microorganisms. Additionally, we performed 16S rRNA gene amplification testing with primer pair E8F (50-GAG TTT GAT CCT GGC TCA G-30) and 1492R (50-ACG GYT ACC TTG TTA CGA CTT-30) (Lane, 1991). For the OSMAC cultivation approach, we performed 24-well-scale and Erlenmeyer flask-scale cultivation. We prepared forty 24-well Duetz system plates with 20 different media (each medium 4 ml) and autoclaved the plates before inoculation (10%) with a preculture grown for 14 days. Applied media were diluted medium 5254 (Khosravi Babadi *et al.*, 2021)(2 g soluble starch, 2 g glycerol, 0.5 g corn steep liquor 1 g peptone, 0.4 g yeast extract, 0.2 g NaCl; pH 5.5, buffered with 1.95 g/L MES), HD (0.5 g casein peptone, 0.1 g glucose, 0.25 g yeast extract; pH 5.0, unbuffered; DSMZ ID 1135), and PSYA 5 (Campanharo *et al.*, 2016)(1.8 g KH_2PO_4, 0.2 g $MgSO_4$ x $7H_2O$, 1 g yeast extract; optionally with addition of sucrose at 0.5% or 3%; pH 5.0, unbuffered) as well as a diversity of VL55-based and R2A-based (1-fold, 2-fold, 4-fold) media generated by optional addition of different carbon sources (Pektin, GlcNAc – 0.05%; Xylan, Cellobiose – 0.5%; and Sucrose 3%; applied to 2-fold R2A)(Table 9) (Oberpaul *et al.*, 2022, 2020). Furthermore, as a surface attachment enabling strategy, we added ceramic beads (hollow cylinders; 3 pcs/well plus 4 ml of media), quartz beads (round; 10 pcs/well plus 4 ml of media), or solidified the respective medium using 0.8% phytagel. The plates were incubated at 28°C, shaking at 120 rpm or with manual shaking once per day. Cultivation in a 300 ml Erlenmeyer flask was executed in 100 mL media by inoculation (10%) with a preculture grown for 7 days. The cultures were incubated at 28°C for 14 days with or without shaking at 75 rpm. We added 20 g of ceramic beads (Eheim mech) to non-shaking cultures. After cultivation, cultures were frozen, freeze-dried and extracted (see below).

For Trp-induced kinetic studies, we used a 1.2-fold concentrated R2A medium buffered with MES at a pH of 6.0, which was supplemented in a 4:1 ratio with 1% sterile-filtered Trp (CAS 73-22-3) to a final concentration of 0.2% Trp and a 1-fold concentrated medium. Kinetic studies were performed in 300 ml flasks at least in duplicates. For sample collection, we transferred 4 ml of culture to 24-deep-well plates for further processing. In parallel, the remaining culture sample was used for quality control by agar plating (R2A-MES, pH 6.0) and 16S rDNA PCR. After the final harvest day, the 24-well plates were freeze-dried and extracted using methanol (SI-extended methods and results).

2.8.2.2 Bioassays

Antimicrobial Screenings of crude extracts were performed using microplate broth dilution assays at the final assay concentration of 0.25-fold, 0.5-fold, one-fold and two-fold (in relation to the culture volume) against bacteria and fungi (*Escherichia coli* ATCC 35218, *Bacillus subtilis* DSM 10, *Mycobacterium smegmatis* ATCC 607, *Candida albicans* FH 2173 and *Septoria tritici* MUCL 45407) as described previously (Oberpaul *et al.*, 2022). Crude extracts showing at least 70% growth inhibition were considered bioactive and were subjected to microfractionation for dereplication of bioactivity-causing metabolites (Oberpaul *et al.*, 2022). Testing of crude extracts on barley seedlings was performed in glass vials filled with 5 ml of SH medium, inoculated with one surface-disinfected barley seed per vial (SI-extended methods and results).

2.8.2.3 Creation and Curation of the Acidobacteriota Genome Dataset

As of March 2023, Acidobacteriota assemblies matching our selection criteria (NCBI:txid, filter: RefSeq annotation) were retrieved from NCBI in .gbff file format. In addition, four Acidobacteriota assemblies, from strains isolated from termite nest material (FHG110202, FHG110214), soil (FHG110511) and a genome generated for *A. borealis* DSM 23886 complemented the dataset (n = 1425). Genome assemblies were selected based on their completeness and contamination using CheckM2(Chklovski *et al.*, 2023a). Only high-quality genome assemblies defined by MIMAG (Bowers *et al.*, 2017) evaluation criteria demonstrating 90-100% completeness and 0-5% contamination were retained for further evaluation (n = 648). For this, redundancies and taxonomic affiliation were checked by GTDB-Tk (Parks *et al.*, 2022). In total, 618 curated genome assemblies were processed for metadata evaluation, BGC potential, and PGPT mining (SI-extended methods and results for details - 6.2.1, page 158). To correlate the genomic data to the taxonomic affiliation, a whole-genome sequence-based phylogenetic tree of the GTDB-generated multiple-sequence alignment file was visualised using iTOL v4 (Letunic and Bork, 2019).

2.8.2.4 Genomic Analysis of PGPT Traits

The coding sequences from the previously described genomic dataset were extracted using a custom Python pipeline built with BioPython (SeqIO). To generate FASTA documents for further processing, we retrieved the gene name, product description and translated protein sequence for each CDS. For pre-existing FASTA files, the extraction of CDSs can be skipped. To document the number of CDSs extracted per genome, a log file was generated. The PLABASE PGPT database (Patz *et al.*, 2021) (PGPT_BASE_nr_Aug2021n_ul_1.dmnd) was used as a reference for a similarity search against the extracted protein sequences using DIAMOND v2. The script supports various sensitivity modes, enabling users to select the most suitable mode based on their specific input. The parameters and applied evaluation thresholds used in this study are described in the Extended Methods and Results section of the supplementary information.

2.8.2.5 Mass Spectrometry

Ultra-high-performance liquid chromatography (UHPLC) coupled with quadrupole time-of-flight (QTOF) high-resolution mass spectrometry (HR-MS) and tandem MS (MS/MS) was performed using an Agilent 1290 Infinity LC system connected to a maXis II QTOF mass spectrometer (Bruker Daltonics) equipped with an electrospray ionisation (ESI) source. Separation of samples was achieved on an Acquity UPLC BEH C18 column (130 Å pore size, 1.7 µm particle size, 2.1 × 100 mm) fitted with a matching pre-column (2.1 × 5 mm, 1.7 µm, 130 Å) at a temperature of 45°C. Chromatographic separation was achieved using a linear gradient elution of solvent A (water with 0.1% formic acid) and solvent B (acetonitrile with 0.1% formic acid) at a constant flow rate of 0.6 ml/min. The gradient program was as follows: 95% A at 0.00–0.30 min, gradually decreasing to 4.75% A by 18.00 min, then to 0% A at 18.10 min, held until 22.50 min, returned to 95% A at 22.60 min, and equilibrated until 25.00 min. Mass spectra were recorded in positive mode over a mass range of m/z 50–2000 with a spectra rate of 1 Hz. MS/MS acquisition was performed at a 6 Hz scan rate, targeting the top five most intense ions per full mass spectrum for collision-induced dissociation (CID) using nitrogen, as per the table published by Eichberg *et al.* (Eichberg *et al.*, 2024). Precursors were excluded after two spectra, with a release duration of

0.5 min, and re-evaluated if their signal intensity increased by ≥1.5-fold. All data analysis was performed with DataAnalysis 4.4 (Bruker, Billerica, MA, USA).

2.8.2.6 Chemotype-Barcoding Matrix, Multivariate Analysis and Statistics

Raw data processing was performed using DataAnalysis 5.2 (Bruker) as previously published, and the complete data set was copied and processed with both line spectra thresholds (here: 5,000 and 10,000) (Brinkmann *et al.*, 2022a). Bucketing was performed on both sets at the same time, resulting in one table containing all buckets deemed identical for every sample (under both thresholds) and by applying the data curation measures as described in the extended methods and results of the supplementary information. This analysis revealed a final table of 219 samples with 3,092 buckets (OSMAC) and 48 samples with 2,759 buckets (Trp-Induction), which were used for the respective barcode generation and for categorising. The generated bucket tables were used as (i) a binary presence/absence matrix (result 1) or (ii) a quantitative intensity matrix (result 5). For PCoA on the binary matrix, (i) Jaccard dissimilarities were computed (skbio.diversity.beta_diversity, metric="jaccard"), and PCoA (skbio.stats.ordination.pcoa) was done using scikit-bio (v0.7.0). PERMANOVA (skbio.stats.distance.permanova) assessed group differences with 999 permutations. We tested the effects of medium, cultivation size (small-scale vs. large-scale), and bead condition (Quartz/Ceramic/None) in a small-scale subset. For PCA on quantitative data (ii), we log2-transformed peak area intensities using scikit-learn (v1.3) in Python. Before PCA, undetected features were defined as zero, and samples were normalised to the injection volume wherever indicated. All visualisations were generated using Matplotlib, with attributes changed based on metadata information through shape, colour, and border style to reflect replicates and experimental conditions.

2.8.3 Results

2.8.3.1 Diverse Cultivation Strategies Reveal the Production of Phytohormones

Extensive metagenomic data from globally distributed bioresources indicate an underexplored biosynthetic and ecological potential of the Acidobacteriota phylum. During our ongoing efforts to isolate and bioprospect bacterial diversity for its potential in the production of natural products and related environmental functions, we retrieved three Acidobacteriota strains from termite nest material (Oberpaul *et al.*, 2020) and one strain from a German soil sample (Oberpaul *et al.*, 2022). To evaluate their biosynthetic potential, we isolated their total DNA together with that of the type strain *A. borealis* DSM 23886 and performed whole-genome sequencing. The generated genomic data underwent quality check using CheckM2 (Chklovski *et al.*, 2023a). By applying recognised quality criteria (completion >90% and contamination <5% (Bowers *et al.*, 2017)), one assembly (FHG110206) was disqualified for further analysis, while the other three assemblies were used for genome-based taxonomic affiliation by the genome taxonomy database (GTDB) (Parks *et al.*, 2022) (Supplementary Table 1, published along with the preprint – page 64). This positioned the termite nest-associated strains as members of the genera *Acidobacterium* (FHG110202) and *Terracidiphilus* (FHG110214), while the soil-associated strain FHG110511 was identified to be an *Edaphobacter*. All three strains belong to the family of *Acidobacteriaceae* (collectively referred to as FHG *Acidobacteriaceae* in this manuscript). Thus, we selected further available strains of this family (*Edaphobacter aggregans* DSM 19364*, Acidobacterium sp.* S8*, Silvibacterium bohemicum* S15*, Acidicapsa borealis* DSM 23886*, Granulicella rosea* DSM 18704, and *Bryocella*

elongata DSM 22489)(Table 10) and performed an extensive OSMAC approach (Bode *et al.*, 2002), conceptualised to increase the likelihood of covering a favourable condition for the strains to proliferate and promote gene expression. To this end, we modified the carbon sources, vessel size, and shaking speed, and applied solidification and surface texture methods using quartz or ceramic beads, potentially enabling the surface attachment of the cells. Cultivations were performed on a 100 ml scale using 300 ml Erlenmeyer flasks or at a 4 ml scale using 24 well Duetz system plates (Table 9). Samples were taken on days 5, 7, 10, 12, and 14 of the respective cultivations. The generated samples were dried by lyophilisation and extracted with methanol, resulting in a total of 892 organic extracts. An aliquot of each extract was reconstituted in dimethyl sulfoxide (DMSO) and used to determine the antimicrobial bioactivity by screening against the bacterial indicator strains *Escherichia coli* ATCC 35218, *Bacillus subtilis* DSM 10, and *Mycobacterium smegmatis* ATCC 607, as well as the fungus *Candida albicans* FH2173. This, however, did not reveal any reproducible antimicrobial effect enabling traceability of a causative agent (data not shown). Furthermore, we conducted a bioactivity-independent and metabolomics-driven bioprospecting campaign. Initially, we focused on the metabolic profiling of strains FHG110202 and FHG110206 and assessed the metabolic composition of their generated extracts (n = 244, including media controls) analytically by performing ultra-high-performance liquid chromatography coupled to high-resolution mass spectrometry (UHPLC-QTOF-HR-MS). Molecular features (*m/z*, retention time) were detected and aligned into buckets. Combining timepoints for each strain-media combination and presence-absence analysis led to 72 conditions (including media controls) with 20285 features present and distributed into 2747 buckets (Figure S2-1). To assess the effect of the medium on the metabolome of the strains, we filtered medium-derived features (n=1795). Pairwise Jaccard distances between samples were ordinated by principal coordinates analysis (PCoA) for strains FHG110206 and FHG110202 (Figure 15A). Metabolomic profiles clustered mainly by medium (colour code) and vessel size (top, left cluster) rather than by strain. We also observe a distinct cluster for medium containing quartz beads (bottom, right cluster). A PERMANOVA analysis confirmed this observation, suggesting a strong OSMAC response. The medium identity accounts for 87% of the variance among strain extracts ($F=7.09$, $p=0.001$), while the strain identity within the same medium explains a much smaller share (1.7%, $F=0.77$, $p=0.001$). Within the small-scale set, the bead condition explained 29.2% of the differences ($F = 7.64$, $p = 0.001$), consistent with the observed quartz-bead cluster. Thus, metabolomes are primarily influenced by the medium and vessel size, with only a small strain-level difference. We also plotted unique features of strain-produced metabolites for individual media conditions (Figure 15), and found that in both strains (FHG110202:FHG110206), media F (87:125), U (65:64), W (86:86), and X (197:116) are enriched in unique features compared to the average unique feature number across all media (22.2:20.7). This indicates that the large-scale cultivation size in 300 ml Erlenmeyer flasks yields a more distinct metabolome compare to the small-scale cultivation in 24-well plates.

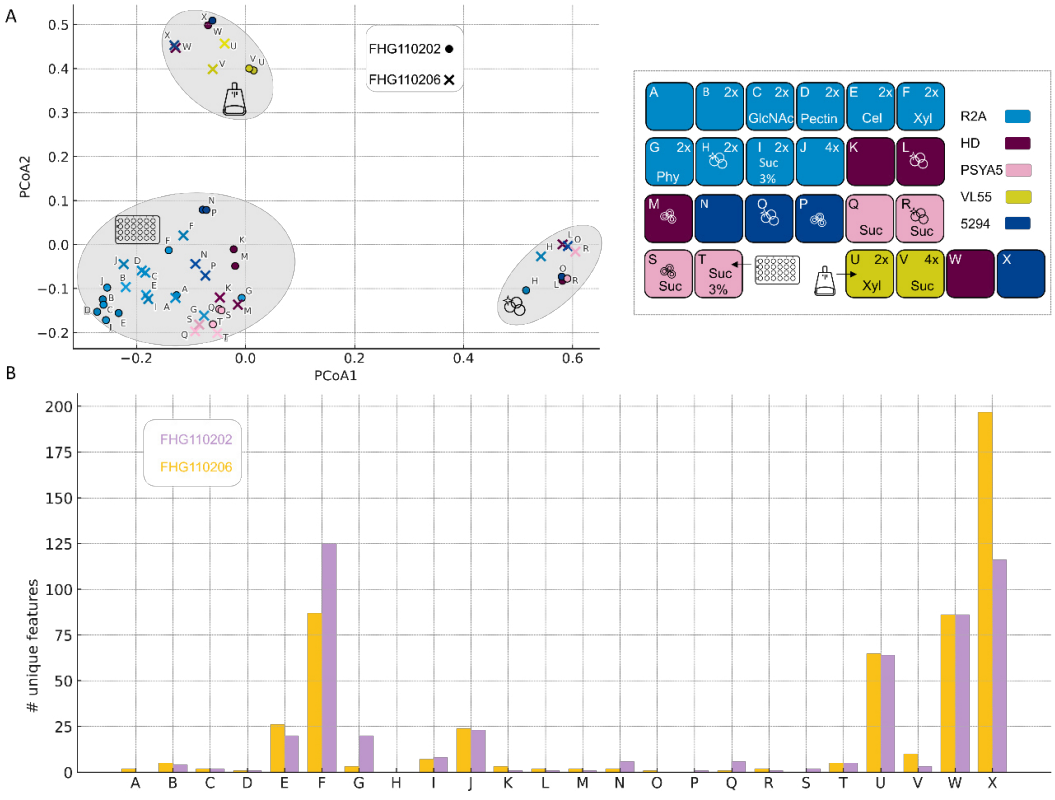

Figure 15: OSMAC metabolome variation by medium, strain, and cultivation scale. The legend details the media compositions (A-X), based on base-media colour, concentration, additives, and bead status. Samples labelled A-T were cultivated in 24-well plates (small-scale), while U-X were grown in 300 ml Erlenmeyer flasks. If not stated otherwise, additives were supplied in concentrations of GlcNAc, N-acetyl-glucosamine (0.05%); Cel, cellobiose (0.5%); Xyl, xylan (0.5%); Phy, phytagel (0.8%); Suc, sucrose (0.5%). (A) PCoA of control-filtered LC/MS presence/absence fingerprints using Jaccard distance. Samples are labelled by the medium letter (A-X) and coloured by base-medium (see legend). Samples from FHG110202 are marked with circles, and FHG110206 with crosses. Light grey ellipses highlight visually distinct clusters, and icons indicate their common features: large-scale flask cultivation (flask), small-scale 24-well cultivation (24-well plate), and quartz beads containing samples (circles). (B) Number of unique features per medium (A-X) and strain as clustered bars. Strains are indicated by colour: FHG110202 (light purple), FHG110206 (yellow). Unique features are defined as not being present in any medium control and only present in the respective medium, but not in any other medium.

The detected metabolite features were matched against our in-house reference compound MS database, which contained approximately 1300 structurally characterised microbial metabolites at the time of data processing. By comparison of mass to charge ratios and retention time, we were able to annotate the phytohormone IAA in the extracts of our FHG *Acidobacteriaceae* strains with a confidence level 1 (Identified metabolite: as proposed by the Chemical Analysis Working Group of the Metabolomics Standards Initiative (Sumner *et al.*, 2007)). The production of the phytohormone IAA has been reported for members of the *Acidobacteriaceae* (Kielak *et al.*, 2016b). To systematically describe the cultivation conditions that favour IAA production, we visualised the peak area of IAA in FHG110202 as an example across all tested media and setups (Figure S2-2). While IAA was not produced under conditions leading to the isolation of the strains (VL-55 base media), it became detectable in more complex media. Among these, R2A-based cultivation, with agitation, consistently showed the best IAA production. To further analyse the phytohormone production of FHG110202, we included

additional standards: Kinetin (CAS 525-79-1), Trans-zeatin (CAS 1637-39-4), Indole-3 Butyric Acid (CAS 133-32-4), Brassinolide (CAS 72962-43-7), jasmonic acid methyl ester (CAS 39924-52-2), jasmonic acid (CAS 77026-92-7), abscisic acid (CAS 21293-29-8), indole-3-propionic acid (CAS 830-96-6), 1,2-Diphenylurea (CAS 102-07-8) and *N6*-(*Δ2*-Isopentenyl)adenine (iP)(CAS 2365-40-4) to our *in-house* reference database and explicitly checked for their presence in the generated crude extracts. This revealed the presence of the cytokinin iP at a confidence level of 1. The detection of cytokinin iP as a biosynthetic product of members of the Acidobacteriota phylum is a new finding.

2.8.3.2 Biosynthetic Potential of the Acidobacteriota Phylum

To complement the metabolomic workflow, we evaluated the genomic potential of our FHG *Acidobacteriaceae* and comparatively aligned it to the broader Acidobacteriota landscape. Therefore, we initially searched for publicly available genomes matching the criteria Query: "Acidobacteriota" NOT "anomalous" in the NCBI database. This yielded 1428 genomes that underwent quality control using CheckM2 (Chklovski *et al.*, 2023a) (Figure S2-3), applying the above-defined quality criteria. This resulted in a final dataset comprising 618 genome assemblies. Overall, most of the analysed genomes were retrieved from metagenomes (n = 533; 86%) (Supplementary Table 1). A detailed bioinformatic analysis of this dataset is reported in the Extended Methods and Results section of the supplementary information. The taxonomic affiliation was established by GTDB (Parks *et al.*, 2022) and the generated multiple-sequence alignment was used to build a phylogenetic tree. The bioinformatically predicted features of each strain were assigned and visualised within the tree using iTOL (Letunic and Bork, 2019). The genomes in the dataset span 12 classes (Figure 16), dominated by the *Acidobacteriae* (n = 242), *Blastocatellia* (n = 130) and *Vicinamibacteria* (n = 98), which also represent the majority of today's strains brought to culture (76/86-88%). To evaluate the biosynthetic potential of the Acidobacteriota phylum, we predicted BGCs using the standalone version of antiSMASH (Blin *et al.*, 2021b). In total, 3865 BGCs were detected in the genome dataset. The most abundant BGC class in the dataset is ribosomally produced and post-translationally modified peptides (RiPPs) (n = 1152, 30%), followed by non-ribosomal peptide synthetases (NRPSs) (n = 978, 26%) and terpenes (n = 777, 13%). The total average cluster number per class in our acidobacterial dataset is 6.2 ± 4.4 BGCs. The *Acidobacteriae* class exhibits the highest overall number of BGCs, with an average of 7.4 ± 3.7 BGCs per genome, followed by UBA6911 and Mor1, with 6.3 ± 2.5 and 5.8 ± 2.1 BGCs, respectively. Dedicated analysis of the FHG *Acidobacteriaceae* revealed an average of 9 ± 3.46 BGCs, slightly higher than for the overall phylum and the *Acidobacteriae* class value, but comparable to the overall family value (7.9 ± 3.0). Again, the most represented group of BGCs is the RIPPs (sum: 10), followed by the terpenes (sum: 7) and the NRPS (sum: 4) BGCs. Notably, the *Holophagae* class, despite having an average BGC number of 5.1, exhibits a high standard deviation of 10.5, indicating a high potential for some of its representatives. This value can be accounted for by the *Acanthopleuribacteraceae* (Z-score: 3.5), showing high BGC counts of >40, seven times higher than the average Acidobacteriota genome. Intriguingly, the first acidobacterial NPs recently isolated also come from this family (Leopold-Messer *et al.*, 2023). While these talented strains have been isolated from marine animals, Acidobacteriota are known to be habitat generalists with a wide distribution, primarily found in soil ecosystems. They are characterised by a variety of extracellular products, such as BGCs, which are more extensive in soil-preferring lineages compared to their non-soil-preferring relatives, including the *Holophagae*. Although analysing biosynthetic potential at the class level is

useful, our analysis demonstrates the advantages of examining family and genus levels. This approach helps identify the most promising taxonomic hotspots for discovering new natural products (NP).

To complement the evaluation of the Acidobacteriota phylum, we performed sequence similarity calculations on the total of 3865 BGCs in our dataset using BiG-SCAPE, including the MIBiG reference BGCs, and applied a cutoff of 0.5. 1651 BGCs remained as singletons, contrasted by the largest GCF that consists of 25 members, all from the *Acidobacteriae* class. Twenty-five GCFs are specific to one respective taxonomic class (Figure S2-8); most of these GCFs contain more than 10 representatives (16/25), indicating clear taxonomic conservation. This taxonomic conservation is further illustrated by comparing the GCF composition of the different strains using cosine distance and visualising this similarity as uniform manifold approximation and projection (UMAP) (Figure S2-9). Here, strains belonging to the same taxonomic classes cluster within the visualisation, indicating overall conservation between phylogenetic groups. However, sequence similarity calculations revealed that the FHG *Acidobacteriaceae* do not share any of their clusters (Figure S2-8). Further, alignments with MiBiG Reference BGCs revealed a co-clustering of 11 acidobacterial BGCs in six GCFs (Figure 16), showing compositional BGC similarity to, e.g. the myxochromide and the occidiofungin A-reference BGCs (SI-extended methods and results). Intriguingly, the latter is a match to our strain FHG110511, harbouring a PKS/NRPS hybrid BGC that clusters in GCF1959 with the occidiofungin A reference BGC (MiBiG BGC0001711) at a pairwise identity of 70.4% (Figure S2-11). Occidiofungin A is a *Burkholderia contaminans*-produced NP that exhibits broad antifungal activity and is in early clinical development (gel OCF001, phase 1) for use against multidrug-resistant vaginal yeast (*Candida* spp.) infections (Cothrell *et al.*, 2023; Gu *et al.*, 2011). The comparative evaluation of the NRPS and NRPS/PKS hybrid core gene architecture revealed a highly conserved module arrangement with certain discrepancies in predicted A-domain specificity. Testing of the strain's organic extracts against *C. albicans,* however, showed no antifungal activity.

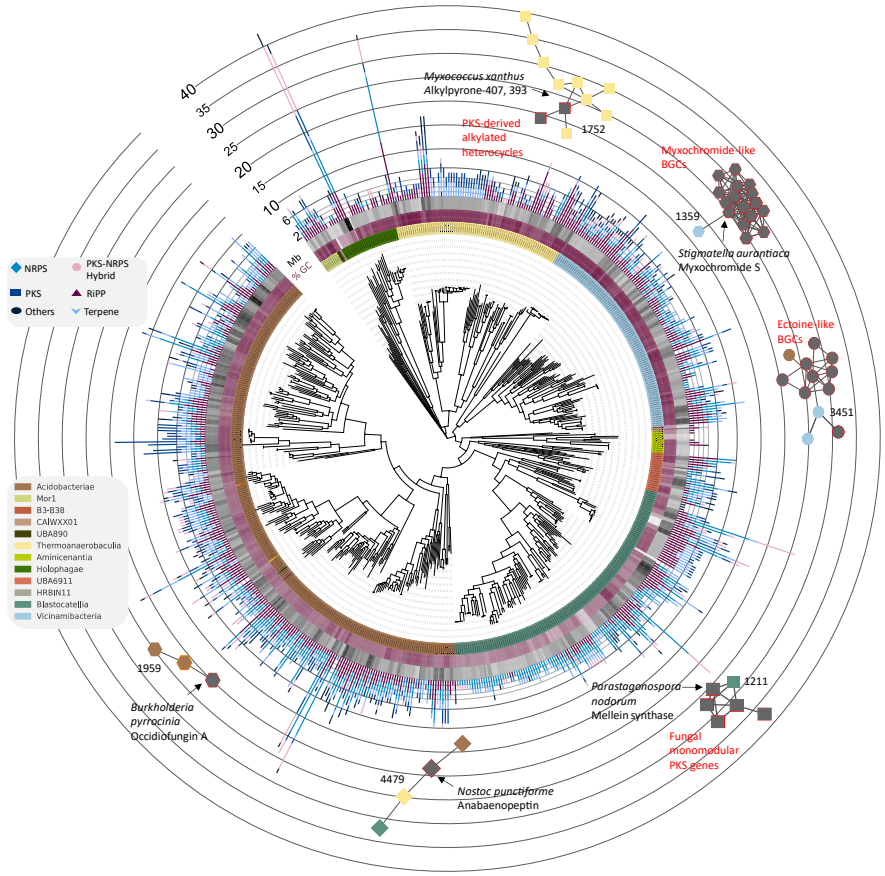

Figure 16: BGC composition of the Acidobacteriota Phylum and BiG-SCAPE Network of acidobacterial BGCs clustering with MIBiG reference clusters. Multiple Sequence Alignments generated by GTDB were used to create a phylogenetic tree of 618 representatives of the Acidobacteriota. Twelve acidobacterial classes were identified by GTDB and are indicated by different colours (ring 1). FHG strains are colour-coded as follows: FHG110214 (green), A. borealis DSM 23886(purple), FhG110202 (yellow), and FhG110511 (orange). Ring two indicates the genome size of the individual representatives, ranging from black (larger genomes) to grey (smaller genomes). Bar charts indicate the total number of antiSMASH-detected BGCs, differentiated into different BGC types.

The occidiofungin-like BGC represents a singleton hit in respect to multimodular NRPS, PKS or hybrid systems within the genomes of the FHG *Acidobacteriaceae*, indicating a generally underdeveloped talent of our strains to produce structurally complex NPs and a possible reason for the lack of antimicrobial activity observations during our initial bioprospecting campaign.

2.8.3.3 Tryptophan Supplementation Increases IAA but Decreases iP Production in Acidobacteriaceae without Enhancing Barley Seedling Biomass

During the metabolomic-based inspection of the OSMAC experiment, we identified the biosynthetic potential of our FHG *Acidobacteriaceae* to produce the phytohormones IAA and iP. To further profile the IAA and iP production of our FHG *Acidobacteriaceae* and the reference strain *A. borealis* DSM 23886, we cultivated those strains in media with and without supplementation of tryptophan (Trp), commonly known to increase IAA production in microorganisms (Spaepen *et al.*, 2007). We performed a kinetics study to analyse the influence of 0.2% Trp on the relative IAA and iP production over 14 days following the methanolic extraction of the cultures. Overall, the amounts of IAA and iP increased over time, reaching their highest levels on days 10 and 12 of growth (Figure 17and Figure S2-12). Indeed, the supplementation of Trp enhanced IAA production significantly in all strains. In contrast, supplementation of Trp consistently suppressed iP production in the analysed strains except for FhG110202 and *A. borealis* DSM 23886, where this effect is less pronounced (Figure 17A). This initial profiling revealed that strain FHG110214 was the most productive iP producer, while exhibiting comparable productivity in IAA to *A. borealis* DSM 23886, FHG110202, and FHG110206. Based on this production profile, we selected FHG110214 and *A. borealis* DSM 23886 to perform an absolute quantification of phytohormones. When supplemented with Trp, *A. borealis* DSM 23886 produced IAA to a final concentration of 0.326 ± 0.063 µg/ml, whereas FHG110214 produced IAA at a concentration of 0.104 ± 0.021 µg/ml. Instead, without supplementation of Trp, the IAA titres dropped to 0.124 ± 0.029 µg/ml (3-fold) and 0.015 ± 0.003 µg/ml (7-fold) in *A. borealis* DSM 23886 and FHG110214, respectively. This production profile is reversed for iP, where Trp supplementation resulted in decreased production for both strains FHG110214 (With Trp: 0.0058 ±0.0012 µg/ml; without Trp: 0.0143 ± 0.0024 µg/ml) and *A. borealis* DSM 23886 (With Trp: 0.005 ±0.0004 µg/ml; Without Trp: 0.0108 ± 0.0008 µg/ml). These production titres are in a comparable range to quantities previously reported for plant-growth-promoting bacteria (PGPB) (Lovecká *et al.*, 2023; Schmidt *et al.*, 2018) and, e.g. 10-fold lower compared to those reported for *Pseudomonas*, *Streptomyces*, and *Enterobacter* species, which have been shown to produce 10–150 µg/ml under optimised conditions. The production of the cytokinin iP has not been described so far for the Acidobacteriota phylum. We quantified its production to 0.003 µg/ml in the most productive strain of our study. This is in alignment with the literature, which reports, for example, 17.5 pmol/ml (translating to 0.00356 µg/ml) in *Achromobacter xylosoxidans* (Santana *et al.*, 2022) and lower concentrations such as 0.0003 µg/ml (1.5 nM) in *Paenibacillus polymyxa* (Environment Canada and Health Canada, 2015).

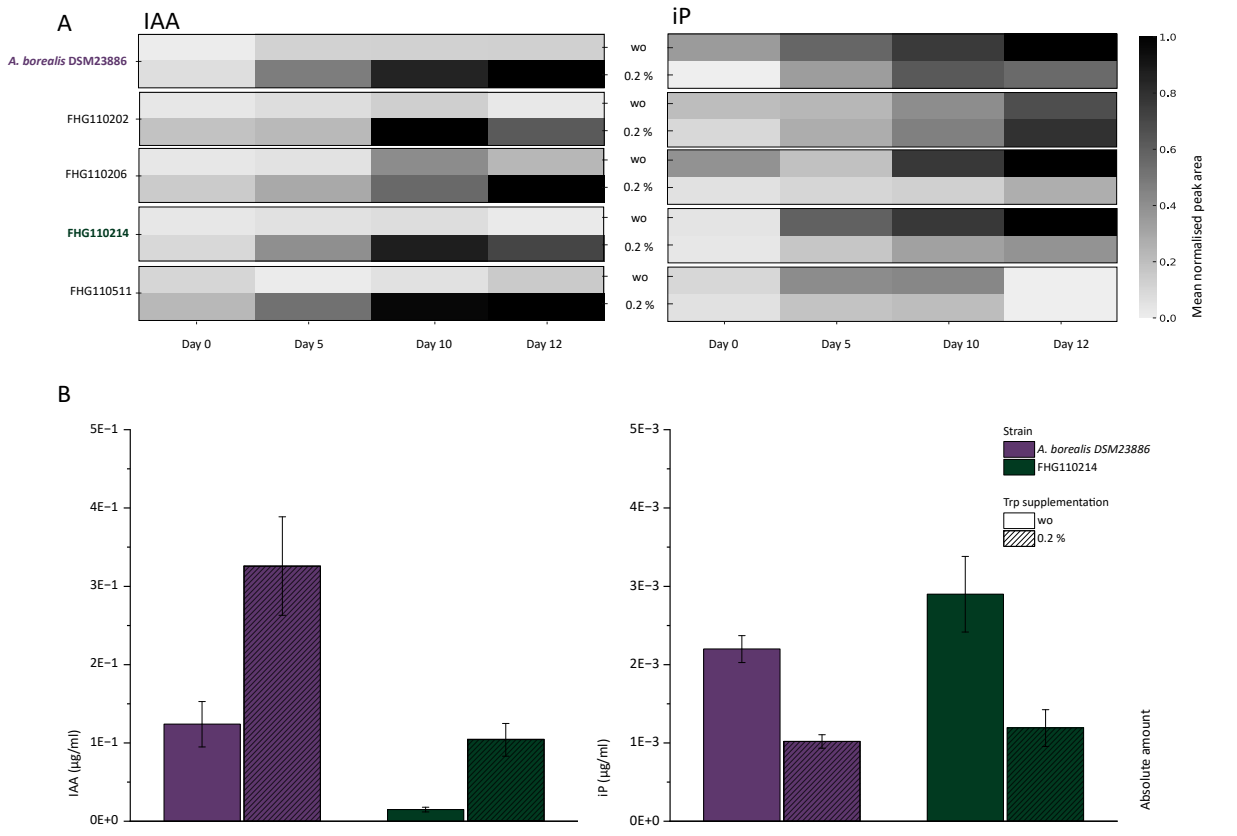

Figure 17: Relative quantification of IAA and iP in FHG Acidobacteriaceae and A. borealis DSM 23886 as normalised peak area (A) and absolute quantification of these phytohormones in the top producers FHG110214 (iP) and A. borealis DSM 23886 (IAA)(B). A Raw peak areas were normalised against the largest peak area within one biological kinetic study for IAA and iP separately. Mean values were calculated and plotted as a heatmap to indicate temporal and dynamic induction patterns between strains and supplementation status. Trp supplementation results in higher IAA production but generally has a negative impact on iP production. This effect is less pronounced in FHG110202 and A. borealis DSM 23886. B Absolute concentrations were derived from a calibration curve of IAA and iP, correlating peak area with given concentrations. Cultivations were repeated in triplicate, and the average of the calculated concentrations is shown for A. borealis DSM 23886 (green) and FHG110214 (purple).

Previously, it has been shown that culture broths of IAA-producing *Acidobacteriaceae* promoted root and shoot growth in *Arabidopsis thaliana* (Gonçalves *et al.*, 2024; Kielak *et al.*, 2016b) and growth and chlorophyll content in duckweed (Yoneda *et al.*, 2021). Furthermore, cytokines have also been shown to have a beneficial impact on plant growth and defence response against pathogens and pests (Akhtar *et al.*, 2020). To evaluate the plant growth-promoting potential of our extracts, we applied them in concentrations matching the culture concentration on barley seedlings. Additionally, we included a media extract control, as well as IAA and iP at defined concentrations, for testing. Briefly, extracts were spotted into glass vials containing 5 ml SH-medium, and the solvent was evaporated before one barley seed was added per vial. After an initial phase of darkness for germination, a day/night cycle incubation was performed for 6 days (Figure S2-13). Afterwards, the plants were extracted from the medium, cleaned, and separated into root and shoot before being lyophilised for dry biomass determination (Figure 18 A and B). All samples showed a consistently higher biomass if the sample contained Trp. The pure compound controls showed growth comparable to that of the medium with IAA at a concentration of 0.442 µM, resulting in a biomass of 11 mg (similar to the medium control containing

0.2% Trp). Compared to the controls, extracts from FHG110214 and *A. borealis* DSM 23886 decreased plant biomass in the shoots (Figure 18A) and roots (Figure 18B) with and without added Trp. However, the addition of Trp to the cultivations decreased the adverse growth effect of the extract itself. Overall, the crude extracts of *A. borealis* and FHG110214 did not significantly promote plant growth of barley seeds in the experimental setup tested. It has been demonstrated that IAA can promote plant growth when bacterial colonies are in direct contact with roots (Kielak *et al.*, 2016b). Our experimental setup, in contrast, was executed using cell-free organic extracts to provide a controlled sample. This, however, may not replicate localised, high-concentration phytohormone effects occurring in natural rhizosphere environments. Furthermore, applying cells or cell-free supernatants may provide further traits, such as enzymes with a beneficial effect on plant growth, not covered by organic extracts. In future studies, live bacterial cultures or co-cultivation systems with plants would provide more ecologically relevant data, albeit with increased challenges associated with cultivation.

Interestingly, our findings, along with those of Kielak *et al.*, (Kielak *et al.*, 2016b) and Gonçalves *et al.* (Gonçalves *et al.*, 2024), suggest that phytohormone production is common among Acidobacteriota, particularly among strains isolated from decaying wood or lignocellulosic substrates. This raises intriguing questions about the ecological roles of Acidobacteriota strains producing compounds like IAA and iP in natural systems (Fu *et al.*, 2015).

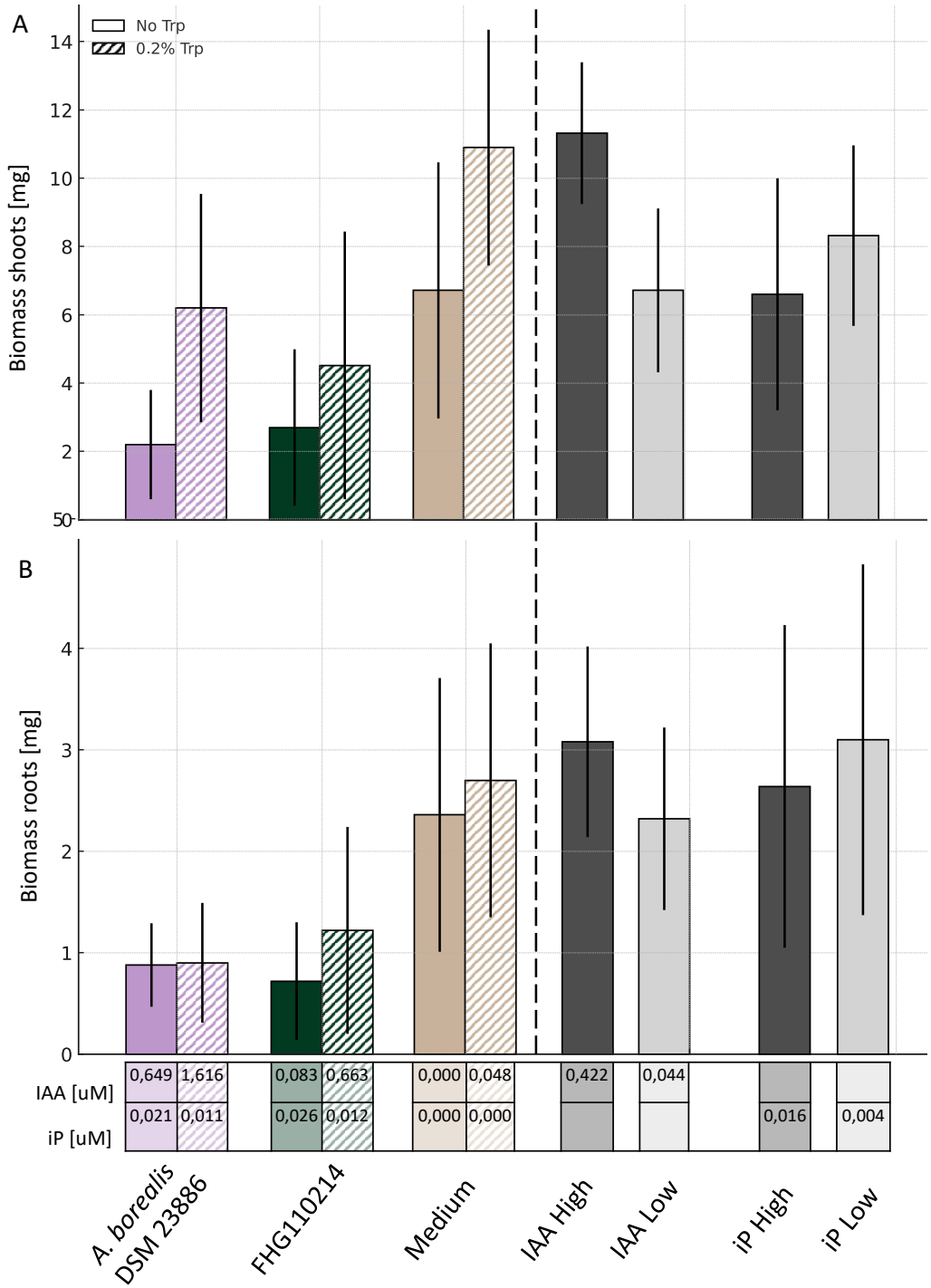

Figure 18: Dry Biomass of barley shoots (A) and roots (B) treated with extracts of strains *A. borealis* DSM 23886 (purple), FHG110214 (green). Striped bars indicate extracts from Trp-supplemented cultivations, while solid bars indicate that extracts were not supplemented with Trp. Media controls (brown), as well as IAA and iP as pure compounds(grey), were also tested. The table indicates the final concentrations of phytohormone applied to the barley seedlings. In general, supplementation with Trp results in increased biomass, especially in the shoots. While pure compound controls show comparable growth to the media control, the strain extracts inhibit plant growth. This negative effect is diminished upon supplementation with Trp, especially in the shoots.

2.8.3.4 In-Depth Bioinformatic Analysis of PGP Traits in Acidobacteriota and FHG *Acidobacteriaceae*

Taken together, these results confirm that *Acidobacteriaceae* produce phytohormones in relevant amounts, but their role in PGP remains unresolved. Because iP can arise from the degradation of prenylated tRNAs in many bacteria (Koenig *et al.*, 2002), we asked whether *Acidobacteriaceae* also encode dedicated cytokinin pathways (involving lonely guy (LOG) enzymes) and whether these co-occur with additional PGPTs. Furthermore, we were interested in the biosynthetic mechanisms responsible for producing phytohormones. For batch analysis of the 618 genomes previously analysed, we employed a DIAMOND workflow against the PLABASE database (see materials and methods) and summarised hits as coverage (% of PGPT families detected). For the analysis of taxonomic patterns, we aggregated the results by GTDB-Tk family status, only considering acidobacterial families that have been scientifically described or have more than 20 representative genomes available (n = 391, 64.2% of the total dataset). Figure 19 represents the percentage coverage of PGPTs categorised by PLABASE at "level two". Values were hierarchically clustered to visualise similarity between different families and PGPTs. The analysis separated the families into two distinct clusters.

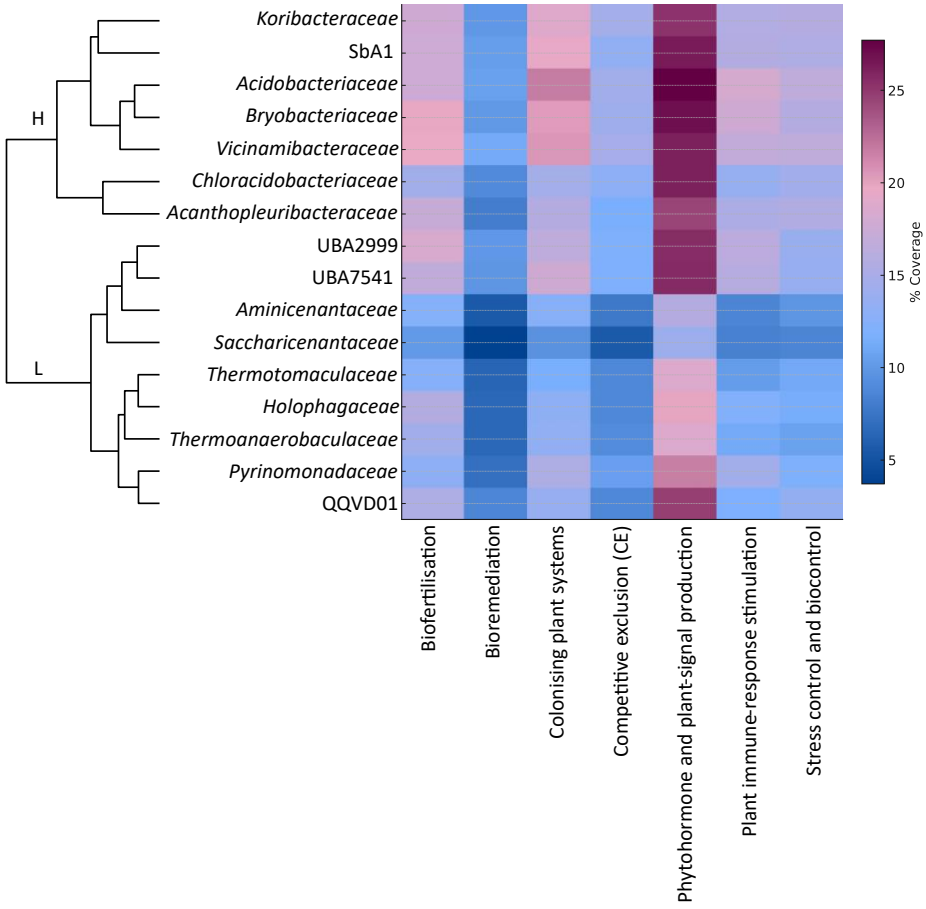

Figure 19: PGPT coverage across Acidobacteriota families. The heatmap displays the mean % coverage of PLABASE PGPT categories (level 2) for families with formal names or more than 20 genomes (n = 16). Coverage values were calculated as the proportion of detected PGPT families per functional category, relative to the total number of families defined in the PLABASE ontology. Rows and columns are hierarchically clustered. Two distinct groups emerge, indicating lower (L, bottom) and higher (H, top) PGPT coverage. The latter exhibits higher potential, particularly in the production of phytohormones and plant signals.

The nine families in the lower-coverage cluster (L) (UBA2999, UBA7541, *Aminicenantaceae*, *Saccharicenantaceae*, *Thermotomaculaceae*, *Holophagaceae*, *Thermoanaerobaculaceae*, *Pyrinomonadaceae*) show a comparatively low average PGPT potential (12.7±1.2%) with uniformly modest values across categories such as biofertilisation: 14.3 ± 1.4%; colonising plant system: 13.7 ± 1.4%; stress-control-biocontrol: 11.6 ± 1.0% and especially low bioremediation: 7.2 ± 0.9%. In contrast, families in the higher-coverage cluster H (*Koribacteraceae*, SbA1, *Acidobacteriaceae*, *Bryobacteraceae*, *Vicinamibacteraceae*, *Chloracidobacteriaceae*, *Acanthopleuribacteraceae*) showed consistently higher coverage with representative means of 26.1 ± 2.6% for plant signal production, 18.8 ± 2.4% for colonising plant system and 13.7 ± 1.7% competitive exclusion (overall range ~17–26% across categories. Across both clusters, the bioremediation category is underrepresented, while the phytohormone and plant signalling group ranks top among both clusters. Overall, the families of *Bryobacteriaceae*, *Acidobacteriaceae*, and *Vicinamibacteriaceae* exhibit the highest levels of PGPTs. Notably, the *Acidobacteriaceae* perform best in the colonising plant system group (e.g., covering EPS production) and the phytohormone/plant signal production category.

We analysed DIAMOND outputs for enzymatic equipment supporting IAA production in our FHG *Acidobacteriaceae* strains. In bacteria, three pathways are described (Spaepen *et al.*, 2007): indole-3-pyruvic acid (IPyA), indole-3-acetamide (IAM) and tryptamine (TAM). In the IPyA route, Trp is transaminated to indole-3-pyruvate, decarboxylated to indole-3-acetaldehyde, and oxidised to IAA; the IAM pathway converts Trp to indole-3-acetamide, then hydrolyses it to IAA; the TAM pathway decarboxylates Trp to tryptamine, which is subsequently oxidised via intermediates such as indole-3-acetaldehyde to IAA. All FHG *Acidobacteriaceae* include the key enzymes to synthesise Trp (trpA-E) and terminal enzymes for IAA production, such as an *amiE* homolog for the hydrolysis of indole-3-acetamide, and two aldehyde-dehydrogenases that may catalyse the final step of IAA production in both the TAM and the IPyA pathways. However, the automatic setup did not detect intermediate enzymes. Metabolomics using reference standards identified indole-pyruvic acid (CAS: 392-12-1) upon Trp supplementation, whereas tryptamine (CAS: 61-54-1) was not detected in FHG110202, FHG110214 and *A. borealis* DSM23886. FHG110511 showed neither intermediate. These observations support IPyA as the predominant route over TAM, while IAM remains a possible but unconfirmed option.

The production of iP is mediated by two distinct pathways; the t-RNA isopentenyltransferase (IPT) pathway involves the isoprenylation of tRNAs by t-RNA IPTs (*miaA*) and the context-dependent generation of iP through the degradation of these tRNAs (Frébortová and Frébort, 2021; Wei *et al.*, 2023). The pathway is found ubiquitously in bacteria. Instead, the second pathway is less common and is directly associated with regulated cytokinin output in bacteria, which usually interact with plants. Here, the adenylate IPT modifies adenosine nucleotides using dimethylallyldiphosphate (Sugawara *et al.*, 2008), thereby producing iP-like nucleotides further activated by lonely guy (LOG) enzymes (Kuroha *et al.*, 2009; Seo *et al.*, 2016). Our genomic survey revealed that FHG110511, as well as FHG110214 and *A. borealis* DSM 23886, contain the tRNA-IPT (*miaA*) and accessory genes, while FHG110202 additionally contains a LOG (lonely guy) enzyme. Instead, the adenylate IPT could not be detected in any of the FHG *Acidobacteriaceae*. As indicated by the literature, the *miaA* gene is omnipresent in our acidobacterial dataset, while the LOG enzymes are less present (477/618) but still highly abundant in various acidobacterial classes. Notably, the Trp-responsive iP production aligned with our pathway content. In the strains harbouring the tRNA IDT pathway alone, we determined a reduced iP

production upon Trp supplementation. This might be explained by an overall decrease in tRNA turnover due to the high availability of Trp, resulting in a subsequent reduction in global translation and, consequently, lower cytokine release through tRNA degradation mechanisms. Instead, FHG110202, the only inspected strain that harbours a LOG enzyme, does not show a Trp-dependent decrease in iP production, thus possibly less prone to global protein biosynthesis and tRNA turnover rates.

We also specifically looked for the presence of vital PGPT factors that could be tested in wet lab experiments. This included plate assays indicating phosphate solubilisation, nitrogen fixation and ACC-deaminase activity. The enzyme ACC deaminase (AcdS) breaks down 1-aminocyclopropane-1-carboxylic acid into ethylene, thereby playing a vital role in the plant defence mechanism. We did not detect *acdS* homologs in our FHG *Acidobacteriaceae* strains. Consistently, we could not detect any growth on plates containing 1-aminocyclopropane-1-carboxylic acid (data not shown). The enzyme AcdS, correlated to PGPT family number 7345, was also searched for in the entire dataset, but was not detected in any strain. The ability to solubilise phosphate from phytate was evaluated on medium containing phytate (data not shown). However, even though we did not detect any growth on agar plates, we identified genes related to phosphate solubilisation through genome mining. *A. borealis* DSM 23886(1.02e-119), FHG110214 (4.91e-107), and FHG110511 (1.57e-148) contained homologs of *appA* (PGPT2560), a phytase known to release Pi from organic phosphorus via a phytase or nucleotidase reaction. We also searched for phytase homologs belonging to PGPT families 2560, 2580, and 2585 in the acidobacterial dataset. Phytases were present in 117 genomes, but only in the classes of *Acidobacteriae* and *Blastocatellia*. The ability to fix nitrogen has been described for the Acidobacteriota before. Recently, Gonçalves *et al.* (Gonçalves *et al.*, 2024) described the presence of nif clusters in members of the *Holophagaceae* and *Acidobacteriaceae* families. In our genome analysis, we confirmed that members of the *Holophagaceae (Holophagaceae bacterium isolate CTSoil_007)* and *Acidobacteriaceae* (Acidobacteria bacterium isolate SZAS TMP-7) contained nitrogenase clusters encompassing all required *nif* genes A-F (PGPT families PGPT1-30). Besides the *Holophagaceae*, *nif* genes were also detected in *Bryobacteraceae* TMP-7, but not in any of the FHG *Acidobacteriaceae*.

Compared to the large-scale metagenomic analysis of Gonçalves *et al.*, our PGPT analysis shows both shared and divergent trends across the acidobacterial families. Some families, such as *Acidobacteriaceae*, *Thermoanaerobaculum*, *Aminicenantesaceae*, *Holophagaceae*, and *Koribacteraceae,* are grouped similarly in terms of their overall equipment, either uniformly high or low; other families show significant discrepancies. Specifically, our results diverge for the families *Bryobacteriaceae, Vicinamibacteraceae, Pyrinomonadaceae,* and *Chloracidobacteriaceae,* where the PGP potential differs substantially. These differences likely arise from the limited overlap in representative genomes between the two datasets. Overall, 9.2% of our dataset is also covered by the Goncalves dataset. Additionally, our analysis includes seven additional Acidobacteriota families. Notably, the dataset coverage is particularly low for most families with detected discrepancies, including *Bryobacteriaceae* (1/74), *Vicinamibacteraceae* (2/6), and *Pyrinomonadaceae* (2/87). The overlap of the *Chloracidobacteriaceae* (10/14) is relatively high; however, in addition to the dataset composition, methodological differences likely also contribute to the differences. Gonçalves *et al.* focused on 91 curated PGPTs, whereas our genome mining approach utilised the complete PLABASE database, encompassing over 1,300 plant growth-promoting gene families. This broader functional scope may capture additional trait diversity and help explain the differential PGPT evaluation and distribution observed across families.

The PGPT profiles of the FHG *Acidobacteriaceae* are consistent with both our broader family-level analysis (Figure 19) and the patterns reported in the large-scale metagenomic study by Goncalves *et al.* (2024). In all cases, *Acidobacteriaceae* show a high frequency of PGPTs, indicating a conserved functional repertoire within this lineage. Notably, Gonçalves *et al.* also identified a similar PGPT pattern in this family, including genes associated with phytohormone biosynthesis and phytase activity. These congruent findings suggest that *Acidobacteriaceae* may play a robust and stable role in plant–microbe interactions across diverse environments.

2.8.3.5 Effects of Tryptophan on Metabolite Production and Phytopathogen Inhibition

Trp is abundant in the root exudates of many plants, and is described to be used by bacteria to synthesise auxins (Noor *et al.*, 2023). Indeed, the synthesis of phytohormones by rhizosphere bacteria often depends on the presence of Trp precursors in the root exudates (Kravchenko *et al.*, 2004). Thus, this metabolic shift might reflect an environmental response of the strains towards the Trp-enriched plant environment (Eze and Amuji, 2024). To investigate whether supplementation with Trp induces a global metabolic shift or stimulates the production of phytohormones, and to potentially describe the regulated metabolites, we performed time-resolved untargeted metabolomics analysis. Extracts cultivated with and without Trp were analysed with ultra-high-performance liquid chromatography–quadrupole time-of-flight high-resolution mass spectrometry (UHPLC-QTOF-HRMS), and the calculated feature table was filtered to remove features identified as noise. Subsequent principal component analysis (PCA) of log2-transformed feature intensities (Figure S2-14) revealed that the samples did not exhibit a strict separation by strain or cultivation day, but instead were separated by supplementation status, indicating a broad metabolomic shift associated with Trp supplementation. Furthermore, we observed temporal metabolic shifts that were more distinct for *A. borealis* DSM 23886, especially if supplemented with Trp.

To further investigate Trp-dependent metabolomic responses, we filtered and categorised features based on their presence-absence and fold-change behaviour in response to Trp supplementation. Briefly, features were retained if they were present in both bucketing calculations. Following filtering, feature intensities were compared between the uninduced and induced conditions to derive fold changes. Based on those, we assigned biological categories from “extinguished” to “switched-on”. Individual features were then plotted and coloured to indicate their biological response (Figure S2-15). The Barcoding plot confirmed the metabolic shift in response to Trp supplementation. We observed differences in media consumption (left side) and changes in features that were clearly strain-dependent (right side). Overall, there is an apparent change in the metabolism indicated by the number of features that change upon Trp supplementation. For *A. borealis* DSM 23886, 165 features are detected on day 14 without Trp, while 684 features (4.1x) are present if Trp is supplemented. The same trend is observed in FHG110214. Here, the feature number increases from 109 (without Trp supplementation) to 337 (3.1x) (with Trp supplementation). We independently verified this result in a second cultivation (Figure S2-15, bottom half). Here, a similar fold change is observed. While the feature number increases 3.2-fold in *A. borealis* DSM 23886, it exhibits a similar behaviour in FHG110214 (2.8-fold). Those numbers already indicate a strong metabolic activation. This is supported by the fact that, considering all

features present under both conditions, 80% of the features are at least lightly upregulated, while 19% are downregulated, and only 1% remain stable.

To further investigate Trp-dependent metabolomic responses considering media and strain-specific features, we classified Trp-activated uptake as referring to features initially present in the medium that were consumed only in Trp-induced cultures. In contrast, Trp-inhibited uptake describes the opposite: features consumed only in uninduced cultures. The rest of the classifications remained the same. Features that were stable in the medium and did not change in intensity were removed from the analysis. Figure 20 (top bar chart) displays the percentage distribution of features across these categories. The "switched on" class accounts for the largest fraction in both strains (*A. borealis* DSM 23886: 56.8%; FHG110214: 45.1%). Specifically, 10.4 and 16.9% of features exhibit an "increased production" profile, while 7.5 and 6.3%, respectively, are "switched off." Small proportions (below 10%) represent different metabolisation patterns of media components under induced and uninduced conditions.

Given the pronounced metabolic changes observed in Trp-induced extracts, we were encouraged to re-enter our initial bioactivity discovery approach, now particularly testing the generated extracts from FHG110214 and *A. borealis* DSM 23886 for activity against the phytopathogenic fungus *Septoria tritici* MUCL45407, a residue-borne wheat pathogen causing *Septoria tritici* blotch (STB), one of the most significant diseases of wheat (Fones and Gurr, 2015). By testing our extracts in a microbroth dilution assay, we indeed determined a strong growth-inhibitory effect, exclusively in the samples that underwent Trp supplementation. For the identification of activity-causing compounds, one representative active extract of FHG110214 and *A. borealis* DSM 23886 was microfractionated and re-screened against *S.tritici* MUCL 45407. Based on the molecular formula and the observed MS/MS fragmentation pattern, several Trp-related compounds were dereplicated within the active fractions of both strains, with a confidence level 2 (putatively annotated metabolites (Sumner *et al.*, 2007)). This enabled the annotation of hydroxyglyantrypine (Peng *et al.*, 2013) as well as malassezindoles A, B (Mayser *et al.*, 2007) and putative yet undescribed derivatives thereof (m/z: 148.1349, 376.1293, 360.1348, 362.1498). Notably, malassezindoles production is known to be strongly dependent on Trp as the sole nitrogen source in their fungal production strains (Mayser *et al.*, 1998). This observation aligns well with our cultivation conditions and experimental setup. However, since the malassezindoles have not been described for bacteria yet, we have expanded our quality control by including regular agar plating, which is conducted concurrently with sampling at the beginning and end of cultivation. Additionally, we continued to perform 16S/18S sequencing from liquid culture samples. Agar plates were incubated for at least 14 days, documented, and routinely inspected using a binocular microscope. Neither measure raised any concerns about handling fungal-contaminated samples. Besides malassezindoles, we annotated the β-carbolines acetyl-beta-carboline, pityriacitrin B (Irlinger *et al.*, 2005; Mayser *et al.*, 2007, 2002) and a methylated derivative thereof. Pityriacitrins are produced by fungi and bacteria (D. Huang *et al.*, 2022) and recognised for their UV-protective properties (Machowinski *et al.*, 2006). Furthermore, the β-carbolines scaffold has been recently reviewed in the context of antifungal activity (Dai *et al.*, 2018). Indeed, pityriacitrin A, B, and their methyl esters have been described to inhibit fungal growth (Gaitanis *et al.*, 2019; D. Huang *et al.*, 2022). Furthermore, pityriacitrin B was shown to have protective activity against fungal colonisation of plant hosts (D. Huang *et al.*, 2022). In bacterial systems, beta-carboline cores, such as those found in pityriacitrin, are formed via Pictet-Spenglerase enzymes (Li *et al.*, 2023; Ueda *et al.*, 2018; Wang *et al.*, 2016). Instead, biosynthesis of the Trp-

dependent metabolites, closely related to the malassezindoles, has been described for the smut fungus *Ustilago maydis*, where a single enzymatic reaction, catalysed by the enzyme Tam1, is essential for their formation, converting Trp to indolepyruvate (Zuther *et al.*, 2008). Although the biosynthetic pathways for these compounds have not been elucidated in bacteria, a similar mechanism can be inferred. However, a BLAST search for analogous proteins within our acidobacterial genome dataset did not yield any significant hits. An additional signal observed in the active fraction, co-eluting with malassezindoles under the applied chromatographic conditions, suggested the presence of a peptide, with an m/z value of 519.1655 [M+H]+ corresponding to the molecular formula $C_{30}H_{22}N_4O_5$.

We focused our isolation efforts on compounds previously reported to exhibit antifungal properties. Preparative and semi-preparative isolation methods were employed on the fractions containing malassezindoles and their derivatives. Isolation and structural elucidation of antifungal compounds, including malassezindoles and related derivatives, were carried out by collaborators. Preparative and semi-preparative chromatography yielded two pure compounds, identified by NMR spectroscopy as pityriacitrin B and its methyl ester. Structural assignments were consistent with previous reports (Irlinger *et al.*, 2005; Liew *et al.*, 2014). To examine the features previously identified to be bioactive, we mapped them onto our categorised feature dataset. This allowed us to assess which metabolic response classes these metabolites belong to. Additionally, we explored other Trp-dependent masses by molecular networking, aiming to identify novel or derivative compounds using database annotations.

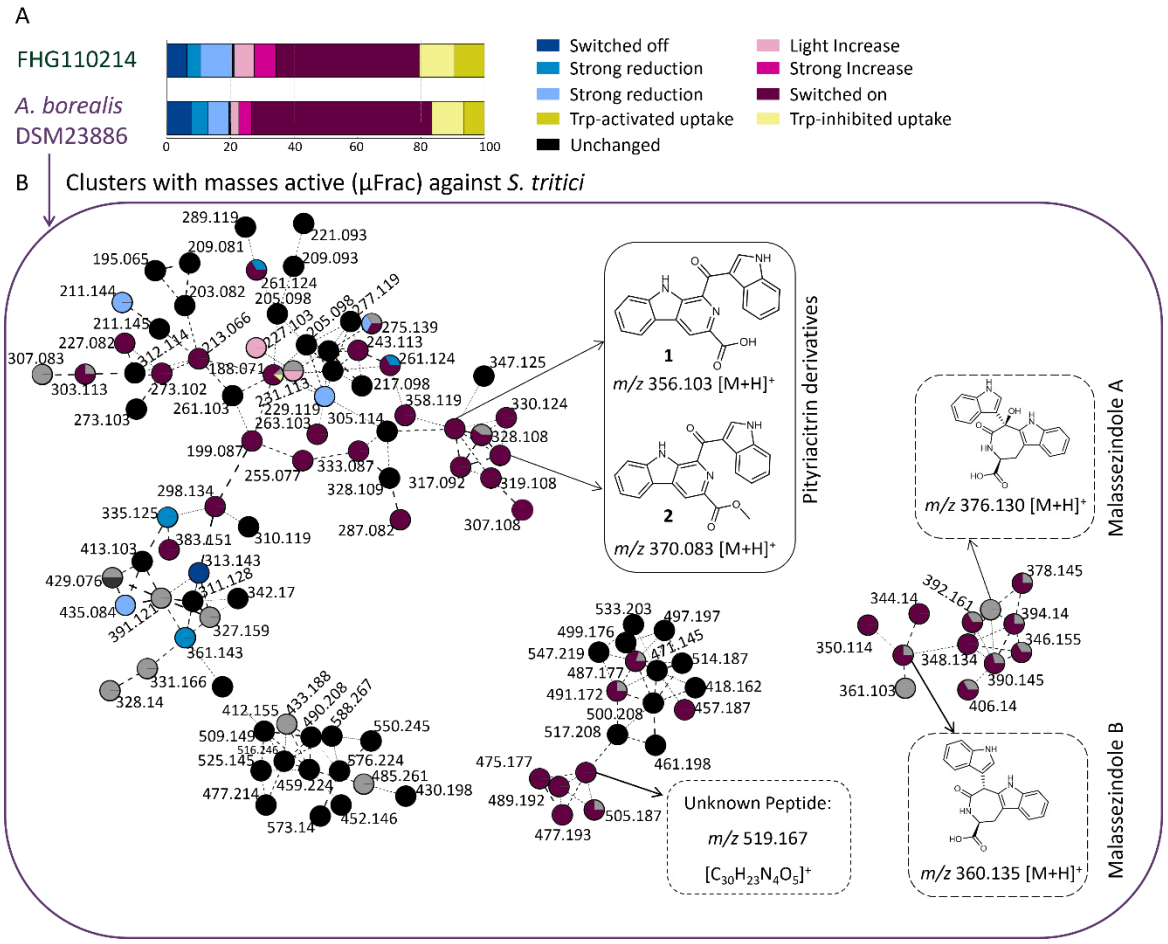

Figure 20: Overall metabolic response of cultivation with Trp in strains FHG110214 and *A. borealis* DSM 23886 (A). Molecular network of masses active against *S. tritici* MUCL 45407 as detected by µFrac-coupled screening of *A. borealis* DSM 23886 extracts (B). **(A)** After feature detection, the relative presence of masses with and without Trp was compared for categorisation. Categories were defined as follows: "switched-off" (dark-blue, no feature detected upon Trp supplementation) and "switched-on" (dark-red, Feature only present upon Trp supplementation). Media components that were metabolised differently upon Trp supplementation are shown in yellow. **(B)**Molecular Networking of MS/MS data generated from a Trp-induced *A. borealis* DSM 23886 culture, limited to clusters containing masses that showed antifungal activity against *S. tritici* MUCL 45407. Active masses were all specifically "switched on" when supplemented with Trp. While some of them remained unknown (e.g. 519.167m/z peptide), others were dereplicated by comparison with database standards (malassezindole A, B) or isolated and structure elucidated (pityriacitrin B and pityriacitrin B methyl ester).

The observed changes in activity profiles and the Trp-dependent modulation of extracts underscore the impact of Trp supplementation on the production of specific, bioactivity-causing compounds. Having already linked several of these induced features to bioactivity through microfractionation, we next turned to molecular networking to explore their structural context within the overall metabolome. This approach enabled us to visualise whether the bioactive compounds belong to larger molecular families, identify structurally related analogues, and determine whether induction leads to the coordinated activation of specific metabolic clusters. By integrating activity data with MS/MS-based network topology, we aimed to prioritise entire subnetworks of interest and assess how broadly Trp supplementation influences distinct chemical families across strains.

Acidobacteriota often dominate soil bacterial communities, particularly in acidic and oligotrophic environments, yet their functional roles remain underexplored. Our findings contribute to growing evidence that members of this phylum possess underestimated biosynthetic and ecological potential, including the ability to produce phytohormones and possibly bioactive metabolites with the potential to antagonise fungal growth.

Beyond our cultivated FHG *Acidobacteriaceae*, genome-mining of 618 high-quality Acidobacteriota genomes offers a complementary perspective to the large-scale MAG-based analysis by Gonçalves *et al.* (2024), which analysed 758 MAGs. Both studies identify *Acidobacteriaceae*, *Bryobacteriaceae*, and *Koribacteraceae* as families with enriched PGPT potential, while families such as *Holophagaceae* and *Vicinamibacteraceae* appear comparatively under-equipped. However, notable discrepancies emerged, likely reflecting differences in the underlying datasets and the methodological approach.

Despite the current technical limitations, exploring Acidobacteriota as sources of novel bioactive compounds or as contributors to plant–soil health remains a promising avenue. A considerable limitation, however, is their challenging growth in laboratory settings. If obtained as isolates, their total cell biomass remained low, even when using optimised media conditions, and some strains even show aggregated growth forms. Given the intriguing biosynthetic potential predicted from the genome, more efforts are needed to characterise and exploit this phylum, including, for example, the development of genetic tools for in situ activation, overexpression, and deletion of specific genes, as well as the establishment of a genetically tractable heterologous expression system in Acidobacteriota-hosts. Together, this could lead to a better understanding and utilisation of Acidobacteriota, aiming to discover new methods for their role in sustainable agriculture, bioremediation, and natural product discovery.

In Chapter 2.7, our team employed a high-throughput microplate cultivation approach using termite-nest material to maintain community diversity and selectively isolate Acidobacteriaceae, including putative new species in *Terracidiphilus* and *Acidobacterium*, thereby providing tractable strains for downstream work. In section 2.8, we combined genome mining with metabolomics to map the distribution of PGPTs and biosynthetic gene clusters across the phylum. We demonstrate that selected *Acidobacteriaceae* produce phytohormones (IAA and iP), and that Trp supplementation alters the metabolome, boosting IAA production while reducing iP. Furthermore, antifungal-linked metabolites pityriacitrin and malassezindoles were detected. Together, these studies provide new isolates, map genetic potential, and link it to measurable metabolites—laying the groundwork to explore Acidobacteriota chemistry and their ecological roles in PGP and beyond.

2.9 HEL, A Multi-Host Synthetic Biology Platform for BGC Activation

2.9.1 Background

Since the first description of microbial dark matter, researchers have been eager to discover and describe these previously overlooked microorganisms to understand their ecological context and role (Rinke *et al.*, 2013), as well as to utilise their genetic potential for various biotechnological applications. Large metagenomic databases and bioinformatic tools, steadily advanced by AI, form the foundation for unravelling their biosynthetic potential; however, accessing them presents additional challenges. When the cultivation of novel isolates fails, the direct cloning of genes of interest, such as enzymes or complex BGCs, is the only viable option for characterising and isolating putative NPs. The first successful expression of a BGC dates back to 1984 when the actinorhodin cluster was transferred from its original producer *Streptomyces coelicolor,* to another *Streptomyces* species (Malpartida and Hopwood, 1984). The first cross-phyla heterologous expression was reported in 2001 when the DEBS (6-deoxyerythronolide B), the polyketide core of erythromycin, was successfully produced in *E. coli* (Pfeifer *et al.*, 2001). In 2002, the first metagenomic DNA was heterologously expressed in *E. coli*, leading to the discovery of the turbomycins (Gillespie *et al.*, 2002).

Cloning of BGCs is most difficult because these clusters are often large, multimodular, and complex (Bloudoff and Schmeing, 2017; Khosla *et al.*, 2009). These characteristics result in significant challenges regarding amplification by PCR (high GC-content, repetitive sequences) and commonly used assembly methods, such as Gibson cloning (Gibson and *et al.*, 2009). For this reason, large and complex BGCs usually require recombination methods that rely on homologous recombination, phage-based systems, or CRISPR-Cas (Wan *et al.*, 2023); The yeast-recombination-based system TAR (Transformation-Associated Recombination) is being most frequently used (Yamanaka *et al.*, 2014). Here, 40-80 bps homologous regions are cloned into a yeast-*E. coli* shuttle bacterial artificial chromosome (BAC) to capture the desired BGC upon co-transformation with the BGC-containing DNA using the yeast's recombination system Since the discovery of the CRISPR-Cas system, this method is also being utilised to specifically cleave and capture BGCs from DNA (Enghiad and *et al.*, 2021; Jiang *et al.*, 2015). In the λ-phage-based recombination system, BGC are captured and stored by phages in their phage head (Casjens and Hendrix, 2015). While this is a very efficient process often used to generate metagenomic or genomic libraries in fosmids (Epicentre, pCC2Fos Library Kit), it has the limitation that the capacity of the phage head only reaches around 40 kb, which is insufficient for very large BGCs. Fosmids are vector systems that rely on the F-factor replication and partition system and are usually combined with λ-phage-based capture. Therefore, they typically only carry the aforementioned 40 kb. The F-factor replication system is inherently stable and has a low copy number (1-2 copies per cell) (Kline, 1985; Tsutsui and Matsubara, 1981). Thus, they usually contain a second origin of replication, oriV, which responds to the initiation protein TrfA, resulting in a higher copy number of the fosmid (10-50 copies/cell) (*CopyControl Fosmid Library Production Kit: pCC1FOS system (inducible oriV/TrfA copy control)*, 2006). Another F-based vector is the BAC system, which is also very stable and has a low copy number. Inserts are introduced by ligation, which is less efficient compared to phage-based systems, but not limited in size (Shizuya and *et al.*, 1992; Tao and Zhang, 1998). Consequently, inserts up to 300 kb can be stably maintained in BAC vectors (Shizuya and *et al.*, 1992). The Cosmids contain different

replication origins depending on their type, often of plasmid origin (e.g., ColE1). They also include λ-phage cos sites for in vitro phage packaging (Hohn and Collins, 1988), making library creation very efficient but again, size-dependent (Agilent Technologies, 2015). Due to their plasmid origin, they usually exist in a medium copy number within the cell but are generally less stable than BAC or Fosmid-like systems.

An essential prerequisite that has yet to be addressed is the selection of suitable heterologous hosts. Due to the overrepresentation of actinomycetes-derived BGCs (Blin *et al.*, 2019), the *Streptomyces* host is an obvious choice (Lasch *et al.*, 2025). However, due to their genetic tractability and ease of use, *E. coli* has also been widely utilised and modified for BGC expression (see above). Other chassis are less frequently employed and are typically specific to specialised niches within various laboratories. Among these, Gram-negative hosts include *Pseudomonas putida*, which contains a rich pool of precursors and cofactors (Loeschcke and Thies, 2015; Nikel and de Lorenzo, 2018), and *Myxococcus xanthus*, favoured for myxobacterial non-ribosomal peptide and polyketide clusters, but also usable as a universal host (Stevens *et al.*, 2010; Wenzel *et al.*, 2005). As Gram-positive hosts, *Bacillus subtilis* strains are commonly chosen due to their robust secretion capabilities and the availability of genetic tools for BGC expression (Radeck *et al.*, 2013). Notably, they encode the sfp phosphopantetheinyl transferase (PPTase), a broad-range PPTase that activates the acyl carrier protein (ACP) (a polyketide synthase (PKS) domain) and a peptidyl carrier protein (PCP) (a non-ribosomal peptide synthetase (NRPS) domain) in many organisms (Put *et al.*, 2024). In addition to bacterial expression hosts, fungal hosts offer the inherent benefit of eukaryotic expression machinery, providing access to plant, algal, and fungal pathways (Meng *et al.*, 2022). Typically, the genus *Aspergillus* (Chiang *et al.*, 2013; Sheng *et al.*, 2025) and *Saccharomyces cerevisiae* (Cautereels *et al.*, 2024; Harvey *et al.*, 2018) are utilised as chassis. The use of different heterologous hosts can also prove advantageous, as first demonstrated by (Craig *et al.*, 2009), who transferred a metagenomic soil cosmid library to six distinct proteobacterial hosts (*Agrobacterium tumefaciens*, *Burkholderia graminis*, *Caulobacter vibrioides*, *E. coli*, *P. putida*, and *Ralstonia metallidurans*) to compare their functional capabilities. Interestingly, the overlap between expressed BGC and produced NPs was minimal across the investigated hosts, underscoring the various genetic capacities and thus encouraging the use of multi-chassis expression for successful and differential BGC activation/expression.

Genetic factors affecting heterologous expression are diverse and have been examined in many studies. Firstly, expression often fails due to promoter or sigma factor mismatches or the lack of necessary transcriptional activators. This issue is usually resolved through promoter swaps (Chiang *et al.*, 2009) or heterologous complementation of transcriptional activators (Arias *et al.*, 1999; Bok and Keller, 2004; B. Wang *et al.*, 2019). Even after successful expression, translation can remain inefficient because of weak ribosome binding or mismatched codon usage. To improve translational efficiency, BGCs can be codon-optimised, ribosomal binding site (RBS) sequences can be altered or supplied through RBS libraries (Rao *et al.*, 2024), and rare tRNAs (Zdanovsky and Zdanovskaia, 2000) can be added. For ribosomal-independent pathways, such as NRPS and PKS systems , specific proteins and enzymes are typically needed; notably, phosphopantetheinylation by phosphopantetheinyl transferase (PPTase) enzymes like sfp (see above) is essential for biosynthesis (Kim *et al.*, 2018). Additionally, A-domains require MbtH-like proteins (Felnagle *et al.*, 2010); other tailoring enzymes such as P450/monooxygenases (Hu *et al.*, 2023) or halogenases, and possibly chaperones (Nishihara *et al.*, 1998), may be needed to facilitate proper modification and folding. These gene products can be supplied

externally when heterologous expression faces difficulties. However, even with the most efficient transcription and translation, expression will be ineffective without essential precursors; these are often externally supplied via cultivation or produced by the bacterial enzymatic machinery that provides necessary extender units like (methyl)malonyl-CoA, propionyl-CoA, or ethylmalonyl-CoA (Chan *et al.*, 2008; Klass *et al.*, 2025). Lastly, heterologous hosts often need resistance cassettes or efflux pumps to counteract the effects of toxic expression products, such as antimicrobial NPs (Severi and Thomas, 2019; Xu *et al.*, 2012). Taken together, this demonstrates the complex network required for successful heterologous expression, which is typically a tedious process that involves multiple rounds of cloning to find the optimal combination of conditions.

The metabolic rewiring, along with the standardised, combinatorial assembly of genetic parts for circuits, is a common strategic asset in synthetic biology. Synthetic biology employs iterative design-build-test-learn cycles to construct and refactor biological systems using standardised, modular genetic parts (Endy, 2005). This is exemplified, for instance, by programmable genetic circuits such as the synthetic 3-gene oscillator by Elowitz and Leibler (Elowitz and Leibler, 2000), which functions similarly to electrical circuits but in *E. coli*. These genetic circuits, along with larger metabolic reprogramming approaches, are advancing due to the establishment of the iGEM competition and their efforts to define, characterise, and document a repository of genetic parts – including promoters, reporters, and other functional genetic elements (Kelwick *et al.*, 2015; Shetty *et al.*, 2008). These functional parts are then integrated into novel cloning pipelines like MoClo (Modular Cloning)(Weber *et al.*, 2011), a Golden-Gate assembly-based method that equips functional parts with standardised fusion sites created by Type IIS restriction enzymes, which cut outside their recognition sequences to facilitate custom-designed fusion sites. Other Golden-Gate derivatives include goldenBraid, PhytoBricks, and CIDAR- MoClo (Iverson *et al.*, 2016; Patron *et al.*, 2015; Sarrion-Perdigones *et al.*, 2013), the latter two being compatible with the MoClo system. Additional cloning techniques, such as homology-based methods like Gibson (Gibson, 2011) and SLiCE,(Zhang *et al.*, 2012) as well as workflows assisted by CRISPR and site-specific recombination techniques like Gateway (Hartley *et al.*, 2000), enable recombinase-mediated, multi-entry cloning, cassette exchange, or orientation flips (e. g., Cre/loxP and Flp/FRT) (Cox, 1983; Hoess and Abremski, 1984). The overall repertoire of cloning methods is then employed to reprogram the entire metabolism of various cellular systems; Notably, the Keasling lab engineered *Saccharomyces cerevisiae* to produce artemisinic acid (2006)(Ro *et al.*, 2006), subsequently optimising strains to achieve industrial-level yields, thereby enabling semi-synthetic production of artemisinin, a plant-derived traditionally extracted from the original plant source (Paddon and *et al.*, 2013). At the opposite end of the engineering spectrum, researchers developed a fully autotrophic *E. coli* strain capable of synthesising all biomass carbon from CO_2, by rewiring the central metabolism and applying adaptive evolution over one year (Gleizer and *et al.*, 2019).

Using a strict definition of refactoring BGCs with modular, standardised parts (or modular CRISPR-Cas systems), and applying an explicit engineering-style design-build-test loop, only a few studies have incorporated synthetic biology approaches to heterologous expression of NPs. Most frequently, BGCs were refactored to include modular promoter/RBS cassettes to decouple them from native regulation and transfer them into heterologous hosts. This was demonstrated with the streptophenanzine BGC, which yielded over 100 congeners of the NP (Bauman *et al.*, 2019). More recently, a standardised toolbox for refactoring multigene circuits and BGCs in *Streptomyces*, based on the MoClo cloning techniques, was developed by Massicard *et al.*, 2024. Finally, CRISPR-Cas, combined with plug-and-play sgRNA cassettes for activating silent fungal BGCs and

allowing predictable control, was utilised to express the otherwise silent macrophirin BGCs of *Penicillium rubens* (Mózsik and *et al.*, 2021). One reason for the limited use of standardised parts for NP expression could be the intrinsic incompatibility of restriction-enzyme-dependent standardised parts with large, sometimes sequence-unknown BGCs that may contain restriction sites.

Here, we introduce HEL (heterologous expression library), a cloning pipeline that combines λ-fosmid creation with modular, standardised, restriction enzyme-dependent expression optimisation. HEL retrofits BGCs with expression and chassis-specific elements, thereby decoupling activation and modulation from host transfer during fosmid creation. Expression libraries and host cassettes are assembled using MoClo for reusability, while coupling to the BGC-containing vector is performed independently via the in vitro λ-phage Gateway reaction. This enables individual parts or functional cassettes to be clonally amplified and reused across different BGCs or libraries. By allowing library-style combinations of promoters, regulators, and constant chassis modules for heterologous host transfer, HEL enhances the chances of finding productive configurations for BGC activation.

2.9.2 Materials and Methods

Tables describing strains, vectors and building blocks are collected in the supplementary information (Table 10-Table 12). Materials were used as indicated below. All strains were routinely maintained in LB medium (per 1 l: 10 g tryptone; 5 g yeast extract; 5 g NaCl; adjust to pH 7.0; in ddH_2O). For expression of GFP and darobactin in *B. subtilis* SCK6 we also used Spizizen medium as described by (Anagnostopoulos and Spizizen, 1961). *Pseudomonas putida* was also maintained in M9 supplemented with 0.4% glycerol (per 1 l: 10.5 g M9 Minimal Salts; 4 ml glycerol; 2 ml 1 M $MgSO_4$; 1 ml 0.1 M $CaCl_2$; in ddH_2O) to test for expression of GFP and darobactin. Antibiotics were used at the following final concentrations: chloramphenicol (Cm), 12.5 µg/ml; erythromycin, 1 µg/ml; kanamycin, 25 µg/ml; ampicillin, 100 µg/ml; and apramycin, 50 µg/ml. Inducers were used at the following concentrations: IPTG: 6.4 mM, arabinose: 5.3mM, N-hexanoyl-DL-homoserine Lactone (AHL): 25nM.

2.9.2.1 Standard Molecular Biology Techniques

For amplification of DNA fragments, we performed PCR using a high-fidelity DNA polymerase in combination with the supplied buffer system. Cycling parameters were adjusted according to the recommendation provided with the enzyme. DNA fragments were resolved by agarose gel electrophoresis and purified on silica-spin columns according to the manufacturer's instructions (Jena Biosciences). *Escherichia coli* strains were made chemically competent by the RbCl method (Sambrook et al) and were routinely stored at -80°C until further use. For transformation, 100 – 500 ng of DNA was mixed with competent *E. coli* cells, heat shocked, and recovered in SOC medium (37°C, 1 h). Cell suspensions were then concentrated and plated on selective agar. Putative transformants were screened by colony PCR using Taq DNA polymerase and vector- or insert-specific primers. PCR products were resolved by agarose gel electrophoresis.

2.9.2.2 Cloning Techniques

2.9.2.2.1 Annealing Reactions

Short modular building blocks were created by annealing complementary oligonucleotides containing the required recognition sites for Type IIS restriction enzymes. Reaction mixtures were prepared with corresponding primer pairs and annealed as described (Zumkeller *et al.*, 2018), with elongation times adjusted to match the expected fragment length. Annealed products were purified using a silica column-based PCR purification kit, according to the manufacturer's instructions (Jena Bioscience).

2.9.2.2.2 Gibson Cloning

For assembly of larger fragments and to create the required vector set, Gibson cloning was used according to (Gibson, 2011). Required fragments were either generated by restriction enzyme digestion or PCR.

2.9.2.2.3 Modular Cloning (MoClo)

MoClo was used to equip BBs with desired fusion sites (BsaI-dependent) and to assemble heterologous expression and host compatibility cassettes (Sap1-dependent). For each assembly, vector and insert(s) were combined at an equimolar 1:1 ratio. Molar amounts were calculated from DNA length using an average nucleotide pair mass of 660 $g \cdot mol^{-1} \cdot bp^{-1}$. Reaction components were combined as a standard MoClo Type IIS restriction–ligation mix (Weber *et al.*, 2011), and incubated in a thermocycler using a cycling program suitable for iterative digestion and ligation (Zumkeller *et al.*, 2018).

2.9.2.2.4 Gateway Cloning

Gateway LR recombination (LR Clonase II Enzyme Mix, ThermoFisher Scientific) was used to transfer MoClo-assembled parts into the preassembled HEL-fosmids. Reaction mixtures were set up and treated according to the manufacturer's instructions.

2.9.2.3 Electroporation (Sucrose) of *Pseudomonas putida* mt-KT2440

A single colony of *P. putida* KT2440 was grown in LB at 30 °C and 200 rpm until reaching mid-log phase (OD_{600} about 0.6 to 0.8). The cultures were placed on ice for 10 min, then spun down at 4,000 × g for 10 min at 4 °C. The resulting pellets were washed three times with ice-cold 300 mM sucrose (10 ml per 50 ml of culture) and resuspended in the same solution to about 1:100 of the original volume. Aliquots of 35–50 µl were kept on ice for immediate use or mixed 1:1 with sterile 40% glycerol for short-term storage at −80 °C. For each transformation, 20 to 100 ng of plasmid DNA were added to 35–50 µl of electrocompetent cells, mixed gently, and transferred to a pre-chilled cuvette with a 1 mm gap. Electroporation was performed at 1.6 kV, 25 µF, and 200 Ω, giving a time constant of about 4 to 5 ms. After pulsing, 1 ml of room-temperature SOC or LB medium was added right away. The cells were then recovered at 30 °C with shaking at 200 rpm for 60 to 120 min, plated on LB agar with the appropriate antibiotics, and incubated at 30 °C for 24 to 48 h.

2.9.2.4 Induction of Natural Competence (Xylose) of *Bacillus subtilis* SCK6

Bacillus subtilis SCK6 was streaked on LB with erythromycin (1 µg/ml). A single colony was picked to inoculate 5 ml of LB and grown overnight at 37 °C shaking 200 rpm. The culture was then diluted with fresh LB to an OD_{600} of about 0.1 to 0.2 and induced with 1% (w/v) D-xylose to activate P_xylA-comK. The cells were further

incubated at 37 °C and 200 rpm for 1.5–2 h until the OD_{600} reached around 1.0. Induced cells (100–200 µl) were then mixed with plasmid DNA (usually 100–500 ng) and incubated at 37 °C for 30 to 90 minutes. The cultures were then diluted with 1 ml of LB and incubated for another 3 h at 37 °C. Finally, the cells were spread onto selective LB agar and incubated at 37 °C overnight.

2.9.2.5 Flow Cytometry

Cells were pelleted at 4,000 rpm for 4 min, and the supernatant was decanted. The pellet was then gently vortexed to resuspend it in ~1.5 µl of 1× PBS. After a second spin at 4,000 rpm for 4 min, the PBS was removed, and the cells were resuspended in 1.5 µl of 1× PBS. Samples were acquired at a low flow rate. Instrument settings were: FSC, voltage E02, amp gain 1.00; SSC, 580 V, amp gain 1.00; FL1, 700 V, amp gain 1.00. The acquisition threshold was set to FSC-H (primary) with SSC-H (secondary) at a value of 100 to acquire green fluorescence of GFP.

2.9.2.6 Plate Measurement of Fluorescence and OD

Growth assays were performed using a VICTOR X3 multilabel plate reader (PerkinElmer). Strains from overnight precultures were diluted 1:1000 into 150 µl liquid medium in 96-well microplates. To minimise evaporation, 70 µl mineral oil was layered onto each inoculated well. At least two biological replicates were included per strain. The OD_{600} and green fluorescence (settings) were recorded every 5 min for the desired duration (18 h total). Data were processed and visualised in R and Microsoft Excel.

2.9.2.7 Extraction of Darobactin

Culture broth was centrifuged at 4000 rpm for 10 min to separate the supernatant from the cells. Secreted metabolites were purified from the supernatant by incubation with Amberlite XAD-16N (non-ionic polystyrene–divinylbenzene resin; typically, 20–60 mesh) at ~10% (v/v) under gentle agitation to bind the peptide. The resin was collected, rinsed with water, and eluted with methanol to yield a crude supernatant extract. The cell pellet was extracted with methanol, and soluble fractions were combined with the resin eluate. The extract was then concentrated by lyophilisation, and desalted on C18 solid-phase cartridges (water wash, elution with increasing organic solvent containing acid).

2.9.2.8 Analytical Procedures

Concentrated and purified extracts were analysed by UPLC–HRMS on a C18 column with an acidified water/acetonitrile gradient, using low-µL injections at standard analytical flow. Detection used positive-ion ESI, and darobactins were assigned by exact mass and diagnostic MS/MS fragmentation patterns.

2.9.3 Results

2.9.3.1 Technological Background, Design and Cloning of the HEL-Cloning Pipeline

We set out to build a fast-operating, modular cloning system that enables heterologous expression of BGCs by coupling expression libraries and the subsequent activation of the desired product, initially in *E. coli* but subsequently across additional expression chassis. Strategically, BGC cloning and capture had to be separated from expression optimisation, to protect large, sequence-unknown inserts (BGCs or (meta)genomic libraries)

from restriction-enzyme digest that are in return, required for the assembly of the expression-tuning library of genetic factors (promoters, ribosome-binding sites, regulators, helper modules). For heterologous expression cassettes, we needed a standardised, modular system that assembles DNA parts position-specifically while also enabling library generation by swapping alternative parts at the same position. Each expression library should include regulatory elements (e.g. RBS and promoters) and optional tuning factors (e.g. transcriptional regulators, and rare tRNAs) alongside constant host-specific factors (*oriT*, selection markers, and modifying enzymes) (Figure 21, light grey vector). We implement these using MoClo, a hierarchical Golden-Gate framework that facilitates library-style assembly by reusing position-coded functional parts. This pipeline is based on Type IIS restriction enzymes, which cut outside their recognition sites. When recognition sites are placed in inverted orientation at the fragment ends, they are removed during assembly, leaving user-defined fusion sites that promote directional, multifragment assemblies. Parts can be provided as PCR products (primers adding fusion sites and Type IIS sites), and/or circular plasmids, and are joined in a single digest-ligate reaction, including the destination vector, Type IIS enzyme, and T4 DNA ligase. If functional parts are equipped with the same fusion sites, mixing them in one reaction automatically generates combinatorial libraries. A built-in ccdB Toxin, counter screens unsuccessful assemblies, preventing carry-over of the circularised template backbone. Together, these features offer an ideal framework for generating randomised, modular, and standardised expression cassettes for BGC activation.

BGC-carrying vectors are either sourced from (meta)genomic libraries or from a defined BGC of interest. Because large-cluster capture methods vary in different labs, we kept the backbone choice flexible; in this study, we used the CopyControl fosmid pCC2Fos (Epicentre, Madison, WI, USA). Due to the inherent λ-phage head space limitation, the use of this system restricts the insert size to approximately 40kb. This size is well-suited for smaller BGCs but not for larger multimodular NRPS and PKS clusters. The same backbone can, however, be used without phage packaging (standard ligation/recombineering approaches), thereby circumventing the 40 kb limit. The final vector is simple and stable, containing only the BGC (or library) in a low-copy plasmid, conserving space and preventing unwanted edits before activation.

Furthermore, we required a method to merge the BGC-containing fosmid with the HEL-cassette (Figure 21), which did not depend on insert size or restriction sites, and aimed to avoid leaving large scars between the HEL-cassette and the BGC. Although in vivo homologous recombination in yeast or *E. coli* could achieve scarless fusions, co-transformation of large vectors is often inefficient and can introduce bias due to the insert size. Therefore, we prioritised in vitro strategies. While Transposon-based systems, such as mariner, enable broad applicability due to minimal sequence requirements, the integration events are random and non-directional. The λ-phage Gateway® recombination platform transfers defined inserts flanked by *attL* sites to *attR*-flanked vectors in a targeted manner. This reaction is catalysed by an integrase (Int), an integration host factor, and an excisionase (Xis), and is supplied as a ready-to-use reaction mixture. This system offers deterministic orientation, is restriction-independent, and leaves only small scars (13bps).

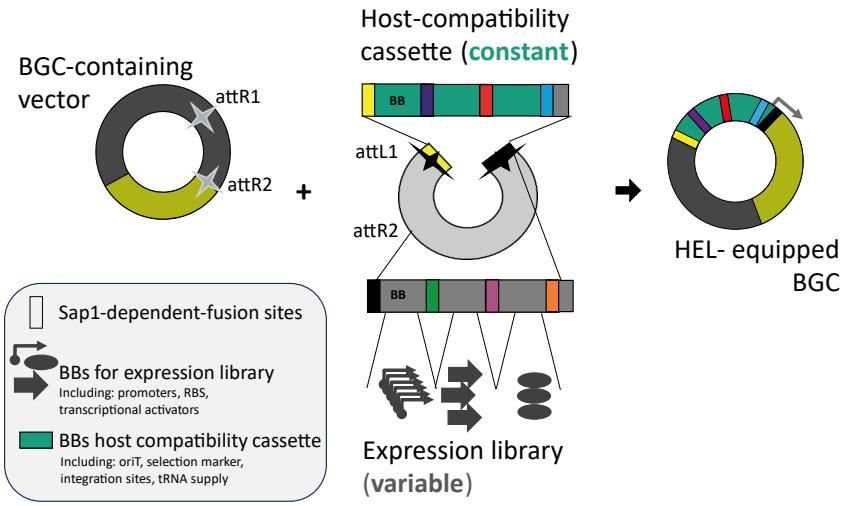

Figure 21: General principle of the HEL-Cloning approach. The BGC containing vector (cloning process not shown) is retrofitted with the HEL-vector containing host-specific constant factors (e.g. origin of transfer, resistance cassette, integration sites) and an expression library, designed to generate a randomised combination of expression factors (promoters, tRNAs, inducers) to increase the chance of inducing expression of the BGC.

To ensure compatibility with various cloning techniques, selected vectors—pUC19 for the HEL-storage vector, pENTR4 for the HEL-assembly vector, and pCC2Fos for the BGC-containing fosmid—were (i) modified to remove internal Type IIS recognition sites used by MoClo (BsaI and SapI), (ii) equipped with *ccdB* negative selection cassettes to counter-select for template backbones, and (iii) fitted with att sites where necessary to generate attachment sites for retrofitting the BGC-containing fosmid. All cloning was performed in *E. coli* standard cloning strains. Constructs carrying *ccdB* were maintained in the *ccdB*-resistant strain *E. coli* DB3.1. Building blocks were introduced via PCR amplification, gene synthesis, or oligo annealing, depending on the size and availability of the template. All assemblies were initially subjected to diagnostic digests before being verified by Sanger sequencing. Full cloning procedures, structure of backbones, and descriptions of building blocks are detailed in the Materials and Methods section. The vector set is available upon request.

In summary, the HEL-pipeline proceeds through three stages (Figure 22). First, functional parts (Figure 22A, Table) are equipped with fusion sites (coloured rectangles), either via PCR or using a pre-cloning step with universal fusion sites (white) into HEL-storage vectors, providing the necessary fusion sites later. Second, assembly of HEL-storage requires a BsaI-dependent MoClo process. Third, PCR products or BBs present in storage vectors are then assembled SapI-dependently in a second MoClo reaction (Figure 22B), utilising HEL-bridging vectors whenever not all positions are filled by BBs. This reaction produces the final expression cassette, which is compatible with the host and then used for retrofitting the BGC-containing vector (Figure 22C) by Gateway assembly. The cloning hierarchy also relies on changing selection markers (AmpR → KanR → CmR) at various cloning stages, allowing for the counterselection of the original vectors. If a BGC-containing fosmid of interest is ready, the cloning of the expression and host-compatibility cassettes typically takes about

three days, and an additional three days are needed for Gateway assembly and verification of the constructs (if required). Subsequently, the retrofitted HEL-fosmid can be transferred into the desired heterologous host.

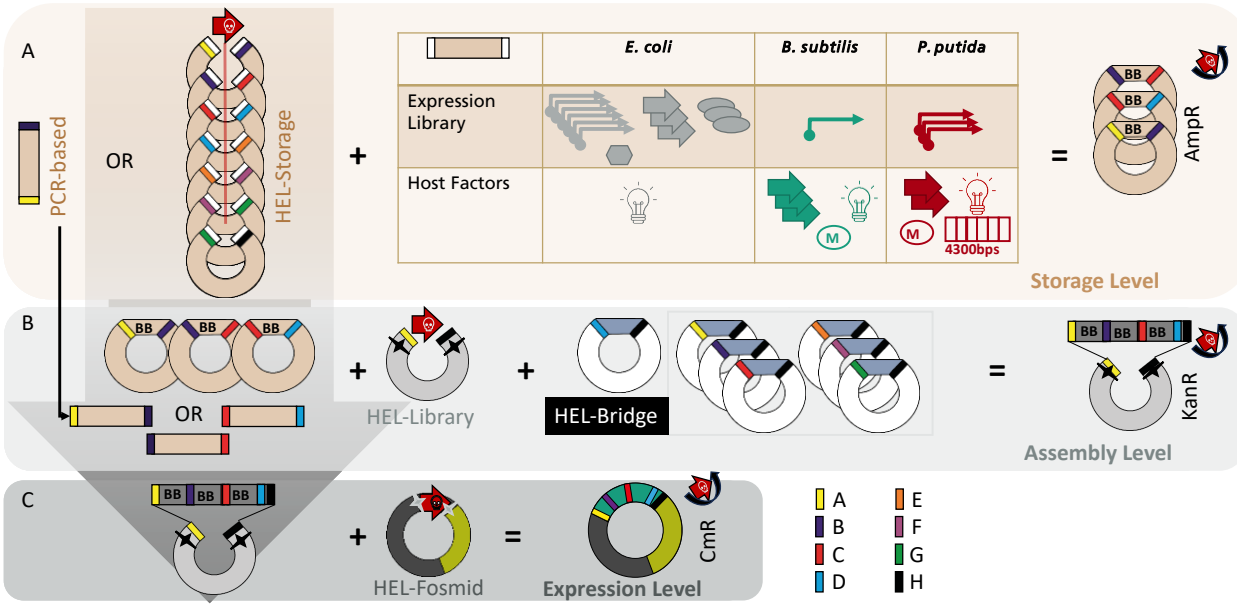

Figure 22: HEL vector set and cloning procedure. (A) BBs are created either by BsaI-dependent cloning or by PCR (beige). (B) The expression and host-compatibility cassette are assembled from PCR parts or BBs using Sap1-dependent MoClo and HEL-bridges when positions are not fully taken (light grey). (C) The pre-assembled BGC-containing fosmid (dark grey) is retrofitted by Gateway cloning. All target vectors include counterselection with the *ccdB* toxin (red arrow). If assembly fails, the *ccdB* toxin remains in the vector and kills future transformants.

This design achieves three practical outcomes. First, BGC assembly remains restriction-independent, making it ideal for sequence-unknown or GC-rich DNA. Second, expression modification becomes modular and reusable: the same set of parts can be mixed and matched for optimisation in *E. coli* and then redeployed on other BGCs without the need for rebuilding. Third, compatibility is directly included: host compatibility cassettes standardise the transfer to additional hosts, so the same captured BGC can be screened across multiple backgrounds without the need for reassembly.

2.9.3.2 Validation of the HEL workflow with a GFP Reporter Fosmid

To verify the functionality of the cloned vector set and procedure end-to-end, we replaced the BGC space on the HEL-fosmid with GFP and assembled a HEL-library vector carrying constitutive promoters of defined strengths (weak: BBa_J23117, medium: BBa_J23106, strong: BBa_J23100) with PCR-Parts positioned at slot A-H (yellow-black). PCR parts were supplied in a 1:1:1:1 ratio to evaluate potential cloning bias (Figure 23A). *Escherichia coli* Epi300 was transformed with the retrofitted HelFos_GFP, and the resulting transformants were arrayed individually in a 96-deep-well plate and conserved as a pool by retrieving whole-cell mass from the agar plate. After isolating plasmid DNA from the individual transformants, the plasmid DNA was sequenced to evaluate the sequence correctness. From 135 colonies, 100 were sequenced, 98 returned high-quality reads, and 97 were correctly assembled. This indicates a relatively high cloning fidelity. The representation of different promoters was biased towards the medium strength promoter (n=46), followed by the weak promoter (=34) and the strong promoter (n=17). From these clones, we selected identified representatives of

the individual promoter strengths as a control for subsequent flow cytometry analysis (Figure 23C). We also included a clone that does not carry any promoter (green) for conclusive assessment of promoter distribution and assembly bias. Single-clone measurements separated distinctly according to promoter strength. Still, a large fraction of cells within each clonal population was non-fluorescent, consistent with known cell-to-cell variability, especially when using plate-grown inoculum that may represent different heterogeneous physiological states. In the pooled library (Figure 23D), fluorescence corresponding to low/medium/low-strong promoters was evident, but a highly fluorescent "strong" peak matching the strong-promoter control was absent. Taken together, this indicates a composition bias during the cloning procedure while still demonstrating the overall functionality of the pipeline.

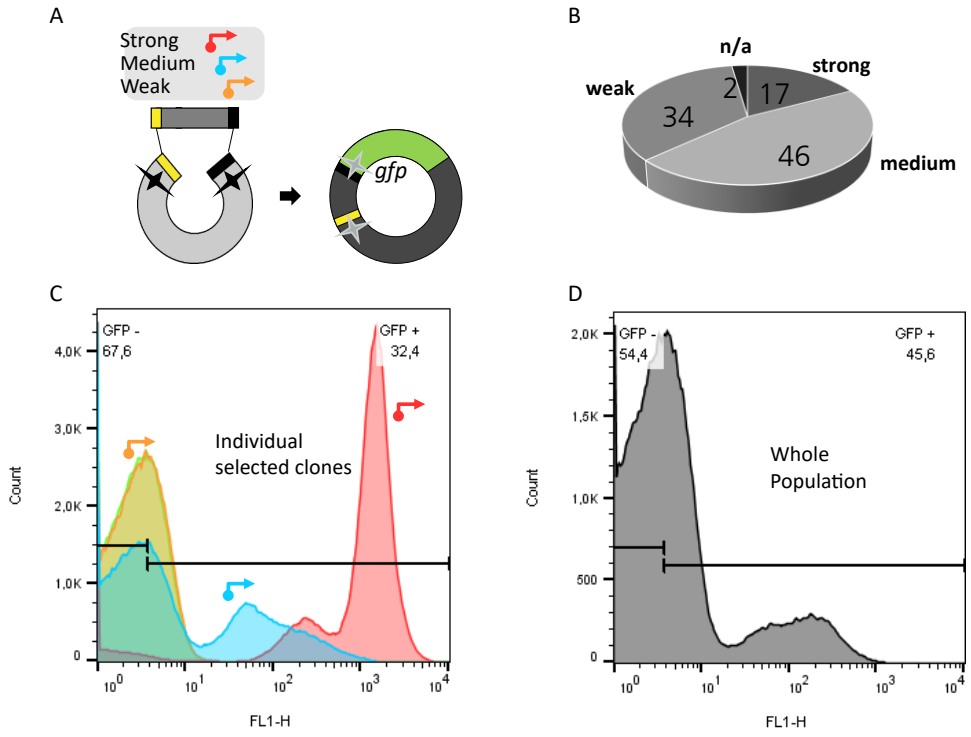

Figure 23: HEL cloning assessed with a constitutive promoter library and GFP reporter. (A) The promoter library was cloned into the HEL vector and recombined with a GFP fosmid. (B) Promoter types in 98 clones were identified by sequencing. (C) Flow cytometry of individual selected promoter clones shows that different promoters drive varying GFP signals, with some overlap and heterogeneity. (D) Most clones exhibit medium to strong GFP expression, although overall signals are lower than those in selected individual clones.

We next tested inducible promoters to (i) expand part size/complexity, (ii) verify functional responsiveness of retrofitted HEL-fosmids, and (ii) reevaluate cloning bias. First the HEL-library vector was assembled using pLUX (AHL-inducible), pBAD (arabinose-inducible), and pLAC (IPTG-inducible) at an input ratio 5:3:2. Although MoClo and Gateway reactions yielded many colonies, GFP induction was observed only with arabinose supplementation and in low numbers (Figure S3 1). Sequencing of ten HEL-library clones showed that eight successful sequencing reactions carried pBAD, one carried pLAC, and pLUX was absent, despite having the highest input. Varying promoter inputs across four mixes–(i) six pBAD: one pLAC: three pLUX; (ii)

1:1:1; (iii) 3:6:1; (iv) 1:6:3, still produced overwhelming pBAD recovery in (i)–(iii), and (iv) yielded one pLAC within nine pBAD clones. These data indicate a strong bias for pBAD PCR products during library assembly.

We hypothesised that PCR-derived parts contributed to bias via inaccurate quantification and occasional fusion site errors. To stabilise inputs, we introduced an intermediate HEL-storage level (Figure 22B). Each functional part was first equipped with universal fusion sites by PCR before being inserted (BsaI-dependent MoClo) into a HEL-storage plasmid with desired fusion sites (BB) for the next assembly level. Using HEL-storage-derived BBs, we rebuilt the inducible-promoter library with an equimolar 1:1:1 mix of pBAD, pLAC, and pLUX. The subsequent sequencing of 96 HEL-library clones revealed that all three promoters were now recovered, with relative frequencies of approximately 2.1 pBAD:1 pLAC:0.2 pLUX. This indicates a substantial improvement over the PCR-based workflow, though pLUX remained under-represented.

We noticed that Gateway assemblies carrying inducible promoters (larger cassettes) yielded more correct clones than those with small constitutive promoters, suggesting a size effect of selected BBs. To quantify this, we rebuilt constitutive-promoter libraries by adding spacer BBs to generate cassettes of 0, 333, 500, 1,333, and 2,500 bp, upstream of the respective promoter. We compared two input types: (i) PCR-derived parts and (ii) HEL-storage BBs. After the full HEL procedure, individual clones were arrayed in 96-well plates alongside defined promoter controls, grown overnight, and measured for GFP fluorescence and growth by optical density (OD_{600}). Fluorescence was normalised to OD and plotted as a violin plot for different spacer sizes (x-axis) (final data handling is described in Methods).

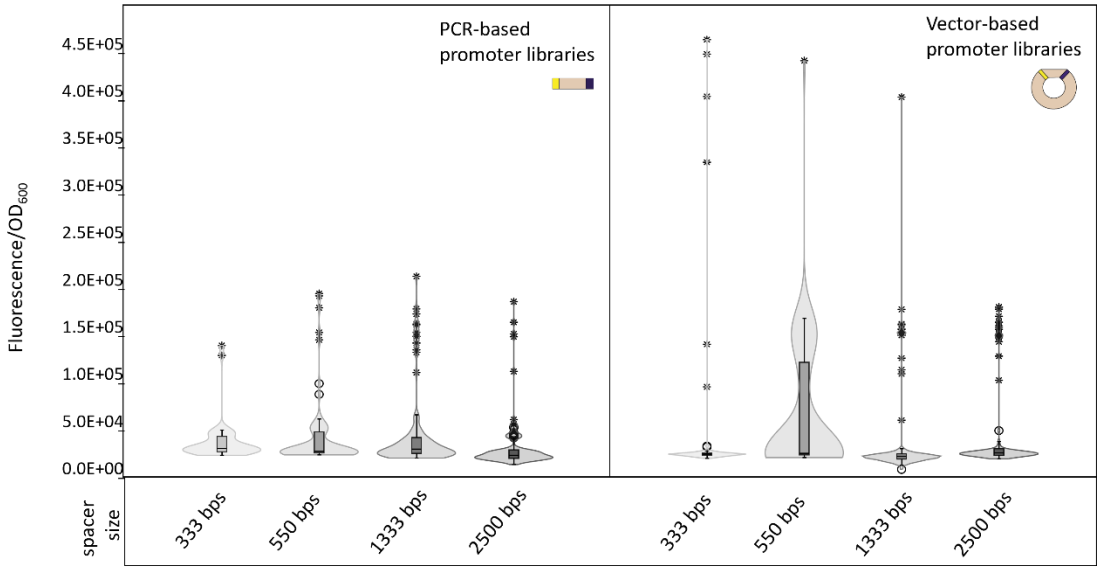

Figure 24: GFP Fluorescence in clones built using constitutive promoters and different spacer sizes, comparing size effects by two distinct methods: PCR-based promoter parts and vector-based BBs. Fluorescence was measured for 88 clones in 96-well plates after overnight incubation. Fluorescence values were normalised to OD and presented as a violin plot to illustrate the signal range and diversity. Results show a higher range and diversity from the vector-based BBs method compared to the PCR-based parts method. The size of the individual parts appears to have only a minor effect, especially for the PCR-based parts.

For PCR-derived parts (Figure 24, left), fluorescence diversity clearly increased with cassette size: libraries Figure 24, right) with larger spacers displayed a range of up to ~13-fold between the lowest and highest signals,

versus ~2-fold without spacers. In contrast, HEL-storage BBs showed consistently broader fluorescence distributions across 333–1333 bp cassettes (≈10-fold spread), and higher overall signal than PCR-based parts; performance dropped at 2500 bp, even though the method was used for larger assemblies for this study (see below) and is documented to work for fragments of up to 150 kb (Thermo Fisher Scientific, 2025). Together, these results suggest that pre-cloned BBs enhance overall robustness and library diversity. Additionally, they can be clonally amplified and stably stored in cells for longer than linear PCR-based parts.

2.9.3.3 Generation of a Strain-Independent High-Copy Fosmid Induction

The HEL-fosmid is harboured as single-copy vector within its respective *E. coli* strain. However, the fosmid system provides a copy-number induction that relies on *E. coli* Epi300, which expresses TrfA from a chromosomal locus regulated by the pBAD promoter. TrfA activates *oriV* (RK2-derived), a second replication origin present on the fosmid. This makes the copy control dependent on a single strain, complicating work with specialist chassis (e.g., BAP1 for NRPS/PKS). We therefore engineered a vector-encoded system to confer strain-independent copy-number induction.

We built a HEL BB that places *trfA* (position BC) under pBAD control (position AB, arabinose-inducible), and an IPTG-inducible T7 promoter at position CD to drive GFP expression (Figure 25A); the gap was bridged using HEL-Bridge-DH. We anticipated that this vector-based control could lead to cell stress and ultimately to self-toxicity due to metabolic overload (Figure 25B). To test this effect, we measured the growth of the high-copy phenotype at different concentrations of kanamycin (resistance marker, 50 µg/ml - 1x, 10x, 100x and 1000x) in comparison to the low-copy vector phenotype (Figure 25C). One strain was induced with arabinose (5.3 mM), while the other was left untreated, and growth was observed for 24 h. At 1×, growth of the induced high-copy strain closely matched the low-copy phenotype (no prolonged lag). Notably, the low-copy strain grew only at 1×, whereas the high-copy strain also grew at 10×, consistent with greater resistance-gene dosage under copy induction. We then quantified GFP after 6 h of induction with arabinose (5.3 mM), IPTG (6.4 mM), or both; OD_{600} and fluorescence were measured, and GFP was normalised to OD. Arabinose induction alone produced a slight increase in GFP (probably reflecting promoter leakiness). IPTG yielded an approximate 2-fold increase over uninduced, and IPTG plus arabinose produced a further ~1.6-fold increase relative to IPTG alone. This observation matches the expected behaviour of copy-induction amplifying the T7-driven GFP expression. Thus, the cassette behaves as designed: IPTG controls transcriptional activation, while arabinose modulates plasmid dosage, additively boosting expression.

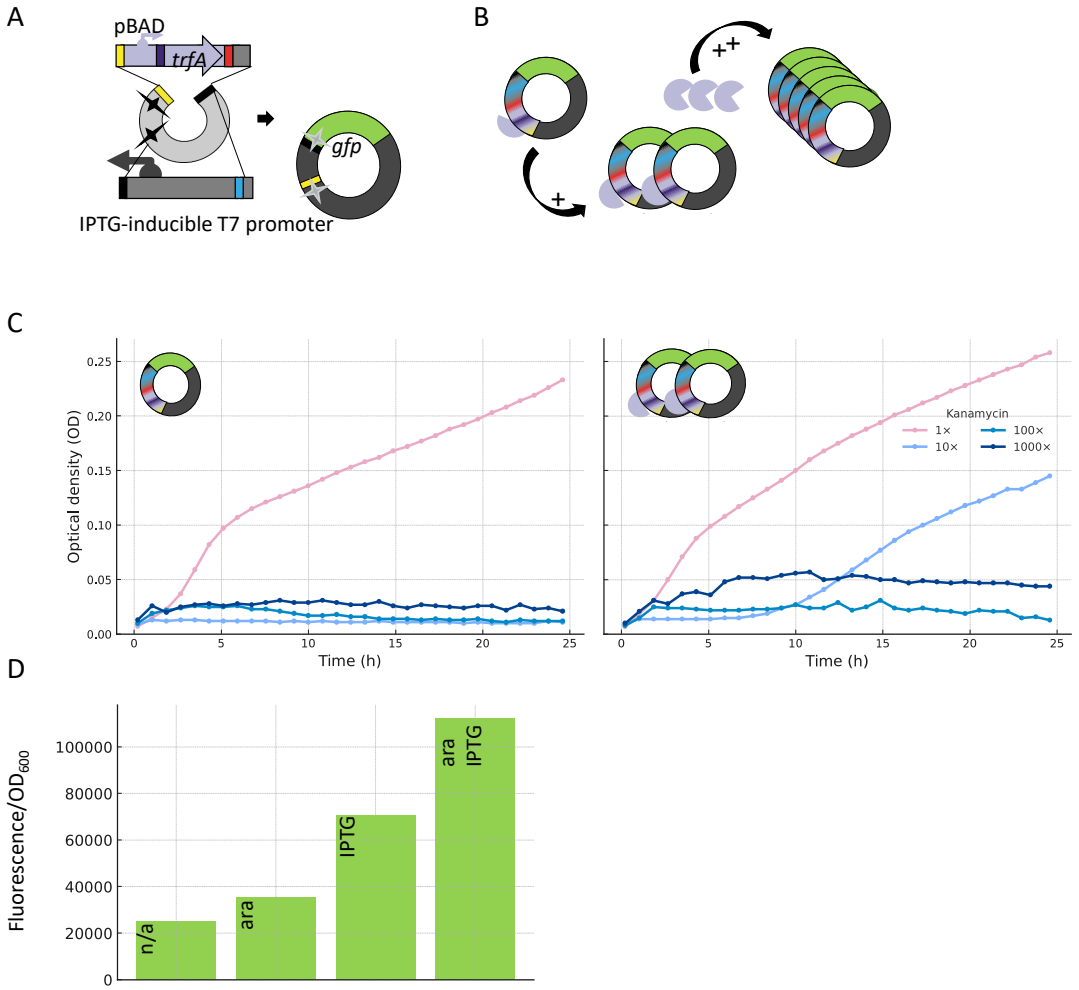

Figure 25: Evaluation of the vector-encoded, high-copy cassette in strain Rosetta/HELFos-*trfA*GFP. (A) Cloning design showing the arabinose-inducible *trfA* replication initiation protein (purple). It is placed upstream of the IPTG-inducible T7 promoter, which drives GFP expression after Gateway recombination with the HELFos-GFP vector (dark grey). (B) TrfA-dependent replication of HELFos-GFP triggers self-initiation mechanisms, leading to a high vector copy number. (C) Growth kinetics are shown over 24 h in 96-well plates. Strains were left uninduced (low-copy phenotype) and induced (high-copy phenotype) with 5.3 mM arabinose. Growth curves reflect different conditions at varying kanamycin concentrations (1x - 50 µg/ml, 10x, 100x, 1000x). The high-copy phenotype does not cause growth defects compared to the low-copy phenotype. It also shows growth at 10x kanamycin concentration, indicating increased resistance due to higher gene dosage. (D) Fluorescence values (normalised to optical density) are presented for each induction condition. Upon arabinose induction (high-copy phenotype only), there is a slight increase in fluorescence, likely from background expression. With IPTG induction (GFP expression), a clear GFP signal appears. The signal further increases upon high-copy induction.

This data indicates that plasmid-encoded TrfA provides effective, strain-independent copy-number control that does not inflict growth impairment due to self-toxicity. The module offers two practical advantages: (i) tuneable amplification of HEL-driven expression independent of the *E. coli* host strain, and (ii) a diagnostic measure of product burden/toxicity by comparing induced versus uninduced states (copy-up ± promoter induction) for any unknown BGC or insert of interest.

2.9.3.4 Expression of a Known BGC (Darobactin) using HEL

To validate HEL with a characterised pathway, we cloned the darobactin biosynthetic gene cluster into the HEL-fosmid. Darobactin is a ribosomally produced peptide that selectively kills Gram-negative pathogens by targeting the BamA protein in the outer membrane. It is encoded on a 10kb BGC, which has already been successfully expressed in *E. coli* using the established pRSF-Duet vector system. We reused the HEL construct with the IPTG-inducible T7 expression cassette and the plasmid-encoded TrfA/*oriV* module for retrofitting of the darobactin-containing fosmid (HELFos-daroV1). This expression cassette closely resembles the pRSFDuet-based, cloning setup that also utilises the strong P7 promoter. Initial transformation into *E. coli* EPI300 yielded no detectable darobactin after extraction. Given the prior success in *E. coli* Rosetta, we updated the HEL host cassette to include a compatible selection marker and transformed *E. coli* Rosetta DarR (darobactin-resistant version) with the new HELFos-daroV2 (including kanamycin resistance). In this strain context, UHPLC–MS analysis detected darobactin in culture extracts, confirming pathway functionality on the HEL-fosmid.

We next compared growth (± arabinose copy induction; 24 h, 96-well format) and production (flask scale) across three conditions: (i) HELFos-daroV2 low copy phenotype (uninduced), (ii) HELFos–daro high copy (arabinose-induced TrfA), and (iii) the pRSF-Duet positive control (Figure 26). The growth of the HEL-transformed strains was similar in the first ~4 h; thereafter, the high-copy strain exhibited a slower increase in OD_{600}, a phenotype not observed with the GFP reporter and therefore consistent with a product-associated burden (auto-toxification). Relative quantification by UHPLC–MS revealed that pRSF-Duet outperformed both fosmid-based constructs by more than 10-fold. Within the HELFos-daro set, high copy produced a larger peak area than low copy, although this is partially offset by lower biomass in the induced culture. Normalising product to biomass (OD or dry weight) would allow a more accurate estimate of specific productivity.

These results demonstrate that HEL supports the expression of a bioactive, antimicrobial BGC, but also underscore the importance of host selection: expression failed in EPI300 yet succeeded in Rosetta, likely reflecting the rare-codon tRNA supply from the Rosetta plasmid. For future projects, a rare-tRNA BB could be incorporated into the HEL library to decouple expression from specialised strains. Finally, assembling a HEL expression library (promoter/RBS/regulator variants ± helper modules) around the darobactin locus will enable systematic tuning and could narrow the performance gap with the pRSF-Duet benchmark expression system.

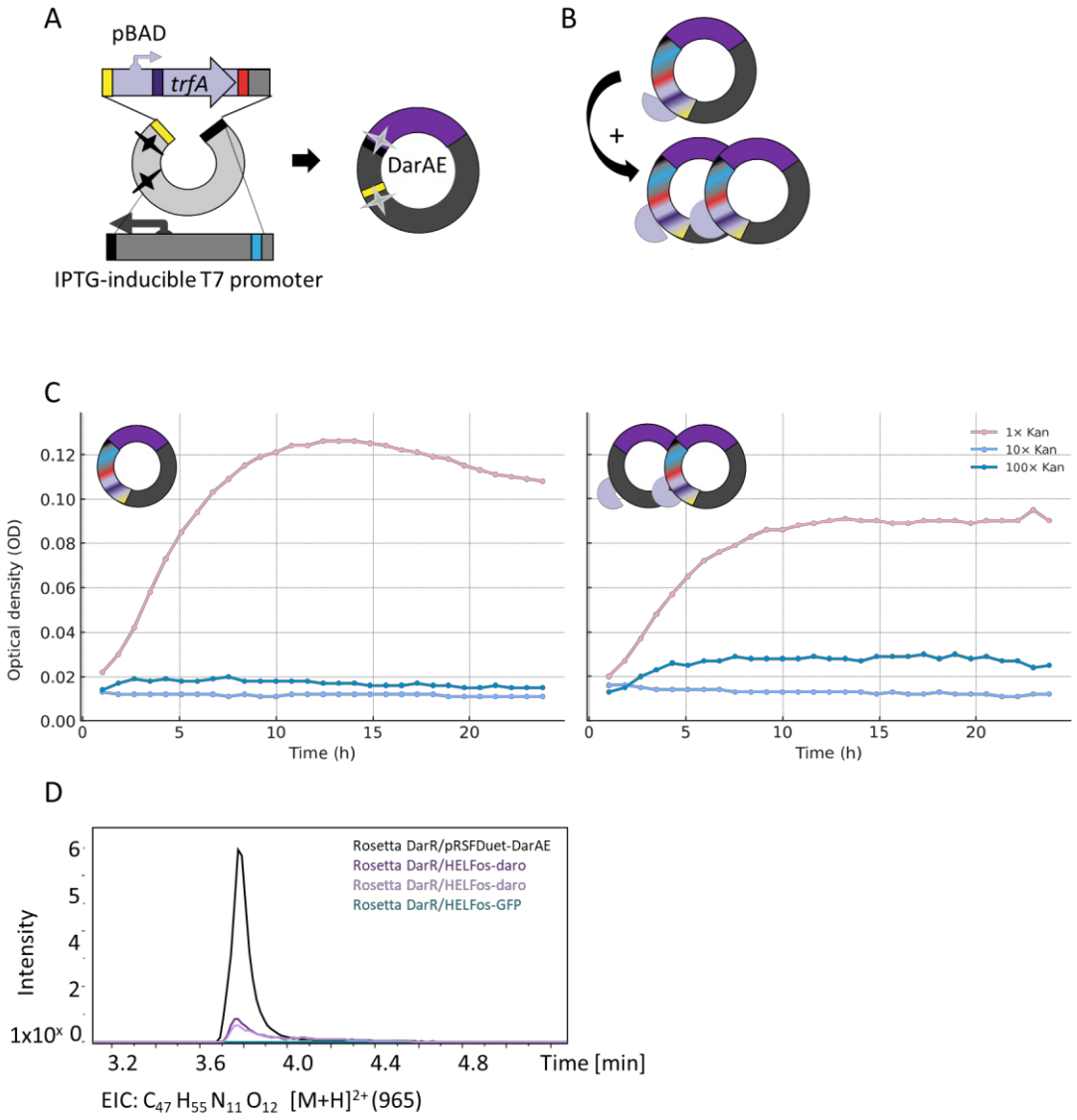

Figure 26: Expression of darobactin using the HEL-cloning system. (A) Cloning design of the darobactin-HEL-fosmid, which includes a high-copy cassette induced by arabinose from pBAD and the IPTG-inducible P7 promoter routinely used in our laboratory for darobactin expression. (B) Growth of *E. coli* Rosetta/FosHel_daro strain with (orange) and without high-copy induction. Growth is impaired in the high-copy phenotype compared to the low-copy phenotype. (C) EIC of darobactin extracted from *E. coli* Rosetta/pRSFDue-daro (red) and from both high-copy and low-copy phenotypes of the strain containing the darobactin HEL-fosmid. The peak area for the established pRSFDuet system is 10-fold higher compared to the HEL system. While the high-copy phenotype shows a larger peak area, this could be due to the lower cell mass observed in the growth experiments.

2.9.3.5 Retrofitting HEL-Fosmids for Non-*E. coli* Hosts

We next examined whether HEL-fosmids can be adapted with host-specific cassettes to enable transfer and expression in hosts beyond *E. coli*. As a proof of principle, we targeted two widely used chassis—*Pseudomonas putida* KT2440 and *Bacillus subtilis* SCK16—to supplement existing *Streptomyces* protocols in our laboratory. Both hosts are natural NP producers and already encode key genetic traits necessary for NP expression. Both chassis are genetically tractable, with established genetic toolkits (Martínez-García *et al.*, 2023; Popp *et al.*,

2017). They are both genetically engineered to have a relatively clean metabolic background, and each offers powerful secretion and efflux capacities. Notably, *P. putida* is known to be tolerant to xenobiotics, therefore making it suitable for robust production of complex and toxic scaffolds. We aimed to express GFP and darobactin in the hosts. To facilitate chromosomal insertion in *B. subtilis*, the HEL-fosmid backbone needed to be complemented with an *amyE* homology arm, enabling the integration of the entire construct into the host genome at the *amyE*, (coding for an amylase) gene locus. For *P. putida*, the fosmid backbone remained unchanged, allowing HELFos-darobactin, previously used for *E. coli*, to be reused for host retrofitting. In addition to darobactin expression, we aimed for a simpler reporter readout and created fosmids with host-specific, already established GFP reporters (*P. putida*: BBa_E0040, *B. subtilis*: Overkamp *et al.*, 2013) . For each host, we designed a minimal cassette to provide replication/integration, selection, and expression. In *B. subtilis*, the cassette carried another *amyE* homology region (complementary to the on the vector), a kanamycin resistance marker, and the strong, constitutive, *Bacillus*-specific promoter pVEG (parts described in Radeck *et al.*, 2013) (Figure 27Ai). In *P. putida*, the RK2-based origin *oriV* works reliably, so we included *oriT* (BBa_J01003) for optional conjugation, an apramycin (*aaC*) resistance marker, the high-copy induction cassette and the Xylose-responsive promoter Pm (Martínez-García *et al.*, 2023) (Figure 27Bi). We used electroporation, as described previously, to transfer fully assembled fosmids into *P. putida* KT2440 and induced natural competence in *B. subtilis* SCK6 using xylose. Transformants were selected on the appropriate antibiotics and restreaked before being tested for their strain identity and potential integration events. For *P. putida,* colonies were streaked out on test strips, indicating oxidase-positive (e.g., *P. putida*, Ox-pos vs. *E. coli*, Ox-neg) strains by a blue colour (Figure 27Bii). For *B. subtilis*, strains were grown for two days on agar plates containing starch. After incubation, the plates were stained with Lugol's solution, which stains the undigested starch. Wild-type *Bacillus* strains typically carry the amylase gene (*amyE*) for starch digestion. In such cases, Lugol's solution can be washed out. When the desired insert is successfully integrated into the *amyE* locus (as described above), the starch can no longer be processed, and the solution colours the remaining starch in the plates (Figure 27Aii). Plates of wild-type *B. subtilis* SCK6 do not show dark purple dye after rinsing of the plate. The transformant selected, potentially carrying HELFos_GFP, exhibits a dark purple colouration, indicating the presence of remaining starch in the medium, likely due to the disruption of the *amyE* locus. Oxidase-positive and amylase-negative strains were boiled to obtain lysates for colony PCRs of product cassette/BGC products (Figure 27ABiii). For both hosts, the lysates yielded the expected PCR products of approximately 2000 bp for *B. subtilis* SCK6 and around 2500 bp for *P. putida* KT2440. In contrast, the controls were negative, confirming the presence of the constructs. These verified clones were stored and subsequently used for expression assays with GFP and darobactin. Despite testing multiple media (LB, Spizizen, M9), vessel formats (96-well plates, Erlenmeyer flasks), incubation regimes (shaking adjusted, samples taken on day 1 (exponential growth), 3 and 5), and readouts (flow cytometry, microscopy, and plate bioassay), we did not reliably detect GFP in either host, and UHPLC–MS failed to detect darobactin. However, culture broths of *B. subtilis* carrying the darobactin construct exhibited visible changes in growth/morphology, which were absent in the controls (Figure S3 2).

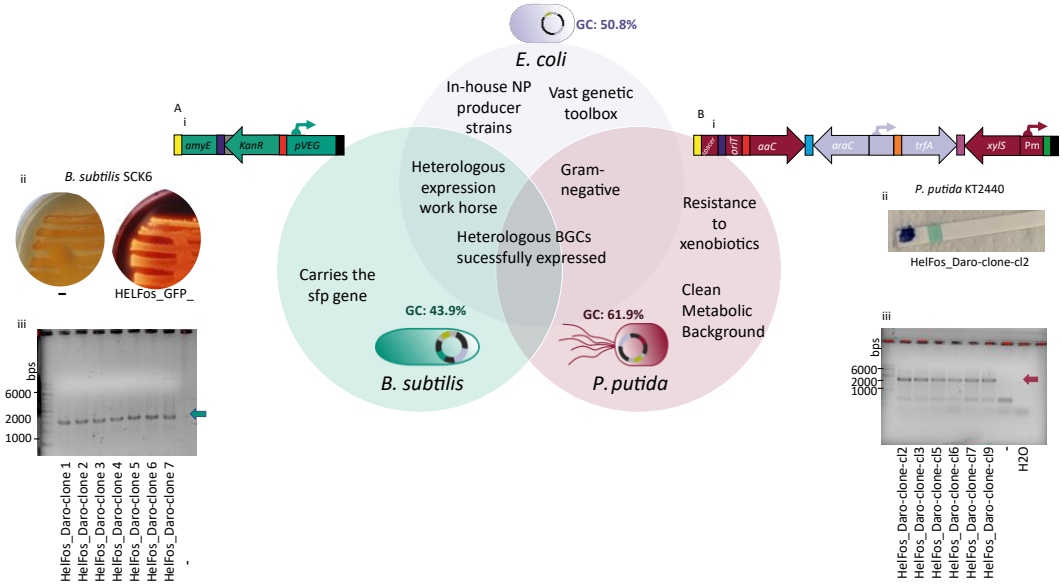

Figure 27: Transfer of HEL-fosmids into heterologous hosts *B. subtilis* SCK 6 (green) and *P. putida* KT2440 (red). The Venn diagram illustrates the characteristics shared and distinct among three heterologous expression hosts (including *E. coli* strains, represented in purple). Host-specific,c cassettes are shown (ABi), strain identity was confirmed (Bii), and integration was tested by starch assay (testing for amylase absence) (Aii) or presence (ABiii) as confirmed by colony PCR.

Several factors may cause a lack of expression. First, the genetic architecture may not be suitable for expression; for example, the fosmid may be maintained at a low copy number in *P. putida*, resulting in low expression. Second, in *B. subtilis*, integration at *amyE* can be unstable or incomplete, resulting in mixed populations and consequently low production. Third, transcription and translation of products can often be mismatched in heterologous hosts. While this may be true for darobactin, with no reports of heterologous expression apart from in *E. coli*, the GFP cassettes used were both optimised and have been previously tested in the respective hosts. To address these issues in upcoming experiments, we could verify copy number and genomic context through Southern blot or targeted sequencing of PCR products and measure transcription with RT-qPCR. In parallel, we could introduce helper modules (rare-tRNA sets), test alternative expression factors (promoters/RBSs), and evaluate the high-copy phenotype to adjust episomal dosage in *P. putida*, a shown requirement also in *E. coli* to succeed in Darobactin-expression.

In summary, we established the reliable transfer and genetic presence of HEL-fosmids in *P. putida* and *B. subtilis*; however, functional expression has not yet been achieved. This emphasises the core HEL idea, that host retrofitting and expression tuning must be co-optimised. Accordingly, we will apply the HEL library in these chassis—systematically varying promoters, RBSs, regulators, and helper building blocks—and reassess GFP for rapid readout and darobactin for NP output. Furthermore, we plan to reassess expression using BGCs that are natively maintained by the respective chassis (e.g. pseudopyronin in *P. putida* and rhizocticin in *B. subtilis*). Analysing and establishing transcription will be the immediate milestones, followed by optimisation of expression strength to demonstrate HEL's usability in non-*E. coli* hosts. Upon successful

expression, the required combination of expression and host factors can give insights into the functional requirement of successful BGC expression, revealing basic biological prerequisites and mechanisms.

3 DISCUSSION

3.1 Summary

This thesis combines genome mining, cultivation, metabolomics, and heterologous expression approaches to explore the NP reservoir of two underexplored bacterial phyla, the Bacteroidota and the Acidobacteriota, and introduces a host-optimised heterologous expression system for the expression of yet unexplored BGCs. A standardised genomic workflow was established for identifying BGC hotspots and prioritising strains for experimental follow-up. In the **Bacteroidota**, we identified the genus *Pedobacter* in general and especially the *P. cryoconitis* clade as particularly enriched in multimodular NRPS gene clusters. Furthermore, we isolated 21 novel *Pedobacter* strains from amphibian skin and integrated them into the analysis. While culture extracts of investigated strains did not show antimicrobial activity, the genomics-guided workflow and the establishment of the genetic tools allowed the isolation of structurally diverse lipopeptides cryopeptin and derivatives thereof, while at the same time linking the BGC of interest to those NPs. This represents one of the first systematic demonstrations of how genome mining in *Pedobacter* can be experimentally validated, highlighting its untapped biosynthetic versatility. In general, the similarity clustering of the overall BGC potential did not indicate many similarities with known reference clusters, further highlighting the novel BGC space present in the *Pedobacter*. Complementary studies expanded the genomic diversity of marine Bacteroidota through sequencing isolates of *Sinomicrobium*, *Algoriphagus*, *Roseivirga*, and *Galbibacter*. These genomes revealed flexirubin-like and pinensin-like clusters, which expand the chemical framework for studying Bacteroidota in natural product discovery. Together, these findings confirm Bacteroidota as not only abundant and ecologically significant microorganisms but also as a promising reservoir of structurally unique NPs.

Members of the **Acidobacteriota** phylum are known to be difficult to culture due to their slow growth rates. Selective cultivation strategies were used by (Oberpaul *et al.*, 2022)Oberpaul *et al.* to isolate four novel isolates that were subsequently characterised, cultivated and profiled by untargeted genomics and metabolomics. This uncovered the production of phytohormones such as indole-3-acetic acid (IAA) and isopentenyl adenine (iP). Tryptophan supplementation induced broad metabolic reprogramming, including the biosynthesis of indole-derived compounds previously described to be produced by marine *Paracoccus sp.*, *Burkholderia sp.* and various fungal species. Although we could not confirm a plant growth promotion effect of the strains on barley seedlings, these findings highlight the strong dynamic response the plant's exudate Trp has on the overall metabolome and hint towards possible interactions between plants and Acidobacteriota.

Finally, the **HEL heterologous expression platform** was designed and evaluated to contribute to the development of a synthetic biology tool to bridge the gap between genomic predictions and metabolite discovery. This toolkit enables the modular retrofitting of captured BGCs with expression and chassis-specific elements and provides a versatile tool for future access to cryptic BGCs. Together with genome mining and cultivation-dependent approaches, HEL represent a third axis of discovery if the BGC is not accessible in the native host.

3.2 Literature Context and Ecological Functions

3.2.1 Genome Mining of Bacteroidota

Genome mining across the Bacteroidota recognises the phylum as an underexplored reservoir of specialised NPs (Brinkmann *et al.*, 2022a; Covas *et al.*, 2023), with *Pedobacter* emerging as a biosynthetic hotspot. The most recent Bacteroidota NPs were isolated by (Brinkmann *et al.*, 2022c)who reported the chitinopeptins and the pentacitidins (Brinkmann *et al.*, 2022c), the latter closely related to the previously described falcitidins (Somanadhan *et al.*, 2013). In their review, Brinkmann *et al.* (2022a/b) list around 34 discrete NP structures, including macrolides, peptides, quinolines and β-lactams, produced mainly by *Chitinophaga*, *Cytophaga*, *Flavobacteria* and *Flexibacter*. This breadth underscores the chemical range accessible within the phylum and highlights also *Pedobacter* as a particularly promising genus for novel NP utilised for human needs.

Within *Pedobacter*, previously described NPs were essentially limited to the (iso)pedopeptins, yet multiple genome and metabolome studies have hinted at additional BGC potential (Bjerketorp *et al.*, 2021). Building on that, we isolated and structure elucidated the cryopeptins, unique linear peptides that were investigated in parallel by Figueiredo *et al.*, 2022 using MS-based approaches. Beyond the cryopeptins, we also isolated novel siderophores from *Pedobacter* (Liu *et al.*, manuscript in preparation). Complementing these metabolic findings, our genomic analysis further identified a BGC putatively encoding a monobactam in *Pedobacter psychrotolerans* DSM 103236. This cluster (SLWO01000004.1 - Region 2) contains a SulN (AOZ21321.1) homologue within an NRPS BGC. It co-occurs with predicted *sulG*/*sulH* homologues for L-2,3-DAP supply as well as a carbamoyltransferase homologue of the caphamycin C BGC (BGC0000319) of *Streptomyces clavuligerus* ATCC 27064. The overall architecture is consistent with the BGC of an SQ28332 scaffold, a NP produced by a *Flexibacter* species. Taken together, these features suggest that *P. psychrotolerans* DSM103236 carries the genetic potential to produce an SQ 28332-like scaffold (Singh *et al.*, 1983), justifying current efforts into the isolation of this NP in our lab. SQ28332 shows weak activity against gram-positive bacteria and is resistant to β-lactamases P99 and TEM-2. Due to the similarity of the BGC to the already described BGC, we can speculate that the strain is also producing an antimicrobially active NP that could help compete under the otherwise harsh, low-temperature conditions from which the strain was originally isolated.

Pedobacter genomes are enriched in AMR genes, especially β-lactamases (Viana *et al.*, 2018a), that hydrolyse β-lactam scaffolds. Among 21 *Pedobacter* species isolated from amphibian specimens, we observed a high β-lactamase gene load (mean 5.6 ± 2.5 per strain). Notably, we also identified β -lactamase genes putatively belonging to class D (Bush and Jacoby, 2010; Zumkeller *et al.*, 2024), an observation novel for the genus. Amphibian-associated *Pedobacter* strains form part of the cutaneous microbiome (Loudon *et al.*, 2014) community known to experience strong iron limitation and intense microbe–microbe competition. From the 21 *Pedobacter* species isolated from amphibian specimens, two isolates carried a desferrioxamine/woodybactin-like NRPS-independent siderophore cluster. A likely function of such NPs is securing iron, improving self-persistence, while indirectly starving fungal competitors. One strain *P. sp*. DEU302 also contained a lanthipeptide that showed some similarity to microcin and pinensin (Caetano *et al.*, 2020b). Additionally, the BGC is flanked by a TonB-dependent receptor. Such combinations are known to exploit a Trojan-horse strategy where antimicrobial peptides use siderophores to be actively transported into

the target cell (Massip and Oswald, 2020). Taken together, both gene products may help shape the community structure, potentially inhibiting the growth of fungi, thereby benefiting the amphibian host. Except for these BGCs, the remaining amphibian-isolated *Pedobacter* genomes are relatively under-equipped for NP biosynthesis, and the predicted clusters do not show similarity to known BGC families, pointing to a lifestyle characterised by a small, specialised metabolism, which allows for survival in niches where competitors are scarce.

Extending beyond terrestrial setting, marine Bacteroidota isolates link BGC diversity to specific niches (e.g sediment, sponge, and coastal seawater). Flexirubin production in *Galbibacter*, pinensin-like clusters in *Sinomicrobium*, and carotenoid pathways in *Algoriphagus* are all consistent with the literature and with adaptations to marine microbial communities, where pigments, protective metabolites, and antifungals mediate competition and symbiosis (Mussagy *et al.*, 2025). The *Sinomicrobium* isolate also encodes a BGC that is similar to the microviridin J cluster (M. N. Ahmed *et al.*, 2017; Ziemert *et al.*, 2008b), a NP family with potent protease inhibition (M. N. Ahmed *et al.*, 2017), disrupting the moulting process of *Daphnia* (Ziemert *et al.*, 2008b). Described initially from cyanobacteria and thought to act as an antifeedant (Rohrlack *et al.*, 2004), their presence and similarity suggest a similar compound in the *Sinomicrobium* isolate could have similar effects and reduce predation.

Within this broader landscape, the *Pedobacter cryoconitis* clade stands out as a hotspot of BGC potential. *P. cryoconitis* DSM14825 (Margesin *et al.*, 2003) was first isolated from organic aggregates on an alpine glacier, and a related strain was later recovered from Antarctic soil on King George's Island (Wong *et al.*, 2013). Masnoddin *et al.*, (2022) expressed three proteins from *P. cryoconitis* BG5, that retained enzymatic activity at low temperatures, consistent with adaptation to an extremophilic environment. The cryopeptins we isolated from this clade showed no biological activity in the screenings that we performed. They contain unusual dehydrovalines (Siodłak, 2015) residues known to stabilise peptides against degradation (Joaquin *et al.*, 2020; Wang *et al.*, 2024) and also found in the soil-derived antibiotic myxovalargin. Currently, we cannot infer or responsibly speculate about the biological function of this NP.

Overall, the genomic potential of the Bacteroidota by far outperforms the number of isolated compounds. To our knowledge, since (Brinkmann *et al.*, 2022c), no new NP from the phylum has been isolated and structurally elucidated, despite the phylum – and *Pedobacter* – being amenable to cultivation-based discovery. Relative to the few previously known *Pedobacter* metabolites (pedopeptins and isopedopeptins), the isolation of 14 cryopeptins substantially expands the genus's chemical space. Together with the prospective monobactam in *P. psychrotolerans* DSM103236 and the genetic toolkit enabling knock-outs and, prospectively, promoter knock-ins for cluster activation in *Pedobacter*, these results show that cultivation-based discovery in this phylum is practical and set to accelerate. Notably, cryopeptins would likely have been missed by strictly activity-guided discovery, underscoring the value of genome-informed, metabolomics-coupled pipelines for Bacteroidota.

3.2.2 Profiling of Novel Acidobacteriota

Acidobacteriota were long portrayed as slow-growing oligotrophs because they are more common in low-carbon soils, and many early isolates exhibited slow growth and low nutrient needs (Eichorst *et al.*, 2007b; Kielak *et al.*, 2016a). Genomic traits such as a few rRNA operons (Rawat *et al.*, 2012; Ward *et al.*, 2009) also suggest a resource-efficient lifestyle (Klappenbach *et al.*, 2000). However, with more cultured isolates and MAGs available, recent work highlights the phylum's metabolic and phylogenetic diversity, indicating condition-specific metabolic reprogramming (Eichorst *et al.*, 2018; Kielak *et al.*, 2016a). These studies showed that Acidobacteriota adapt their metabolism dynamically to environmental stimuli. Examples include *Acidobacterium capsulatum* switching to a respiro-fermentative mode under low O_2 (Trojan *et al.*, 2024), thereby producing acetate and ethanol, and a *Granulicella sp.* upregulating stress-response genes when carbon becomes abundant (Costa *et al.*, 2020). Beyond central metabolism, Acidobacteriota also adjust their cell envelopes to match environmental stress; they can modify their membranes to match habitat conditions by inserting iso-diabolic acid (isoDA) (Sinninghe Damsté *et al.*, 2011) and hopanoids. Iso-DA is a major membrane-spanning lipid found in Acidobacteriota SD1,3,4 (Damsté *et al.*, 2017; Sinninghe Damsté *et al.*, 2014), which is believed to stabilise membranes across various pH, oxygen, and temperature gradients. Hopanoids are bacterial equivalents of sterols (Sáenz *et al.*, 2015) that alter membrane rigidity and permeability during stress responses (Welander *et al.*, 2009). Adding to this picture, we are, to our knowledge, the first laboratory study investigating the metabolome of two strains in response to Trp, a ubiquitous plant exudate (Spaepen and Vanderleyden, 2011). We observe a significant activation of metabolism, resulting in 3-4 times more detected metabolomic features upon Trp supplementation, dominated by indole compounds, such as pityriacitrin and malassezindoles, as well as the phytohormone IAA. Interestingly, acidobacterial strains are increasingly investigated for plant-microbe interactions and potential beneficial effects for phytosanitary (Kalam *et al.*, 2020). Early experimental research showed that SD1 isolates can act as plant-growth-promoting bacteria in an *Arabidopsis thaliana* model system (Kielak *et al.*, 2016b). In addition, (Yoneda *et al.*, 2021) tested actinobacterial co-cultivation with duckweed, resulting in increased growth and chlorophyll content across six species. The observed increases in root and shoot biomass were contextualised with the strain's potential to produce IAA and to support iron mobilisation. Here, we extend this by quantifying the IAA production and, to the best of our knowledge),the first report of the cytokinin phytohormone iP as an acidobacterial product. Tryptophan supplementation caused an increase IAA titres and opposed decreases in iP production. It has been shown that rhizobacteria boost IAA output when exogenous Trp is available (Liu *et al.*, 2016; Noor *et al.*, 2023). Tryptophan can be abundant in protein-rich microhabitats, including termite nests, the source of our strains. Here, high Trp levels can arise via termite gut passage and excretion (Weihrauch and O'Donnell, 2021), proteolysis in faecal deposits, fungal associates (e.g., *Termitomyces*) (Schmidt *et al.*, 2021), and the accumulation of nitrogenous wastes (Sapountzis *et al.*, 2016), eventually creating local amino-acid hotspots (Nandika *et al.*, 2021). Consistent with this lignocellulose-rich substrate ecology (Vesala *et al.*, 2025) of our strains, previously reported IAA-producing Acidobacteriota had been isolated from decaying wood (Kielak *et al.*). Tryptophan is an energy-intensive amino acid, whose de novo synthesis via the shikimate pathway is among the most costly in bacteria (Akashi and Gojobori, 2002); preliminary investing in IAA and indole compound productions when exogenous Trp is available, respectively provided (by plants).

The presence of phytohormones in our acidobacterial extracts prompted us to explore their biosynthetic mechanisms and the potential presence of other known PGPT, relevant to plant-microbe interactions. In 2024, Gonçalves *et al.* conducted a comprehensive genomic analysis, revealing enrichment of PGP genes in several Acidobacteriota families, including *Acidobacteriaceae*, *Bryobacteraceae*, and *Koribacteraceae*. To contextualise our strains phenotypes at the genome scale, we performed PGPT mining, complementing Gonçalves' study by using a different dataset (only 9% overlap) and a broader bait gene set (6900 versus 91). This approach likewise highlights *Acidobacteriaceae* and the *Bryobacteraceae* and additionally includes the *Vicinamibacteraceae*, as most well-equipped families, while categorising, for example, the *Pyrinomonadaceae* as less promising in terms of PGPT equipment. We specifically searched for genes linked to plant growth promotion (characteristics such as phytases, ACC-deaminase, and nitrogen fixation, cf. 6.3.1) and confirmed previous literature reports of *nif* clusters (Reji and Zhang, 2022) in the *Holophagaceae* (Reji and Zhang, 2022) and *Acidobacteriaceae* groups (Gonçalves *et al.*, 2024; Kapili *et al.*, 2020). In brief, phytases liberate inorganic phosphate from phytate, increasing phosphorus bioavailability to roots; additional phosphatases and organic-acid secretion can further mobilise sparingly soluble phosphates (Dai *et al.*, 2020). Exopolysaccharide (EPS) promotes rhizosphere colonisation and biofilm formation (Naseem *et al.*, 2018), improves soil aggregation and water retention, and can buffer desiccation and ionic stresses, traits also associated with drought resilience (Kalam *et al.*, 2020). Phytohormones such as IAA remodel root architecture by stimulating lateral roots and root hairs (Etesami and Glick, 2024), thereby enlarging the absorptive surface; however, effects are dose-dependent and context-specific. Siderophores chelate iron(III) at low pH or under iron limitation, enhancing plant iron nutrition and by competitive sequestration, limiting pathogen access to iron (Gu *et al.*, 2020; Haas and Défago, 2005; Vansuyt *et al.*, 2007). Nitrogen fixation (e.g. *nif* gene clusters) provides reduced nitrogen (ammonium) to the plant–microbe system under Nitrogen-limiting conditions. 1-Aminocyclopropane-1-carboxylate ACC deaminase (*acdS*)(Glick, 2014), when present in plant-associated bacteria, lowers stress-ethylene levels by degrading the precursor 1-aminocyclopropane-1-carboxylate, sustaining root elongation under salinity, drought, or heavy-metal stress. To date, we have found no peer-reviewed evidence of the ACC deaminase gene (*acdS*) in Acidobacteriota (Singh *et al.*, 2015); one study's 16S-based functional inference suggested that Acidobacteriota might possess *acdS*, but gene-targeted sequencing only recovered Proteobacteria (Manter *et al.*, 2023), indicating that the Acidobacteriota signal was likely an inference artefact. Our genome mining also failed to detect any *acdS* gene, supporting this conclusion. In contrast, other PGPT have already been experimentally characterised in Acidobacteriota, including IAA production, iron acquisition, and EPS production (Kalam *et al.*, 2020). Genomically, Acidobacteriota are known to harbour genes for phosphorus solubilisation (Gonçalves *et al.*, 2024) and nitrogen fixation (Gonçalves *et al.*, 2024; Kalam *et al.*, 2020; Kapili *et al.*, 2020). Consistent with such reports, we documented the presence of *appA*, a phytase gene, but also the presence of *LOG* genes (Kuroha *et al.*, 2009) for cytokinin production in the Acidobacteriota species FHG110202, a feature not previously reported for this phylum (Frébortová and Frébort, 2021). *Nif* clusters were identified across different taxonomic lineages, corroborating previous reports. However, the genomic PGPTs present did not translate into observable growth on standard plates (testing for phytase) or in the plant experiments conducted. Functionally, despite PGPT evidence testing of strain extracts with quantitatively adjusted phytohormone levels in barley assays did not show significant growth promotion in our experimental set-up. Previously, Kielak *et al.* (2016) reported positive growth effects on *A. thaliana* from three SD1

Acidobacteriota strains. Deviating from this study we tested organic extracts instead of living cells. Thus, it might be plausible that a combination of molecules, absent from extracts (e.g. EPS), or additional enzyme-mediated activity requiring live cells, is needed to produce such growth effects in the laboratory. Thus, we also performed greenhouse experiments, incubating barley seeds with acidobacterial cultures before sowing (data not shown), however neither causing significantly increased plant growth, possibly attributed to higher complexity in this soil-based assay or the use of barley.

In the course of our comparative metabolic profiling studies, we also identified pityriacitrin and malassezindoles, exclusively produced in presence of Trp, and as causative agents provoking the suppression of fungal growth. This is consistent with previous reports describing their antifungal activities (Gaitanis *et al.*, 2019). Pityriacitrin was first identified from a marine *Paracoccus* (Nagao *et al.*, 1999) and subsequently reported in further bacterial and fungal genera (*Burkholderia* (Xu *et al.*, 2019)*, Bacillus* (Anh *et al.*, 2022), and the yeast *Malassezia furfur* (Mayser *et al.*, 2002) indicating a general occurrence across different domains and habitats. The malassezindoles are produced by *M. furfur*, along with very similar indole compounds (Magiatis *et al.*, 2013) that are also produced by *Ustilago maydis* (Zuther *et al.*, 2008). They are known to inhibit tyrosinases (Zolghadri *et al.*, 2019), thereby potentially affecting the depigmentation processes (Gaitanis *et al.*, 2012). Their biosynthesis in *U. maydis* relies on a one-enzyme conversion involving *tam1* for the conversion of Trp to indole pyruvate, from which indole pigments form non-enzymatically. Most likely, this enzyme is also responsible for the production of malassezindole in *M. furfur*. By analogy, a similar route could operate in our strains, although we found no *tam1* homologues in our *Acidobacteriaceae*. Since *M. furfur* is part of our skin microbiome, the production of simple indolic compounds with antifungal activity for competitive purposes is a plausible, though hypothetical, explanation. Because some Trp-linked scaffolds, as described above, can form via short conversion cascades, convergent evolution across bacteria and fungi inhabiting similar environments and exposed to similar molecules is likely. Consequently, they may represent parallel solutions to challenges such as UV/oxidative stress and microbial competition, specific to their respective niches (e.g skin, phyllosphere, and surface soils). While the link between Trp and increased IAA production, appears functionally conclusive, the ecological drivers behind the activation of other Trp-upregulated features, such as pityriacitrins and malassezindoles, remain unclear. Speculatively, these compounds could represent a plant-exudate funded defence against pathogenic fungi. Notably, most microbial NPs, suppressing fungal growth are complex, BGCs-encoded metabolites such as cyclic lipopeptides that are also used commercially and bio fungicides.

Thus, we complemented our Acidobacteriota profiling study by a specific examination of their biosynthetic potential for the production of such structurally complex NPs. The most advanced insights into the predicted NP reservoir of the Acidobacteriota come from metagenomics. Early work by Parsley *et al.*, (2011) discovered NRPS/PKS genes of acidobacterial origin in a soil metagenomic library, providing the first gene-level evidence for secondary metabolism in this phylum. Building on that, Crits-Christoph *et al.* (2018) and Waschulin *et al.* (2022) reconstructed global soil MAGs (Crits-Christoph *et al.*, 2018b; Waschulin *et al.*, 2022) and identified complex BGC-rich genomes that showed metatranscriptomic activity, indicating in situ expression of these clusters. Despite these advances, the first NRPS/PKS-derived NPs were isolated only in 2023 from marine *Acanthopleuribacteraceae* (Leopold-Messer *et al.*, 2023) strains, finally confirming that members of this

phylum indeed produce complex NPs. The ecological roles of these, however, remained unsolved. Consequently, the NP research community has begun to carefully investigate these groups, with particular focus on marine Acidobacteriota (personal communication). Our study shows that BGC number did not correlate with the assembly size, a phenomenon typically observed across different taxonomic levels of well-established NP producing bacteria (Adamek *et al.*, 2018; Brinkmann *et al.*, 2022a; Männle *et al.*, 2020). Furthermore, the observed BGCs accumulation in only a few genomes and in specific taxonomic lineages such as the *Acanthopleuribacteriaceae* (Leopold-Messer *et al.*, 2023) and the genus *Candidatus Angelobacter* (Crits-Christoph *et al.*, 2022) hints towards a largely uneven distribution of the NP-related biosynthetic potential within the Acidobacteriota phylum. This shows the need to take a carefully executed and data-driven strain selection process in order to avoid miss-allocation of resources in the lab. Our study provided phylum-level context (618 genomes) and recovered over 3,834 BGCs, dominated by RiPPs, NRPS, and terpenes. They revealed multiple MiBIG-proximal GCFs, including myxochromide-like, anabaenopeptin-like, ectoine-like, and a PKS–NRPS hybrid that co-clusters with the occidiofungin A reference BGC (BGC 0001711) (Gu *et al.*, 2011), present in *A. capsulatum* and our *Edaphobacter* isolate FHG110511. Initially isolated from *Burkholderia contaminans*, occidiofungin A exhibits broad antifungal activity (Lu *et al.*, 2009) by disrupting higher-order actin functions (Hansanant and Smith, 2022). Indeed, this first-in-class and resistance-breaking compound is currently in early clinical development (gel OCF001, phase1) for use against multidrug-resistant vaginal yeast infections (Cothrell *et al.*, 2023). Screening of culture extracts from *Edaphobacter sp.* FHG110511 however did not reveal any antifungal activity yet. While limitations inherent in cultivation-dependent approaches may explain this (see technical limitations), this might also be due to the silence of the BGC remains under the chosen growth condition. Current efforts aim to test this hypothesis experimentally.

The close architectural similarity of some detected acidobacterial BGCs across phyla or even domains boarders suggest HGT events. The BGCs are often embedded in genomic islands flanked by mobile elements (Dobrindt *et al.*, 2004; Penn *et al.*, 2009) that facilitate horizontal gene transfer. In soil and rhizosphere environments—rich in DNA (Ku *et al.*, 2021) and phage (Pratama and van Elsas, 2018) activity—HGT offers a quick route to acquire competitive traits (e.g. antifungals, siderophores, and pigments) (Dobrindt *et al.*, 2004). Studying the genomic characteristics of these BGCs might provide insights into whether the BGC originated from HGT events. Distinguishing HGT from convergence is operationally important. While convergent developments arise under shared ecological pressures, they are biologically functional and adapted for different hosts. The HGT-derived BGCs often come with a regulatory mismatch, explaining why some clusters remain silent in new hosts. This could also clarify why the occidiofungin-like BGC, which matches the well-studied antifungal compound from Proteobacteria, might stay silent in their new genomic context.

Therefore, HGT-acquired BGC could be a candidate for heterologous expression, most likely involving promoter refitting, regulator swaps, and chassis modifications. Where *Burkholderia*-like NRPS/PKS logic is suspected, such as for occidiofungin, a genetically tractable *Burkholderia* chassis could be used to improve the chances of successful expression. Commonly, *B. thailandensis* E264 (mini-Tn7/BAC integration, broad-host-range plasmids) can serve as a compatible heterologous host, although *P. putida* remains a widely used alternative.

3.2.3 HEL, A Multi-Host Synthetic Biology Platform for BGC Activation

The occidiofungin-like target demonstrates, as many have before, that heterologous expression can be crucial for unknown BGC expression or optimisation strategies. When cultivation-dependent expression or discovery fails, our HEL platform offers a route to (i) retrofit the BGC, (ii) systematically test chassis, and (iii) bypass native regulation through promoter exchange or regulatory rewiring. Previously, several complementary approaches have been shown to be effective for BGC activation.

Yamanaka *et al.* (2014) captured the silent taromycin BGC from the marine actinomycete *Saccharomonospora sp.* CBQ-490, utilising TAR and after removing regulators, successfully transferred it into *Streptomyces coelicolor*, resulting in the expression of the large lipopeptide taromycin A. In fungi, the two-PKS BGC asperfuranone was successfully activated in the original host *Aspergillus nidulans* (Chiang *et al.*, 2009) by exchanging the native promoter with strong, regulated promoters. Promoter swaps are commonly used for BGC activation, and more recently, CRISPR-Cas is used for promoter knock-ins and has successfully been used for activation of five cryptic BGCs in *Streptomyces spp*. (M. M. Zhang *et al.*, 2017). Beyond promoter exchange, transcription-level activation was employed by Wang *et al.* (2019) who used vectors that titrate inhibiting transcription factors to activate otherwise silenced BGCs. Next, once expression is activated, metabolic context often becomes the challenge. Different studies have shown that overexpression of precursor supply or supplementation of required precursor optimises or even activates silent BGCs (Gummerlich *et al.*, 2021). Collectively, these expression factors and others including, promoters/sigma-factors, GC/codon usage and transcriptional burden, cofactor/precursor pools (e.g., malonyl-/methylmalonyl-CoA, unusual amino acids), post-translational machinery (PPTases), efflux/self-resistance; and specialized tailoring enzymes or redox partners are intertwined and influence successful expression. As chassis differ in such starting conditions and expression machinery, BGC output varies from host to host.

This diversifying effect has been demonstrated for a metagenomic soil library, revealing a minimal overlap of one NP that was expressed in a broad set of heterologous hosts (six proteobacteria: *Agrobacterium*, *Burkholderia*, *Caulobacter*, *E. coli*, *P. putida*, *Cupriavidus*) (Craig *et al.*, 2009). Commonly, these aspects have motivated researchers to design so-called shuttle vectors (J. J. Zhang *et al.*, 2017) to transfer BGC between selected hosts (e.g. (Troeschel *et al.*, 2012). Other approaches target a broader host spectrum, where Cre/loxP or mariner Transposons (G. Wang *et al.*, 2019) are used for the integration of BGCs in a wide range of γ-Proteobacteria, or demonstrate that even within the *Streptomyces* genus, the expression and production of BGCs vary strongly (Iqbal *et al.*, 2016). Consequently, methods aim to increase the breadth of delivery systems to reach different hosts, a challenge given the often large, multimodular BGCs.

The transfer of such BGCs requires complex cloning techniques that typically rely on recombination methods also employed in natural systems, such as yeast homologous recombination systems, CRISPR-Cas, or phage/transposon-mediated integration. Owing to their size, BGCs demand homologous-based (Fu *et al.*, 2008), restriction-enzyme independent approaches, whereas classical synthetic biology broadly pursues standardised, modular systems that frequently depend on restriction enzymes. This conflict is reflected in the functionality of synthetic biology toolboxes that have been developed for BGC expression; TREX (Loeschcke *et al.*, 2013) and yTREX (Domröse *et al.*, 2017) utilised yeast recombinational cloning, combined with Tn5-based transposition, to capture BGCs and randomly insert them into different heterologous hosts. This strategy has

been successfully used to activate the phenazine and prodigiosin pathways in *P. putida*. Another study employs synthetic biology to refactor the streptophenazine BGC by engineering and inserting promoter/RBS cassettes, thereby decoupling the BGC from its native regulation and allowing heterologous expression with expanded chemical diversity. However, both approaches require multiple cloning steps and modifications to be transferred to other BGCs. Addressing reusability, a recent study adapted a modular toolkit based on the MoClo standard for *Streptomyces* strains, a crucial step toward applying synthetic biology principles in the most widely used NP producer group and heterologous chassis.

HEL attempts to couple homology-based BGC cloning with the strengths of modular, standardised, restriction-enzyme-based expression optimisation approaches. It is designed for modular retrofitting of BGCs with expression and chassis-specific elements, decoupling activation/modification and host transfer from BGC capture and cloning. To achieve this, expression libraries and host-specific cassettes are assembled using MoClo, ensuring reusability and modularity. The coupling with the BGC-carrying vector is performed independently, sing the λ-phage Gateway in vitro Gateway reaction. This allows individual parts or functional cassettes to be clonally amplified and reused to express other BGCs. Another advantage is the opportunity to assemble expression libraries, creating a diverse set of different expression factor combinations. This increases the likelihood of identifying an effective configuration for BGC activation. Such cloning methods are typically refined over consecutive rounds until processes are optimised. The section exemplifies such iteration, as seen in the evaluation of the required size and origin of the building blocks (also addressed in chapter 2.9). Ultimately, the strategic benefit of these approaches is maximised in projects with poor genetic tractability and/or demanding cultivation conditions; This is particularly true for underexplored taxa such as the Acidobacteriota. Building on this framework, HEL will be used and further developed in a funded DFG project (545372354) aimed at expressing unknown RiPPs that are modified by DarE-like radical SAM enzymes, the enzyme also responsible for darobactin maturation.

3.3 Limitations and Challenges

While the established genomic workflow successfully guided experimental work for this thesis (Section "Genome Mining of Bacteroidota") and beyond (β-lactam PhD project) it remains semi-automatic. Currently, it relies on different environments (e.g Docker and Conda) to run discrete tasks but lacks end-to-end execution, from raw reads to genome mining reports. Currently, efforts are focused on developing a fully automated pipeline with a graphical interface for untrained users. A complementary interface displaying high-level strain information as a strain portfolio card, for our strain collection, should also make it more attractive to industry partners. Achieving this requires not only an automated workflow but also digitising existing analogue data scattered across various formats, along with a significant sequencing effort to access the blueprint of potential NPs for individual strains. Incorporating machine learning and/or AI tools could provide further insights, for example, by identifying BGCs that do not follow traditional patterns. While this information is invaluable for research and industry, any in silico prioritisation ultimately requires laboratory validation, which often becomes the main bottleneck.

For the *Pedobacter* study, only one BGC-metabolite linkage was fully resolved, and many predicted clusters remain uncharacterised. We also failed to identify the biological function of the cryopeptins despite extensive

screening efforts. The same applies to our amphibian *Pedobacter* isolates, which were not examined for their functional roles within the salamanders' cutaneous microbiome. Future work could combine cocultivation and combined transcriptomic analysis to provide insights into expressed genes (e.g., chitinase) and potential interactions. The predicted lanthipeptide, situated near a TonB-dependent receptor, also presents an intriguing target for functional investigation, suggesting a possible Trojan-horse strategy in antimicrobial activity.

While Bacteroidota are easily cultivated and genetically tractable, extensive effort has been dedicated to cultivating and characterising Acidobacteriota for this project. Experiments were conducted and optimised for pH (with MES buffer supplementation), with inoculation volumes increased to support the continuous growth of main cultures (10% v/v). The shaking speed was adjusted to 75 rpm, and strict quality control measures were implemented to ensure pure cultures. Despite these efforts, the biomass of Acidobacteriota remained relatively low. This was evidenced not only by visual inspection of the culture broth but also by the low peak height in the base peak chromatograms and the modest number of features detected of the extracts in the untargeted metabolomics approach. While contamination could be easily monitored and assessed, maintaining consistent inoculum across replicates proved difficult because cultures formed stable aggregates that resisted dispersion by shaking, medium additives, or physical disruption. Consequently, optical density could not reliably estimate viable cell mass. In addition, attempts were made to genetically modify *A. borealis* DSM 26883 to disrupt a carotenoid likely responsible for the strain's pink colouration. However, the transformation failed because we were unable to concentrate the cell mass by centrifugation or filtration. Consequently, alternative methods should be employed to achieve higher cell masses for improved genetic accessibility or higher concentrations of metabolites, facilitating subsequent screening or isolation purposes.

These growth limitations of Acidobacteria likely contributed to the difficulty of translating in vitro metabolite detection or in silico prediction into functional or ecological outcomes. While generated extracts cannot fully reflect the complex interactions in living systems, testing isolated strains poses its own challenges for slow-growing, oligotrophic organisms, as seen in the greenhouse experiments performed with live cells. Although we did not test for acidobacterial persistence in soil after the cultivation period was over (e.g., by FISH experiments or amplicon sequencing), it is possible that cells were outcompeted by other microorganisms remaining in the soil. Notably, studies reporting a positive effect on plant growth used medium-based, otherwise axenic setups, where microbial competition was excluded. While growth promotion was clearly detectable in these studies, the underlying reasons for the increase in biomass remain hypothetical. Production of phytohormones, EPS, and iron acquisition could be beneficial, but does not definitively exclude the influence of other small molecules or enzymatic reactions. Novel setups based on small chips could enable the physical separation of plants and bacteria, allowing the growth of difficult-to-cultivate isolates while investigating the effects of secreted small molecules on the plants. Furthermore, it would be valuable to examine the expression patterns of PGPT during co-cultivation to determine whether relevant functions are induced. Perhaps more complex environmental conditions and microbial interactions are necessary for the successful expression and promotion of plant growth.

As noted, several genome mining leads remain to be validated experimentally. An intriguing target, given its potential antifungal activity and its *Burkholderia* origin, is the occidiofungin-like BGC in two acidobacterial strains. In this context, more targeted cultivation-based strategies could be employed before progressing to heterologous expression methods. Heterologous expression comes with its own set of challenges. First of all (i), the occidiofungin-like BGC is relatively large, encompassing over 70kb. Second (ii), the transfer and choice of the heterologous host is challenging since a BGC of acidobacterial origin has not been heterologously expressed before.

(i) The pCC2Fosmid system used to create the HEL-fosmid typically accommodates DNA fragments of 40kb, due to the λ-phage "headful packaging" limit. The vector itself is based on the F-factor replicon of *E. coli*, which has a size of ~100 kb (Koraimann, 2018). The same segregation/replication system is also used by BAC vectors, which carry even larger DNA fragments stably up to about 300 kb (Blaas *et al.*, 2009). From this, we conclude that the fosmid system does not inherently dictate the upper limit and that the challenges of cloning the occidofungin-like cluster will likely be similar to those for other vectors, such as BACs, which require sophisticated cloning techniques and most likely multiple rounds of refactoring.

(ii) Although the ~70kb cluster is challenging to clone, insights gained from the successful expression of the original occidiofungin cluster might assist in activating this cluster, potentially even within a *Burkholderia* heterologous host. While this cluster is a promising candidate for HEL establishment, the system still requires optimisation, particularly regarding its transfer into different heterologous hosts. Despite a successful transfer into *P. putida* mt-KT2440 and *B. subtilis* SCK6, we were unable to detect any of the expected heterologous products (e.g. GFP and darobactin). For darobactin, robust heterologous expression outside *E. coli* has not been demonstrated. Other efforts in our laboratory to express the darobactin BGC in other hosts also failed, consistent with the efforts taken using HEL. However, the *gfp* genes used for the *Pseudomonas* and *Bacillus* equipped vectors were specifically selected, due to their proven functionality in former studies. Since colony PCR has confirmed the presence of plasmid, and various cultivation approaches have ensured sufficient oxygen levels for GFP maturation, the remaining issues likely include plasmid stability, copy number, and expression efficiency. Expanding the HEL-library with more established BBs will be necessary to improve expression (variety) in *E. coli* and other hosts.

4 CONCLUSION & OUTLOOK

This dissertation aimed to develop and validate a genome-guided discovery workflow for the systematic exploration of biosynthetic potential in underexplored bacterial phyla. Additionally, it established a synthetic biology toolkit for modular activation and transfer of biosynthetic gene clusters (BGCs). By combining genomic, metabolomic, and molecular methodologies, this thesis addressed the challenge of connecting genomic potential to detectable chemical and functional phenotypes in Bacteroidota and Acidobacteriota, which are promising yet largely untapped sources of natural products.

To address the first research question, "*From Genomes to Targets*", a standardised workflow for assembled genomes was developed. This pipeline integrates quality control, taxonomic classification, BGC prediction, and similarity clustering into a reproducible workflow that supports strain description and target-specific prioritisation within large strain collections. The application of this workflow to Bacteroidota isolates resulted in the description of multiple novel isolates, including *Pedobacter* strains from amphibian specimens, thereby expanding the genomic reference diversity of the phylum. Comparative genome mining also identified a biosynthetically rich *Pedobacter cryoconitis* branch within publicly available genomes, which was experimentally validated as a hotspot for NPs. Metabolomics-guided isolation and genetic manipulation linked a previously uncharacterized NRPS cluster to the novel lipopeptide family cryopeptins. This outcome demonstrates the workflow's capacity to connect computational predictions with specific molecular products.

The second and third research questions, "*From Culture Isolates to Natural Products" and "From Omic Traits to Ecological Function",* were addressed by investigating the biosynthetic and ecological potential of Acidobacteriota. Novel strains were selectively isolated from termite nests and soil samples using high-throughput cultivation. These isolates were profiled for antimicrobial activity and metabolomic novelty using an OSMAC cultivation approach. Metabolic mining identified two phytohormones, IAA and iP, which were quantified and shown to be regulated by tryptophan. Global metabolome profiling under tryptophan supplementation revealed a significant metabolic shift, including the production of antifungal metabolites. A comprehensive genomic analysis of over 600 previously curated Acidobacteriota genomes mapped the distribution of PGPTs across the phylum. The effect of phytohormone production on barley seedlings was tested, but it did not result in measurable biomass increases under the tested conditions. This finding suggests that PGP of Acidobacteriota may depend on specific strain or environmental interactions. Overall, these results refine the understanding of Acidobacteriota as a metabolically responsive and ecologically relevant lineage with latent biotechnological potential.

The fourth research question, "*From DNA of Interest to Natural Product*" was addressed by designing, creating, and evaluating the HEL platform. This modular cloning system enables the retrofitting of BGC-containing vectors with exchangeable expression libraries for expression variation and host-specific compatibility cassettes for transfer between hosts. The HEL platform bridges the gap between genome mining, heterologous expression, and synthetic biology by enabling rational design and comparative testing of BGCs across multiple

genetic backgrounds. It thereby lays the groundwork for scalable heterologous expression and synthetic refactoring of BGCs from otherwise genetically intractable taxa. In line with the synthetic biology's build–test–learn cycle approach, this workflow can in reverse help identify the regulatory, metabolic, and host factors that support expression of a given BGC, providing functional insights into regulation.

Collectively, the outcomes of this thesis demonstrate that a genome-guided, cultivation-dependent approach can transform the study of underexplored microbial diversity into a systematic discovery pipeline. Key results include a reproducible bioinformatics workflow for strain prioritisation and biosynthetic gene cluster (BGC) analysis, genomic and experimental characterisation of novel Bacteroidota and Acidobacteriota isolates, discovery of cryopeptin lipopeptides as new metabolites linked to specific BGCs, identification and quantification of phytohormones and antifungal-associated metabolites in Acidobacteriota, a global genomic map of plant growth–promoting traits, and the establishment of a modular expression platform for synthetic biology applications.

While these outcomes mark significant progress, continued optimisation and validation of the workflows remain essential. Having that in mind, several directions emerge:

- ◊ While systematic bioprospecting campaigns can continue to yield biosynthetic-rich taxa, they more importantly require an automatic genomic prioritisation workflow. The established platform must then be made available for unskilled users, so that the generated data can be interpreted and tested experimentally.
- ◊ Implementation of metabologenomics tools will allow us to prioritise clusters for investigation of BGC-metabolite linkages. The implementation of AI tools can guide experimental workflows, identifying clusters that putatively produce natural products with a mode of action of interest to researchers. Collective in silico predictions can help guide in situ or co-cultivation studies for ecological validation of BGCs and NPs of interest.
- ◊ Novel (co-)cultivation techniques will allow the study of members of the Acidobacteriota for functional analysis of their ecological potential.
- ◊ Further optimisation of HEL, especially regarding the transfer into heterologous hosts or the establishment of similar toolkits, will be critical to keep up with experimental validation of in silico predicted (cryptic) BGCs.

5 REFERENCES

Adamek, M., Alanjary, M., Sales-Ortells, H., Goodfellow, M., Bull, A.T., Winkler, A., Wibberg, D., Kalinowski, J., Ziemert, N., 2018. Comparative genomics reveals phylogenetic distribution patterns of secondary metabolites in Amycolatopsis species. BMC Genomics 19, 426. https://doi.org/10.1186/s12864-018-4809-4

Agilent Technologies, Inc., 2015. SuperCos 1 Cosmid Vector Kit: Instruction Manual. Agilent Technologies, Inc.

Ahmed, M.N., Reyna-González, E., Schmid, B., Wiebach, V., Süssmuth, R.D., Dittmann, E., Fewer, D.P., 2017. Phylogenomic Analysis of the Microviridin Biosynthetic Pathway Coupled with Targeted Chemo-Enzymatic Synthesis Yields Potent Protease Inhibitors. ACS Chemical Biology 12, 1538–1546. https://doi.org/10.1021/acschembio.7b00124

Ahmed, Y., Rebets, Y., Tokovenko, B., Brötz, E., Luzhetskyy, A., 2017. Identification of butenolide regulatory system controlling secondary metabolism in Streptomyces albus J1074. Sci Rep 7, 9784. https://doi.org/10.1038/s41598-017-10316-y

Akashi, H., Gojobori, T., 2002. Metabolic efficiency and amino acid composition in the proteomes of Escherichia coli and Bacillus subtilis. Proceedings of the National Academy of Sciences of the United States of America 99, 3695–3700. https://doi.org/10.1073/pnas.062526999

Akhtar, S.S., Mekureyaw, M.F., Pandey, C., Roitsch, T., 2020. Role of Cytokinins for Interactions of Plants With Microbial Pathogens and Pest Insects. Front. Plant Sci. 10. https://doi.org/10.3389/fpls.2019.01777

Alanjary, M., Steinke, K., Ziemert, N., 2019. AutoMLST: an automated web server for generating multi-locus species trees highlighting natural product potential. Nucleic Acids Res 47, W276–W282. https://doi.org/10.1093/nar/gkz282

Alcock, B.P., Raphenya, A.R., Lau, T.T.Y., Tsang, K.K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A.-L.V., Cheng, A.A., Liu, S., Min, S.Y., Miroshnichenko, A., Tran, H.-K., Werfalli, R.E., Nasir, J.A., Oloni, M., Speicher, D.J., Florescu, A., Singh, B., Faltyn, M., Hernandez-Koutoucheva, A., Sharma, A.N., Bordeleau, E., Pawlowski, A.C., Zubyk, H.L., Dooley, D., Griffiths, E., Maguire, F., Winsor, G.L., Beiko, R.G., Brinkman, F.S.L., Hsiao, W.W.L., Domselaar, G.V., McArthur, A.G., 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 48, D517–D525. https://doi.org/10.1093/nar/gkz935

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of Molecular Biology 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Ameruoso, A., Villegas Kcam, M.C., Cohen, K.P., Chappell, J., 2022. Activating natural product synthesis using CRISPR interference and activation systems in Streptomyces. Nucleic Acids Res 50, 7751–7760. https://doi.org/10.1093/nar/gkac556

Anagnostopoulos, C., Spizizen, J., 1961. Requirements for transformation in Bacillus subtilis. Journal of Bacteriology 81, 741–746.

Anh, C.V., Kang, J.S., Lee, H.-S., Trinh, P.T.H., Heo, C.-S., Shin, H.J., 2022. New Glycosylated Secondary Metabolites from Marine-Derived Bacteria. Marine Drugs 20, 464. https://doi.org/10.3390/md20070464

Aoi, Y., Kinoshita, T., Hata, T., Ohta, H., Obokata, H., Tsuneda, S., 2009. Hollow-Fiber Membrane Chamber as a Device for In Situ Environmental Cultivation. Applied and Environmental Microbiology 75, 3826–3833. https://doi.org/10.1128/AEM.02542-08

Ara, T., Kodama, Y., Tokimatsu, T., Fukuda, A., Kosuge, T., Mashima, J., Tanizawa, Y., Tanjo, T., Ogasawara, O., Fujisawa, T., Nakamura, Y., Arita, M., 2024. DDBJ update in 2023: the MetaboBank for metabolomics data and associated metadata. Nucleic Acids Res 52, D67–D71. https://doi.org/10.1093/nar/gkad1046

Arias, P., Fernández-Moreno, M.A., Malpartida, F., 1999. Characterization of the Pathway-Specific Positive Transcriptional Regulator for Actinorhodin Biosynthesis inStreptomyces coelicolor A3(2) as a DNA-

Binding Protein. Journal of Bacteriology 181, 6958–6968. https://doi.org/10.1128/jb.181.22.6958-6968.1999

Arkin, A.P., Cottingham, R.W., Henry, C.S., Harris, N.L., Stevens, R.L., Maslov, S., Dehal, P., Ware, D., Perez, F., Canon, S., Sneddon, M.W., Henderson, M.L., Riehl, W.J., Murphy-Olson, D., Chan, S.Y., Kamimura, R.T., Kumari, S., Drake, M.M., Brettin, T.S., Glass, E.M., Chivian, D., Gunter, D., Weston, D.J., Allen, B.H., Baumohl, J., Best, A.A., Bowen, B., Brenner, S.E., Bun, C.C., Chandonia, J.-M., Chia, J.-M., Colasanti, R., Conrad, N., Davis, J.J., Davison, B.H., DeJongh, M., Devoid, S., Dietrich, E., Dubchak, I., Edirisinghe, J.N., Fang, G., Faria, J.P., Frybarger, P.M., Gerlach, W., Gerstein, M., Greiner, A., Gurtowski, J., Haun, H.L., He, F., Jain, R., Joachimiak, M.P., Keegan, K.P., Kondo, S., Kumar, V., Land, M.L., Meyer, F., Mills, M., Novichkov, P.S., Oh, T., Olsen, G.J., Olson, R., Parrello, B., Pasternak, S., Pearson, E., Poon, S.S., Price, G.A., Ramakrishnan, S., Ranjan, P., Ronald, P.C., Schatz, M.C., Seaver, S.M.D., Shukla, M., Sutormin, R.A., Syed, M.H., Thomason, J., Tintle, N.L., Wang, D., Xia, F., Yoo, H., Yoo, S., Yu, D., 2018. KBase: The United States Department of Energy Systems Biology Knowledgebase. Nat Biotechnol 36, 566–569. https://doi.org/10.1038/nbt.4163

Baltz, R.H., 2008. Renaissance in antibacterial discovery from actinomycetes. Current Opinion in Pharmacology, Anti-infectives/New technologies 8, 557–563. https://doi.org/10.1016/j.coph.2008.04.008

Barka, E.A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H.-P., Clément, C., Ouhdouch, Y., van Wezel, G.P., 2015. Taxonomy, Physiology, and Natural Products of Actinobacteria. Microbiology and Molecular Biology Reviews 80, 1–43. https://doi.org/10.1128/mmbr.00019-15

Barona-Gómez, F., Wong, U., Giannakopulos, A.E., Derrick, P.J., Challis, G.L., 2004. Identification of a Cluster of Genes that Directs Desferrioxamine Biosynthesis in Streptomyces coelicolor M145. J. Am. Chem. Soc. 126, 16282–16283. https://doi.org/10.1021/ja045774k

Bauman, K.D., Li, J., Murata, K., Mantovani, S.M., Dahesh, S., Nizet, V., Luhavaya, H., Moore, B.S., 2019. Refactoring the Cryptic Streptophenazine Biosynthetic Gene Cluster Unites Phenazine, Polyketide, and Nonribosomal Peptide Biochemistry. Cell Chem Biol 26, 724-736.e7. https://doi.org/10.1016/j.chembiol.2019.02.004

Beckmann, A., Hüttel, S., Schmitt, V., Müller, R., Stadler, M., 2017. Optimization of the biotechnological production of a novel class of anti-MRSA antibiotics from Chitinophaga sancti. Microbial Cell Factories 16, 143. https://doi.org/10.1186/s12934-017-0756-z

Belova, S.E., Ravin, N.V., Pankratov, T.A., Rakitin, A.L., Ivanova, A.A., Beletsky, A.V., Mardanov, A.V., Sinninghe Damsté, J.S., Dedysh, S.N., 2018. Hydrolytic Capabilities as a Key to Environmental Success: Chitinolytic and Cellulolytic Acidobacteria From Acidic Sub-arctic Soils and Boreal Peatlands. Front. Microbiol. 9. https://doi.org/10.3389/fmicb.2018.02775

Berdy, B., Spoering, A.L., Ling, L.L., Epstein, S.S., 2017. In situ cultivation of previously uncultivable microorganisms using the ichip. Nat Protoc 12, 2232–2242. https://doi.org/10.1038/nprot.2017.074

Bjerketorp, J., Levenfors, J.J., Nord, C., Guss, B., Öberg, B., Broberg, A., 2021. Selective Isolation of Multidrug-Resistant Pedobacter spp., Producers of Novel Antibacterial Peptides. Front. Microbiol. 12. https://doi.org/10.3389/fmicb.2021.642829

Blaas, L., Musteanu, M., Eferl, R., Bauer, A., Casanova, E., 2009. Bacterial artificial chromosomes improve recombinant protein production in mammalian cells. BMC Biotechnology 9, 3–3. https://doi.org/10.1186/1472-6750-9-3

Bletz, M.C., Myers, J., Woodhams, D.C., Rabemananjara, F.C.E., Rakotonirina, A., Weldon, C., Edmonds, D., Vences, M., Harris, R.N., 2017. Estimating Herd Immunity to Amphibian Chytridiomycosis in Madagascar Based on the Defensive Function of Amphibian Skin Bacteria. Frontiers in Microbiology 8.

Blin, K., Pascal Andreu, V., de los Santos, E.L.C., Del Carratore, F., Lee, S.Y., Medema, M.H., Weber, T., 2019. The antiSMASH database version 2: a comprehensive resource on secondary metabolite biosynthetic gene clusters. Nucleic Acids Res 47, D625–D630. https://doi.org/10.1093/nar/gky1060

Blin, K., Shaw, S., Kloosterman, A.M., Charlop-Powers, Z., van Wezel, G.P., Medema, M.H., Weber, T., 2021a. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res 49, W29–W35. https://doi.org/10.1093/nar/gkab335

Blin, K., Shaw, S., Kloosterman, A.M., Charlop-Powers, Z., van Wezel, G.P., Medema, M.H., Weber, T., 2021b. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res 49, W29–W35. https://doi.org/10.1093/nar/gkab335

Blin, K., Shaw, S., Vader, L., Szenei, J., Reitz, Z.L., Augustijn, H.E., Cediel-Becerra, J.D.D., de Crécy-Lagard, V., Koetsier, R.A., Williams, S.E., Cruz-Morales, P., Wongwas, S., Segurado Luchsinger, A.E., Biermann, F., Korenskaia, A., Zdouc, M.M., Meijer, D., Terlouw, B.R., van der Hooft, J.J.J., Ziemert, N., Helfrich, E.J.N., Masschelein, J., Corre, C., Chevrette, M.G., van Wezel, G.P., Medema, M.H., Weber, T., 2025. antiSMASH 8.0: extended gene cluster detection capabilities and analyses of chemistry, enzymology, and regulation. Nucleic Acids Res 53, W32–W38. https://doi.org/10.1093/nar/gkaf334

Bloudoff, K., Schmeing, T.M., 2017. Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, Biophysics in Canada 1865, 1587–1604. https://doi.org/10.1016/j.bbapap.2017.05.010

Bode, H.B., Bethe, B., Höfs, R., Zeeck, A., 2002. Big effects from small changes: possible ways to explore nature's chemical diversity. Chembiochem 3, 619–627. https://doi.org/10.1002/1439-7633(20020703)3:7<619::AID-CBIC619>3.0.CO;2-9

Bok, J.W., Keller, N.P., 2004. LaeA, a Regulator of Secondary Metabolism in Aspergillus spp. Eukaryotic Cell 3, 527–535. https://doi.org/10.1128/ec.3.2.527-535.2004

Bollmann, A., Lewis, K., Epstein, S.S., 2007. Incubation of Environmental Samples in a Diffusion Chamber Increases the Diversity of Recovered Isolates. Applied and Environmental Microbiology 73, 6386–6390. https://doi.org/10.1128/AEM.01309-07

Bonhomme, S., Contreras-Martel, C., Dessen, A., Macheboeuf, P., 2023. Architecture of a PKS-NRPS hybrid megaenzyme involved in the biosynthesis of the genotoxin colibactin. Structure 31, 700-712.e4. https://doi.org/10.1016/j.str.2023.03.012

Boundy-Mills, K., McCluskey, K., Elia, P., Glaeser, J.A., Lindner, D.L., Nobles, D.R., Normanly, J., Ochoa-Corona, F.M., Scott, J.A., Ward, T.J., Webb, K.M., Webster, K., Wertz, J.E., 2020. Preserving US microbe collections sparks future discoveries. J Appl Microbiol 129, 162–174. https://doi.org/10.1111/jam.14525

Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., Schulz, F., Jarett, J., Rivers, A.R., Eloe-Fadrosh, E.A., Tringe, S.G., Ivanova, N.N., Copeland, A., Clum, A., Becraft, E.D., Malmstrom, R.R., Birren, B., Podar, M., Bork, P., Weinstock, G.M., Garrity, G.M., Dodsworth, J.A., Yooseph, S., Sutton, G., Glöckner, F.O., Gilbert, J.A., Nelson, W.C., Hallam, S.J., Jungbluth, S.P., Ettema, T.J.G., Tighe, S., Konstantinidis, K.T., Liu, W.-T., Baker, B.J., Rattei, T., Eisen, J.A., Hedlund, B., McMahon, K.D., Fierer, N., Knight, R., Finn, R., Cochrane, G., Karsch-Mizrachi, I., Tyson, G.W., Rinke, C., Lapidus, A., Meyer, F., Yilmaz, P., Parks, D.H., Murat Eren, A., Schriml, L., Banfield, J.F., Hugenholtz, P., Woyke, T., 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35, 725–731. https://doi.org/10.1038/nbt.3893

Brinkmann, S., Kurz, M., Patras, M.A., Hartwig, C., Marner, M., Leis, B., Billion, A., Kleiner, Y., Bauer, A., Toti, L., Pöverlein, C., Hammann, P.E., Vilcinskas, A., Glaeser, J., Spohn, M., Schäberle, T.F., 2022a. Genomic and Chemical Decryption of the Bacteroidetes Phylum for Its Potential to Biosynthesize Natural Products. Microbiology Spectrum 10, e02479-21. https://doi.org/10.1128/spectrum.02479-21

Brinkmann, S., Kurz, M., Patras, M.A., Hartwig, C., Marner, M., Leis, B., Billion, A., Kleiner, Y., Bauer, A., Toti, L., Pöverlein, C., Hammann, P.E., Vilcinskas, A., Glaeser, J., Spohn, M., Schäberle, T.F., 2022b. Genomic and Chemical Decryption of the Bacteroidetes Phylum for Its Potential to Biosynthesize Natural Products. Microbiol Spectr 10, e02479-21. https://doi.org/10.1128/spectrum.02479-21

Brinkmann, S., Kurz, M., Patras, M.A., Hartwig, C., Marner, M., Leis, B., Billion, A., Kleiner, Y., Bauer, A., Toti, L., Pöverlein, C., Hammann, P.E., Vilcinskas, A., Glaeser, J., Spohn, M.S., Schäberle, T.F., 2021. Genomic and chemical decryption of the Bacteroidetes phylum for its potential to biosynthesize natural products.

Brinkmann, S., Semmler, S., Kersten, C., Patras, M.A., Kurz, M., Fuchs, N., Hammerschmidt, S.J., Legac, J., Hammann, P.E., Vilcinskas, A., Rosenthal, P.J., Schirmeister, T., Bauer, A., Schäberle, T.F., 2022c. Identification, Characterization, and Synthesis of Natural Parasitic Cysteine Protease Inhibitors: Pentacitidins Are More Potent Falcitidin Analogues. ACS Chem. Biol. 17, 576–589. https://doi.org/10.1021/acschembio.1c00861

Brinkmann, S., Spohn, M.S., Schäberle, T.F., 2022d. Bioactive natural products from Bacteroidetes. Nat. Prod. Rep. 39, 1045–1065. https://doi.org/10.1039/D1NP00072A

Brinkmann, S., Spohn, M.S., Schäberle, T.F., 2022e. Bioactive natural products from Bacteroidetes. Nat. Prod. Rep. 39, 1045–1065. https://doi.org/10.1039/D1NP00072A

Buijs, Y., Bech, P.K., Vazquez-Albacete, D., Bentzon-Tilia, M., Sonnenschein, E.C., Gram, L., Zhang, S.-D., 2019. Marine Proteobacteria as a source of natural products: advances in molecular tools and strategies. Nat. Prod. Rep. 36, 1333–1350. https://doi.org/10.1039/C9NP00020H

Bush, K., Jacoby, G.A., 2010. Updated functional classification of beta-lactamases. Antimicrobial Agents and Chemotherapy 54, 969–976. https://doi.org/10.1128/AAC.01009-09

Bushin, L.B., Covington, B.C., Rued, B.E., Federle, M.J., Seyedsayamdost, M.R., 2020. Discovery and Biosynthesis of Streptosactin, a Sactipeptide with an Alternative Topology Encoded by Commensal Bacteria in the Human Microbiome. J. Am. Chem. Soc. 142, 16265–16275. https://doi.org/10.1021/jacs.0c05546

Butler, M.S., Henderson, I.R., Capon, R.J., Blaskovich, M.A.T., 2023. Antibiotics in the clinical pipeline as of December 2022. J Antibiot 76, 431–473. https://doi.org/10.1038/s41429-023-00629-8

Butler, M.S., Vollmer, W., Goodall, E.C.A., Capon, R.J., Henderson, I.R., Blaskovich, M.A.T., 2024. A Review of Antibacterial Candidates with New Modes of Action. ACS Infect. Dis. 10, 3440–3474. https://doi.org/10.1021/acsinfecdis.4c00218

Caetano, T., van der Donk, W., Mendo, S., 2020a. Bacteroidetes can be a rich source of novel lanthipeptides: The case study of Pedobacter lusitanus. Microbiological Research 235, 126441. https://doi.org/10.1016/j.micres.2020.126441

Caetano, T., van der Donk, W.A., Mendo, S., 2020b. Bacteroidetes can be a rich source of novel lanthipeptides: The case study of Pedobacter lusitanus. Microbiological Research 235, 126441–126441. https://doi.org/10.1016/j.micres.2020.126441

Campanharo, J.C., Kielak, A.M., Castellane, T.C.L., Kuramae, E.E., Lemos, E.G. de M., 2016. Optimized medium culture for Acidobacteria subdivision 1 strains. FEMS Microbiology Letters 363, fnw245. https://doi.org/10.1093/femsle/fnw245

Carroll, L.M., Larralde, M., Fleck, J.S., Ponnudurai, R., Milanese, A., Cappio, E., Zeller, G., 2021. Accurate de novo identification of biosynthetic gene clusters with GECCO. https://doi.org/10.1101/2021.05.03.442509

Casjens, S.R., Hendrix, R.W., 2015. Bacteriophage lambda: Early pioneer and still relevant. Virology, 60th Anniversary Issue 479–480, 310–330. https://doi.org/10.1016/j.virol.2015.02.010

Cautereels, C., Smets, J., Bircham, P., De Ruysscher, D., Zimmermann, A., De Rijk, P., Steensels, J., Gorkovskiy, A., Masschelein, J., Verstrepen, K.J., 2024. Combinatorial optimization of gene expression through recombinase-mediated promoter and terminator shuffling in yeast. Nat Commun 15, 1112. https://doi.org/10.1038/s41467-024-44997-7

CDC, 2025. Antimicrobial Resistance Threats in the United States, 2021-2022 [WWW Document]. Antimicrobial Resistance. URL https://www.cdc.gov/antimicrobial-resistance/data-research/threats/update-2022.html (accessed 9.6.25).

Challis, G.L., Naismith, J.H., 2004. Structural aspects of non-ribosomal peptide biosynthesis. Current Opinion in Structural Biology 14, 748–756. https://doi.org/10.1016/j.sbi.2004.10.005

Chan, Y.A., Podevels, A.M., Kevany, B.M., Thomas, M.G., 2008. Biosynthesis of polyketide synthase extender units. Nat. Prod. Rep. 26, 90–114. https://doi.org/10.1039/B801658P

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chiang, Y.-M., Oakley, C.E., Ahuja, M., Entwistle, R., Schultz, A., Chang, S.-L., Sung, C.T., Wang, C.C.C., Oakley, B.R., 2013. An efficient system for heterologous expression of secondary metabolite genes in Aspergillus nidulans. J Am Chem Soc 135, 7720–7731. https://doi.org/10.1021/ja401945a

Chiang, Y.-M., Szewczyk, E., Davidson, A.D., Keller, N., Oakley, B.R., Wang, C.C.C., 2009. A gene cluster containing two fungal polyketide synthases encodes the biosynthetic pathway for a polyketide, asperfuranone, in Aspergillus nidulans. J Am Chem Soc 131, 2965–2970. https://doi.org/10.1021/ja8088185

Chklovski, A., Parks, D.H., Woodcroft, B.J., Tyson, G.W., 2023a. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Nat Methods 20, 1203–1212. https://doi.org/10.1038/s41592-023-01940-w

Chklovski, A., Parks, D.H., Woodcroft, B.J., Tyson, G.W., 2023b. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Nat Methods 20, 1203–1212. https://doi.org/10.1038/s41592-023-01940-w

Choi, M., Carver, J., Chiva, C., Tzouros, M., Huang, T., Tsai, T.-H., Pullman, B., Bernhardt, O.M., Hüttenhain, R., Teo, G.C., Perez-Riverol, Y., Muntel, J., Müller, M., Goetze, S., Pavlou, M., Verschueren, E., Wollscheid, B., Nesvizhskii, A.I., Reiter, L., Dunkley, T., Sabidó, E., Bandeira, N., Vitek, O., 2020. MassIVE.quant: a community resource of quantitative mass spectrometry–based proteomics datasets. Nat Methods 17, 981–984. https://doi.org/10.1038/s41592-020-0955-0

Cimermancic, P., Medema, M.H., Claesen, J., Kurita, K., Wieland Brown, L.C., Mavrommatis, K., Pati, A., Godfrey, P.A., Koehrsen, M., Clardy, J., Birren, B.W., Takano, E., Sali, A., Linington, R.G., Fischbach, M.A., 2014. Insights into Secondary Metabolism from a Global Analysis of Prokaryotic Biosynthetic Gene Clusters. Cell 158, 412–421. https://doi.org/10.1016/j.cell.2014.06.034

Clardy, J., Fischbach, M.A., Walsh, C.T., 2006. New antibiotics from bacterial natural products. Nat Biotechnol 24, 1541–1550. https://doi.org/10.1038/nbt1266

Coates, J.D., Ellis, D.J., Gaw, C.V., Lovley, D.R., 1999. Geothrix fermentans gen. nov., sp. nov., a novel Fe(III)-reducing bacterium from a hydrocarbon-contaminated aquifer. Int. J. Syst. Bacteriol. 49 Pt 4, 1615–1622. https://doi.org/10.1099/00207713-49-4-1615

Coolahan, M., Whalen, K.E., 2025. A review of quorum-sensing and its role in mediating interkingdom interactions in the ocean. Commun Biol 8, 179. https://doi.org/10.1038/s42003-025-07608-9

CopyControl Fosmid Library Production Kit: pCC1FOS system (inducible oriV/TrfA copy control), 2006. . Epicentre (Illumina).

Corre, C., L. Challis, G., 2009. New natural product biosynthetic chemistry discovered by genome mining. Natural Product Reports 26, 977–986. https://doi.org/10.1039/B713024B

Corsaro, D., Wylezich, C., Walochnik, J., Venditti, D., Michel, R., 2017. Molecular identification of bacterial endosymbionts of Sappinia strains. Parasitol Res 116, 549–558. https://doi.org/10.1007/s00436-016-5319-4

Costa, O.Y.A., Zerillo, M.M., Zühlke, D., Kielak, A.M., Pijl, A., Riedel, K., Kuramae, E.E., 2020. Responses of Acidobacteria Granulicella sp. WH15 to High Carbon Revealed by Integrated Omics Analyses. Microorganisms 8, 244. https://doi.org/10.3390/microorganisms8020244

Cothrell, A., Cao, K., Bonasera, R., Tenorio, A., Orugunty, R., Smith, L., 2023. Intravaginal Gel for Sustained Delivery of Occidiofungin and Long-Lasting Antifungal Effects. Gels 9, 787. https://doi.org/10.3390/gels9100787

Covas, C., Caetano, T., Cruz, A., Santos, T., Dias, L., Klein, G., Abdulmawjood, A., Rodríguez-Alcalá, L.M., Pimentel, L.L., Gomes, A., Freitas, A.C., Garcia-Serrano, A., Fontecha, J., Mendo, S., 2017. Pedobacter lusitanus sp. nov., isolated from sludge of a deactivated uranium mine. International Journal of Systematic and Evolutionary Microbiology 67, 1339–1348. https://doi.org/10.1099/ijsem.0.001814

Covas, C., Figueiredo, G., Gomes, M., Santos, T., Mendo, S., Caetano, T.S., 2023. The Pangenome of Gram-Negative Environmental Bacteria Hides a Promising Biotechnological Potential. Microorganisms 11, 2445. https://doi.org/10.3390/microorganisms11102445

Cox, M.M., 1983. The FLP protein of the yeast 2-micron plasmid: expression of a eukaryotic genetic recombination system in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America 80, 4223–4227. https://doi.org/10.1073/pnas.80.14.4223

Craig, J.W., Chang, F.-Y., Brady, S.F., 2009. Natural Products from Environmental DNA Hosted in Ralstonia metallidurans. ACS Chem. Biol. 4, 23–28. https://doi.org/10.1021/cb8002754

Crits-Christoph, A., Diamond, S., Al-Shayeb, B., Valentin-Alvarado, L., Banfield, J.F., 2022. A widely distributed genus of soil Acidobacteria genomically enriched in biosynthetic gene clusters. ISME Commun 2, 70. https://doi.org/10.1038/s43705-022-00140-5

Crits-Christoph, A., Diamond, S., Butterfield, C.N., Thomas, B.C., Banfield, J.F., 2018a. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. Nature 558, 440–444. https://doi.org/10.1038/s41586-018-0207-y

Crits-Christoph, A., Diamond, S., Butterfield, C.N., Thomas, B.C., Banfield, J.F., 2018b. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. Nature 558, 440. https://doi.org/10.1038/s41586-018-0207-y

Dai, J., Dan, W., Schneider, U., Wang, J., 2018. β-Carboline alkaloid monomers and dimers: Occurrence, structural diversity, and biological activities. Eur J Med Chem 157, 622–656. https://doi.org/10.1016/j.ejmech.2018.08.027

Dai, J., Ouyang, Y., Gupte, R., Liu, X.J.A., Li, Y., Yang, F., Chen, S., Provin, T., Van Schaik, E., Samuel, J.E., Jayaraman, A., Zhou, A., de Figueiredo, P., Zhou, J., Han, A., 2025. Microfluidic droplets with amended culture media cultivate a greater diversity of soil microorganisms. Applied and Environmental Microbiology 91, e01794-24. https://doi.org/10.1128/aem.01794-24

Dai, Z., Liu, G., Chen, H., Chen, C., Wang, J., Ai, S., Wei, D., Li, D., Ma, B., Tang, C., Brookes, P.C., Xu, J., 2020. Long-term nutrient inputs shift soil microbial functional profiles of phosphorus cycling in diverse agroecosystems. ISME J 14, 757–770. https://doi.org/10.1038/s41396-019-0567-9

Damsté, J.S.S., Rijpstra, W.I.C., Dedysh, S.N., Foesel, B.U., Villanueva, L., 2017. Pheno- and Genotyping of Hopanoid Production in Acidobacteria. Front. Microbiol. 8. https://doi.org/10.3389/fmicb.2017.00968

Davies, J., 2013. Specialized microbial metabolites: functions and origins. J Antibiot 66, 361–364. https://doi.org/10.1038/ja.2013.61

Dedysh, S.N., Yilmaz, P., 2018. Refining the taxonomic structure of the phylum Acidobacteria. International Journal of Systematic and Evolutionary Microbiology 68, 3796–3806. https://doi.org/10.1099/ijsem.0.003062

Dinglasan, J.L.N., Otani, H., Doering, D.T., Udwary, D., Mouncey, N.J., 2025. Microbial secondary metabolites: advancements to accelerate discovery towards application. Nat Rev Microbiol 1–17. https://doi.org/10.1038/s41579-024-01141-y

Dobrindt, U., Hochhut, B., Hentschel, U., Hacker, J., 2004. Genomic islands in pathogenic and environmental microorganisms. Nat Rev Microbiol 2, 414–424. https://doi.org/10.1038/nrmicro884

Doekel, S., Eppelmann, K., Marahiel, M.A., 2002. Heterologous expression of nonribosomal peptide synthetases in B. subtilis: construction of a bi-functional B subtilis/E coli shuttle vector system. FEMS Microbiol Lett 216, 185–191. https://doi.org/10.1111/j.1574-6968.2002.tb11434.x

Domröse, A., Weihmann, R., Thies, S., Jaeger, K.-E., Drepper, T., Loeschcke, A., 2017. Rapid generation of recombinant *Pseudomonas putida* secondary metabolite producers using yTREX. Synthetic and Systems Biotechnology 2, 310–319. https://doi.org/10.1016/j.synbio.2017.11.001

Donadio, S., Monciardini, P., Sosio, M., 2007. Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. Nat. Prod. Rep. 24, 1073–1109. https://doi.org/10.1039/B514050C

D'Onofrio, A., Crawford, J.M., Stewart, E.J., Witt, K., Gavrish, E., Epstein, S., Clardy, J., Lewis, K., 2010. Siderophores from Neighboring Organisms Promote the Growth of Uncultured Bacteria. Chemistry & Biology 17, 254–264. https://doi.org/10.1016/j.chembiol.2010.02.010

Doroghazi, J.R., Metcalf, W.W., 2013. Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. BMC Genomics 14, 611. https://doi.org/10.1186/1471-2164-14-611

Dyksma, S., Pester, M., 2023. Oxygen respiration and polysaccharide degradation by a sulfate-reducing acidobacterium. Nat Commun 14, 6337. https://doi.org/10.1038/s41467-023-42074-z

Eichberg, J., Oberpaul, M., Hartwig, C., Geißler, A.H., Culmsee, C., Vilcinskas, A., Böttcher-Friebertshäuser, E., Brückner, H., Degenkolb, T., Hardes, K., 2024. Structural characterization and bioactivity profiling of the fungal peptaibiotic tolypin reveal protective effects against influenza viruses. Archiv der Pharmazie 357, e2400384. https://doi.org/10.1002/ardp.202400384

Eichorst, S.A., Breznak, J.A., Schmidt, T.M., 2007a. Isolation and Characterization of Soil Bacteria That Define Terriglobus gen. nov., in the Phylum Acidobacteria. Appl. Environ. Microbiol. 73, 2708–2717. https://doi.org/10.1128/AEM.02140-06

Eichorst, S.A., Breznak, J.A., Schmidt, T.M., 2007b. Isolation and Characterization of Soil Bacteria That Define Terriglobus gen. nov., in the Phylum Acidobacteria. Applied and Environmental Microbiology 73, 2708–2717. https://doi.org/10.1128/AEM.02140-06

Eichorst, S.A., Kuske, C.R., Schmidt, T.M., 2011. Influence of Plant Polymers on the Distribution and Cultivation of Bacteria in the Phylum Acidobacteria. Applied and Environmental Microbiology 77, 586–596. https://doi.org/10.1128/AEM.01080-10

Eichorst, S.A., Trojan, D., Roux, S., Herbold, C., Rattei, T., Woebken, D., 2018. Genomic insights into the Acidobacteria reveal strategies for their success in terrestrial environments. Environ Microbiol 20, 1041–1063. https://doi.org/10.1111/1462-2920.14043

Eldjárn, G.H., Ramsay, A., Hooft, J.J.J. van der, Duncan, K.R., Soldatou, S., Rousu, J., Daly, R., Wandy, J., Rogers, S., 2021. Ranking microbial metabolomic and genomic links in the NPLinker framework using complementary scoring functions. PLOS Computational Biology 17, e1008920. https://doi.org/10.1371/journal.pcbi.1008920

Elowitz, M.B., Leibler, S., 2000. A synthetic oscillatory network of transcriptional regulators (Repressilator). Nature 403, 335–338. https://doi.org/10.1038/35002125

Endy, D., 2005. Foundations for engineering biology. Nature 438, 449–453. https://doi.org/10.1038/nature04342

Enghiad, B., *et al.*, 2021. Cas12a-assisted precise targeted cloning using in vivo Cre-lox recombination (CAPTURE). Nature Communications 12, 1171. https://doi.org/10.1038/s41467-021-21275-4

Enghiad, B., Huang, C., Guo, F., Jiang, G., Wang, B., Tabatabaei, S.K., Martin, T.A., Zhao, H., 2021. Cas12a-assisted precise targeted cloning using in vivo Cre-lox recombination. Nat Commun 12, 1171. https://doi.org/10.1038/s41467-021-21275-4

Environment Canada, Health Canada, 2015. Final Screening Assessment for Paenibacillus polymyxa ATCC 842; Paenibacillus polymyxa ATCC 55407; Paenibacillus polymyxa 13540-4 (Final Screening Assessment No. En14- 230/2015E- PDF), Canadian Environmental Protection Act, 1999 (CEPA 1999) Screening Assessment. Government of Canada, Ottawa.

Etesami, H., Glick, B.R., 2024. Bacterial indole-3-acetic acid: A key regulator for plant growth, plant-microbe interactions, and agricultural adaptive resilience. Microbiological Research 281, 127602. https://doi.org/10.1016/j.micres.2024.127602

Ewels, P., Magnusson, M., Lundin, S., Käller, M., 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Eze, M.O., Amuji, C.F., 2024. Elucidating the significant roles of root exudates in organic pollutant biotransformation within the rhizosphere. Sci Rep 14, 2359. https://doi.org/10.1038/s41598-024-53027-x

Fautz, E., Reichenbach, H., 1980. A simple test for flexirubin-type pigments. FEMS Microbiology Letters 8, 87–90.

Feher, M., Schmidt, J.M., 2003. Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. J Chem Inf Comput Sci 43, 218–227. https://doi.org/10.1021/ci0200467

Felnagle, E.A., Barkei, J.J., Park, H., Podevels, A.M., McMahon, M.D., Drott, D.W., Thomas, M.G., 2010. MbtH-Like Proteins as Integral Components of Bacterial Nonribosomal Peptide Synthetases. Biochemistry 49, 8815–8817. https://doi.org/10.1021/bi1012854

Felske, A., De Vos, W.M., Akkermans, A.D.L., 2000. Spatial distribution of 16S rRNA levels from uncultured acidobacteria in soil. Lett. Appl. Microbiol. 31, 118–122. https://doi.org/10.1046/j.1365-2672.2000.00780.x

Fernández-Gómez, B., Richter, M., Schüler, M., Pinhassi, J., Acinas, S.G., González, J.M., Pedrós-Alió, C., 2013. Ecology of marine Bacteroidetes: a comparative genomics approach. ISME J 7, 1026–1037. https://doi.org/10.1038/ismej.2012.169

Figueiredo, G., Gomes, M., Covas, C., Mendo, S., Caetano, T., 2022. The Unexplored Wealth of Microbial Secondary Metabolites: the Sphingobacteriaceae Case Study. Microb Ecol 83, 470–481. https://doi.org/10.1007/s00248-021-01762-3

Firn, R.D., Jones, C.G., 2003. Natural products – a simple model to explain chemical diversity. Nat. Prod. Rep. 20, 382–391. https://doi.org/10.1039/B208815K

Flieder, M., Buongiorno, J., Herbold, C.W., Hausmann, B., Rattei, T., Lloyd, K.G., Loy, A., Wasmund, K., 2021. Novel taxa of Acidobacteriota implicated in seafloor sulfur cycling. ISME J 1–22. https://doi.org/10.1038/s41396-021-00992-0

F. Little, R., Hertweck, C., 2022. Chain release mechanisms in polyketide and non-ribosomal peptide biosynthesis. Natural Product Reports 39, 163–205. https://doi.org/10.1039/D1NP00035G

Foesel, B.U., Rohde, M., Overmann, J., 2013. Blastocatella fastidiosa gen. nov., sp. nov., isolated from semiarid savanna soil – The first described species of Acidobacteria subdivision 4. Systematic and Applied Microbiology 36, 82–89. https://doi.org/10.1016/j.syapm.2012.11.002

Fones, H., Gurr, S., 2015. The impact of Septoria tritici Blotch disease on wheat: An EU perspective. Fungal genetics and biology : FG & B 79, 3–7. https://doi.org/10.1016/j.fgb.2015.04.004

Fox, J.L., 2014. Fraunhofer to mine Sanofi microbial collection. Nature Biotechnology 32, 305–305. https://doi.org/10.1038/nbt0414-305a

Frébortová, J., Frébort, I., 2021. Biochemical and Structural Aspects of Cytokinin Biosynthesis and Degradation in Bacteria. Microorganisms 9, 1314. https://doi.org/10.3390/microorganisms9061314

Frenk, J., Gómez-Dantés, O., Knaul, F.M., 2011. Globalization and Infectious Diseases. Infectious Disease Clinics of North America, Global Health, Global Health Education, and Infectious Disease: the New Millennium, Part II 25, 593–599. https://doi.org/10.1016/j.idc.2011.05.003

Fu, J., Wenzel, S.C., Perlova, O., Wang, J., Gross, F., Tang, Z., Yin, Y., Stewart, A.F., Müller, R., Zhang, Y., 2008. Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. Nucleic Acids Res 36, e113–e113. https://doi.org/10.1093/nar/gkn499

Fu, S.-F., Wei, J.-Y., Chen, H.-W., Liu, Y.-Y., Lu, H.-Y., Chou, J.-Y., 2015. Indole-3-acetic acid: A widespread physiological code in interactions of fungi with other organisms. Plant Signal Behav 10, e1048052. https://doi.org/10.1080/15592324.2015.1048052

Gaitanis, G., Magiatis, P., Hantschke, M., Bassukas, I.D., Velegraki, A., 2012. The Malassezia Genus in Skin and Systemic Diseases. Clinical Microbiology Reviews 25, 106–141. https://doi.org/10.1128/CMR.00021-11

Gaitanis, G., Magiatis, P., Mexia, N., Melliou, E., Efstratiou, M.A., Bassukas, I.D., Velegraki, A., 2019. Antifungal activity of selected Malassezia indolic compounds detected in culture. Mycoses 62, 597–603. https://doi.org/10.1111/myc.12893

Genilloud, O., 2019. Natural products discovery and potential for new antibiotics. Current Opinion in Microbiology, Antimicrobials 51, 81–87. https://doi.org/10.1016/j.mib.2019.10.012

Gerlt, J.A., Bouvier, J.T., Davidson, D.B., Imker, H.J., Sadkhin, B., Slater, D.R., Whalen, K.L., 2015. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1854, 1019–1037. https://doi.org/10.1016/j.bbapap.2015.04.015

Gibson, D.G., 2011. Enzymatic assembly of overlapping DNA fragments, in: Methods in Enzymology. pp. 349–361. https://doi.org/10.1016/B978-0-12-385120-8.00015-2

Gibson, D.G., *et al.*, 2009. Assembly of large DNA molecules using Gibson assembly. Nature Methods 6, 343–345. https://doi.org/10.1038/nmeth.1318

Giguere, A.T., Eichorst, S.A., Meier, D.V., Herbold, C.W., Richter, A., Greening, C., Woebken, D., 2021. Acidobacteria are active and abundant members of diverse atmospheric H2-oxidizing communities detected in temperate soils. ISME J 15, 363–376. https://doi.org/10.1038/s41396-020-00750-8

Gillespie, D.E., Brady, S.F., Bettermann, A.D., Cianciotto, N.P., Liles, M.R., Rondon, M.R., Clardy, J., Goodman, R.M., Handelsman, J., 2002. Isolation of Antibiotics Turbomycin A and B from a Metagenomic Library of Soil Microbial DNA. Applied and Environmental Microbiology 68, 4301–4306. https://doi.org/10.1128/AEM.68.9.4301-4306.2002

Gleizer, S., *et al.*, 2019. Conversion of Escherichia coli to generate all biomass carbon from CO_2. Cell 179, 1255–1263. https://doi.org/10.1016/j.cell.2019.11.009

Glick, B.R., 2014. Bacteria with ACC deaminase can promote plant growth and help to feed the world. Microbiol Res 169, 30–39. https://doi.org/10.1016/j.micres.2013.09.009

Goecks, J., Nekrutenko, A., Taylor, J., The Galaxy Team, 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biology 11, R86. https://doi.org/10.1186/gb-2010-11-8-r86

Gomez-Escribano, J.P., Bibb, M.J., 2011. Engineering Streptomyces coelicolor for heterologous expression of secondary metabolite gene clusters. Microbial Biotechnology 4, 207–215. https://doi.org/10.1111/j.1751-7915.2010.00219.x

Gonçalves, O.S., Fernandes, A.S., Tupy, S.M., Ferreira, T.G., Almeida, L.N., Creevey, C.J., Santana, M.F., 2024. Insights into plant interactions and the biogeochemical role of the globally widespread Acidobacteriota phylum. Soil Biology and Biochemistry 192, 109369. https://doi.org/10.1016/j.soilbio.2024.109369

Gross, H., Loper, J.E., 2009. Genomics of secondary metabolite production by Pseudomonas spp. Nat. Prod. Rep. 26, 1408–1446. https://doi.org/10.1039/B817075B

Gu, G., Smith, L., Liu, A., Lu, S.-E., 2011. Genetic and Biochemical Map for the Biosynthesis of Occidiofungin, an Antifungal Produced by Burkholderia contaminans Strain MS14 ▿. Appl Environ Microbiol 77, 6189–6198. https://doi.org/10.1128/AEM.00377-11

Gu, S., Wei, Z., Shao, Z., Friman, V.-P., Cao, K., Yang, T., Kramer, J., Wang, X., Li, M., Mei, X., Xu, Y., Shen, Q., Kümmerli, R., Jousset, A., 2020. Competition for iron drives phytopathogen control by natural rhizosphere microbiomes. Nature Microbiology 5, 1002–1010. https://doi.org/10.1038/s41564-020-0719-8

Gu, Y., Lu, M., Wang, Z., Wu, X., Chen, Y., 2017. Expanding the Catalytic Promiscuity of Heparinase III from Pedobacter heparinus. Chemistry – A European Journal 23, 2548–2551. https://doi.org/10.1002/chem.201605929

Gummerlich, N., Manderscheid, N., Rebets, Y., Myronovskyi, M., Gläser, L., Kuhl, M., Wittmann, C., Luzhetskyy, A., 2021. Engineering the precursor pool to modulate the production of pamamycins in the heterologous host *S. albus* J1074. Metabolic Engineering 67, 11–18. https://doi.org/10.1016/j.ymben.2021.05.004

Gustafsson, C., Govindarajan, S., Minshull, J., 2004. Codon bias and heterologous protein expression. Trends in Biotechnology 22, 346–353. https://doi.org/10.1016/j.tibtech.2004.04.006

Haas, D., Défago, G., 2005. Biological control of soil-borne pathogens by fluorescent pseudomonads. Nature Reviews Microbiology 3, 307–319. https://doi.org/10.1038/nrmicro1129

Halamka, T.A., Garcia, A., Evans, T.W., Schubert, S., Younkin, A., Hinrichs, K.-U., Kopf, S., 2024. Occurrence of ceramides in the Acidobacterium Solibacter usitatus: implications for bacterial physiology and sphingolipids in soils. Front. Geochem. 2. https://doi.org/10.3389/fgeoc.2024.1400278

Hannigan, G.D., Prihoda, D., Palicka, A., Soukup, J., Klempir, O., Rampula, L., Durcak, J., Wurst, M., Kotowski, J., Chang, D., Wang, R., Piizzi, G., Temesi, G., Hazuda, D.J., Woelk, C.H., Bitton, D.A., 2019. A deep learning genome-mining strategy for biosynthetic gene cluster prediction. Nucleic Acids Res 47, e110. https://doi.org/10.1093/nar/gkz654

Hansanant, N., Smith, L., 2022. Occidiofungin: Actin Binding as a Novel Mechanism of Action in an Antifungal Agent. Antibiotics 11, 1143. https://doi.org/10.3390/antibiotics11091143

Hartley, J.L., Temple, G.F., Brasch, M.A., 2000. DNA cloning using in vitro site-specific recombination (Gateway). Genome Research 10, 1788–1795. https://doi.org/10.1101/gr.143000

Harvey, A.L., Edrada-Ebel, R., Quinn, R.J., 2015. The re-emergence of natural products for drug discovery in the genomics era. Nature Reviews Drug Discovery 14, 111–129. https://doi.org/10.1038/nrd4510

Harvey, C.J.B., Tang, M., Schlecht, U., Horecka, J., Fischer, C.R., Lin, H.-C., Li, J., Naughton, B., Cherry, J.A., Miranda, M., Li, Y.F., Chu, A.M., Hennessy, J.R., Vandova, G.A., Inglis, D., Aiyar, R.S., Steinmetz, L.M., Davis, R.W., Medema, M.H., Sattely, E., Khosla, C., St. Onge, R.P., Tang, Y., Hillenmeyer, M.E., 2018. HEx: A heterologous expression platform for the discovery of fungal natural products. Science Advances 4, eaar5459. https://doi.org/10.1126/sciadv.aar5459

Hausmann, B., Pelikan, C., Herbold, C.W., Köstlbacher, S., Albertsen, M., Eichorst, S.A., Rio, T.G. del, Huemer, M., Nielsen, P.H., Rattei, T., Stingl, U., Tringe, S.G., Trojan, D., Wentrup, C., Woebken, D., Pester, M., Loy, A., 2018. Peatland Acidobacteria with a dissimilatory sulfur metabolism. ISME J 12, 1729–1742. https://doi.org/10.1038/s41396-018-0077-1

He, X., Li, N., Chen, X., Zhang, Y., Zhang, X., Song, X., 2020. Pedobacter indicus sp. nov., isolated from deep-sea sediment. Antonie van Leeuwenhoek 113, 357–364. https://doi.org/10.1007/s10482-019-01346-9

Herbst, D.A., Boll, B., Zocher, G., Stehle, T., Heide, L., 2013. Structural Basis of the Interaction of MbtH-like Proteins, Putative Regulators of Nonribosomal Peptide Biosynthesis, with Adenylating Enzymes. J Biol Chem 288, 1991–2003. https://doi.org/10.1074/jbc.M112.420182

Herrmann, J., Fayad, A.A., Müller, R., 2017. Natural products from myxobacteria: novel metabolites and bioactivities. Nat. Prod. Rep. 34, 135–160. https://doi.org/10.1039/C6NP00106H

Hibbing, M.E., Fuqua, C., Parsek, M.R., Peterson, S.B., 2010. Bacterial competition: surviving and thriving in the microbial jungle. Nat Rev Microbiol 8, 15–25. https://doi.org/10.1038/nrmicro2259

Hirota-Takahata, Y., Kozuma, S., Kuraya, N., Fukuda, D., Nakajima, M., Ando, O., 2014. Pedopeptins, novel inhibitors of LPS: Taxonomy of producing organism, fermentation, isolation, physicochemical properties and structural elucidation. J Antibiot 67, 243–251. https://doi.org/10.1038/ja.2013.122

Hoess, R.H., Abremski, K., 1984. Interaction of the bacteriophage P1 recombinase Cre with the loxP site. Proceedings of the National Academy of Sciences of the United States of America 81, 1026–1030. https://doi.org/10.1073/pnas.81.4.1026

Hohn, B., Collins, J., 1988. Ten years of cosmids. Trends in Biotechnology 6, 293–298. https://doi.org/10.1016/0167-7799(88)90022-4

Hong, J., 2014. Natural Product Synthesis at the Interface of Chemistry and Biology. Chemistry – A European Journal 20, 10204–10212. https://doi.org/10.1002/chem.201402804

Hu, B., Zhao, X., Wang, E., Zhou, J., Li, J., Chen, J., Du, G., 2023. Efficient heterologous expression of cytochrome P450 enzymes in microorganisms for the biosynthesis of natural products. Critical Reviews in Biotechnology 43, 227–241. https://doi.org/10.1080/07388551.2022.2029344

Hu, E.-Z., Lan, X.-R., Liu, Z.-L., Gao, J., Niu, D.-K., 2022. A positive correlation between GC content and growth temperature in prokaryotes. BMC Genomics 23, 110. https://doi.org/10.1186/s12864-022-08353-7

Huang, D., Zhang, Z., Li, Y., Liu, F., Huang, W., Min, Y., Wang, K., Yang, J., Cao, C., Gong, Y., Ke, S., 2022. Carboline derivatives based on natural pityriacitrin as potential antifungal agents. Phytochemistry Letters 48, 100–105. https://doi.org/10.1016/j.phytol.2022.02.010

Huang, G., Qu, Q., Wang, M., Huang, M., Zhou, W., Wei, F., 2022. Global landscape of gut microbiome diversity and antibiotic resistomes across vertebrates. Sci Total Environ 838, 156178. https://doi.org/10.1016/j.scitotenv.2022.156178

Huang, J., Peng, X., Qin, K., Liu, Y., Niu, J., Liu, J., Dong, J., Zhang, Y., Peng, F., 2022. Pedobacter mucosus sp. nov., isolated from a soil sample of glacier foreland in Austre Lovénbreen, Arctic. International Journal of Systematic and Evolutionary Microbiology 72, 005448. https://doi.org/10.1099/ijsem.0.005448

Huber, F., Ridder, L., Verhoeven, S., Spaaks, J.H., Diblen, F., Rogers, S., van der Hooft, J.J.J., 2021a. Spec2Vec: Improved mass spectral similarity scoring through learning of structural relationships. PLoS Comput Biol 17, e1008724. https://doi.org/10.1371/journal.pcbi.1008724

Huber, F., van der Burg, S., van der Hooft, J.J.J., Ridder, L., 2021b. MS2DeepScore: a novel deep learning similarity measure to compare tandem mass spectra. Journal of Cheminformatics 13, 84. https://doi.org/10.1186/s13321-021-00558-4

Huber, K.J., Pester, M., Eichorst, S.A., Navarrete, A.A., Foesel, B.U., 2022. Editorial: Acidobacteria – Towards Unraveling the Secrets of a Widespread, Though Enigmatic, Phylum. Frontiers in Microbiology 13.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., Hernsdorf, A.W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D.A., Finstad, K.M., Amundson, R., Thomas, B.C., Banfield, J.F., 2016. A new view of the tree of life. Nat Microbiol 1, 16048. https://doi.org/10.1038/nmicrobiol.2016.48

Huo, Y.-Y., Xu, L., Wang, C.-S., Yang, J.-Y., You, H., Oren, A., Xu, X.-W. 2013, 2013. Fabibacterpacificus sp. nov., a moderately halophilic bacterium isolated from seawater. International Journal of Systematic and Evolutionary Microbiology 63, 3710–3714. https://doi.org/10.1099/ijs.0.051276-0

Hutchings, M.I., Truman, A.W., Wilkinson, B., 2019. Antibiotics: past, present and future. Current Opinion in Microbiology, Antimicrobials 51, 72–80. https://doi.org/10.1016/j.mib.2019.10.008

Intergovernmental Panel on Climate Change, 2023. Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (No. 978-92-9169-164–7). IPCC, Geneva. https://doi.org/10.59327/IPCC/AR6-9789291691647

Iqbal, H.A., Low-Beinart, L., Obiajulu, J.U., Brady, S.F., 2016. Natural Product Discovery through Improved Functional Metagenomics in Streptomyces. J Am Chem Soc 138, 9341–9344. https://doi.org/10.1021/jacs.6b02921

Irlinger, B., Bartsch, A., Krämer, H.-J., Mayser, P., Steglich, W., 2005. New Tryptophan Metabolites from Cultures of the Lipophilic Yeast Malassezia furfur. Helvetica Chimica Acta 88, 1472–1485. https://doi.org/10.1002/hlca.200590118

Iverson, S.V., Haddock, T.L., Beal, J., Densmore, D.M., 2016. CIDAR MoClo: Improved MoClo Assembly Standard and New E. coli Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. ACS Synthetic Biology 5, 99–103. https://doi.org/10.1021/acssynbio.5b00124

Jiang, W., Zhao, X., Gabrieli, T., Lou, C., Ebenstein, Y., Zhu, T.F., 2015. Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. Nat Commun 6, 8101. https://doi.org/10.1038/ncomms9101

Joaquin, D., Lee, M.A., Kastner, D.W., Singh, J., Morrill, S.T., Damstedt, G., Morris, J., Mason, J.C., Downey, C.W., Lewis, J.C., 2020. Impact of Dehydroamino Acids on the Structure and Stability of Incipient 310-Helical Peptides. The Journal of Organic Chemistry 85, 1601–1613. https://doi.org/10.1021/acs.joc.9b02981

Jones, R.T., Robeson, M.S., Lauber, C.L., Hamady, M., Knight, R., Fierer, N., 2009. A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. The ISME Journal 3, 442–453. https://doi.org/10.1038/ismej.2008.127

Julien, B., Shah, S., 2002. Heterologous Expression of Epothilone Biosynthetic Genes in Myxococcus xanthus. Antimicrobial Agents and Chemotherapy 46, 2772–2778. https://doi.org/10.1128/aac.46.9.2772-2778.2002

Kaeberlein, T., Lewis, K., Epstein, S.S., 2002. Isolating "Uncultivable" Microorganisms in Pure Culture in a Simulated Natural Environment. Science 296, 1127–1129. https://doi.org/10.1126/science.1070633

Kalam, S., Basu, A., Ahmad, I., Sayyed, R.Z., El-Enshasy, H.A., Dailin, D.J., Suriani, N.L., 2020. Recent Understanding of Soil Acidobacteria and Their Ecological Significance: A Critical Review. Frontiers in Microbiology 11, 2712. https://doi.org/10.3389/fmicb.2020.580024

Kalkreuter, E., Kautsar, S.A., Yang, D., Bader, C.D., Teijaro, C.N., Fluegel, L.L., Davis, C.M., Simpson, J.R., Lauterbach, L., Steele, A.D., Gui, C., Meng, S., Li, G., Viehrig, K., Ye, F., Su, P., Kiefer, A.F., Nichols, A., Cepeda, A.J., Yan, W., Fan, B., Jiang, Y., Adhikari, A., Zheng, C.-J., Schuster, L., Cowan, T.M., Smanski, M.J., Chevrette, M.G., Carvalho, L.P.S. de, Shen, B., 2024. The Natural Products Discovery Center: Release of the First 8490 Sequenced Strains for Exploring Actinobacteria Biosynthetic Diversity. https://doi.org/10.1101/2023.12.14.571759

Kane, J.F., 1995. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. Current Opinion in Biotechnology 6, 494–500. https://doi.org/10.1016/0958-1669(95)80082-4

Kapili, B.J., Sipes, K., Bailey, J.V., Orphan, V.J., 2020. Evidence for phylogenetically and catabolically diverse active diazotrophs in deep-sea sediment. The ISME Journal. https://doi.org/10.1038/s41396-019-0584-8

Karsch-Mizrachi, I., Arita, M., Burdett, T., Cochrane, G., Nakamura, Y., Pruitt, K.D., Schneider, V.A., of the International Nucleotide Sequence Database Collaboration, on behalf, 2025. The international nucleotide sequence database collaboration (INSDC): enhancing global participation. Nucleic Acids Res 53, D62–D66. https://doi.org/10.1093/nar/gkae1058

Katz, L., Baltz, R.H., 2016. Natural product discovery: past, present, and future. J. Ind. Microbiol. Biotechnol. 43, 155–176. https://doi.org/10.1007/s10295-015-1723-5

Kautsar, S.A., Blin, K., Shaw, S., Navarro-Muñoz, J.C., Terlouw, B.R., van der Hooft, J.J.J., van Santen, J.A., Tracanna, V., Suarez Duran, H.G., Pascal Andreu, V., Selem-Mojica, N., Alanjary, M., Robinson, S.L., Lund, G., Epstein, S.C., Sisto, A.C., Charkoudian, L.K., Collemare, J., Linington, R.G., Weber, T., Medema, M.H., 2020. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. Nucleic Acids Research 48, D454–D458. https://doi.org/10.1093/nar/gkz882

Kautsar, S.A., van der Hooft, J.J.J., de Ridder, D., Medema, M.H., 2021. BiG-SLiCE: A highly scalable tool maps the diversity of 1.2 million biosynthetic gene clusters. Gigascience 10, giaa154. https://doi.org/10.1093/gigascience/giaa154

Keatinge-Clay, A.T., 2016. Stereocontrol within Polyketide Assembly Lines. Nat Prod Rep 33, 141–149. https://doi.org/10.1039/c5np00092k

Kelwick, R., Bowater, L., Yeoman, K.H., Bowater, R.P., 2015. Promoting microbiology education through the iGEM synthetic biology competition. FEMS Microbiol Lett 362, fnv129. https://doi.org/10.1093/femsle/fnv129

Khaldi, N., Seifuddin, F.T., Turner, G., Haft, D., Nierman, W.C., Wolfe, K.H., Fedorova, N.D., 2010. SMURF: Genomic mapping of fungal secondary metabolite clusters. Fungal Genetics and Biology 47, 736–741. https://doi.org/10.1016/j.fgb.2010.06.003

Khosla, C., Kapur, S., Cane, D.E., 2009. Revisiting the modularity of modular polyketide synthases. Current Opinion in Chemical Biology, Molecular Diversity 13, 135–143. https://doi.org/10.1016/j.cbpa.2008.12.018

Khosravi Babadi, Z., Ebrahimipour, G., Wink, J., Narmani, A., Risdian, C., 2021. Isolation and identification of Streptomyces sp. Act4Zk, a good producer of Staurosporine and some derivatives. Lett. Appl. Microbiol. 72, 206–218. https://doi.org/10.1111/lam.13415

Kielak, A.M., Barreto, C.C., Kowalchuk, G.A., van Veen, J.A., Kuramae, E.E., 2016a. The Ecology of Acidobacteria: Moving beyond Genes and Genomes. Front. Microbiol. 7. https://doi.org/10.3389/fmicb.2016.00744

Kielak, A.M., Castellane, T.C.L., Campanharo, J.C., Colnago, L.A., Costa, O.Y.A., Corradi da Silva, M.L., van Veen, J.A., Lemos, E.G.M., Kuramae, E.E., 2017. Characterization of novel Acidobacteria exopolysaccharides with potential industrial and ecological applications. Sci Rep 7, 41193. https://doi.org/10.1038/srep41193

Kielak, A.M., Cipriano, M.A.P., Kuramae, E.E., 2016b. Acidobacteria strains from subdivision 1 act as plant growth-promoting bacteria. Arch Microbiol 198, 987–993. https://doi.org/10.1007/s00203-016-1260-2

Kim, J.H., Feng, Z., Bauer, J.D., Kallifidas, D., Calle, P.Y., Brady, S.F., 2010. Cloning large natural product gene clusters from the environment: Piecing environmental DNA gene clusters back together with TAR. Biopolymers 93, 833–844. https://doi.org/10.1002/bip.21450

Kim, J.H., Komatsu, M., Shin-ya, K., Omura, S., Ikeda, H., 2018. Distribution and functional analysis of the phosphopantetheinyl transferase superfamily in Actinomycetales microorganisms. Proceedings of the National Academy of Sciences 115, 6828–6833. https://doi.org/10.1073/pnas.1800715115

Kirst, H.A., 2010. The spinosyn family of insecticides: realizing the potential of natural products research. J Antibiot 63, 101–111. https://doi.org/10.1038/ja.2010.5

Kishimoto, N., Kosako, Y., Tano, T., 1991. Acidobacterium capsulatum gen. nov., sp. nov.: An acidophilic chemoorganotrophic bacterium containing menaquinone from acidic mineral environment. Current Microbiology 22, 1–7. https://doi.org/10.1007/BF02106205

Klappenbach, J.A., Dunbar, J.M., Schmidt, T.M., 2000. rRNA Operon Copy Number Reflects Ecological Strategies of Bacteria. Applied and Environmental Microbiology 66, 1328–1333. https://doi.org/10.1128/AEM.66.4.1328-1333.2000

Klass, S.H., Wesselkamper, M., Cowan, A.E., Lee, N., Lanclos, N., Cheong, S., Wang, Z., Chen, Y., Gin, J.W., Petzold, C.J., Keasling, J.D., 2025. Engineering controllable alteration of malonyl-CoA levels to enhance polyketide production. Nat Chem Biol 21, 1214–1225. https://doi.org/10.1038/s41589-025-01911-6

Kline, B.C., 1985. A review of mini-F plasmid maintenance. Plasmid 14, 1–16. https://doi.org/10.1016/0147-619X(85)90027-7

Koch, I.H., Gich, F., Dunfield, P.F., Overmann, J., 2008. Edaphobacter modestus gen. nov., sp. nov., and Edaphobacter aggregans sp. nov., acidobacteria isolated from alpine and forest soils. International Journal of Systematic and Evolutionary Microbiology 58, 1114–1122. https://doi.org/10.1099/ijs.0.65303-0

Koenig, R.L., Morris, R.O., Polacco, J.C., 2002. tRNA Is the Source of Low-Level trans-Zeatin Production in Methylobacterium spp. Journal of Bacteriology 184, 1832–1842. https://doi.org/10.1128/jb.184.7.1832-1842.2002

Koraimann, G., 2018. Spread and Persistence of Virulence and Antibiotic Resistance Genes: A Ride on the F Plasmid Conjugation Module. EcoSal Plus 8. https://doi.org/10.1128/ecosalplus.ESP-0003-2018

Kozuma, S., Hirota-Takahata, Y., Fukuda, D., Kuraya, N., Nakajima, M., Ando, O., 2014. Screening and biological activities of pedopeptins, novel inhibitors of LPS produced by soil bacteria. J Antibiot 67, 237–242. https://doi.org/10.1038/ja.2013.121

Kravchenko, L.V., Azarova, T.S., Makarova, N.M., Tikhonovich, I.A., 2004. The Effect of Tryptophan Present in Plant Root Exudates on the Phytostimulating Activity of Rhizobacteria. Microbiology 73, 156–158. https://doi.org/10.1023/B:MICI.0000023982.76684.9d

Ku, Y.-S., Wang, Z., Duan, S., Lam, H.-M., 2021. Rhizospheric Communication through Mobile Genetic Element Transfers for the Regulation of Microbe–Plant Interactions. Biology 10, 477. https://doi.org/10.3390/biology10060477

Kumpfmüller, J., Methling, K., Fang, L., Pfeifer, B.A., Lalk, M., Schweder, T., 2016. Production of the polyketide 6-deoxyerythronolide B in the heterologous host Bacillus subtilis. Appl Microbiol Biotechnol 100, 1209–1220. https://doi.org/10.1007/s00253-015-6990-6

Kuroha, T., Tokunaga, H., Kojima, M., Ueda, N., Ishida, T., Nagawa, S., Fukuda, H., Sugimoto, K., Sakakibara, H., 2009. Functional Analyses of LONELY GUY Cytokinin-Activating Enzymes Reveal the Importance of the Direct Activation Pathway in Arabidopsis. Plant Cell 21, 3152–3169. https://doi.org/10.1105/tpc.109.068676

Kuske, C.R., Barns, S.M., Busch, J.D., 1997. Diverse uncultivated bacterial groups from soils of the arid southwestern United States that are present in many geographic regions. Applied and Environmental Microbiology 63, 3614–3621. https://doi.org/10.1128/aem.63.9.3614-3621.1997

Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., Reid, R., Khosla, C., Walsh, C.T., 1996. A new enzyme superfamily — the phosphopantetheinyl transferases. Chemistry & Biology 3, 923–936. https://doi.org/10.1016/S1074-5521(96)90181-7

Lasch, C., Myronovskyi, M., Luzhetskyy, A., 2025. Streptomyces as a versatile host platform for heterologous production of microbial natural products. Natural Product Reports. https://doi.org/10.1039/D5NP00036J

Lauer, A., Simon, M.A., Banning, J.L., André, E., Duncan, K., Harris, R.N., 2007. Common Cutaneous Bacteria from the Eastern Red-Backed Salamander Can Inhibit Pathogenic Fungi. cope 2007, 630–640. https://doi.org/10.1643/0045-8511(2007)2007[630:CCBFTE]2.0.CO;2

Lawrence, J., Sodahlon, Y.K., Ogoussan, K.T., Hopkins, A.D., 2015. Growth, Challenges, and Solutions over 25 Years of Mectizan and the Impact on Onchocerciasis Control. PLoS Negl Trop Dis 9, e0003507. https://doi.org/10.1371/journal.pntd.0003507

Leão, T.F., Wang, M., da Silva, R., Gurevich, A., Bauermeister, A., Gomes, P.W.P., Brejnrod, A., Glukhov, E., Aron, A.T., Louwen, J.J.R., Kim, H.W., Reher, R., Fiore, M.F., van der Hooft, J.J.J., Gerwick, L., Gerwick, W.H., Bandeira, N., Dorrestein, P.C., 2022. NPOmix: A machine learning classifier to connect mass spectrometry fragmentation data to biosynthetic gene clusters. PNAS Nexus 1, pgac257. https://doi.org/10.1093/pnasnexus/pgac257

Lee, I., Ouk Kim, Y., Park, S.-C., Chun, J., 2016a. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol 66, 1100–1103. https://doi.org/10.1099/ijsem.0.000760

Lee, I., Ouk Kim, Y., Park, S.-C., Chun, J. 2016, 2016b. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. International Journal of Systematic and Evolutionary Microbiology 66, 1100–1103. https://doi.org/10.1099/ijsem.0.000760

Lee, S.-H., Ka, J.-O., Cho, J.-C., 2008. Members of the phylum Acidobacteria are dominant and metabolically active in rhizosphere soil. FEMS Microbiol Lett 285, 263–269. https://doi.org/10.1111/j.1574-6968.2008.01232.x

Leibniz Institute DSMZ, L., n.d. Phylum Acidobacteriota [WWW Document]. LPSN — List of Prokaryotic names with Standing in Nomenclature. URL https://lpsn.dsmz.de/phylum/acidobacteriota (accessed 10.3.25).

Leopold-Messer, S., Chepkirui, C., Mabesoone, M.F.J., Meyer, J., Paoli, L., Sunagawa, S., Uria, A.R., Wakimoto, T., Piel, J., 2023. Animal-associated marine Acidobacteria with a rich natural-product repertoire. Chem 9, 3696–3713. https://doi.org/10.1016/j.chempr.2023.11.003

Letunic, I., Bork, P., 2024. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Research 52, W78–W82. https://doi.org/10.1093/nar/gkae268

Letunic, I., Bork, P., 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47, W256–W259. https://doi.org/10.1093/nar/gkz239

Lewis, K., 2020. The Science of Antibiotic Discovery. Cell 181, 29–45. https://doi.org/10.1016/j.cell.2020.02.056

Li, M.H.T., Ung, P.M.U., Zajkowski, J., Garneau-Tsodikova, S., Sherman, D.H., 2009. Automated genome mining for natural products. BMC Bioinformatics 10, 185–185. https://doi.org/10.1186/1471-2105-10-185

Li, W., O'Neill, K.R., Haft, D.H., DiCuccio, M., Chetvernin, V., Badretdin, A., Coulouris, G., Chitsaz, F., Derbyshire, M.K., Durkin, A.S., Gonzales, N.R., Gwadz, M., Lanczycki, C.J., Song, J.S., Thanki, N., Wang, J., Yamashita, R.A., Yang, M., Zheng, C., Marchler-Bauer, A., Thibaud-Nissen, F., 2021. RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. Nucleic Acids Res 49, D1020–D1028. https://doi.org/10.1093/nar/gkaa1105

Li, X.-L., Sun, Y., Yin, Y., Zhan, S., Wang, C., 2023. A bacterial-like Pictet–Spenglerase drives the evolution of fungi to produce β-carboline glycosides together with separate genes. Proceedings of the National Academy of Sciences 120, e2303327120. https://doi.org/10.1073/pnas.2303327120

Liesack, W., Bak, F., Kreft, J.-U., Stackebrandt, E., 1994. Holophaga foetida gen. nov., sp. nov., a new, homoacetogenic bacterium degrading methoxylated aromatic compounds. Arch. Microbiol. 162, 85–90. https://doi.org/10.1007/BF00264378

Liew, L.P.P., Fleming, J.M., Longeon, A., Mouray, E., Florent, I., Bourguet-Kondracki, M.-L., Copp, B.R., 2014. Synthesis of 1-indolyl substituted β-carboline natural products and discovery of antimalarial and cytotoxic activities. Tetrahedron 70, 4910–4920. https://doi.org/10.1016/j.tet.2014.05.068

Lightfield, J., Fram, N.R., Ely, B., 2011. Across Bacterial Phyla, Distantly-Related Genomes with Similar Genomic GC Content Have Similar Patterns of Amino Acid Usage. PLOS ONE 6, e17677. https://doi.org/10.1371/journal.pone.0017677

Ling, L.L., Schneider, T., Peoples, A.J., Spoering, A.L., Engels, I., Conlon, B.P., Mueller, A., Schäberle, T.F., Hughes, D.E., Epstein, S., Jones, M., Lazarides, L., Steadman, V.A., Cohen, D.R., Felix, C.R., Fetterman, K.A., Millett, W.P., Nitti, A.G., Zullo, A.M., Chen, C., Lewis, K., 2015. A new antibiotic kills pathogens without detectable resistance. Nature 517, 455–459. https://doi.org/10.1038/nature14098

Liu, X., Li, F., Sun, T., Guo, J., Zhang, X., Zheng, X., Du, L., Zhang, W., Ma, L., Li, S., 2022. Three pairs of surrogate redox partners comparison for Class I cytochrome P450 enzyme activity reconstitution. Commun Biol 5, 791. https://doi.org/10.1038/s42003-022-03764-4

Liu, Y., Chen, L., Zhang, N., Li, Z., Zhang, G., Xu, Y., Shen, Q., Zhang, R., 2016. Plant-Microbe Communication Enhances Auxin Biosynthesis by a Root-Associated Bacterium, Bacillus amyloliquefaciens SQR9. MPMI 29, 324–330. https://doi.org/10.1094/MPMI-10-15-0239-R

Lloyd, K.G., Steen, A.D., Ladau, J., Yin, J., Crosby, L., 2018. Phylogenetically Novel Uncultured Microbial Cells Dominate Earth Microbiomes. mSystems 3, 10.1128/msystems.00055-18. https://doi.org/10.1128/msystems.00055-18

Loeschcke, A., Markert, A., Wilhelm, S., Wirtz, A., Rosenau, F., Jaeger, K.-E., Drepper, T., 2013. TREX: a universal tool for the transfer and expression of biosynthetic pathways in bacteria. ACS Synthetic Biology 2, 22–33. https://doi.org/10.1021/sb3000657

Loeschcke, A., Thies, S., 2015. Pseudomonas putida—a versatile host for the production of natural products. Appl Microbiol Biotechnol 99, 6197–6214. https://doi.org/10.1007/s00253-015-6745-4

Losey, N.A., Stevenson, B.S., Busse, H.-J., Damsté, J.S.S., Rijpstra, W.I.C., Rudd, S., Lawson, P.A., 2013. Thermoanaerobaculum aquaticum gen. nov., sp. nov., the first cultivated member of Acidobacteria subdivision 23, isolated from a hot spring. International Journal of Systematic and Evolutionary Microbiology 63, 4149–4157. https://doi.org/10.1099/ijs.0.051425-0

Loudon, A.H., Holland, J.A., Umile, T.P., Burzynski, E.A., Minbiole, K.P.C., Harris, R.N., 2014. Interactions between amphibians' symbiotic bacteria cause the production of emergent anti-fungal metabolites. Front Microbiol 5, 441. https://doi.org/10.3389/fmicb.2014.00441

Lovecká, P., Kroneislová, G., Novotná, Z., Röderová, J., Demnerová, K., 2023. Plant Growth-Promoting Endophytic Bacteria Isolated from Miscanthus giganteus and Their Antifungal Activity. Microorganisms 11, 2710. https://doi.org/10.3390/microorganisms11112710

Lu, S.-E., Novak, J., Austin, F.W., Gu, G., Ellis, D., Kirk, M., Wilson-Stanford, S., Tonelli, M., Smith, L., 2009. Occidiofungin, a Unique Antifungal Glycopeptide Produced by a Strain of Burkholderia contaminans. Biochemistry 48, 8312–8321. https://doi.org/10.1021/bi900814c

Ludwig, W., Bauer, S.H., Bauer, M., Held, I., Kirchhof, G., Schulze, R., Huber, I., Spring, S., Hartmann, A., Schleifer, K.H., 1997. Detection and in situ identification of representatives of a widely distributed new bacterial phylum. FEMS Microbiology Letters 153, 181–190. https://doi.org/10.1111/j.1574-6968.1997.tb10480.x

Machowinski, A., Krämer, H.-J., Hort, W., Mayser, P., 2006. Pityriacitrin – a potent UV filter produced by Malassezia furfur and its effect on human skin microflora. Mycoses 49, 388–392. https://doi.org/10.1111/j.1439-0507.2006.01265.x

Magiatis, P., Pappas, P., Gaitanis, G., Mexia, N., Melliou, E., Galanou, M., Vlachos, C., Stathopoulou, K., Skaltsounis, A.L., Marselos, M., Velegraki, A., Denison, M.S., Bassukas, I.D., 2013. Malassezia yeasts produce a collection of exceptionally potent activators of the Ah (dioxin) receptor detected in diseased human skin. J Invest Dermatol 133, 2023–2030. https://doi.org/10.1038/jid.2013.92

Mahler, L., Niehs, S.P., Martin, K., Weber, T., Scherlach, K., Hertweck, C., Roth, M., Rosenbaum, M.A., 2021. Highly parallelized droplet cultivation and prioritization of antibiotic producers from natural microbial communities. eLife 10, e64774. https://doi.org/10.7554/eLife.64774

Malpartida, F., Hopwood, D.A., 1984. Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host. Nature 309, 462–464. https://doi.org/10.1038/309462a0

Männle, D., McKinnie, S.M.K., Mantri, S.S., Steinke, K., Lu, Z., Moore, B.S., Ziemert, N., Kaysser, L., 2020. Comparative Genomics and Metabolomics in the Genus Nocardia. mSystems 5, 10.1128/msystems.00125-20. https://doi.org/10.1128/msystems.00125-20

Manter, D.K., Hamm, A.K., Deel, H.L., 2023. Community structure and abundance of ACC deaminase containing bacteria in soils with 16S-PICRUSt2 inference or direct *acdS* gene sequencing. Journal of Microbiological Methods 211, 106740. https://doi.org/10.1016/j.mimet.2023.106740

Margesin, R., Spröer, C., Schumann, P., Schinner, F., 2003. Pedobacter cryoconitis sp. nov., a facultative psychrophile from alpine glacier cryoconite. International Journal of Systematic and Evolutionary Microbiology 53, 1291–1296. https://doi.org/10.1099/ijs.0.02436-0

Martínez-García, E., Fraile, S., Algar, E., Aparicio, T., Velázquez, E., Calles, B., Tas, H., Blázquez, B., Martín, B., Prieto, C., Sánchez-Sampedro, L., Nørholm, M.H.H., Volke, D.C., Wirth, N.T., Dvořák, P., Alejaldre, L., Grozinger, L., Crowther, M., Goñi-Moreno, A., Nikel, P.I., Nogales, J., de Lorenzo, V., 2023. SEVA 4.0: an update of the Standard European Vector Architecture database for advanced analysis and programming of bacterial phenotypes. Nucleic Acids Res 51, D1558–D1567. https://doi.org/10.1093/nar/gkac1059

Masnoddin, M., Ling, C.M.W.V., Yusof, N.A., 2022. Functional Analysis of Conserved Hypothetical Proteins from the Antarctic Bacterium, Pedobacter cryoconitis Strain BG5 Reveals Protein Cold Adaptation and Thermal Tolerance Strategies. Microorganisms 10, 1654–1654. https://doi.org/10.3390/microorganisms10081654

Massicard, J.-M., Noel, D., Calderari, A., Le Jeune, A., Pauthenier, C., Weissman, K.J., 2024. Modular Cloning Tools for Streptomyces spp. and Application to the De Novo Biosynthesis of Flavokermesic Acid. ACS Synth. Biol. 13, 3354–3365. https://doi.org/10.1021/acssynbio.4c00412

Massip, C., Oswald, E., 2020. Siderophore-Microcins in Escherichia coli: Determinants of Digestive Colonization, the First Step Toward Virulence. Front Cell Infect Microbiol 10, 381. https://doi.org/10.3389/fcimb.2020.00381

Matsuo, Y., Imagawa, H., Nishizawa, M., Shizuri, Y., 2005. Isolation of an Algal Morphogenesis Inducer from a Marine Bacterium. Science 307, 1598–1598. https://doi.org/10.1126/science.1105486

Mayser, P., Schäfer, U., Krämer, H.-J., Irlinger, B., Steglich, W., 2002. Pityriacitrin – an ultraviolet-absorbing indole alkaloid from the yeast Malassezia furfur. Arch Dermatol Res 294, 131–134. https://doi.org/10.1007/s00403-002-0294-2

Mayser, P., Wenzel, M., Krämer, H.-J., Kindler, B.L.J., Spiteller, P., Haase, G., 2007. Production of indole pigments by Candida glabrata. Med Mycol 45, 519–524. https://doi.org/10.1080/13693780701411557

Mayser, P., Wille, G., Imkampe, A., Thoma, W., Arnold, N., Monsees, T., 1998. Synthesis of fluorochromes and pigments in Malassezia furfur by use of tryptophan as the single nitrogen source. Mycoses 41, 265–271. https://doi.org/10.1111/j.1439-0507.1998.tb00336.x

McKELLAR, Q.A., Benchaoui, H.A., 1996. Avermectins and milbemycins. Journal of Veterinary Pharmacology and Therapeutics 19, 331–351. https://doi.org/10.1111/j.1365-2885.1996.tb00062.x

McReynolds, E., Elshahed, M.S., Youssef, N.H., 2025. An ecological-evolutionary perspective on the genomic diversity and habitat preferences of the Acidobacteriota. Microbial Genomics 11, 001344. https://doi.org/10.1099/mgen.0.001344

Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M.A., Weber, T., Takano, E., Breitling, R., 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res 39, W339–W346. https://doi.org/10.1093/nar/gkr466

Medema, M.H., Fischbach, M.A., 2015a. Computational approaches to natural product discovery. Nat Chem Biol 11, 639–648. https://doi.org/10.1038/nchembio.1884

Medema, M.H., Fischbach, M.A., 2015b. Computational approaches to natural product discovery. Nat Chem Biol 11, 639–648. https://doi.org/10.1038/nchembio.1884

Mehta-Kolte, M.G., Bond, D.R., 2012. Geothrix fermentans Secretes Two Different Redox-Active Compounds To Utilize Electron Acceptors across a Wide Range of Redox Potentials. Appl Environ Microbiol 78, 6987–6995. https://doi.org/10.1128/AEM.01460-12

Meier-Kolthoff, J.P., Göker, M., 2019a. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun 10, 2182. https://doi.org/10.1038/s41467-019-10210-3

Meier-Kolthoff, J.P., Göker, M., 2019b. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun 10, 2182. https://doi.org/10.1038/s41467-019-10210-3

Meng, X., Fang, Y., Ding, M., Zhang, Y., Jia, K., Li, Z., Collemare, J., Liu, W., 2022. Developing fungal heterologous expression platforms to explore and improve the production of natural products from fungal biodiversity. Biotechnology Advances 54, 107866. https://doi.org/10.1016/j.biotechadv.2021.107866

Mertz, F.P., Yao, R.C., 1990. Saccharopolyspora spinosa sp. nov. Isolated from Soil Collected in a Sugar Mill Rum Still. International Journal of Systematic Bacteriology 40, 34–39. https://doi.org/10.1099/00207713-40-1-34

Merwin, N.J., Mousa, W.K., Dejong, C.A., Skinnider, M.A., Cannon, M.J., Li, H., Dial, K., Gunabalasingam, M., Johnston, C., Magarvey, N.A., 2020. DeepRiPP integrates multiomics data to automate discovery of novel ribosomally synthesized natural products. Proceedings of the National Academy of Sciences 117, 371–380. https://doi.org/10.1073/pnas.1901493116

Miethke, M., Pieroni, M., Weber, T., Brönstrup, M., Hammann, P., Halby, L., Arimondo, P.B., Glaser, P., Aigle, B., Bode, H.B., Moreira, R., Li, Y., Luzhetskyy, A., Medema, M.H., Pernodet, J.-L., Stadler, M., Tormo, J.R., Genilloud, O., Truman, A.W., Weissman, K.J., Takano, E., Sabatini, S., Stegmann, E., Brötz-Oesterhelt, H., Wohlleben, W., Seemann, M., Empting, M., Hirsch, A.K.H., Loretz, B., Lehr, C.-M., Titz, A., Herrmann, J., Jaeger, T., Alt, S., Hesterkamp, T., Winterhalter, M., Schiefer, A., Pfarr, K., Hoerauf, A., Graz, H., Graz, M., Lindvall, M., Ramurthy, S., Karlén, A., van Dongen, M., Petkovic, H., Keller, A., Peyrane, F., Donadio, S., Fraisse, L., Piddock, L.J.V., Gilbert, I.H., Moser, H.E., Müller, R., 2021. Towards the sustainable discovery and development of new antibiotics. Nat Rev Chem 5, 726–749. https://doi.org/10.1038/s41570-021-00313-1

Miller, B.R., Gulick, A.M., 2016. Structural Biology of Nonribosomal Peptide Synthetases, in: Evans, B.S. (Ed.), Nonribosomal Peptide and Polyketide Biosynthesis: Methods and Protocols. Springer, New York, NY, pp. 3–29. https://doi.org/10.1007/978-1-4939-3375-4_1

Mitchell, A.L., Almeida, A., Beracochea, M., Boland, M., Burgin, J., Cochrane, G., Crusoe, M.R., Kale, V., Potter, S.C., Richardson, L.J., Sakharova, E., Scheremetjew, M., Korobeynikov, A., Shlemov, A., Kunyavskaya, O., Lapidus, A., Finn, R.D., 2020. MGnify: the microbiome analysis resource in 2020. Nucleic Acids Research 48, D570–D578. https://doi.org/10.1093/nar/gkz1035

Mohr, K.I., Volz, C., Jansen, R., Wray, V., Hoffmann, J., Bernecker, S., Wink, J., Gerth, K., Stadler, M., Müller, R., 2015. Pinensins: The First Antifungal Lantibiotics. Angewandte Chemie International Edition 54, 11254–11258. https://doi.org/10.1002/anie.201500927

Mora, C., McKenzie, T., Gaw, I.M., Dean, J.M., von Hammerstein, H., Knudson, T.A., Setter, R.O., Smith, C.Z., Webster, K.M., Patz, J.A., Franklin, E.C., 2022. Over half of known human pathogenic diseases can be aggravated by climate change. Nat. Clim. Chang. 12, 869–875. https://doi.org/10.1038/s41558-022-01426-1

Mori, S., Green, K.D., Choi, R., Buchko, G.W., Fried, M.G., Garneau-Tsodikova, S., 2018. Using MbtH-Like Proteins to Alter the Substrate Profile of a Nonribosomal Peptide Adenylation Enzyme. ChemBioChem 19, 2186–2194. https://doi.org/10.1002/cbic.201800240

Morpheus, n.d. . https://software.broadinstitute.org/morpheus.

Mortensen, A., Aguilar, F., Crebelli, R., Di Domenico, A., Frutos, M.J., Galtier, P., Gott, D., Gundert-Remy, U., Lambré, C., Leblanc, J., Lindtner, O., Moldeus, P., Mosesso, P., Oskarsson, A., Parent-Massin, D., Stankovic, I., Waalkens-Berendsen, I., Woutersen, R.A., Wright, M., Younes, M., Brimer, L., Christodoulidou, A., Lodi, F., Gelgelova, P., Dusemund, B., 2017. Re-evaluation of xanthan gum (E 415) as a food additive. EFSA J 15, e04909. https://doi.org/10.2903/j.efsa.2017.4909

Mózsik, L., *et al.*, 2021. CRISPR-based transcriptional activation tool for silent genes in filamentous fungi (activation of macrophorin BGC in Penicillium rubens). Scientific Reports 11, 1118. https://doi.org/10.1038/s41598-020-80864-3

Mussagy, C.U., Caicedo-Paz, A.V., Giuffrida, D., Mondello, L., Tropea, A., 2025. Flexixanthin: A next-generation rare microbial carotenoid for food applications. Food Bioscience 71, 107252. https://doi.org/10.1016/j.fbio.2025.107252

Nagao, T., Adachi, K., Nishida, F., Nishishima, M., Mochida, K., 1999. New ultraviolet-absorbing substance produced by marine bacteria and its production. JP11-269175.

Nandika, D., Karlinasari, L., Arinana, A., Batubara, I., Sitanggang, P.S., Santoso, D., Witasari, L.D., Rachmayanti, Y., Firmansyah, D., Sudiana, I.K., Hertanto, D.M., 2021. Chemical Components of Fungus Comb from Indo-Malayan Termite Macrotermes gilvus Hagen Mound and Its Bioactivity against Wood-Staining Fungi. Forests 12, 1591. https://doi.org/10.3390/f12111591

Naseem, H., Ahsan, M., Shahid, M.A., Khan, N., 2018. Exopolysaccharides producing rhizobacteria and their role in plant growth and drought tolerance. J Basic Microbiol 58, 1009–1022. https://doi.org/10.1002/jobm.201800309

Nasser, M., Sharma, M., Kaur, G., 2024. Advances in the production of biosurfactants as green ingredients in home and personal care products. Front Chem 12, 1382547. https://doi.org/10.3389/fchem.2024.1382547

Nauwynck, W., Faust, K., Boon, N., 2025. Droplet microfluidics for single-cell studies: a frontier in ecological understanding of microbiomes. FEMS Microbiol Rev 49, fuaf032. https://doi.org/10.1093/femsre/fuaf032

Navarro-Muñoz, J.C., Selem-Mojica, N., Mullowney, M.W., Kautsar, S.A., Tryon, J.H., Parkinson, E.I., De Los Santos, E.L.C., Yeong, M., Cruz-Morales, P., Abubucker, S., Roeters, A., Lokhorst, W., Fernandez-Guerra, A., Cappelini, L.T.D., Goering, A.W., Thomson, R.J., Metcalf, W.W., Kelleher, N.L., Barona-Gomez, F., Medema, M.H., 2020a. A computational framework to explore large-scale biosynthetic diversity. Nat Chem Biol 16, 60–68. https://doi.org/10.1038/s41589-019-0400-9

Navarro-Muñoz, J.C., Selem-Mojica, N., Mullowney, M.W., Kautsar, S.A., Tryon, J.H., Parkinson, E.I., De Los Santos, E.L.C., Yeong, M., Cruz-Morales, P., Abubucker, S., Roeters, A., Lokhorst, W., Fernandez-Guerra, A., Cappelini, L.T.D., Goering, A.W., Thomson, R.J., Metcalf, W.W., Kelleher, N.L., Barona-Gomez, F., Medema,

M.H., 2020b. A computational framework to explore large-scale biosynthetic diversity. Nat Chem Biol 16, 60–68. https://doi.org/10.1038/s41589-019-0400-9

Navarro-Muñoz, J.C., Selem-Mojica, N., Mullowney, M.W., Kautsar, S.A., Tryon, J.H., Parkinson, E.I., De Los Santos, E.L.C., Yeong, M., Cruz-Morales, P., Abubucker, S., Roeters, A., Lokhorst, W., Fernandez-Guerra, A., Cappelini, L.T.D., Goering, A.W., Thomson, R.J., Metcalf, W.W., Kelleher, N.L., Barona-Gomez, F., Medema, M.H., 2020c. A computational framework to explore large-scale biosynthetic diversity. Nat Chem Biol 16, 60–68. https://doi.org/10.1038/s41589-019-0400-9

Newman, D.J., Cragg, G.M., 2020. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. J. Nat. Prod. 83, 770–803. https://doi.org/10.1021/acs.jnatprod.9b01285

Nichols, D., Cahoon, N., Trakhtenberg, E.M., Pham, L., Mehta, A., Belanger, A., Kanigan, T., Lewis, K., Epstein, S.S., 2010. Use of Ichip for High-Throughput In Situ Cultivation of "Uncultivable" Microbial Species. Applied and Environmental Microbiology 76, 2445–2450. https://doi.org/10.1128/AEM.01754-09

Nikel, P.I., de Lorenzo, V., 2018. Pseudomonas putida as a functional chassis for industrial biocatalysis: from native biochemistry to trans-metabolism. Metabolic Engineering 50, 142–155. https://doi.org/10.1016/j.ymben.2018.05.005

Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H., Yura, T., 1998. Chaperone Coexpression Plasmids: Differential and Synergistic Roles of DnaK-DnaJ-GrpE and GroEL-GroES in Assisting Folding of an Allergen of Japanese Cedar Pollen, Cryj2, inEscherichia coli. Applied and Environmental Microbiology 64, 1694–1699. https://doi.org/10.1128/AEM.64.5.1694-1699.1998

Noor, A., Ziaf, K., Naveed, M., Khan, K.S., Ghani, M.A., Ahmad, I., Anwar, R., Siddiqui, M.H., Shakeel, A., Khan, A.I., 2023. L-Tryptophan-Dependent Auxin-Producing Plant-Growth-Promoting Bacteria Improve Seed Yield and Quality of Carrot by Altering the Umbel Order. Horticulturae 9, 954. https://doi.org/10.3390/horticulturae9090954

Nord, C., Bjerketorp, J., Levenfors, J.J., Cao, S., Strömstedt, A.A., Guss, B., Larsson, R., Hughes, D., Öberg, B., Broberg, A., 2020a. Isopedopeptins A–H: Cationic Cyclic Lipodepsipeptides from Pedobacter cryoconitis UP508 Targeting WHO Top-Priority Carbapenem-Resistant Bacteria. ACS Chem. Biol. 15, 2937–2944. https://doi.org/10.1021/acschembio.0c00568

Nord, C., Bjerketorp, J., Levenfors, J.J., Cao, S., Strömstedt, A.A., Guss, B., Larsson, R., Hughes, D., Öberg, B., Broberg, A., 2020b. Isopedopeptins A–H: Cationic Cyclic Lipodepsipeptides from Pedobacter cryoconitis UP508 Targeting WHO Top-Priority Carbapenem-Resistant Bacteria. ACS Chem Biol 15, 2937–2944. https://doi.org/10.1021/acschembio.0c00568

Nordberg, H., Cantor, M., Dusheyko, S., Hua, S., Poliakov, A., Shabalov, I., Smirnova, T., Grigoriev, I.V., Dubchak, I., 2014. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. Nucleic Acids Res 42, D26–D31. https://doi.org/10.1093/nar/gkt1069

Nothias, L.-F., Petras, D., Schmid, R., Dührkop, K., Rainer, J., Sarvepalli, A., Protsyuk, I., Ernst, M., Tsugawa, H., Fleischauer, M., Aicheler, F., Aksenov, A.A., Alka, O., Allard, P.-M., Barsch, A., Cachet, X., Caraballo-Rodriguez, A.M., Da Silva, R.R., Dang, T., Garg, N., Gauglitz, J.M., Gurevich, A., Isaac, G., Jarmusch, A.K., Kameník, Z., Kang, K.B., Kessler, N., Koester, I., Korf, A., Le Gouellec, A., Ludwig, M., Martin H., C., McCall, L.-I., McSayles, J., Meyer, S.W., Mohimani, H., Morsy, M., Moyne, O., Neumann, S., Neuweger, H., Nguyen, N.H., Nothias-Esposito, M., Paolini, J., Phelan, V.V., Pluskal, T., Quinn, R.A., Rogers, S., Shrestha, B., Tripathi, A., van der Hooft, J.J.J., Vargas, F., Weldon, K.C., Witting, M., Yang, H., Zhang, Z., Zubeil, F., Kohlbacher, O., Böcker, S., Alexandrov, T., Bandeira, N., Wang, M., Dorrestein, P.C., 2020. Feature-based molecular networking in the GNPS analysis environment. Nat Methods 17, 905–908. https://doi.org/10.1038/s41592-020-0933-6

Nouioui, I., Carro, L., García-López, M., Meier-Kolthoff, J.P., Woyke, T., Kyrpides, N.C., Pukall, R., Klenk, H.-P., Goodfellow, M., Göker, M., 2018. Genome-Based Taxonomic Classification of the Phylum Actinobacteria. Front. Microbiol. 9. https://doi.org/10.3389/fmicb.2018.02007

Oberpaul, M., Brinkmann, S., Marner, M., Mihajlovic, S., Leis, B., Patras, M.A., Hartwig, C., Vilcinskas, A., Hammann, P.E., Schäberle, T.F., Spohn, M., Glaeser, J., 2022. Combination of high-throughput microfluidics and FACS technologies to leverage the numbers game in natural product discovery. Microbial Biotechnology 15, 415–430. https://doi.org/10.1111/1751-7915.13872

Oberpaul, M., Zumkeller, C.M., Culver, T., Spohn, M., Mihajlovic, S., Leis, B., Glaeser, S.P., Plarre, R., McMahon, D.P., Hammann, P., Schäberle, T.F., Glaeser, J., Vilcinskas, A., 2020. High-throughput cultivation for the

selective isolation of Acidobacteria from termite nests. Front. Microbiol. 11. https://doi.org/10.3389/fmicb.2020.597628

O'Cathail, C., Ahamed, A., Burgin, J., Cummins, C., Devaraj, R., Gueye, K., Gupta, D., Gupta, V., Haseeb, M., Ihsan, M., Ivanov, E., Jayathilaka, S., Kadhirvelu, V., Kumar, M., Lathi, A., Leinonen, R., McKinnon, J., Meszaros, L., Pauperio, J., Pesant, S., Rahman, N., Rinck, G., Selvakumar, S., Suman, S., Sunthornyotin, Y., Ventouratou, M., Waheed, Z., Woollard, P., Yuan, D., Zyoud, A., Burdett, T., Cochrane, G., 2025. The European Nucleotide Archive in 2024. Nucleic Acids Res 53, D49–D55. https://doi.org/10.1093/nar/gkae975

Ochi, K., 2017. Insights into microbial cryptic gene activation and strain improvement: principle, application and technical aspects. J Antibiot 70, 25–40. https://doi.org/10.1038/ja.2016.82

Okada, B.K., Seyedsayamdost, M.R., 2017. Antibiotic dialogues: induction of silent biosynthetic gene clusters by exogenous small molecules. FEMS Microbiol Rev 41, 19–33. https://doi.org/10.1093/femsre/fuw035

Onaka, H., Tabata, H., Igarashi, Y., Sato, Y., Furumai, T., 2001. Goadsporin, a chemical substance which promotes secondary metabolism and morphogenesis in streptomycetes. I. Purification and characterization. J Antibiot (Tokyo) 54, 1036–1044. https://doi.org/10.7164/antibiotics.54.1036

Overkamp, W., Beilharz, K., Detert Oude Weme, R., Solopova, A., Karsens, H., Kovács, Á.T., Kok, J., Kuipers, O.P., Veening, J.-W., 2013. Benchmarking Various Green Fluorescent Protein Variants in Bacillus subtilis, Streptococcus pneumoniae, and Lactococcus lactis for Live Cell Imaging. Applied and Environmental Microbiology 79, 6481–6490. https://doi.org/10.1128/AEM.02033-13

Paddon, C.J., *et al.*, 2013. High-level semi-synthetic production of the potent antimalarial artemisinin. Nature 496, 528–532. https://doi.org/10.1038/nature12051

Pankratov, T.A., Dedysh, S.N., 2010. Granulicella paludicola gen. nov., sp. nov., Granulicella pectinivorans sp. nov., Granulicella aggregans sp. nov. and Granulicella rosea sp. nov., acidophilic, polymer-degrading acidobacteria from Sphagnum peat bogs. International Journal of Systematic and Evolutionary Microbiology, 60, 2951–2959. https://doi.org/10.1099/ijs.0.021824-0

Parks, D.H., Chuvochina, M., Rinke, C., Mussig, A.J., Chaumeil, P.-A., Hugenholtz, P., 2022. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Research 50, D785–D794. https://doi.org/10.1093/nar/gkab776

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., Hugenholtz, P., 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol 36, 996–1004. https://doi.org/10.1038/nbt.4229

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015a. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055. https://doi.org/10.1101/gr.186072.114

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015b. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055. https://doi.org/10.1101/gr.186072.114

Parsley, L.C., Linneman, J., Goode, A.M., Becklund, K., George, I., Goodman, R.M., Lopanik, N.B., Liles, M.R., 2011. Polyketide synthase pathways identified from a metagenomic library are derived from soil Acidobacteria. FEMS Microbiology Ecology 78, 176–187. https://doi.org/10.1111/j.1574-6941.2011.01122.x

Pascal Andreu, V., Augustijn, H.E., van den Berg, K., van der Hooft, J.J.J., Fischbach, M.A., Medema, M.H., 2021. BiG-MAP: an Automated Pipeline To Profile Metabolic Gene Cluster Abundance and Expression in Microbiomes. mSystems 6, 10.1128/msystems.00937-21. https://doi.org/10.1128/msystems.00937-21

Patron, N.J., Orzaez, D., Marillonnet, S., Warzecha, H., Matthewman, C., Youles, M., Raitskin, O., Leveau, A., Farré, G., Rogers, C., 2015. Standards for plant synthetic biology: a common syntax for exchange of DNA parts. New Phytologist 208, 13–19. https://doi.org/10.1111/nph.13532

Patz, S., Gautam, A., Becker, M., Ruppel, S., Rodríguez-Palenzuela, P., Huson, D., 2021. PLaBAse: A comprehensive web resource for analyzing the plant growth-promoting potential of plant-associated bacteria. https://doi.org/10.1101/2021.12.13.472471

Payne, D.J., Gwynn, M.N., Holmes, D.J., Pompliano, D.L., 2007. Drugs for bad bugs: confronting the challenges of antibacterial discovery. Nat Rev Drug Discov 6, 29–40. https://doi.org/10.1038/nrd2201

Peng, J., Lin, T., Wang, W., Xin, Z., Zhu, T., Gu, Q., Li, D., 2013. Antiviral Alkaloids Produced by the Mangrove-Derived Fungus Cladosporium sp. PJX-41. J. Nat. Prod. 76, 1133–1140. https://doi.org/10.1021/np400200k

Penn, K., Jenkins, C., Nett, M., Udwary, D.W., Gontang, E.A., McGlinchey, R.P., Foster, B., Lapidus, A., Podell, S., Allen, E.E., Moore, B.S., Jensen, P.R., 2009. Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. ISME J 3, 1193–1203. https://doi.org/10.1038/ismej.2009.58

Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E., Khosla, C., 2001. Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. Science 291, 1790–1792. https://doi.org/10.1126/science.1058092

Popp, P.F., Dotzler, M., Radeck, J., Bartels, J., Mascher, T., 2017. The Bacillus BioBrick Box 2.0: expanding the genetic toolbox for the standardized work with Bacillus subtilis. Sci Rep 7, 15058. https://doi.org/10.1038/s41598-017-15107-z

Pratama, A.A., van Elsas, J.D., 2018. The 'Neglected' Soil Virome – Potential Role and Impact. Trends in Microbiology 26, 649–662. https://doi.org/10.1016/j.tim.2017.12.004

Price-Whelan, A., Dietrich, L.E.P., Newman, D.K., 2006. Rethinking "secondary" metabolism: physiological roles for phenazine antibiotics. Nat Chem Biol 2, 71–78. https://doi.org/10.1038/nchembio764

Put, H., Gerstmans, H., Capelle, H.V., Fauvart, M., Michiels, J., Masschelein, J., 2024. Bacillus subtilis as a host for natural product discovery and engineering of biosynthetic gene clusters. Natural Product Reports 41, 1113–1151. https://doi.org/10.1039/D3NP00065F

Quaiser, A., Ochsenreiter, T., Lanz, C., Schuster, S.C., Treusch, A.H., Eck, J., Schleper, C., 2003. Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics. Molecular Microbiology 50, 563–575. https://doi.org/10.1046/j.1365-2958.2003.03707.x

Radeck, J., Kraft, K., Bartels, J., Cikovic, T., Dürr, F., Emenegger, J., Kelterborn, S., Sauer, C., Fritz, G., Gebhard, S., Mascher, T., 2013. The Bacillus BioBrick Box: generation and evaluation of essential genetic building blocks for standardized work with Bacillus subtilis. Journal of Biological Engineering 7, 29. https://doi.org/10.1186/1754-1611-7-29

Rao, X., Li, D., Su, Z., Nomura, C.T., Chen, S., Wang, Q., 2024. A smart RBS library and its prediction model for robust and accurate fine-tuning of gene expression in *Bacillus* species. Metabolic Engineering 81, 1–9. https://doi.org/10.1016/j.ymben.2023.11.002

Rawat, S.R., Männistö, M.K., Bromberg, Y., Häggblom, M.M., 2012. Comparative genomic and physiological analysis provides insights into the role of Acidobacteria in organic carbon utilization in Arctic tundra soils. FEMS Microbiology Ecology 82, 341–355. https://doi.org/10.1111/j.1574-6941.2012.01381.x

Reji, L., Zhang, X., 2022. Genome-Resolved Metagenomics Informs the Functional Ecology of Uncultured Acidobacteria in Redox Oscillated Sphagnum Peat. mSystems 7, e00055-22. https://doi.org/10.1128/msystems.00055-22

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.-F., Darling, A., Malfatti, S., Swan, B.K., Gies, E.A., Dodsworth, J.A., Hedlund, B.P., Tsiamis, G., Sievert, S.M., Liu, W.-T., Eisen, J.A., Hallam, S.J., Kyrpides, N.C., Stepanauskas, R., Rubin, E.M., Hugenholtz, P., Woyke, T., 2013. Insights into the phylogeny and coding potential of microbial dark matter. Nature 499, 431–437. https://doi.org/10.1038/nature12352

Ro, D.-K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J., Chang, M.C.Y., Withers, S.T., Shiba, Y., Sarpong, R., Keasling, J.D., 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440, 940–943. https://doi.org/10.1038/nature04640

Rohrlack, T., Christoffersen, K., Kaebernick, M., Neilan, B.A., 2004. Cyanobacterial protease inhibitor microviridin J causes a lethal molting disruption in Daphnia pulicaria. Applied and Environmental Microbiology 70, 5047–5050. https://doi.org/10.1128/AEM.70.8.5047-5050.2004

Romanello, M., Walawender, M., Hsu, S.-C., Moskeland, A., Palmeiro-Silva, Y., Scamman, D., Ali, Z., Ameli, N., Angelova, D., Ayeb-Karlsson, S., 2024. The 2024 report of the Lancet Countdown on health and climate

change: facing record-breaking threats from delayed action. The Lancet 404, 1847–1896. https://doi.org/10.1016/S0140-6736(24)01822-1

Rudolf, J.D., Chang, C.-Y., Ma, M., Shen, B., 2017. Cytochromes P450 for natural product biosynthesis in Streptomyces: sequence, structure, and function. Nat. Prod. Rep. 34, 1141–1172. https://doi.org/10.1039/C7NP00034K

Ruhl, I.A., Sheremet, A., Furgason, C.C., Krause, S., Bowers, R.M., Jarett, J.K., Tran, T.M., Grasby, S.E., Woyke, T., Dunfield, P.F., 2022. GAL08, an Uncultivated Group of Acidobacteria, Is a Dominant Bacterial Clade in a Neutral Hot Spring. Front Microbiol 12, 787651. https://doi.org/10.3389/fmicb.2021.787651

Rutledge, P.J., Challis, G.L., 2015. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. Nat Rev Microbiol 13, 509–523. https://doi.org/10.1038/nrmicro3496

Sack, E.L.W., van der Wielen, P.W.J.J., van der Kooij, D., 2011. Flavobacterium johnsoniae as a Model Organism for Characterizing Biopolymer Utilization in Oligotrophic Freshwater Environments. Applied and Environmental Microbiology 77, 6931–6938. https://doi.org/10.1128/AEM.00372-11

Sáenz, J.P., Grosser, D., Bradley, A.S., Lagny, T.J., Lavrynenko, O., Broda, M., Simons, K., 2015. Hopanoids as functional analogues of cholesterol in bacterial membranes. Proceedings of the National Academy of Sciences 112, 11971–11976. https://doi.org/10.1073/pnas.1515607112

Sait, M., Hugenholtz, P., Janssen, P.H., 2002. Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys. Environmental Microbiology 4, 654–666. https://doi.org/10.1046/j.1462-2920.2002.00352.x

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences 74, 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Santana, M.M., Rosa, A.P., Zamarreño, A.M., García-Mina, J.M., Rai, A., Cruz, C., 2022. Achromobacter xylosoxidans and Enteromorpha intestinalis Extract Improve Tomato Growth under Salt Stress. Agronomy 12, 934. https://doi.org/10.3390/agronomy12040934

Santibáñez, R., Lara, F., Barros, T.M., Mardones, E., Cuadra, F., Thomson, P., 2022. Ocular Microbiome in a Group of Clinically Healthy Horses. Animals (Basel) 12, 943. https://doi.org/10.3390/ani12080943

Sapountzis, P., de Verges, J., Rousk, K., Cilliers, M., Vorster, B.J., Poulsen, M., 2016. Potential for Nitrogen Fixation in the Fungus-Growing Termite Symbiosis. Front Microbiol 7, 1993. https://doi.org/10.3389/fmicb.2016.01993

Sarrion-Perdigones, A., Vazquez-Vilar, M., Palací, J., Castelijns, B., Forment, J., Ziarsolo, P., Blanca, J., Granell, A., Orzaez, D., 2013. GoldenBraid 2.0: a comprehensive DNA assembly framework for plant synthetic biology. Plant Physiology 162, 1618–1631. https://doi.org/10.1104/pp.113.217661

Schaenzer, A.J., Wang, W., Hackenberger, D., Wright, G.D., 2024. Identification and characterization of the siderochelin biosynthetic gene cluster via coculture. mBio 15, e01871-24. https://doi.org/10.1128/mbio.01871-24

Schatz, A., Bugie, E., Waksman, S.A., 1944. Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. Proceedings of the Society for Experimental Biology and Medicine 55, 66–69. https://doi.org/10.3181/00379727-55-14461

Schatz, A., Waksman, S.A., 1944. Effect of streptomycin and other antibiotic substances upon Mycobacterium tuberculosis and related organisms. Proceedings of the Society for Experimental Biology and Medicine 57, 244–248. https://doi.org/10.3181/00379727-57-14769

Schmidt, C.S., Mrnka, L., Frantík, T., Lovecká, P., Vosátka, M., 2018. Plant growth promotion of Miscanthus × giganteus by endophytic bacteria and fungi on non-polluted and polluted soils. World J Microbiol Biotechnol 34, 48. https://doi.org/10.1007/s11274-018-2426-7

Schmidt, M., Lee, N., Zhan, C., Roberts, J.B., Nava, A.A., Keiser, L.S., Vilchez, A.A., Chen, Y., Petzold, C.J., Haushalter, R.W., Blank, L.M., Keasling, J.D., 2023. Maximizing Heterologous Expression of Engineered Type I Polyketide Synthases: Investigating Codon Optimization Strategies. ACS Synth. Biol. 12, 3366–3380. https://doi.org/10.1021/acssynbio.3c00367

Schmidt, S., Kildgaard, S., Guo, H., Beemelmanns, C., Poulsen, M., 2021. The chemical ecology of the fungus-farming termite symbiosis. Natural Product Reports 39, 231–248. https://doi.org/10.1039/d1np00022e

Schober, I., Koblitz, J., Sardà Carbasse, J., Ebeling, C., Schmidt, M.L., Podstawka, A., Gupta, R., Ilangovan, V., Chamanara, J., Overmann, J., Reimer, L.C., 2024. BacDive in 2025: the core database for prokaryotic strain data. Nucleic Acids Res 53, D748–D756. https://doi.org/10.1093/nar/gkae959

Schöner, T.A., Fuchs, S.W., Schönau, C., Bode, H.B., 2014. Initiation of the flexirubin biosynthesis in Chitinophaga pinensis. Microb Biotechnol 7, 232–241. https://doi.org/10.1111/1751-7915.12110

Schorn, M.A., Verhoeven, S., Ridder, L., Huber, F., Acharya, D.D., Aksenov, A.A., Aleti, G., Moghaddam, J.A., Aron, A.T., Aziz, S., Bauermeister, A., Bauman, K.D., Baunach, M., Beemelmanns, C., Beman, J.M., Berlanga-Clavero, M.V., Blacutt, A.A., Bode, H.B., Boullie, A., Brejnrod, A., Bugni, T.S., Calteau, A., Cao, L., Carrión, V.J., Castelo-Branco, R., Chanana, S., Chase, A.B., Chevrette, M.G., Costa-Lotufo, L.V., Crawford, J.M., Currie, C.R., Cuypers, B., Dang, T., de Rond, T., Demko, A.M., Dittmann, E., Du, C., Drozd, C., Dujardin, J.-C., Dutton, R.J., Edlund, A., Fewer, D.P., Garg, N., Gauglitz, J.M., Gentry, E.C., Gerwick, L., Glukhov, E., Gross, H., Gugger, M., Guillén Matus, D.G., Helfrich, E.J.N., Hempel, B.-F., Hur, J.-S., Iorio, M., Jensen, P.R., Kang, K.B., Kaysser, L., Kelleher, N.L., Kim, C.S., Kim, K.H., Koester, I., König, G.M., Leao, T., Lee, S.R., Lee, Y.-Y., Li, X., Little, J.C., Maloney, K.N., Männle, D., Martin H., C., McAvoy, A.C., Metcalf, W.W., Mohimani, H., Molina-Santiago, C., Moore, B.S., Mullowney, M.W., Muskat, M., Nothias, L.-F., O'Neill, E.C., Parkinson, E.I., Petras, D., Piel, J., Pierce, E.C., Pires, K., Reher, R., Romero, D., Roper, M.C., Rust, M., Saad, H., Saenz, C., Sanchez, L.M., Sørensen, S.J., Sosio, M., Süssmuth, R.D., Sweeney, D., Tahlan, K., Thomson, R.J., Tobias, N.J., Trindade-Silva, A.E., van Wezel, G.P., Wang, M., Weldon, K.C., Zhang, F., Ziemert, N., Duncan, K.R., Crüsemann, M., Rogers, S., Dorrestein, P.C., Medema, M.H., van der Hooft, J.J.J., 2021. A community resource for paired genomic and metabolomic data mining. Nat Chem Biol 17, 363–368. https://doi.org/10.1038/s41589-020-00724-z

Schwecke, T., Aparicio, J.F., Molnár, I., König, A., Khaw, L.E., Haydock, S.F., Oliynyk, M., Caffrey, P., Cortés, J., Lester, J.B., 1995. The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. Proceedings of the National Academy of Sciences 92, 7839–7843. https://doi.org/10.1073/pnas.92.17.7839

Schwengers, O., Jelonek, L., Dieckmann, M.A., Beyvers, S., Blom, J., Goesmann, A. 2021, n.d. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. Microbial Genomics 7, 000685. https://doi.org/10.1099/mgen.0.000685

S. Dickschat, J., 2010. Quorum sensing and bacterial biofilms. Natural Product Reports 27, 343–369. https://doi.org/10.1039/B804469B

Seca, A.M.L., Pinto, D.C.G.A., 2019. Biological Potential and Medical Use of Secondary Metabolites. Medicines 6, 66. https://doi.org/10.3390/medicines6020066

Sen, R., Ishak, H.D., Estrada, D., Dowd, S.E., Hong, E., Mueller, U.G., 2009. Generalized antifungal activity and 454-screening of Pseudonocardia and Amycolatopsis bacteria in nests of fungus-growing ants. Proceedings of the National Academy of Sciences 106, 17805–17810. https://doi.org/10.1073/pnas.0904827106

Seo, H., Kim, S., Sagong, H.-Y., Son, H.F., Jin, K.S., Kim, I.-K., Kim, K.-J., 2016. Structural basis for cytokinin production by LOG from Corynebacterium glutamicum. Sci Rep 6, 31390. https://doi.org/10.1038/srep31390

Severi, E., Thomas, G.H., 2019. Antibiotic export: transporters involved in the final step of natural product production. Microbiology 165, 805–818. https://doi.org/10.1099/mic.0.000794

Seyedsayamdost, M.R., 2014. High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. Proceedings of the National Academy of Sciences 111, 7266–7271. https://doi.org/10.1073/pnas.1400019111

Sheng, Y., Qiu, S., Deng, Y., Zeng, B., 2025. Recent Advances in Heterologous Protein Expression and Natural Product Synthesis by Aspergillus. J Fungi (Basel) 11, 534. https://doi.org/10.3390/jof11070534

Shenvi, R.A., 2024. Natural Product Synthesis in the 21st Century: Beyond the Mountain Top. ACS Cent. Sci. 10, 519–528. https://doi.org/10.1021/acscentsci.3c01518

Shetty, R.P., Endy, D., Knight, T.F., 2008. Engineering BioBrick vectors from BioBrick parts. Journal of Biological Engineering 2, 5. https://doi.org/10.1186/1754-1611-2-5

Shizuya, H., *et al.*, 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using BAC vectors. Proceedings of the National Academy of Sciences USA 89, 8794–8797. https://doi.org/10.1073/pnas.89.18.8794

Silva, S.G., Paula, P., da Silva, J.P., Mil-Homens, D., Teixeira, M.C., Fialho, A.M., Costa, R., Keller-Costa, T., 2022. Insights into the Antimicrobial Activities and Metabolomes of Aquimarina (Flavobacteriaceae, Bacteroidetes) Species from the Rare Marine Biosphere. Marine Drugs 20, 423. https://doi.org/10.3390/md20070423

Singh, P.D., Johnson, J.H., Ward, P.C., Wells, J.S., Trejo, W.H., Sykes, R.B., 1983. SQ 28,332, a new monobactam produced by a Flexibacter sp. Taxonomy, fermentation, isolation, structure determination and biological properties. The Journal of Antibiotics (Tokyo) 36, 1245–1251. https://doi.org/10.7164/antibiotics.36.1245

Singh, R.P., Shelke, G.M., Kumar, A., Jha, P.N., 2015. Biochemistry and genetics of ACC deaminase: a weapon to "stress ethylene" produced in plants. Front Microbiol 6, 937. https://doi.org/10.3389/fmicb.2015.00937

Sinninghe Damsté, J.S., Rijpstra, W.I.C., Hopmans, E.C., Foesel, B.U., Wüst, P.K., Overmann, J., Tank, M., Bryant, D.A., Dunfield, P.F., Houghton, K., Stott, M.B., 2014. Ether- and Ester-Bound iso-Diabolic Acid and Other Lipids in Members of Acidobacteria Subdivision 4. Appl Environ Microbiol 80, 5207–5218. https://doi.org/10.1128/AEM.01066-14

Sinninghe Damsté, J.S., Rijpstra, W.I.C., Hopmans, E.C., Weijers, J.W.H., Foesel, B.U., Overmann, J., Dedysh, S.N., 2011. 13,16-Dimethyl Octacosanedioic Acid (iso-Diabolic Acid), a Common Membrane-Spanning Lipid of Acidobacteria Subdivisions 1 and 3. Applied and Environmental Microbiology 77, 4147–4154. https://doi.org/10.1128/AEM.00466-11

Siodłak, D., 2015. α,β-Dehydroamino acids in naturally occurring peptides. Amino Acids 47, 1–17. https://doi.org/10.1007/s00726-014-1846-4

Sipes, K., Buongiorno, J., Steen, A.D., Abramov, A.A., Abuah, C., Peters, S.L., Gianonne, R.J., Hettich, R.L., Boike, J., Garcia, S.L., Vishnivetskaya, T.A., Lloyd, K.G., 2024. Depth-specific distribution of bacterial MAGs in permafrost active layer in Ny Ålesund, Svalbard (79°N). Systematic and Applied Microbiology 47, 126544. https://doi.org/10.1016/j.syapm.2024.126544

Skinnider, M.A., Johnston, C.W., Gunabalasingam, M., Merwin, N.J., Kieliszek, A.M., MacLellan, R.J., Li, H., Ranieri, M.R.M., Webster, A.L.H., Cao, M.P.T., Pfeifle, A., Spencer, N., To, Q.H., Wallace, D.P., Dejong, C.A., Magarvey, N.A., 2020. Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. Nat Commun 11, 6058. https://doi.org/10.1038/s41467-020-19986-1

Smanski, M.J., Zhou, H., Claesen, J., Shen, B., Fischbach, M.A., Voigt, C.A., 2016. Synthetic biology to access and expand nature's chemical diversity. Nat Rev Microbiol 14, 135–149. https://doi.org/10.1038/nrmicro.2015.24

Smith, K., 2013. A Brief History of NCBI's Formation and Growth, in: The NCBI Handbook [Internet]. 2nd Edition. National Center for Biotechnology Information (US).

Somanadhan, B., Kotturi, S.R., Yan Leong, C., Glover, R.P., Huang, Y., Flotow, H., Buss, A.D., Lear, M.J., Butler, M.S., 2013. Isolation and synthesis of falcitidin, a novel myxobacterial-derived acyltetrapeptide with activity against the malaria target falcipain-2. J Antibiot 66, 259–264. https://doi.org/10.1038/ja.2012.123

Spaepen, S., Vanderleyden, J., 2011. Auxin and Plant-Microbe Interactions. Cold Spring Harb Perspect Biol 3, a001438. https://doi.org/10.1101/cshperspect.a001438

Spaepen, S., Vanderleyden, J., Remans, R., 2007. Indole-3-acetic acid in microbial and microorganism-plant signaling. FEMS Microbiology Reviews 31, 425–448. https://doi.org/10.1111/j.1574-6976.2007.00072.x

Staff, N., 2020. Recent enhancements in Genome Workbench version 3.4.1. NCBI Insights. URL https://ncbiinsights.ncbi.nlm.nih.gov/2020/08/06/gbench-3-4-1/ (accessed 10.3.25).

Starcevic, A., Zucko, J., Simunkovic, J., Long, P.F., Cullum, J., Hranueli, D., 2008. ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. Nucleic Acids Research 36, 6882–6892. https://doi.org/10.1093/nar/gkn685

Steele, A.D., Teijaro, C.N., Yang, D., Shen, B., 2019. Leveraging a large microbial strain collection for natural product discovery. Journal of Biological Chemistry 294, 16567–16576. https://doi.org/10.1074/jbc.REV119.006514

Stein, T., Heinzmann, S., Solovieva, I., Entian, K.-D., 2003. Function of *Lactococcus lactis* Nisin Immunity Genes *nisI* and *nisFEG* after Coordinated Expression in the Surrogate Host *Bacillus subtilis* *. Journal of Biological Chemistry 278, 89–94. https://doi.org/10.1074/jbc.M207237200

Steinmetz, H., Gerth, K., Jansen, R., Schläger, N., Dehn, R., Reinecke, S., Kirschning, A., Müller, R., 2011. Elansolid A, a Unique Macrolide Antibiotic from Chitinophaga sancti Isolated as Two Stable Atropisomers. Angewandte Chemie International Edition 50, 532–536. https://doi.org/10.1002/anie.201005226

Stevens, D.C., Henry, M.R., Murphy, K.A., Boddy, C.N., 2010. Heterologous Expression of the Oxytetracycline Biosynthetic Pathway in Myxococcus xanthus. Applied and Environmental Microbiology 76, 2681–2683. https://doi.org/10.1128/AEM.02841-09

Stratton, C.F., Newman, D.J., Tan, D.S., 2015. Cheminformatic comparison of approved drugs from natural product versus synthetic origins. Bioorganic & Medicinal Chemistry Letters, Recent Advances in Medicinal Chemistry and Chemical Biology 25, 4802–4807. https://doi.org/10.1016/j.bmcl.2015.07.014

Stravs, M.A., Dührkop, K., Böcker, S., Zamboni, N., 2022. MSNovelist: de novo structure generation from mass spectra. Nat Methods 19, 865–870. https://doi.org/10.1038/s41592-022-01486-3

Sugawara, H., Ueda, N., Kojima, M., Makita, N., Yamaya, T., Sakakibara, H., 2008. Structural insight into the reaction mechanism and evolution of cytokinin biosynthesis. Proc Natl Acad Sci U S A 105, 2734–2739. https://doi.org/10.1073/pnas.0707374105

Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A., Fan, T.W.-M., Fiehn, O., Goodacre, R., Griffin, J.L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A.N., Lindon, J.C., Marriott, P., Nicholls, A.W., Reily, M.D., Thaden, J.J., Viant, M.R., 2007. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics 3, 211–221. https://doi.org/10.1007/s11306-007-0082-2

Sundar, S., Chakravarty, J., Agarwal, D., Rai, M., Murray, H.W., 2010. Single-Dose Liposomal Amphotericin B for Visceral Leishmaniasis in India. New England Journal of Medicine 362, 504–512. https://doi.org/10.1056/NEJMoa0903627

Süssmuth, R.D., Mainz, A., 2017. Nonribosomal Peptide Synthesis—Principles and Prospects. Angewandte Chemie International Edition 56, 3770–3821. https://doi.org/10.1002/anie.201609079

Tao, L., Yao, H., Kasai, H., Misawa, N., Cheng, Q., 2006. A carotenoid synthesis gene cluster from Algoriphagus sp. KK10202C with a novel fusion-type lycopene beta-cyclase gene. Mol Genet Genomics 276, 79–86. https://doi.org/10.1007/s00438-006-0121-0

Tao, Q., Zhang, H.-B., 1998. Cloning and stable maintenance of DNA fragments over 300 kb in Escherichia coli with conventional plasmid-based vectors. Nucleic Acids Res 26, 4901–4909. https://doi.org/10.1093/nar/26.21.4901

Teng, W., Liao, B., Chen, M., Shu, W., 2022. Genomic Legacies of Ancient Adaptation Illuminate GC-Content Evolution in Bacteria. Microbiology Spectrum 11, e02145-22. https://doi.org/10.1128/spectrum.02145-22

Terlouw, B.R., Blin, K., Navarro-Muñoz, J.C., Avalon, N.E., Chevrette, M.G., Egbert, S., Lee, S., Meijer, D., Recchia, M.J.J., Reitz, Z.L., van Santen, J.A., Selem-Mojica, N., Tørring, T., Zaroubi, L., Alanjary, M., Aleti, G., Aguilar, C., Al-Salihi, S.A.A., Augustijn, H.E., Avelar-Rivas, J.A., Avitia-Domínguez, L.A., Barona-Gómez, F., Bernaldo-Agüero, J., Bielinski, V.A., Biermann, F., Booth, T.J., Carrion Bravo, V.J., Castelo-Branco, R., Chagas, F.O., Cruz-Morales, P., Du, C., Duncan, K.R., Gavriilidou, A., Gayrard, D., Gutiérrez-García, K., Haslinger, K., Helfrich, E.J.N., van der Hooft, J.J.J., Jati, A.P., Kalkreuter, E., Kalyvas, N., Kang, K.B., Kautsar, S., Kim, W., Kunjapur, A.M., Li, Y.-X., Lin, G.-M., Loureiro, C., Louwen, J.J.R., Louwen, N.L.L., Lund, G., Parra, J., Philmus, B., Pourmohsenin, B., Pronk, L.J.U., Rego, A., Rex, D.A.B., Robinson, S., Rosas-Becerra, L.R., Roxborough, E.T., Schorn, M.A., Scobie, D.J., Singh, K.S., Sokolova, N., Tang, X., Udwary, D., Vigneshwari, A., Vind, K., Vromans, S.P.J.M., Waschulin, V., Williams, S.E., Winter, J.M., Witte, T.E., Xie, H., Yang, D., Yu, J., Zdouc, M., Zhong, Z., Collemare, J., Linington, R.G., Weber, T., Medema, M.H., 2023. MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. Nucleic Acids Research 51, D603–D610. https://doi.org/10.1093/nar/gkac1049

Thermo Fisher Scientific, 2025. Gateway Recombination and Seamless Cloning Support — Getting Started [WWW Document]. URL https://www.thermofisher.com/us/en/home/technical-resources/technical-reference-library/cloning-technical-support-center/recombination-based-cloning/recombination-based-cloning-getting-started.html

Traxler, M.F., Watrous, J.D., Alexandrov, T., Dorrestein, P.C., Kolter, R., 2013. Interspecies Interactions Stimulate Diversification of the Streptomyces coelicolor Secreted Metabolome. mBio 4, 10.1128/mbio.00459-13. https://doi.org/10.1128/mbio.00459-13

Troeschel, S.C., Thies, S., Link, O., Real, C.I., Knops, K., Wilhelm, S., Rosenau, F., Jaeger, K.-E., 2012. Novel broad host range shuttle vectors for expression in Escherichia coli, Bacillus subtilis and Pseudomonas putida. J Biotechnol 161, 71–79. https://doi.org/10.1016/j.jbiotec.2012.02.020

Trojan, D., García-Robledo, E., Hausmann, B., Revsbech, N.P., Woebken, D., Eichorst, S.A., 2024. A respiro-fermentative strategy to survive nanoxia in Acidobacterium capsulatum. FEMS Microbiol Ecol 100, fiae152. https://doi.org/10.1093/femsec/fiae152

Tsutsui, H., Matsubara, K., 1981. Replication Control and Switch-Off Function as Observed with a Mini-F Factor Plasmid. Journal of Bacteriology 147, 509–516. https://doi.org/10.1128/jb.147.2.509-516.1981

T. Walsh, C., 2023. Tailoring enzyme strategies and functional groups in biosynthetic pathways. Natural Product Reports 40, 326–386. https://doi.org/10.1039/D2NP00048B

Ueda, S., Kitani, S., Namba, T., Arai, M., Ikeda, H., Nihira, T., 2018. Engineered production of kitasetalic acid, a new tetrahydro-β-carboline with the ability to suppress glucose-regulated protein synthesis. J Antibiot 71, 854–861. https://doi.org/10.1038/s41429-018-0074-7

Ueoka, R., Ise, Y., Ohtsuka, S., Okada, S., Yamori, T., Matsunaga, S., 2010. Yaku'amides A and B, Cytotoxic Linear Peptides Rich in Dehydroamino Acids from the Marine Sponge Ceratopsion sp. J. Am. Chem. Soc. 132, 17692–17694. https://doi.org/10.1021/ja109275z

Ullmann, I.F., Nygaard, A.B., Tunsjø, H.S., Charnock, C., 2020a. Whole genome sequencing and antibiotic diffusion assays, provide new insight on drug resistance in the genus Pedobacter. FEMS Microbiology Ecology 96, fiaa088. https://doi.org/10.1093/femsec/fiaa088

Ullmann, I.F., Nygaard, A.B., Tunsjø, H.S., Charnock, C., 2020b. Whole genome sequencing and antibiotic diffusion assays, provide new insight on drug resistance in the genus Pedobacter. FEMS Microbiol Ecol 96, fiaa088. https://doi.org/10.1093/femsec/fiaa088

van Santen, J.A., Poynton, E.F., Iskakova, D., McMann, E., Alsup, T.A., Clark, T.N., Fergusson, C.H., Fewer, D.P., Hughes, A.H., McCadden, C.A., Parra, J., Soldatou, S., Rudolf, J.D., Janssen, E.M.-L., Duncan, K.R., Linington, R.G., 2022. The Natural Products Atlas 2.0: a database of microbially-derived natural products. Nucleic Acids Res 50, D1317–D1323. https://doi.org/10.1093/nar/gkab941

Vansuyt, G., Robin, A., Briat, J.-F., Curie, C., Lemanceau, P., 2007. Iron acquisition from Fe-pyoverdine by Arabidopsis thaliana. Molecular Plant-Microbe Interactions 20, 441–447. https://doi.org/10.1094/MPMI-20-4-0441

Vesala, R., Ghosh, S., Murphy, R., Poulsen, M., Schmidt, S., Mwangala, L., Wooller, M.J., Rikkinen, J., Arppe, L., 2025. Nutritional dynamics in fungus-growing termites: Integration of biological nitrogen fixation and microbial amino acids. Functional Ecology 39, 2481–2495. https://doi.org/10.1111/1365-2435.70108

Viana, A.T., Caetano, T., Covas, C., Santos, T., Mendo, S., 2018a. Environmental superbugs: The case study of Pedobacter spp. Environmental Pollution 241, 1048–1055. https://doi.org/10.1016/j.envpol.2018.06.047

Viana, A.T., Caetano, T., Covas, C., Santos, T., Mendo, S., 2018b. Environmental superbugs: The case study of Pedobacter spp. Environ Pollut 241, 1048–1055. https://doi.org/10.1016/j.envpol.2018.06.047

Waksman, S.A., Schatz, A., Reynolds, D.M., 1946. Production of antibiotic substances by actinomycetes. Annals of the New York Academy of Sciences 48, 73–86. https://doi.org/10.1111/j.1749-6632.1946.tb31756.x

Walker, A.S., Clardy, J., 2021. A Machine Learning Bioinformatics Method to Predict Biological Activity from Biosynthetic Gene Clusters. J. Chem. Inf. Model. 61, 2560–2571. https://doi.org/10.1021/acs.jcim.0c01304

Walsh, C.T., Fischbach, M.A., 2010. Natural Products Version 2.0: Connecting Genes to Molecules. J. Am. Chem. Soc. 132, 2469–2493. https://doi.org/10.1021/ja909118a

Wan, J., Ma, N., Yuan, H., 2023. Recent advances in the direct cloning of large natural product biosynthetic gene clusters. Engineering Microbiology 3, 100085. https://doi.org/10.1016/j.engmic.2023.100085

Wang, B., Guo, F., Dong, S.-H., Zhao, H., 2019. Activation of silent biosynthetic gene clusters using transcription factor decoys. Nat Chem Biol 15, 111–114. https://doi.org/10.1038/s41589-018-0187-0

Wang, G., Zhao, Z., Ke, J., Engel, Y., Shi, Y.-M., Robinson, D., Bingol, K., Zhang, Z., Bowen, B., Louie, K., Wang, B., Evans, R., Miyamoto, Y., Cheng, K., Kosina, S., De Raad, M., Silva, L., Luhrs, A., Lubbe, A., Hoyt, D.W., Francavilla, C., Otani, H., Deutsch, S., Washton, N.M., Rubin, E.M., Mouncey, N.J., Visel, A., Northen, T., Cheng, J.-F., Bode, H.B., Yoshikuni, Y., 2019. CRAGE enables rapid activation of biosynthetic gene clusters in undomesticated bacteria. Nat Microbiol 4, 2498–2510. https://doi.org/10.1038/s41564-019-0573-8

Wang, H., Fewer, D.P., Holm, L., Rouhiainen, L., Sivonen, K., 2014. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. Proceedings of the National Academy of Sciences 111, 9259–9264. https://doi.org/10.1073/pnas.1401734111

Wang, J., Chen, M., Lv, Y., Jiang, Y., Qiu, L., 2016. Edaphobacter dinghuensis sp. nov., an acidobacterium isolated from lower subtropical forest soil. International Journal of Systematic and Evolutionary Microbiology, 66, 276–282. https://doi.org/10.1099/ijsem.0.000710

Wang, S., Wu, K., Tang, Y.-J., Deng, H., 2024. Dehydroamino acid residues in bioactive natural products. Nat. Prod. Rep. 41, 273–297. https://doi.org/10.1039/D3NP00041A

Ward, N.L., Challacombe, J.F., Janssen, P.H., Henrissat, B., Coutinho, P.M., Wu, M., Xie, G., Haft, D.H., Sait, M., Badger, J., Barabote, R.D., Bradley, B., Brettin, T.S., Brinkac, L.M., Bruce, D., Creasy, T., Daugherty, S.C., Davidsen, T.M., DeBoy, R.T., Detter, J.C., Dodson, R.J., Durkin, A.S., Ganapathy, A., Gwinn-Giglio, M., Han, C.S., Khouri, H., Kiss, H., Kothari, S.P., Madupu, R., Nelson, K.E., Nelson, W.C., Paulsen, I., Penn, K., Ren, Q., Rosovitz, M.J., Selengut, J.D., Shrivastava, S., Sullivan, S.A., Tapia, R., Thompson, L.S., Watkins, K.L., Yang, Q., Yu, C., Zafar, N., Zhou, L., Kuske, C.R., 2009. Three Genomes from the Phylum Acidobacteria Provide Insight into the Lifestyles of These Microorganisms in Soils. Appl. Environ. Microbiol. 75, 2046–2056. https://doi.org/10.1128/AEM.02294-08

Waschulin, V., Borsetto, C., James, R., Newsham, K.K., Donadio, S., Corre, C., Wellington, E., 2022. Biosynthetic potential of uncultured Antarctic soil bacteria revealed through long-read metagenomic sequencing. ISME J 16, 101–111. https://doi.org/10.1038/s41396-021-01052-3

Weber, E., Engler, C., Gruetzner, R., Werner, S., Marillonnet, S., 2011. A Modular Cloning (MoClo) system for standardized assembly of multigene constructs. PLOS ONE 6, e16765. https://doi.org/10.1371/journal.pone.0016765

Weber, T., Rausch, C., Lopez, P., Hoof, I., Gaykova, V., Huson, D.H., Wohlleben, W., 2009. CLUSEAN: a computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. Journal of Biotechnology 140, 13–17. https://doi.org/10.1016/j.jbiotec.2009.01.007

Wei, X., Moreno-Hagelsieb, G., Glick, B.R., Doxey, A.C., 2023. Comparative analysis of adenylate isopentenyl transferase genes in plant growth-promoting bacteria and plant pathogenic bacteria. Heliyon 9, e13955. https://doi.org/10.1016/j.heliyon.2023.e13955

Weihrauch, D., O'Donnell, M.J., 2021. Mechanisms of nitrogen excretion in insects. Current Opinion in Insect Science, Global change biology. Molecular physiology. October (2021) 47, 25–30. https://doi.org/10.1016/j.cois.2021.02.007

Welander, P.V., Hunter, R.C., Zhang, L., Sessions, A.L., Summons, R.E., Newman, D.K., 2009. Hopanoids Play a Role in Membrane Integrity and pH Homeostasis in Rhodopseudomonas palustris TIE-1. J Bacteriol 191, 6145–6156. https://doi.org/10.1128/JB.00460-09

Wendt-Pienkowski, E., Huang, Y., Zhang, J., Li, B., Jiang, H., Kwon, H., Hutchinson, C.R., Shen, B., 2005. Cloning, Sequencing, Analysis, and Heterologous Expression of the Fredericamycin Biosynthetic Gene Cluster from Streptomyces griseus. J. Am. Chem. Soc. 127, 16442–16452. https://doi.org/10.1021/ja054376u

Wenzel, S.C., Gross, F., Zhang, Y., Fu, J., Stewart, A.F., Müller, R., 2005. Heterologous expression of a myxobacterial natural products assembly line in pseudomonads via red/ET recombineering. Chem. Biol. 12, 349–356. https://doi.org/10.1016/j.chembiol.2004.12.012

Wick, R.R., Judd, L.M., Gorrie, C.L., Holt, K.E., 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLOS Computational Biology 13, e1005595. https://doi.org/10.1371/journal.pcbi.1005595

Willey, J.M., Willems, A., Kodani, S., Nodwell, J.R., 2006. Morphogenetic surfactants and their role in the formation of aerial hyphae in Streptomyces coelicolor. Molecular Microbiology 59, 731–742. https://doi.org/10.1111/j.1365-2958.2005.05018.x

Wohlleben, W., Mast, Y., Stegmann, E., Ziemert, N., 2016. Antibiotic drug discovery. Microbial Biotechnology 9, 541–548. https://doi.org/10.1111/1751-7915.12388

Wong, C.M.V.L., Tam, H.K., Ng, W.M., Boo, S.Y., González, M., 2013. Characterisation of a cryptic plasmid from an Antarctic bacterium *Pedobacter cryoconitis* strain BG5. Plasmid 69, 186–193. https://doi.org/10.1016/j.plasmid.2012.12.002

Wong, H.L., Bulzu, P.-A., Ghai, R., Chiriac, M.-C., Salcher, M.M., 2024. Ubiquitous genome streamlined Acidobacteriota in freshwater environments. ISME Commun 4, ycae124. https://doi.org/10.1093/ismeco/ycae124

World Health Organization, 2025. Global Antimicrobial Resistance and Use Surveillance System (GLASS) report. Antibiotic use data for 2022 (No. 978-92-4- 010812–7). World Health Organization, Geneva.

World Health Organization, 2024. 2023 Antibacterial agents in clinical and preclinical development: an overview and analysis (No. 978-92-4- 009400– 0). World Health Organization, Geneva.

World Health Organization, 2023. Elimination of human onchocerciasis: progress report, 2022–2023. Weekly Epidemiological Record 98, 572–582.

World Health Organization, 2020. Ending the neglect to attain the Sustainable Development Goals: a road map for neglected tropical diseases 2021–2030 (No. 978-92-4- 001035–2). World Health Organization, Geneva.

Xie, F., Zhao, H., Liu, J., Yang, X., Neuber, M., Agrawal, A.A., Kaur, A., Herrmann, J., Kalinina, O.V., Wei, X., Müller, R., Fu, C., 2024. Autologous DNA mobilization and multiplication expedite natural products discovery from bacteria. Science 386, eabq7333. https://doi.org/10.1126/science.abq7333

Xu, M., Wang, W., Waglechner, N., Culp, E.J., Guitor, A.K., Wright, G.D., 2020. GPAHex-A synthetic biology platform for Type IV–V glycopeptide antibiotic production and discovery. Nat Commun 11, 5232. https://doi.org/10.1038/s41467-020-19138-5

Xu, T., Shi, L., Zhang, Y., Wang, K., Yang, Z., Ke, S., 2019. Synthesis and biological evaluation of marine alkaloid-oriented β-carboline analogues. European Journal of Medicinal Chemistry 168, 293–300. https://doi.org/10.1016/j.ejmech.2019.02.060

Xu, Y., Tian, X.-P., Liu, Y.-J., Li, J., Kim, C.-J., Yin, H., Li, W.-J., Zhang, S. 2013, n.d. Sinomicrobium oceani gen. nov., sp. nov., a member of the family Flavobacteriaceae isolated from marine sediment. International Journal of Systematic and Evolutionary Microbiology 63, 1045–1050. https://doi.org/10.1099/ijs.0.041889-0

Xu, Y., Willems, A., Au-yeung, C., Tahlan, K., Nodwell, J.R., 2012. A Two-Step Mechanism for the Activation of Actinorhodin Export and Resistance in Streptomyces coelicolor. mBio 3, 10.1128/mbio.00191-12. https://doi.org/10.1128/mbio.00191-12

Xu, Z., Wang, Y., Chater, K.F., Ou, H.-Y., Xu, H.H., Deng, Z., Tao, M., 2017. Large-Scale Transposition Mutagenesis of Streptomyces coelicolor Identifies Hundreds of Genes Influencing Antibiotic Biosynthesis. Applied and Environmental Microbiology 83, e02889-16. https://doi.org/10.1128/AEM.02889-16

Yamanaka, K., Reynolds, K.A., Kersten, R.D., Ryan, K.S., Gonzalez, D.J., Nizet, V., Dorrestein, P.C., Moore, B.S., 2014. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. Proceedings of the National Academy of Sciences 111, 1957–1962. https://doi.org/10.1073/pnas.1319584111

Yan, D., Zhou, M., Adduri, A., Zhuang, Y., Guler, M., Liu, S., Shin, H., Kovach, T., Oh, G., Liu, X., Deng, Y., Wang, X., Cao, L., Sherman, D.H., Schultz, P.J., Kersten, R.D., Clement, J.A., Tripathi, A., Behsaz, B., Mohimani, H., 2024. Discovering type I cis-AT polyketides through computational mass spectrometry and genome mining with Seq2PKS. Nat Commun 15, 5356. https://doi.org/10.1038/s41467-024-49587-1

Yang, C., Li, Y., Guo, Q., Lai, Q., Zheng, T., Tian, Y., 2013. Algoriphagus zhangzhouensis sp. nov., isolated from mangrove sediment. Int J Syst Evol Microbiol 63, 1621–1626. https://doi.org/10.1099/ijs.0.044271-0

Yoneda, Y., Yamamoto, K., Makino, A., Tanaka, Y., Meng, X.-Y., Hashimoto, J., Shin-ya, K., Satoh, N., Fujie, M., Toyama, T., Mori, K., Ike, M., Morikawa, M., Kamagata, Y., Tamaki, H., 2021. Novel Plant-Associated

Acidobacteria Promotes Growth of Common Floating Aquatic Plants, Duckweeds. Microorganisms 9, 1133. https://doi.org/10.3390/microorganisms9061133

Zaburannyi, N., Rabyk, M., Ostash, B., Fedorenko, V., Luzhetskyy, A., 2014. Insights into naturally minimised Streptomyces albus J1074 genome. BMC Genomics 15, 97. https://doi.org/10.1186/1471-2164-15-97

Zdanovsky, A.G., Zdanovskaia, M.V., 2000. Simple and Efficient Method for Heterologous Expression of Clostridial Proteins. Appl Environ Microbiol 66, 3166–3173. https://doi.org/10.1128/aem.66.8.3166-3173.2000

Zdouc, M.M., Blin, K., Louwen, N.L.L., Navarro, J., Loureiro, C., Bader, C.D., Bailey, C.B., Barra, L., Booth, T.J., Bozhüyük, K.A.J., Cediel-Becerra, J.D.D., Charlop-Powers, Z., Chevrette, M.G., Chooi, Y.H., D'Agostino, P.M., de Rond, T., Del Pup, E., Duncan, K.R., Gu, W., Hanif, N., Helfrich, E.J.N., Jenner, M., Katsuyama, Y., Korenskaia, A., Krug, D., Libis, V., Lund, G.A., Mantri, S., Morgan, K.D., Owen, C., Phan, C.-S., Philmus, B., Reitz, Z.L., Robinson, S.L., Singh, K.S., Teufel, R., Tong, Y., Tugizimana, F., Ulanova, D., Winter, J.M., Aguilar, C., Akiyama, D.Y., Al-Salihi, S.A.A., Alanjary, M., Alberti, F., Aleti, G., Alharthi, S.A., Rojo, M.Y.A., Arishi, A.A., Augustijn, H.E., Avalon, N.E., Avelar-Rivas, J.A., Axt, K.K., Barbieri, H.B., Barbosa, J.C.J., Barboza Segato, L.G., Barrett, S.E., Baunach, M., Beemelmanns, C., Beqaj, D., Berger, T., Bernaldo-Agüero, J., Bettenbühl, S.M., Bielinski, V.A., Biermann, F., Borges, R.M., Borriss, R., Breitenbach, M., Bretscher, K.M., Brigham, M.W., Buedenbender, L., Bulcock, B.W., Cano-Prieto, C., Capela, J., Carrion, V.J., Carter, R.S., Castelo-Branco, R., Castro-Falcón, G., Chagas, F.O., Charria-Girón, E., Chaudhri, A.A., Chaudhry, V., Choi, H., Choi, Y., Choupannejad, R., Chromy, J., Donahey, M.S.C., Collemare, J., Connolly, J.A., Creamer, K.E., Crüsemann, M., Cruz, A.A., Cumsille, A., Dallery, J.-F., Damas-Ramos, L.C., Damiani, T., de Kruijff, M., Martín, B.D., Sala, G.D., Dillen, J., Doering, D.T., Dommaraju, S.R., Durusu, S., Egbert, S., Ellerhorst, M., Faussurier, B., Fetter, A., Feuermann, M., Fewer, D.P., Foldi, J., Frediansyah, A., Garza, E.A., Gavriilidou, A., Gentile, A., Gerke, J., Gerstmans, H., Gomez-Escribano, J.P., González-Salazar, L.A., Grayson, N.E., Greco, C., Gomez, J.E.G., Guerra, S., Flores, S.G., Gurevich, A., Gutiérrez-García, K., Hart, L., Haslinger, K., He, B., Hebra, T., Hemmann, J.L., Hindra, H., Höing, L., Holland, D.C., Holme, J.E., Horch, T., Hrab, P., Hu, J., Huynh, T.-H., Hwang, J.-Y., Iacovelli, R., Iftime, D., Iorio, M., Jayachandran, S., Jeong, E., Jing, J., Jung, J.J., Kakumu, Y., Kalkreuter, E., Kang, K.B., Kang, S., Kim, W., Kim, G.J., Kim, H., Kim, H.U., Klapper, M., Koetsier, R.A., Kollten, C., Kovács, Á.T., Kriukova, Y., Kubach, N., Kunjapur, A.M., Kushnareva, A.K., Kust, A., Lamber, J., Larralde, M., Larsen, N.J., Launay, A.P., Le, N.-T.-H., Lebeer, S., Lee, B.T., Lee, K., Lev, K.L., Li, S.-M., Li, Y.-X., Licona-Cassani, C., Lien, A., Liu, J., Lopez, J.A.V., Machushynets, N.V., Macias, M.I., Mahmud, T., Maleckis, M., Martinez-Martinez, A.M., Mast, Y., Maximo, M.F., McBride, C.M., McLellan, R.M., Bhatt, K.M., Melkonian, C., Merrild, A., Metsä-Ketelä, M., Mitchell, D.A., Müller, A.V., Nguyen, G.-S., Nguyen, H.T., Niedermeyer, T.H.J., O'Hare, J.H., Ossowicki, A., Ostash, B.O., Otani, H., Padva, L., Paliyal, S., Pan, X., Panghal, M., Parade, D.S., Park, J., Parra, J., Rubio, M.P., Pham, H.T., Pidot, S.J., Piel, J., Pourmohsenin, B., Rakhmanov, M., Ramesh, S., Rasmussen, M.H., Rego, A., Reher, R., Rice, A.J., Rigolet, A., Romero-Otero, A., Rosas-Becerra, L.R., Rosiles, P.Y., Rutz, A., Ryu, B., Sahadeo, L.-A., Saldanha, M., Salvi, L., Sánchez-Carvajal, E., Santos-Medellin, C., Sbaraini, N., Schoellhorn, S.M., Schumm, C., Sehnal, L., Selem, N., Shah, A.D., Shishido, T.K., Sieber, S., Silviani, V., Singh, G., Singh, H., Sokolova, N., Sonnenschein, E.C., Sosio, M., Sowa, S.T., Steffen, K., Stegmann, E., Streiff, A.B., Strüder, A., Surup, F., Svenningsen, T., Sweeney, D., Szenei, J., Tagirdzhanov, A., Tan, B., Tarnowski, M.J., Terlouw, B.R., Rey, T., Thome, N.U., Torres Ortega, L.R., Tørring, T., Trindade, M., Truman, A.W., Tvilum, M., Udwary, D.W., Ulbricht, C., Vader, L., van Wezel, G.P., Walmsley, M., Warnasinghe, R., Weddeling, H.G., Weir, A.N.M., Williams, K., Williams, S.E., Witte, T.E., Rocca, S.M.W., Yamada, K., Yang, Dong, Yang, Dongsoo, Yu, J., Zhou, Z., Ziemert, N., Zimmer, L., Zimmermann, A., Zimmermann, C., van der Hooft, J.J.J., Linington, R.G., Weber, T., Medema, M.H., 2025. MIBiG 4.0: advancing biosynthetic gene cluster curation through global collaboration. Nucleic Acids Res 53, D678–D690. https://doi.org/10.1093/nar/gkae1115

Zengler, K., Toledo, G., Rappé, M., Elkins, J., Mathur, E.J., Short, J.M., Keller, M., 2002. Cultivating the uncultured. PNAS 99, 15681–15686. https://doi.org/10.1073/pnas.252630999

Zhang, H., Wang, Y., Wu, J., Skalina, K., Pfeifer, B.A., 2010. Complete Biosynthesis of Erythromycin A and Designed Analogs Using *E. coli* as a Heterologous Host. Chemistry & Biology 17, 1232–1240. https://doi.org/10.1016/j.chembiol.2010.09.013

Zhang, J.J., Tang, X., Moore, B.S., 2019. Genetic platforms for heterologous expression of microbial natural products. Nat. Prod. Rep. 36, 1313–1332. https://doi.org/10.1039/C9NP00025A

Zhang, J.J., Tang, X., Zhang, M., Nguyen, D., Moore, B.S., 2017. Broad-Host-Range Expression Reveals Native and Host Regulatory Elements That Influence Heterologous Antibiotic Production in Gram-Negative Bacteria. mBio 8, e01291-17. https://doi.org/10.1128/mBio.01291-17

Zhang, M.M., Wong, F.T., Wang, Y., Luo, S., Lim, Y.H., Heng, E., Yeo, W.L., Cobb, R.E., Enghiad, B., Ang, E.L., Zhao, H., 2017. CRISPR–Cas9 strategy for activation of silent Streptomyces biosynthetic gene clusters. Nat Chem Biol 13, 607–609. https://doi.org/10.1038/nchembio.2341

Zhang, W., Li, Y., Tang, Y., 2008. Engineered biosynthesis of bacterial aromatic polyketides in Escherichia coli. Proceedings of the National Academy of Sciences 105, 20683–20688. https://doi.org/10.1073/pnas.0809084105

Zhang, Y., Werling, U., Edelmann, W., 2012. SLiCE: a novel bacterial cell extract-based DNA cloning method. Nucleic Acids Res 40, e55. https://doi.org/10.1093/nar/gkr1288

Zhu, B., Ni, F., Ning, L., Yao, Z., Du, Y., 2018a. Cloning and biochemical characterization of a novel κ-carrageenase from newly isolated marine bacterium *Pedobacter hainanensis* NJ-02. International Journal of Biological Macromolecules 108, 1331–1338. https://doi.org/10.1016/j.ijbiomac.2017.11.040

Zhu, B., Ni, F., Ning, L., Yao, Z., Du, Y., 2018b. Cloning and biochemical characterization of a novel κ-carrageenase from newly isolated marine bacterium Pedobacter hainanensis NJ-02. International Journal of Biological Macromolecules 108, 1331–1338. https://doi.org/10.1016/j.ijbiomac.2017.11.040

Ziemert, N., Ishida, K., Liaimer, A., Hertweck, C., Dittmann, E., 2008a. Ribosomal Synthesis of Tricyclic Depsipeptides in Bloom-Forming Cyanobacteria. Angewandte Chemie International Edition 47, 7756–7759. https://doi.org/10.1002/anie.200802730

Ziemert, N., Ishida, K., Liaimer, A., Hertweck, C., Dittmann, E., 2008b. Ribosomal synthesis of tricyclic depsipeptides in bloom-forming cyanobacteria. Angewandte Chemie International Edition 47, 7756–7759. https://doi.org/10.1002/anie.200802730

Ziemert, N., Lechner, A., Wietz, M., Millán-Aguiñaga, N., Chavarria, K.L., Jensen, P.R., 2014. Diversity and evolution of secondary metabolism in the marine actinomycete genus Salinispora. Proceedings of the National Academy of Sciences of the United States of America 111, E1130–E1139. https://doi.org/10.1073/pnas.1324161111

Ziemert, N., Podell, S., Penn, K., Badger, J.H., Allen, E., Jensen, P.R., 2012. The Natural Product Domain Seeker NaPDoS: A Phylogeny Based Bioinformatic Tool to Classify Secondary Metabolite Gene Diversity. PLOS ONE 7, e34064. https://doi.org/10.1371/journal.pone.0034064

Zolghadri, S., Bahrami, A., Khan, M.T.H., Munoz-Munoz, J., Garcia-Molina, F., Garcia-Canovas, F., Saboury, A.A., 2019. A comprehensive review on tyrosinase inhibitors. Journal of Enzyme Inhibition and Medicinal Chemistry 34, 279–309. https://doi.org/10.1080/14756366.2018.1545767

Zumkeller, C., Schindler, D., Waldminghaus, T., 2018. Modular Assembly of Synthetic Secondary Chromosomes, in: Dame, R.T. (Ed.), Bacterial Chromatin: Methods and Protocols, Methods in Molecular Biology. Springer, New York, NY, pp. 71–94. https://doi.org/10.1007/978-1-4939-8675-0_5

Zumkeller, C.M., Bletz, M.C., Rakotoarison, A., Sabino-Pinto, J., Reiter, S., Spohn, M., Schwengers, O., Goesmann, A., Vences, M., Mihajlovic, S., Schäberle, T.F., 2024. Draft genome sequences of 21 Pedobacter strains isolated from amphibian specimens. Microbiology Resource Announcements 13, e01185-23. https://doi.org/10.1128/mra.01185-23

Zuther, K., Mayser, P., Hettwer, U., Wu, W., Spiteller, P., Kindler, B.L.J., Karlovsky, P., Basse, C.W., Schirawski, J., 2008. The tryptophan aminotransferase Tam1 catalyses the single biosynthetic step for tryptophan-dependent pigment synthesis in Ustilago maydis. Molecular Microbiology 68, 152–172. https://doi.org/10.1111/j.1365-2958.2008.06144.x

6 SUPPLEMENTARY INFORMATION

6.1 Genome Mining of Bacteroidota

6.1.1 Python Scripts

SI 1:individualClusterCounter-CMZ.py

```
#!/usr/bin/python3

import os

import re

import sys


# Function to print progress

def print_progress(current, total):

    progress = (current / total) * 100

    print(f"Progress: {progress:.2f}% ({current}/{total})")


# Get input directory from command-line argument, or exit if not provided

if len(sys.argv) < 2:

    print("Usage: python3 script.py <input_directory>
[<output_directory>]")

    sys.exit(1)


input_dir = os.path.expanduser(sys.argv[1])  # Get input directory from
argument

if not os.path.isdir(input_dir):

    print(f"Error: The provided input directory '{input_dir}' does not
exist.")

    sys.exit(1)
```

```
# Optional: Get output directory from the second argument, or use input directory by default
output_dir = os.path.expanduser(sys.argv[2]) if len(sys.argv) > 2 else input_dir
os.chdir(output_dir)

# Initialize variables
dict_cluster = {}
dict_genome_cluster_counter = {}
index_html_count = 0
troublemaker = []
files_with_index_html = 0
list_of_match_list_dict = []
header_list = ["Folder name"]

# Get a list of all directories and subdirectories within the input directory
dir_list = []
for root, dirs, files in os.walk(input_dir):
    for d in dirs:
        dir_list.append(os.path.join(root, d))

# Sort the directories
dir_list.sort()

# Print the total number of directories to process
print(f"Total number of folders to process: {len(dir_list)}\n")

# Processing folders
for i, singleFolder in enumerate(dir_list):
```

```
    os.chdir(singleFolder)
    if not os.path.isfile("index.html"):
        troublemaker.append(singleFolder)
        index_html_count += 1
    else:
        with open("index.html", "r") as file:
            match = []
            files_with_index_html += 1
            for line in file:
                match.append(re.findall('<div class="regbutton [A-z]+',
line ))
            match_list = list(filter(None, match))
            match_list_dict = {"Folder name": singleFolder}
            for m in match_list:
                match_split = str(m).split()
                match_split[-1] = match_split[-1][0:-2]
                if match_split[-1] not in header_list:
                    header_list.append(match_split[-1])
                if match_split[-1] in match_list_dict:
                    match_list_dict[match_split[-1]] += 1
                else:
                    match_list_dict[match_split[-1]] = 1
            list_of_match_list_dict.append(match_list_dict)

    # Print progress
    print_progress(i + 1, len(dir_list))

# Print header list
print("\nHeader_list:", header_list)
```

```
# Change back to the output directory
os.chdir(output_dir)


# Writing results to output file
output_file = "table_individualClusterCounter"
print(f"\nWriting results to {output_file}...\n")
with open(output_file, "w") as table:
    # Write header
    table.write("\t".join(header_list) + "\n")


    # Write data for each folder
    for match_list_dict in list_of_match_list_dict:
        for header in header_list:
            value = str(match_list_dict.get(header, 0))
            table.write(value + "\t")
        table.write("\n")


# Final result printing
print(f"Files with 'index.html': {files_with_index_html}")
print(f"Number of files without 'index.html': {index_html_count}")
print("Folders missing 'index.html':", troublemaker)
print(f"Results saved in: {output_file}")



# Print the total number of directories to process
print(f"Total number of folders to process: {len(dir_list)}\n")
```

SI 2: RenameFilesWithDirectory.py

```
  GNU nano 6.2
RenameFileswithDirectory.py
import os
import sys


# Check if a directory was provided as an argument
if len(sys.argv) != 2:
    print("Usage: python rename_gbk_files.py <directory_path>")
    sys.exit(1)


# Get the root directory from the command-line argument
root_directory = os.path.expanduser(sys.argv[1])


# Check if the directory exists
if not os.path.isdir(root_directory):
    print(f"Error: Directory '{root_directory}' does not exist.")
    sys.exit(1)


# Loop through all subdirectories
for subdir, _, files in os.walk(root_directory):
    # Get the name of the current subdirectory
    dir_name = os.path.basename(subdir)
    print(f"Processing directory: {subdir}")


    # Loop through all files in the subdirectory
    for file_name in files:
        # Skip hidden files and ensure only .gbk files are renamed
        if file_name.startswith('.') or not file_name.endswith('.gbk'):
            continue
```

```
        # Create the new file name with the directory name as a prefix
        new_file_name = f"{dir_name}_{file_name}"

        # Create the full paths for renaming
        old_path = os.path.join(subdir, file_name)
        new_path = os.path.join(subdir, new_file_name)

        # Ensure that the file exists before renaming
        if os.path.exists(old_path):
            os.rename(old_path, new_path)
            print(f"Renamed {old_path} to {new_path}")
        else:
            print(f"File {old_path} not found!")
```

6.1.2 Figures

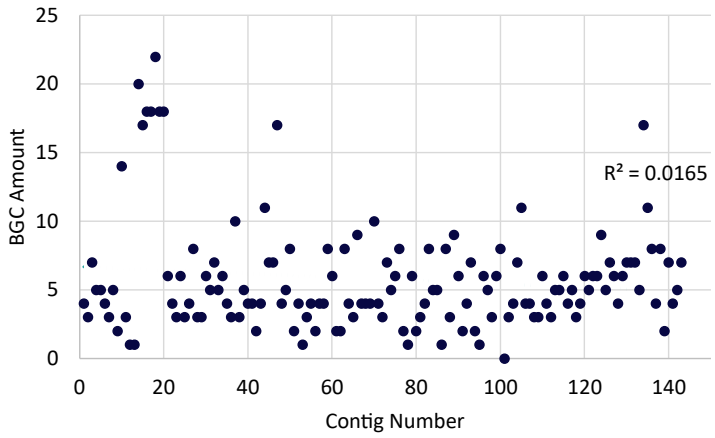

Figure S1-1. Plot showing the number of antiSMASH detected BGCs plotted against the number of contigs per respective genome.

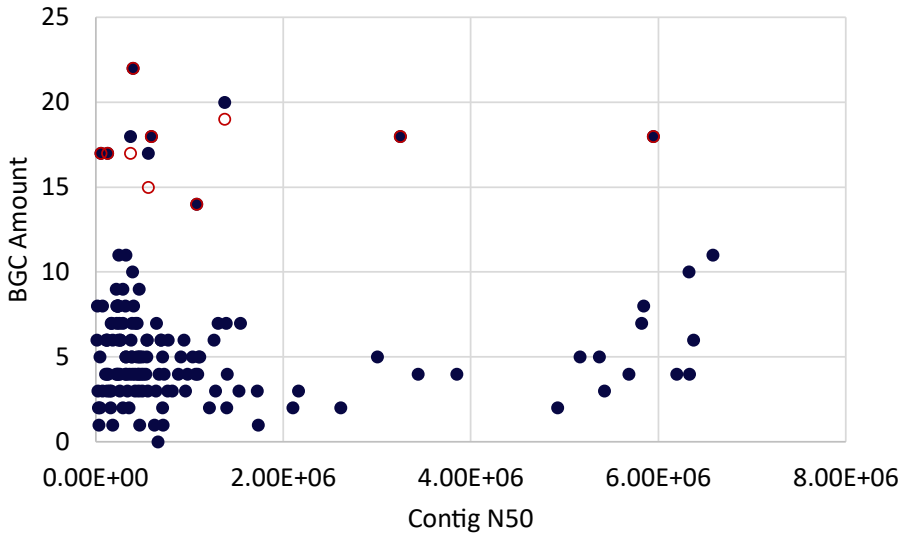

Figure S1 2-. Plot showing the N50 values plotted against the BGC amount. Low N50 genomes containing many BGCs were manually curated to remove broken cluster fragments. Three strains contained broken NRPS clusters, resulting in a lower final BGC count (red circles).

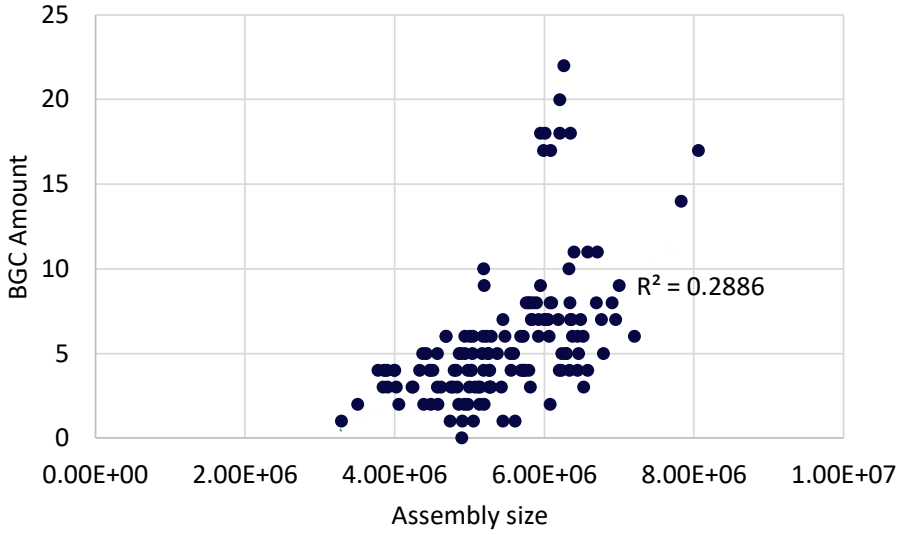

Figure S1-3. Plot showing the relationship between the number of detected BGCs and the total assembly size of each respective genome.

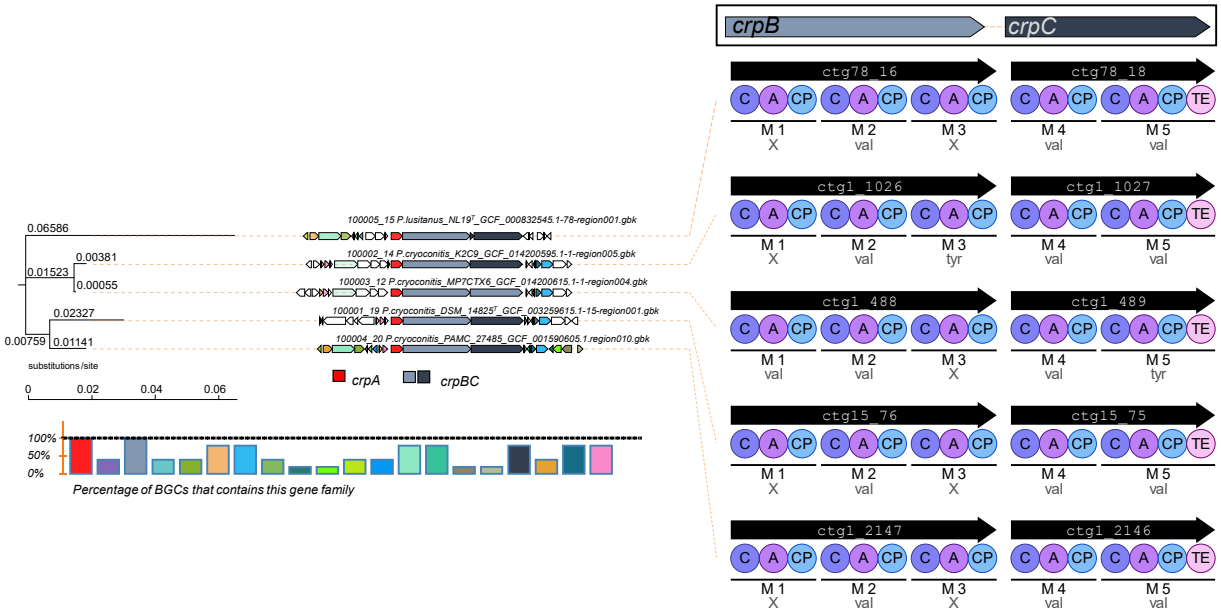

Figure S1-4. Corason Alignment of BiG-SCAPE identified GCF NRPS - 4 Cryopeptin. Corason was run in default settings, with *crpA* (red) selected as the query gene. Homologous genes predicted by Corason have the same color. The lower left panel of the Corason output indicates the presence of the individual genes of the Cryopeptin BGC in the different genomes (100% - present in all strains). The right side shows the antiSMASH predicted a-domain specificities of the different Cryopeptin BGCs. The *crpC* gene of *P. cryoconitis* MP7CTX6 has a predicted difference (indicated by the color (=*crpB)*) in a-domain specificity (tyr instead of val) but shares the same number of modules (n=2; M4, M5) and domain classes (C4: LCL and C5: modAA as predicted by NaPDoS).

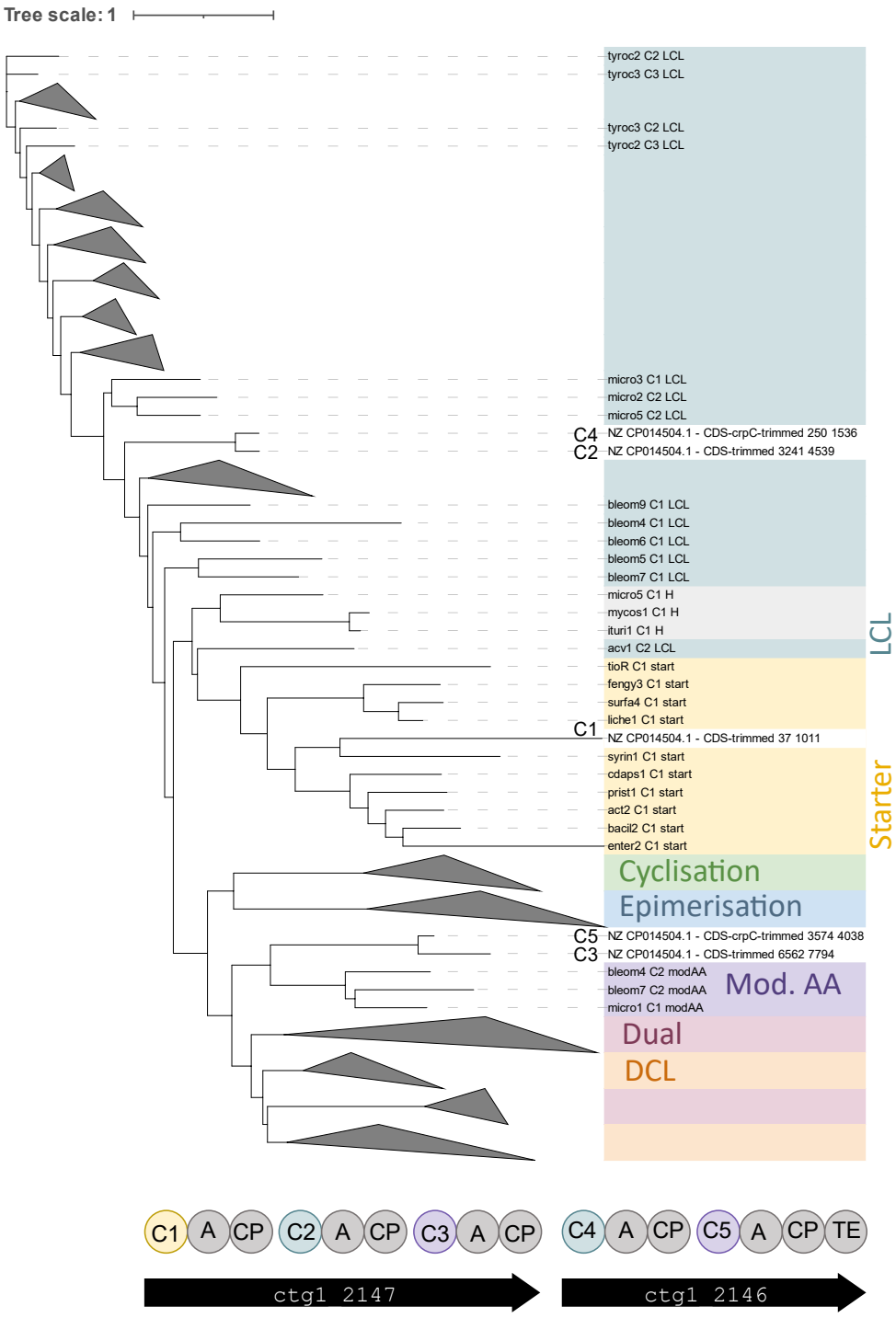

Figure S1-5. NaPDoS analysis of all C-domains predicted for the *crp* BGC. Colored clades indicate the different domain classes. Some clades are collapsed to reduce the figure size (iTOL). Below the tree the NRPS genes *crpB* (2147) and *crpC* (2146) are indicated including the different domains. C-domains are numbered and colored depending on the domain class that they are most closely related to. C1 clusters with the "starter" domain class, while C2 and C5 are predicted to be LCL-type domain classes. C3 and C5 are most closely related to the "modified AA" class, which is known to be involved in modifications of the incorporated amino acids such as dehydration processes

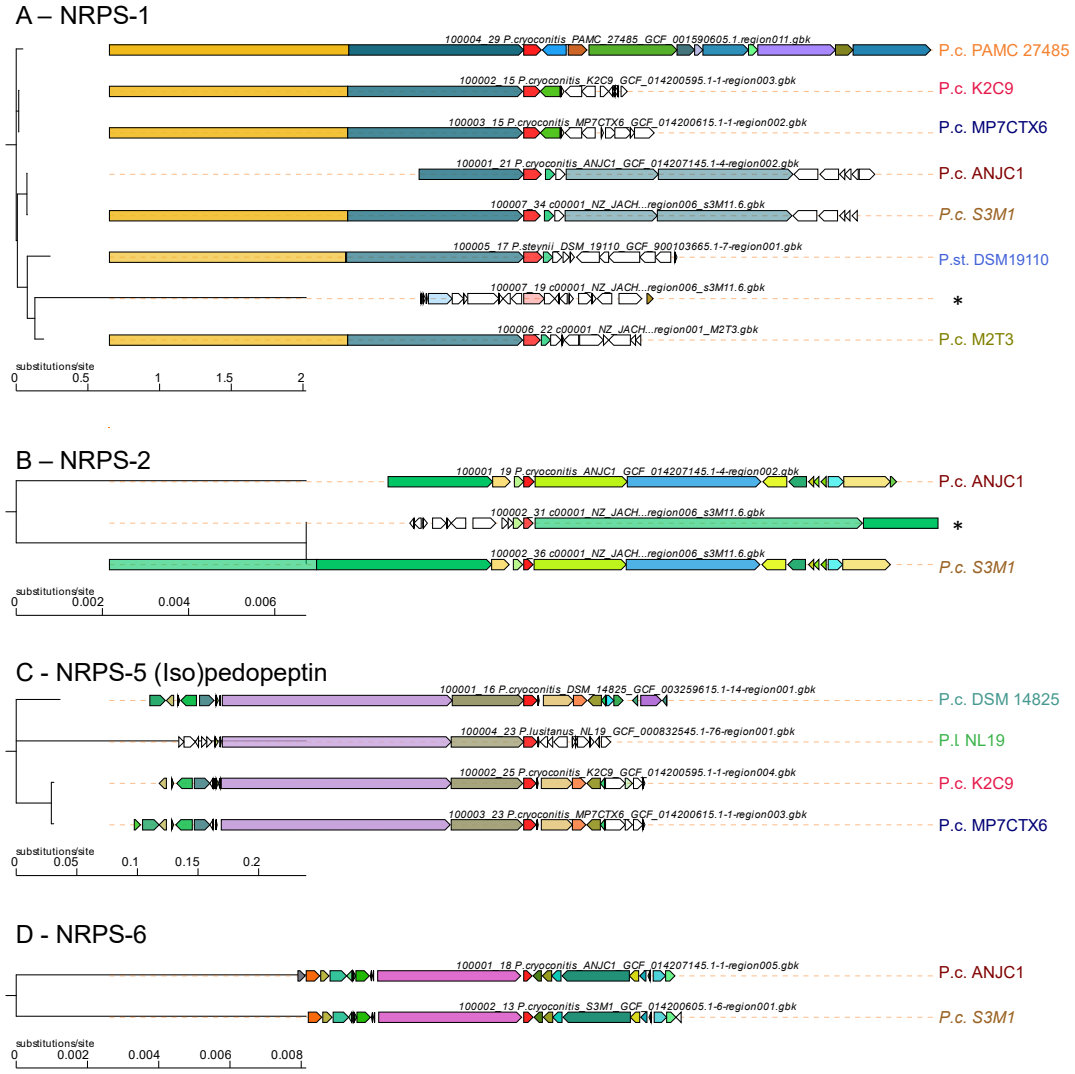

Figure S1-6. Corason Alignments of multimodular NRPS BGCs present in multiple strains. Corason was run in default settings, query genes are marked in red. BGC-containing strains are labeled at the ends of the clusters. NRPS-1 and NRPS-2 (A, B) Show fragments of clusters resulting from the fused, collectively detected BGCs by antiSMASH (marked with *). (C) Corason Alignment of Pedopeptin and putative Isopedopeptin clusters. The Pedopeptin BGC of *P. lusitanus* NL19^{T} is less similar to the others. Coloured genes indicate more modifying genes in the *P. cryoconitis* strains.

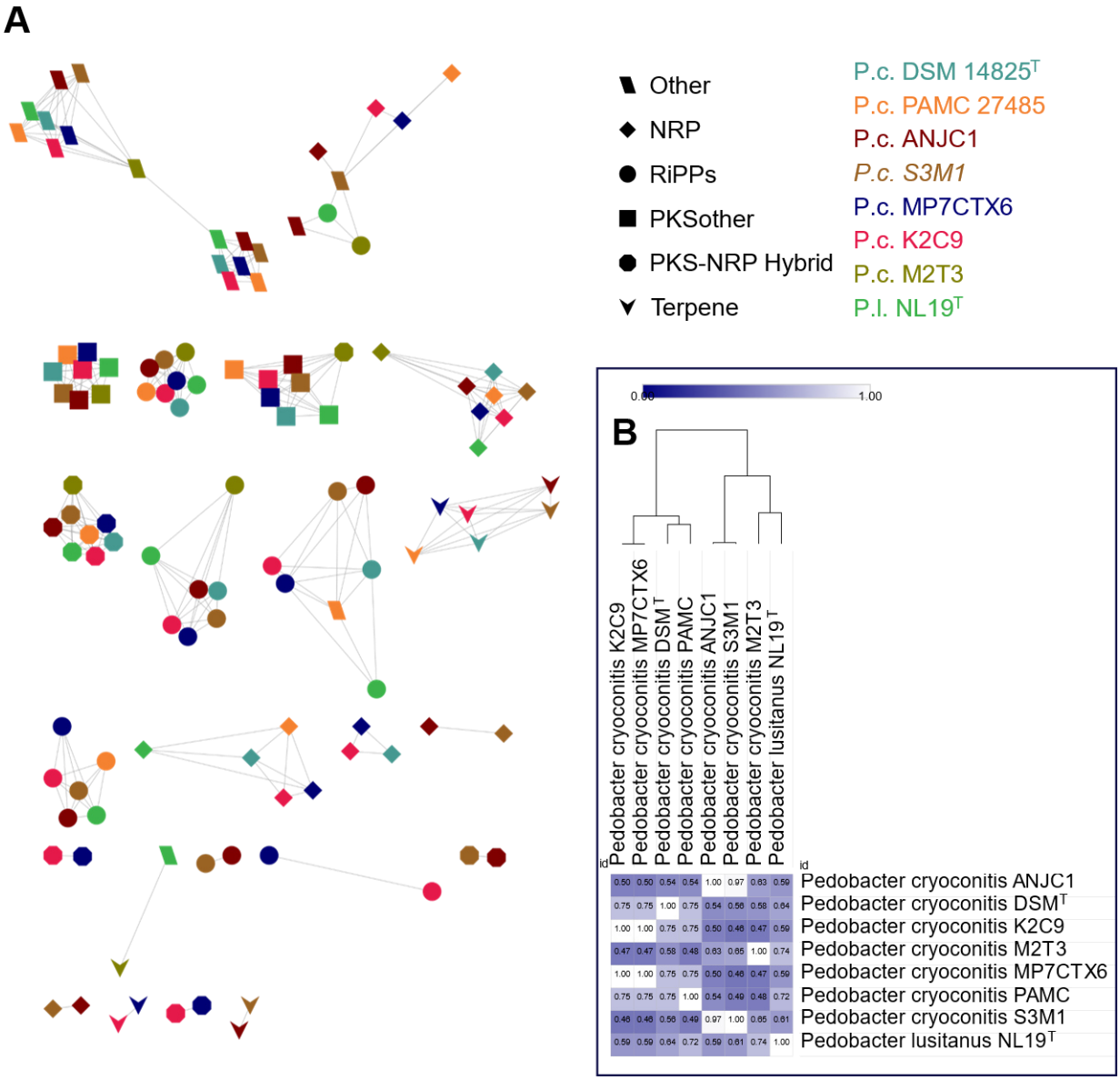

Figure S1-7. A BiG-SCAPE analysis of anti-SMASH detected BGCS in the 8 strains associated to the *P. cryoconitis* clade (Cutoffs for GCFs: 0.6). The node shape indicates the BGC type, and the node color is the strain in the BGC that was detected. Numbers at nodes describe the Big-SCAPE-determined gene cluster family. Besides the shown clusters, 48 singletons were determined. No BGC automatically clusters with any known MiBiG Reference cluster. B Heat Map illustrating the cosine similarity of the overall BGC composition in the *P. cryoconitis* branch. The calculated tree is based on that similarity and highlights the similar BGC composition in the groups containing for one K2C9 and MP7CTX6 and for another ANJC1 and S3M1.

6.2 Profiling of Novel Acidobacteriota

6.2.1 Extended Materials and Methods

6.2.1.1 Analysis of Assemblies and their Biosynthetic Gene Cluster Potential

The 618 curated genome assemblies were processed to comparatively evaluate metadata such as assembly size, GC-content and N50, monitored by Geneious (v11.1.5), and to predict the BGC potential. BGC prediction was performed using antiSMASH (v6.1.1); afterwards, predicted BGCs were analysed using BiG-SCAPE (Navarro-Muñoz *et al.*, 2020a) at a cutoff of 0.5, including MiBIG reference clusters in the analysis. The sequence similarity network was visualised using Cytoscape (v3.10.2). UMAP was used to reduce the high-dimensional biosynthetic similarity data reported by BiG-SCAPE into two dimensions, allowing for the visualisation of overall differences in BGC composition between different strains in our acidobacterial dataset. The binary presence/absence matrix (BiG-SCAPE output) was used to compute pairwise cosine similarities using scikit-learn (version 1.3). The matrix of cosine distances was then projected in two dimensions using UMAP (UMAP-learn, v0.5.5, metric='cosine', random_state=42) and visualised using matplotlib (v3.7). Strains were coloured according to their taxonomic class to reflect phylogenetic relationships.

Our analysis revealed that the metagenome-assembled genomes (MAGs) are, on average, slightly smaller (4.88 ± 1.46 Mb) than isolate genomes (5.28 ± 1.71 Mb) (Figure S4). This difference is statistically modest with the Mann–Whitney U test yielding p = 0.04, which is significant at the 5% level but does not meet a more stringent 1% threshold. All assemblies not affiliated to a genus by GTDB are MAGs (n=85; 13.7%). Contig numbers ranged from 1 to 1357, with an average of 200 contigs per genome. The BGC number did not significantly correlate with the contig number, indicating no significant overestimation of BGCs due to fragmented clusters (Figure S7). Interestingly, the BGC number also did not correlate with the assembly size, a phenomenon typically observed across different taxonomic levels (Adamek *et al.*, 2018; Brinkmann *et al.*, 2022a; Männle *et al.*, 2020).

The GC content of the analysed genome set ranged from 34.3% to 73.8% (Figure S5). This is substantially broader than previously proposed for the Acidobacteriota phylum (McReynolds *et al.*, 2025). This broad GC range exceeds the typical GC distribution reported for most bacterial phyla, with comparable broadness covered only by the *Pseudomonadota*. Notably, from our dataset, 37 genomes exhibited a GC content ≥70%, all belonging to the three classes Mor1 (average GC% 68.0 ± 4.2, n9), *Vicinamibacteria* (average GC% 67.5 ± 2.7, n98), and *Thermoanaerobaculia* (average GC% 67.9 ± 3.2, n79) (Figure S6). GC-rich bacterial genomes are considered an adaptation to high temperatures, with a demonstrated positive correlation between the optimal growth temperature of bacteria (Hu *et al.*, 2022; Lightfield *et al.*, 2011; Teng *et al.*, 2022) and their GC content. Indeed, the *Thermoanaerobaculia* class has been generally associated with high-temperature environments, and its single isolated strain, *Thermoanaerobaculum aquaticum* DSM 24856, exhibits optimal growth at 60°C (Losey *et al.*, 2013). Considering the wide range of GC-content of the Acidobacteriota phylum, we evaluated whether the GC-content correlates with the number of BGCs detected, but did not find any correlation (Figure S5; 0.03; p = 0.482). Indeed, the most enriched classes in terms of GC content (Mor1, *Vicinamibacteria*, *Thermoanaerobaculia*; Figure S6) are rather under-equipped in BGC load, and only Mor1 shows a moderate

correlation between GC content and BGC number. Instead, the *Acidobacteriae* class exhibits a strong negative correlation, with lower GC-content genomes harbouring a larger number of BGCs. The most recognised bacterial NP producing taxa, the Actinomycetes class of the Actinomycetota phylum, is well known for its large genomes (up to 10 Mbps) and high GC content (Barka *et al.*, 2015). Within this phylum, genome size and GC content appear to be phylogenetically conserved and positively correlated (Nouioui *et al.*, 2018). Furthermore, genome size and number of secondary metabolite biosynthetic gene clusters are positively correlated in the Actinomycetes class, indicating that larger genomes can accommodate more gene clusters devoted to secondary metabolism (Doroghazi and Metcalf, 2013). The most talented orders for natural product biosynthesis (e.g. *Streptomycetales*, *Pseudonocardiales*, *Streptosporangiales* and *Micromonosporales*) carry the biggest genomes (average > 7.5 kb) and highest GC content (average >70%), outcompeting orders with lower biosynthetic potential (e.g. *Corynebacteriales*, *Propionibacteriales*, *Micrococcales*, and *Bifidobacteriales*) characterised by comparatively low genome sizes (average < 5.1 kb) and lower GC content (average < 70%) (Nouioui *et al.*, 2018).

The evaluation of the Gene Cluster Families (GCFs) generated by BiG-SCAPE (Navarro-Muñoz *et al.*, 2020a), enabled conclusions on their distributions within the dataset and their correlation to BGCs, encoding the production of known NPs. Filtering revealed GCFs shared among different acidobacterial classes (n = 23; Figure S10). Most of the GCFs are shared between the *Acidobacteriae* and the *Vicinamibacteria* (n = 7). There is no clear trend in BGC cluster type; instead, NRP, RiPP, Terpenes, and PKS1 cluster types are shared. In general, the class of *Acidobacteriae* shares GCFs with every class except for UBA890 and Mor1. The most diverse GCFs when it comes to classes are GCF3545 and GCF5459, both RiPP clusters shared between three different classes (*Vicinamibacteria* and *Acidobacteriae* and *Blastocatellia* or UBA6911, respectively). GCF4479 is an NRP cluster shared among *Acidobacteriaceae*, *Blastocatellia*, and *Thermoanaerobacterales*. The last one is GCF5622, a PKS1 cluster harboured by *Acidobacteriae*, UBA6911 and *Vicinamibacteria*.

GCF alignment to the MiBiG Reference BGCs revealed several matches. In total, 11 acidobacterial clusters, grouped into six different GCFs, showed similarities to known MIBiG reference clusters. These clusters span various compound classes and taxonomic origins. GCF1211 appears to encode a PKS1 system involved in the biosynthesis of aromatic or benzenoid structures, which are known to be produced by fungi. GCF1752 was associated with alkylpyrone-407 and alkylpyrone-393, and a similar cluster, present in the *Thermoanaerobaculia* class, suggesting the production of small, polyketide-derived alkylated heterocycles. Besides PKS systems, we identified three NRP GCFs with related, known MIBiG reference clusters. GCF1359 contains different cluster types, all of which produce myxochromide D and S, linked to an NRP cluster from a representative of the *Chloracidobacteriales*. GCF 4479 contains three input genomes from different classes (*Blastocatellia*, *Thermoanaerobaculia*, *Acidobacteriae*) that are linked to anabaenopeptin, a cyanobacterial NRP cluster. One additional correlation hit was found within GCF1959, covering a PKS-NRPS hybrid of *Acidobacterium capsulatum* and the soil isolate FHG110511, which clusters with the occidiofungin A-reference BGC (MiBiG ID: BGC0001711) (Terlouw *et al.*, 2023). We also found three input genomes, belonging to the classes *Vicinamibacteria* and *Acidobacteriae*, that clustered with the reference clusters for ectoine.

6.2.1.2 Genomic Analysis of PGPT Traits

DIAMOND was run as follows:

> diamond blastp -q <input.fasta> -d <PGPT_DB> -o <output.tsv>

--outfmt 6 qtitle qseqid sseqid pident length mismatch gapopen qstart qend sstart send evalue bitscore

--evalue 1e-5 --threads 8 --<sensitivity_mode>

DIAMOND results were further processed to include relevant metadata and to minimise redundant hits. First (1.) we retained hits with E-values ≤1E-50 (RawHits). Second, we filtered for the best hit per plant-growth-promoting-trait (PGPT) family for each CDS. Finally, we reduced the dataset to contain only the best overall PGPT hit per CDS based on the lowest e-value (Output labelled with prefix "best_"). This output was annotated by merging the filtered hits with the PGPT ontology file, thereby matching the PGPT family numbers with their hierarchical classifications and functions. The output was used to analyse gene-specific content of individual strains. An overall summary of PGPT families per analysed genome was logged in PGPT-summary.txt. To get insights into taxonomic distributions of PGPTs, we merged the results with taxonomic metadata previously generated from GTDB-Tk. Based on taxonomic affiliation, the PGPT output can be optionally grouped by specifying a rank (e.g., family, order).

To evaluate the PGPT potential at the genome level, we computed the functional coverage across the different PGPT hierarchy levels. For each genome or taxonomic batch, we considered the number of PGPT families detected, the total possible number of PGPT families present in the database per hierarchy level, to calculate:

$$Coverage = \left(\frac{Detected\ PGPTs_i}{Total\ Possible\ PGPTs}\right) \times 100$$

Where i is the genome or taxonomy group.

For taxonomic batch comparison, the individually calculated coverage per genome is averaged across all genomes within the batch.

$$Coverage\ = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{Detected\ PGPTs_i}{Total\ PGPTs} \times 100\right)$$

The values were exported to an Excel file, where the different sheets cover different PGPT levels.

All analysis steps were performed in Python 3.9 using pandas, matplotlib, seaborn, and BioPython. DIAMOND v2 was used for protein alignment.

6.2.1.3 Analytical Procedures

The generated high-resolution mass spectrometry (HR-MS) data sets were copied and processed with both line spectra thresholds (5,000 and 10,000). Bucketing was performed on both thresholds at the same time, resulting in one table representing the OSMAC process and a second table representing the Trp-induction experiments, with both tables containing all buckets deemed identical for every sample (under both thresholds). These tables were subsequently curated: Buckets were only deemed present if they were detected in the 10,000

samples (4,406 (OSMAC) / 3,673 buckets (Trp-Induction) did not meet this criterion and were deleted). To avoid "uniqueness" due to compounds being just above the detection threshold, entries from the 5,000 set were used for these buckets; the 10,000 entries were deleted. 37 (OSMAC) / 54 (Trp-Induction) buckets were not filled in the respective tables and were subsequently deleted, resulting in two final tables representing 219 samples with 3,092 buckets (OSMAC) and 48 samples with 2,759 buckets (Trp-Induction), respectively. For OSMAC-barcoding, the conditions of each strain-media combination were combined. Each feature was deemed present when it occurred in two (grey) or more (black) extracts of the strain-media combination. Features that occurred only once were ignored as noise. Media control extracts were produced twice for each chemical composition (except for two media, where n = 1). For each chemical composition, a feature was deemed present when it occurred once (grey) or twice (black). The final table contained 2747 buckets, which were used for barcode generation.

Instead, for the Trp-Induction, a logical analysis was performed. Buckets filled in samples without tryptophan were coloured grey if present. For samples spiked with tryptophan, the sample was compared to its non-spiked counterpart and each bucket was categorised and assigned as Extinguished; Strong increase; Strong reduction; Light reduction; Present; Light increase; Strong increase:

Non-Spiked	**Spiked**	**Ratio Spiked/Non-Spiked**	**Category**
Present	Absent		Extinguished
Absent	Present		Strong increase
Present	Present	> 0 and < 0.5	Strong reduction
Present	Present	≥ 0.5 and < 0.8	Light reduction
Present	Present	≥ 0.8 and < 1.25	Present
Present	Present	≥ 1.25 and < 2.0	Light increase
Present	Present	≥ 2.0	Strong increase

To achieve a readable representation, the buckets in the input table were sorted "decreasing" according to presence/absence for each sample, then according to "sum present" and subsequently for "present in media". Readers need to be aware that the colouration of buckets categorised as "extinguished" does not mean they are present anymore in the Trp-spiked sample.

6.2.2 Tables

Table 8 Bacterial strains and sources

Strains	Source	Genome Information
Edaphobacter aggregans DSM 19364	DSMZ	GCF_000745965.1; High contamination value – not in data set
Acidobacterium sp. S8	From Petr Baldarian	GCF_009765985.1; High contamination value – not in dataset
Silvibacterium bohemicum S15	From Petr Baldarian	GCF_001006305.1, quality fine, maybe should be included in the set
Acidicapsa borealis DSM 23886	DSMZ	In-house genome
Granulicella rosea DSM 18704	DSMZ	GCF_900188085.1, low completeness level
Bryocella elongata DSM 22489	From Svetlana Dedysh	GCF_900108185.1, good quality genome, included in analysis
FhG110202: Acidobacterium	In-house strain, termite nest	In-house genome
FhG110214:Terracidiphilus	In-house strain, termite nest	In-house genome
FhG110511:Edaphobacter	In-house strain, soil	In-house genome

Table 9 OSMAC media combinations

Letter	Cultivation Size	Media composition	Media addition 1	Media addition 2
A	Small Scale	R2A pH 5.5		
B	Small Scale	2x R2A pH 5.5		
C	Small Scale	2x R2A pH 5.5		0.05% GlcNAc
D	Small Scale	2x R2A pH 5.5		0.05% Pektin
E	Small Scale	2x R2A pH 5.5		0.5% Cellobiose
F	Small Scale	2x R2A pH 5.5		0.5% Xylan
G	Small Scale	2x R2A pH 5.5		0.8% Phytagel
H	Small Scale	2x R2A pH 5.5	10 pcs Quartz beads	
I	Small Scale	2x R2A pH 5.5		3% Sucrose
J	Small Scale	4x R2A pH 5.5		
K	Small Scale	HD pH 5.0		
L	Small Scale	HD pH 5.0	10 pcs Quartz beads	
M	Small Scale	HD pH 5.0	3 pcs ceramic beads	
N	Small Scale	5294 pH 5.5		

O	Small Scale	5294 pH 5.5	10 pcs Quartz beads	
P	Small Scale	5294 pH 5.5	3 pcs ceramic beads	
Q	Small Scale	PSYA5 pH 5.0		0.5% Sucrose
R	Small Scale	PSYA5 pH 5.0	10 pcs Quartz beads	0.5% Sucrose
S	Small Scale	PSYA5 pH 5.0	3 pcs ceramic beads	0.5% Sucrose
T	Small Scale	PSYA5 pH 5.0		3% Sucrose
U	Large Scale	VL55		0.5% Xylane
V	Large Scale	VL55		0.5% Sucrose
W	Large Scale	HD		
X	Large Scale	5294 modified		

6.2.3 Figures

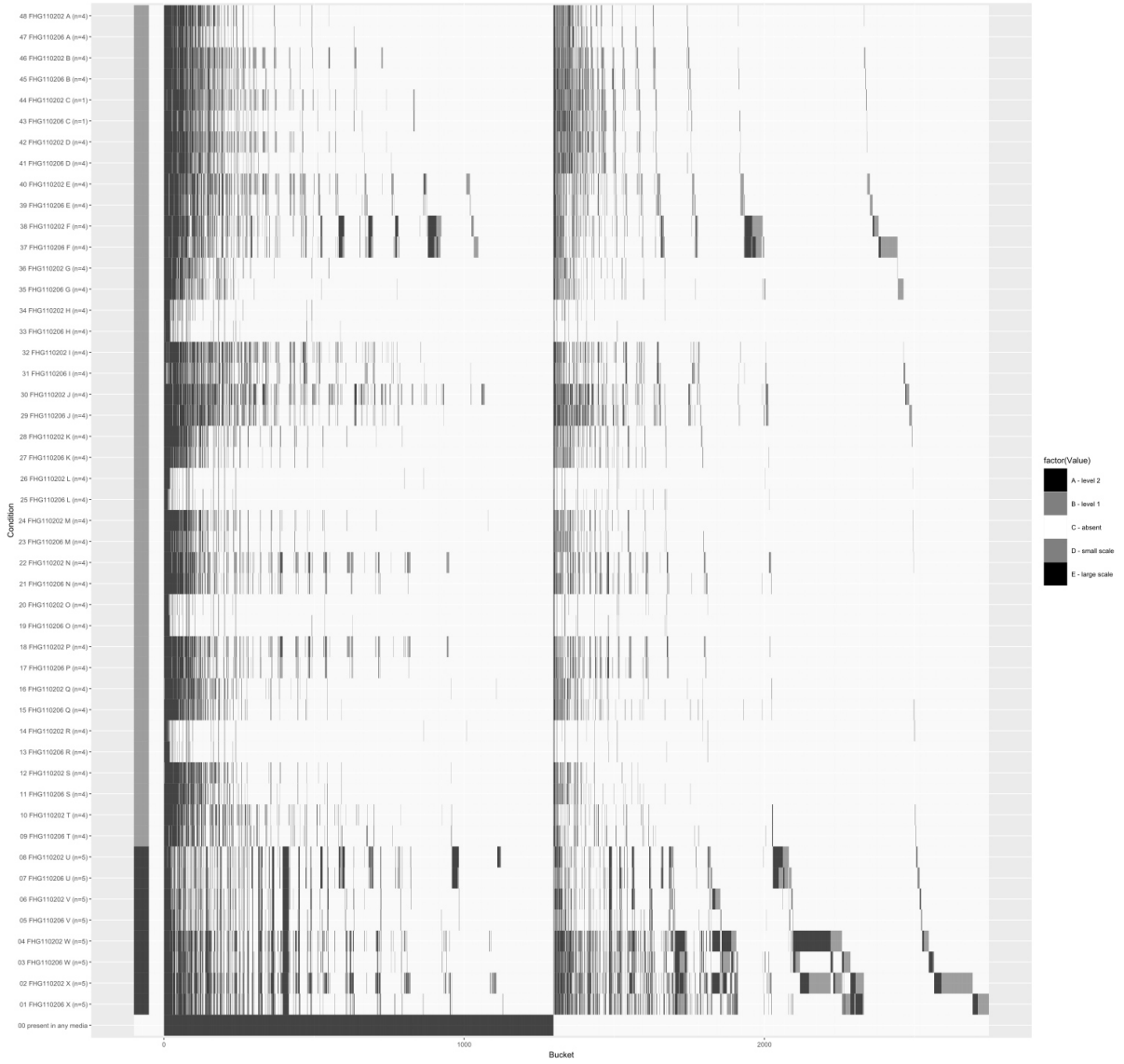

Figure S2-1: UHPLC–QTOF–HR–MS profiles from 244 extracts of strains FHG110202 and FHG110206 (incl. media controls). Profiles were aligned and bucketed into 2,747 features across 72 conditions. For strain–medium combinations, a bucket is considered present if it is observed in ≥2 extracts (grey = twice; black = ≥3); singletons were excluded. For media controls (columns, left, black label) (typically n=2), presence is scored once (grey) or twice (black). Quartz-beads (L, O, R, H) cultivations contained fewer buckets than the other conditions, while the large-scale cultivations (bottom, dark grey rows), showed more buckets compared to the small-scale cultivations (top, light grey rows).

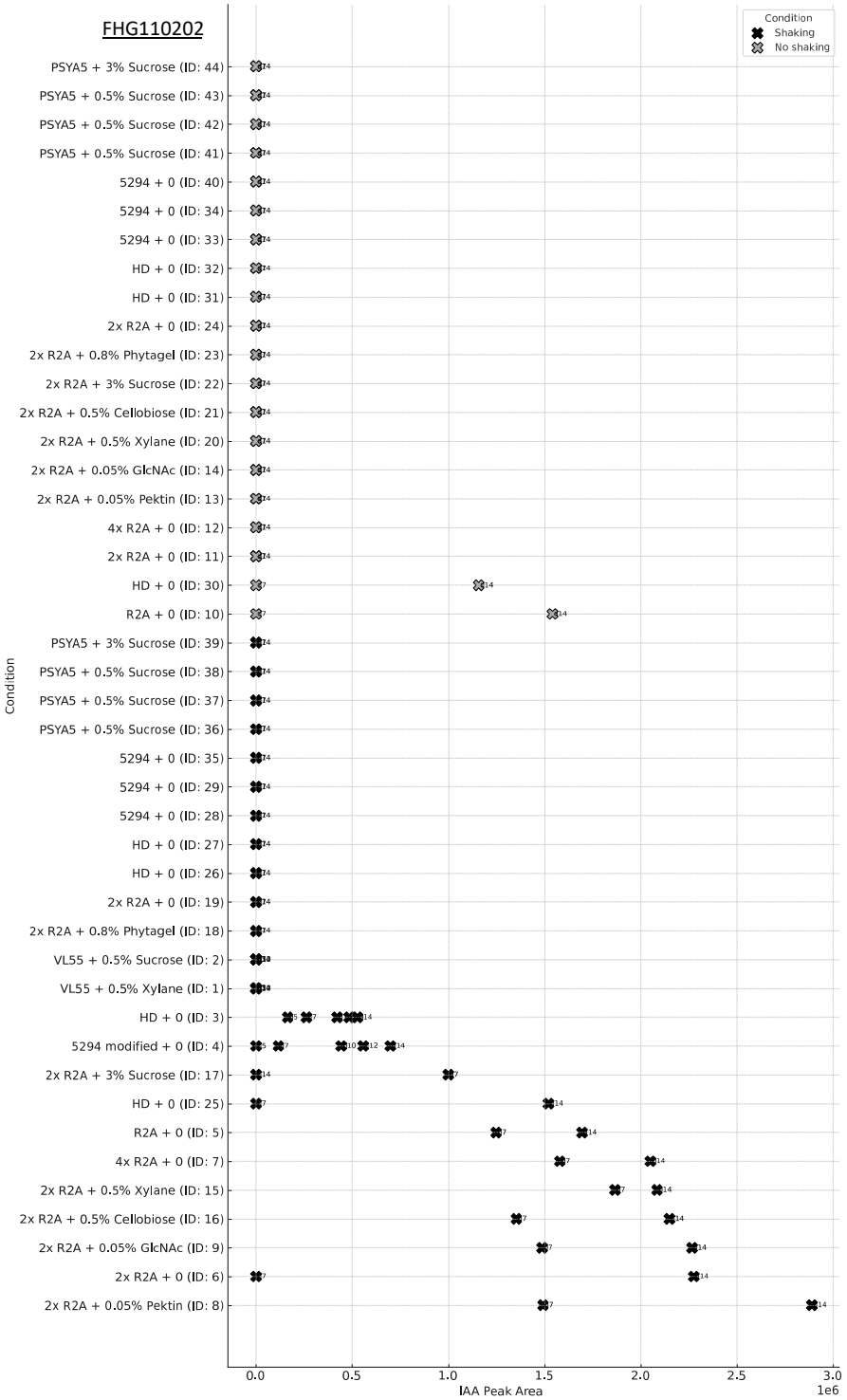

Figure S2-2: Exemplary Production Profile of IAA (Raw Peak Area) in Strain FHG110202. Feature 292.4s: 176.107m/z was extracted from the OSMAC feature table and tracked over time and condition (n =44, cf. OSMAC condition table for exact ID combinations, y-axis) to evaluate potential media composition and kinetics for further studies of IAA production (peak area, x-axis). Peak area values are plotted as crosses, with colour indicating agitation (light grey: No shaking; black: Shaking). Cultivation days are labelled beside crosses. In general, IAA production increases with cultivation time until day 14. The strain FHG110202 produced IAA in media base R2A, HD, and modified 5294, but not in PSYA5 and the original isolation medium VL55. Better production is observed with shaking rather than standing cultures and in R2A base medium.

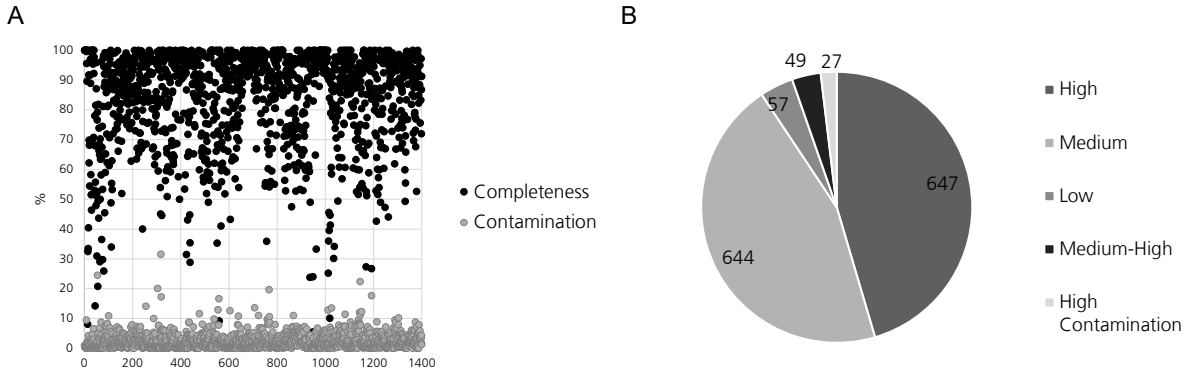

Figure S2-3: CheckM2 Quality Control Data of NCBI-retrieved genomes labelled as belonging to the phylum Acidobacteriota. A) Graph showing data from individual genomes indicating completeness (black) and contamination (grey) values in per cent. B) MIMAG evaluations of retrieved genomes. From 1428 genomes, we retained 647 high-quality genomes for further analysis.

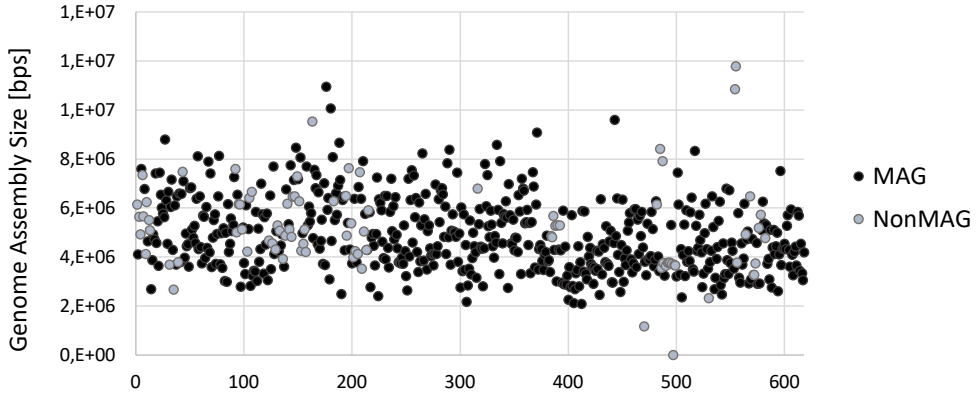

Figure S2-4: Genome Assembly Size of Metagenome-assembled Genomes (black) and Isolate Genomes (Non-MAG, grey). Despite being slightly smaller on average, the genome size distribution of MAGs compared to that of isolate genomes is not statistically significant.

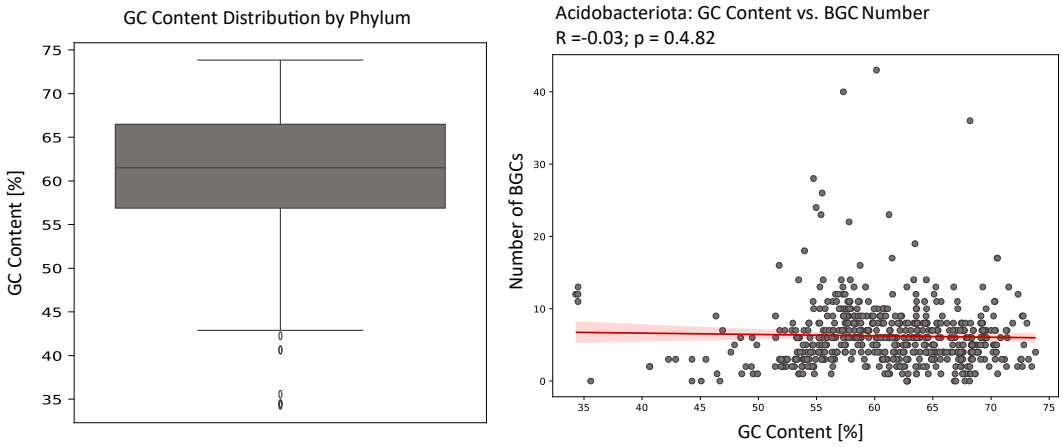

Figure S2-5: Distribution of GC Content (left) and Correlation of GC Content and BGC Number (right) in the Acidobacteriota.

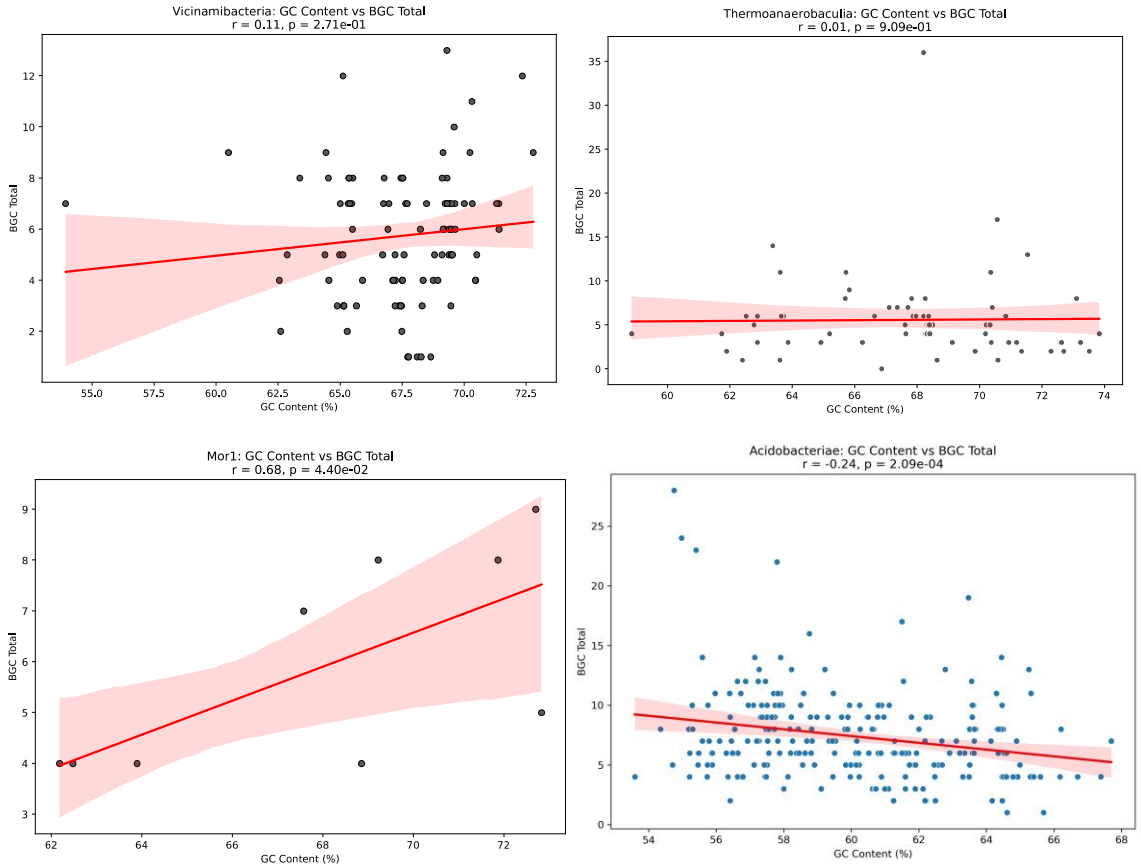

Figure S2-6: Correlation of GC Content and BGC Number in acidobacterial classes (Mor1, *Vicinamibacteria*, *Thermoanaerobaculia* reaching 70% GC content and *Acidobacteriacea*. The Mor1 class shows higher BGC numbers with increasing GC content (p = 0.044), but exhibits an overall average BGC number compared to the rest of the phylum. The BGC numbers are higher in *Vicinamibacteria* and *Thermoanaerobaculia*, but there is no correlation between those and the GC content. In comparison, the class Acidobacteriaceae exhibits a clear negative correlation with genomes characterised by lower GC content, which harbour more BGCs.

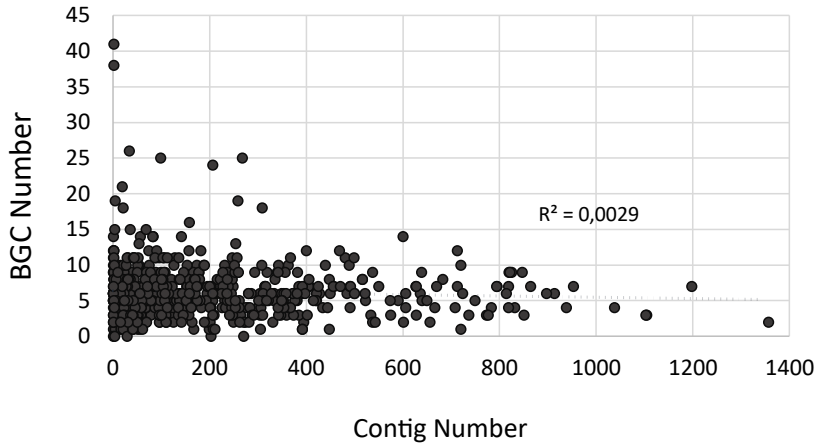

Figure S2-7: BGC Number plotted against the Contig Number of Acidobacteriota Genomes. There is no increase in BGC number with more fragmented genome assemblies (higher contig numbers), indicating that the total number of BGCs is not overestimated due to the presence of fragmented genome assemblies.

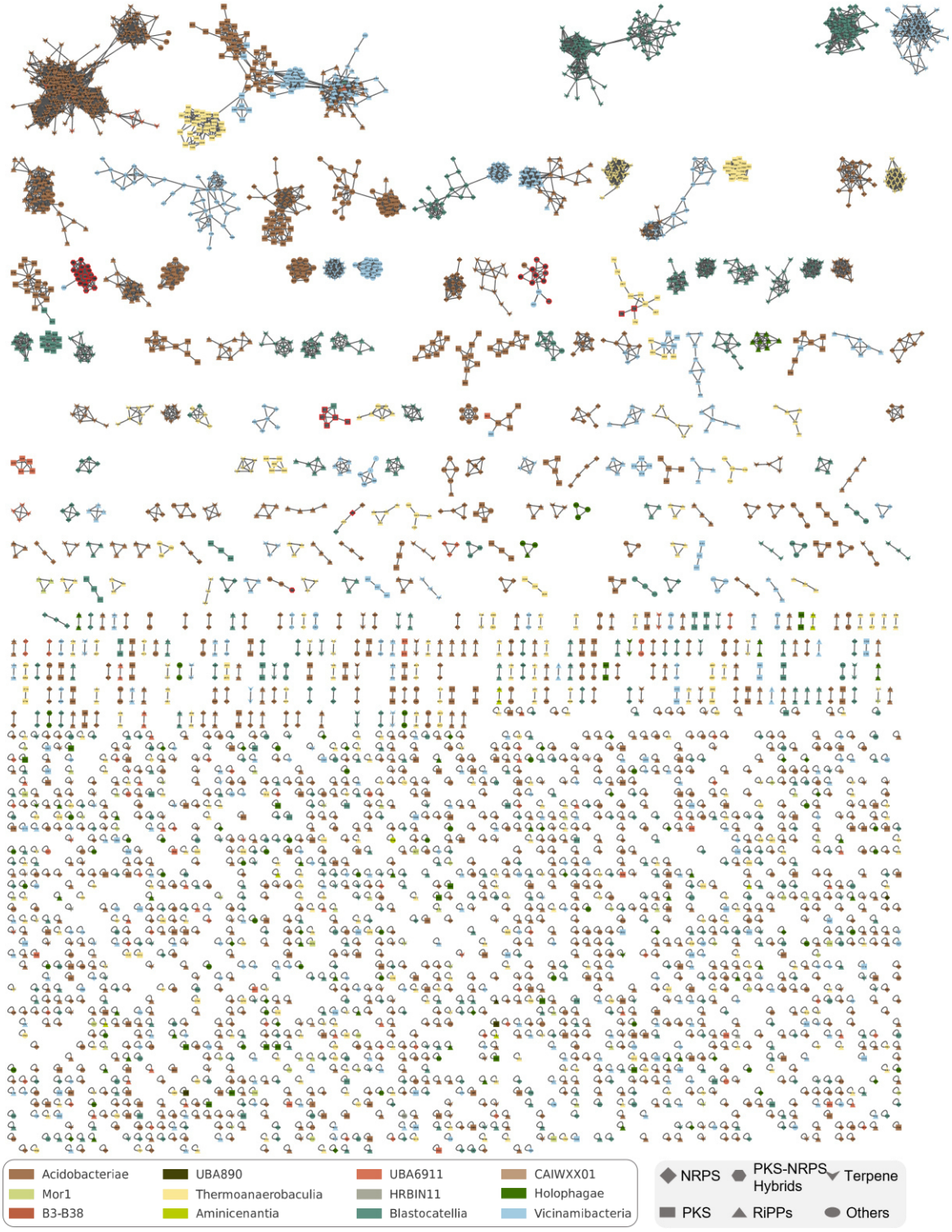

Figure S2-8: BiG-SCAPE Similarity Network of antiSMASH-detected BGCs in the Acidobacteriota dataset at a cutoff value of 0.5. Node colours indicate the taxonomic origin of the detected BGC at the class level (see legend, bottom left side). Node shapes indicate the BGC type assigned by (see legend, bottom right side). Nodes outlined in red correspond to MiBIG reference clusters.

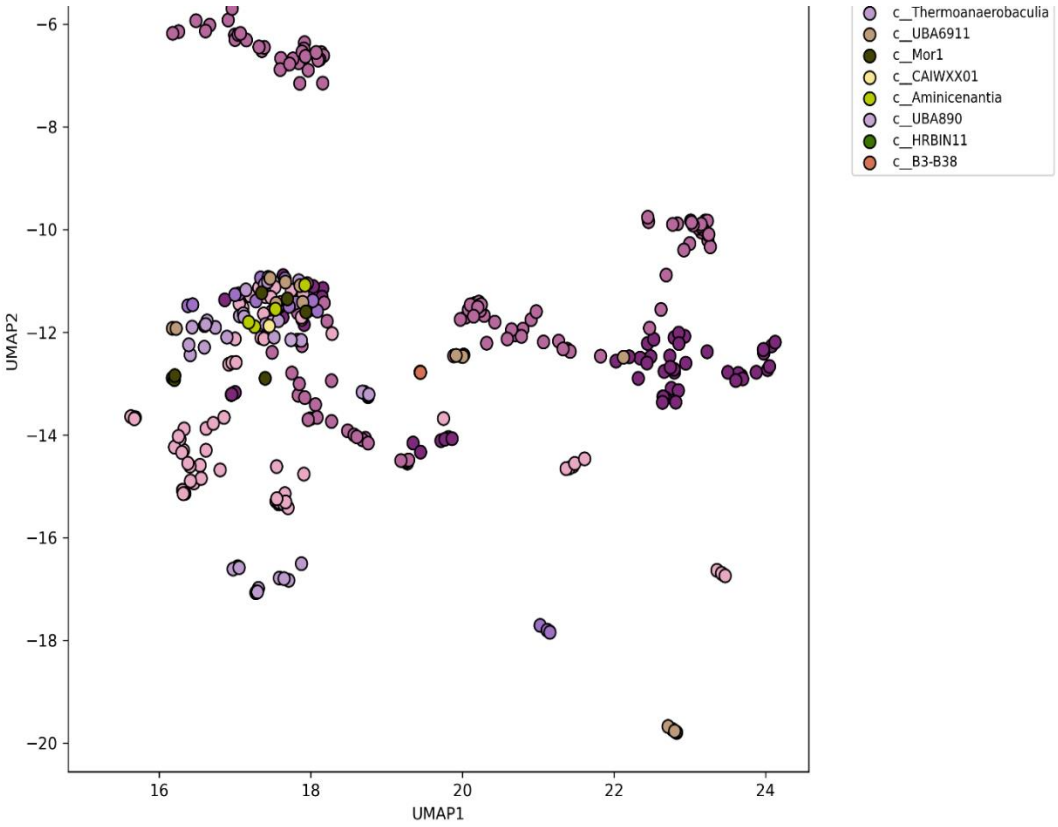

Figure S2-9: UMAP Projection of Biosynthetic Gene Cluster (BGC) Profiles across Acidobacteriota Strains. Each point represents a single microbial strain, positioned based on the cosine similarity of its biosynthetic gene cluster (BGC) composition using Uniform Manifold Approximation and Projection (UMAP). Strains are colour-coded by their taxonomic class to highlight biosynthetic diversity and lineage-specific clustering patterns. The plot reveals distinct groupings, suggesting taxonomically coherent BGC profiles, with some overlap potentially indicating shared ecological niches or horizontal gene transfer.

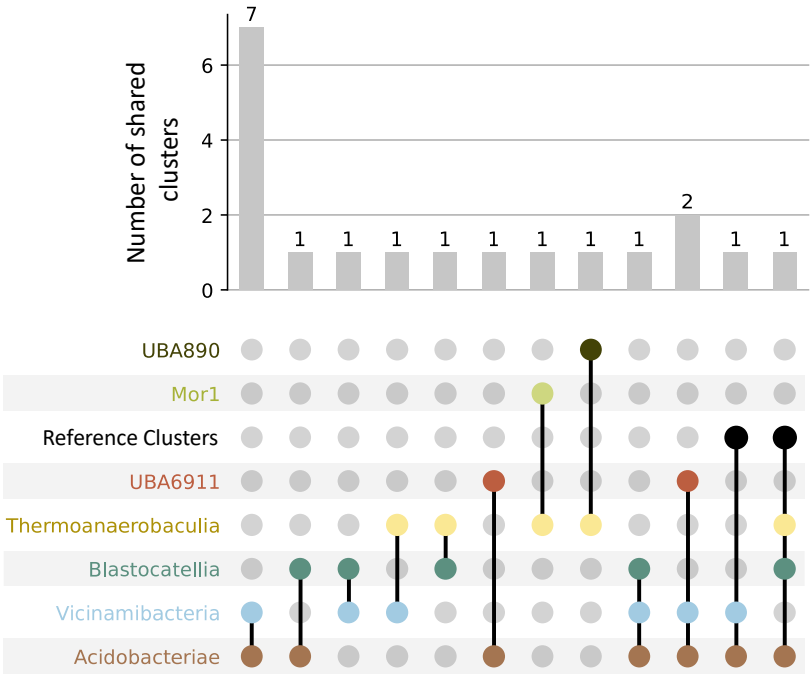

Figure S2-10: Gene Cluster Families present in different Acidobacterial Classes. Bottom: Circles indicate one or more GCFs present in the respective class (left side), lines connect GCFs that are shared between classes. Top: The bar chart indicates the number of GCFs described by the class combinations. The classes *Vicinamibacteria* (blue) and *Acidobacteriae* share the most GCFs (n = 7). Both classes also share two GCFs with the UBA6911 family. Two GCFs are shared between three classes. Both classes are RiPP clusters (data not shown).

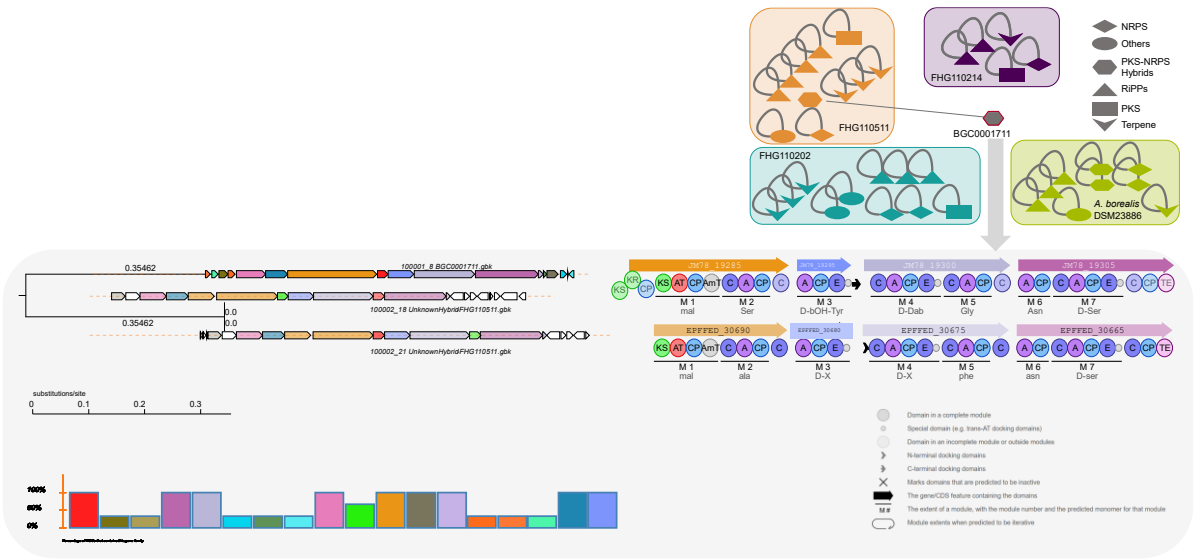

Figure S2-11: BiG-SCAPE Similarity Network of antiSMASH-detected BGCs of the FHG Acidobacteriaceae and *A. borealis* DSM 23886 at a cutoff value of 0.5. MIBiG reference clusters were included to dereplicate similar clusters that had already been described in the literature. The four analysed strains do not share any BGCs and mostly remained as singletons. FHG110511 harbours an NRPS-PKS cluster that shows similarity to the occidiofungin cluster (MIBiG-Reference BGC00001711). Both clusters are PKS-NRPS hybrids, sharing an annotated β-lactamase gene (red) at a central position within the cluster (pairwise identity: 70.4%). Downstream, both clusters harbour a PKS-NRPS hybrid gene, containing the same number of modules, but not every module matches the a-domain selectivity. Upstream, both clusters contain three multimodular NRPS genes, each harbouring the same number of modules. However, only one gene is predicted to share the amino acid selectivity of the A-domains between the two strains. FHG110511 encodes an additional metallohydrolase within the cluster. At the Upstream border of the cluster, both strains harbour a glycosyl transferase, while downstream we find an additional NRPS-like gene involved in fatty acid synthesis.

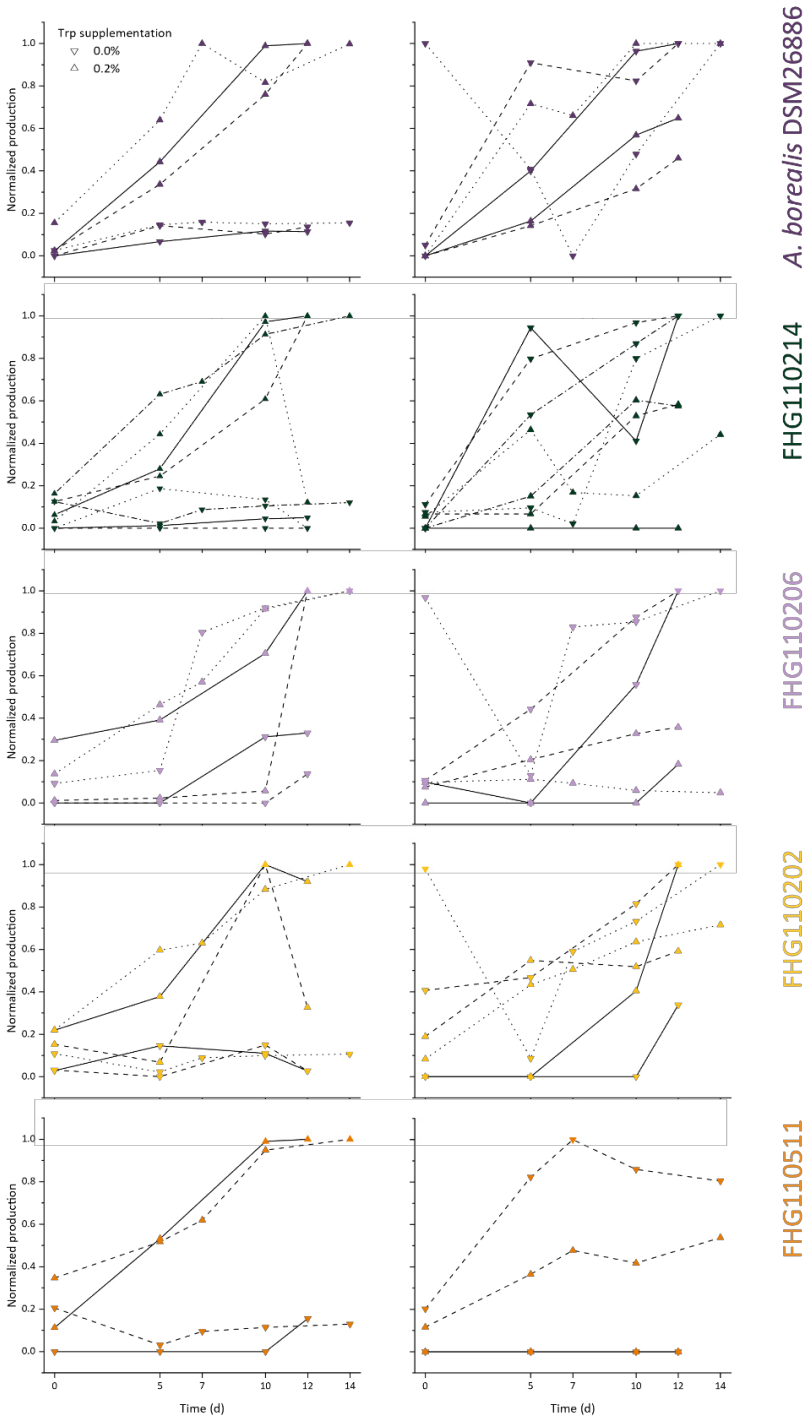

Figure S2-12: IAA (left) and iP (right) Production in Investigated Strains Over Time with (triangle) and without Tryptophan (turned triangle). Available replicates are shown individually to indicate trends. Raw peak areas were normalised within each biological replicate against the highest production of the respective phytohormone. The production of both phytohormones increases over time and typically peaks on the last day of cultivation. While IAA production is consistently increased upon Tryptophan supplementation, the kinetics of iP are less predictable. In general, supplementation with Trp results in a decrease in production, except for strains *A. borealis* DSM 23886 and FHG110202, where this trend is less pronounced.

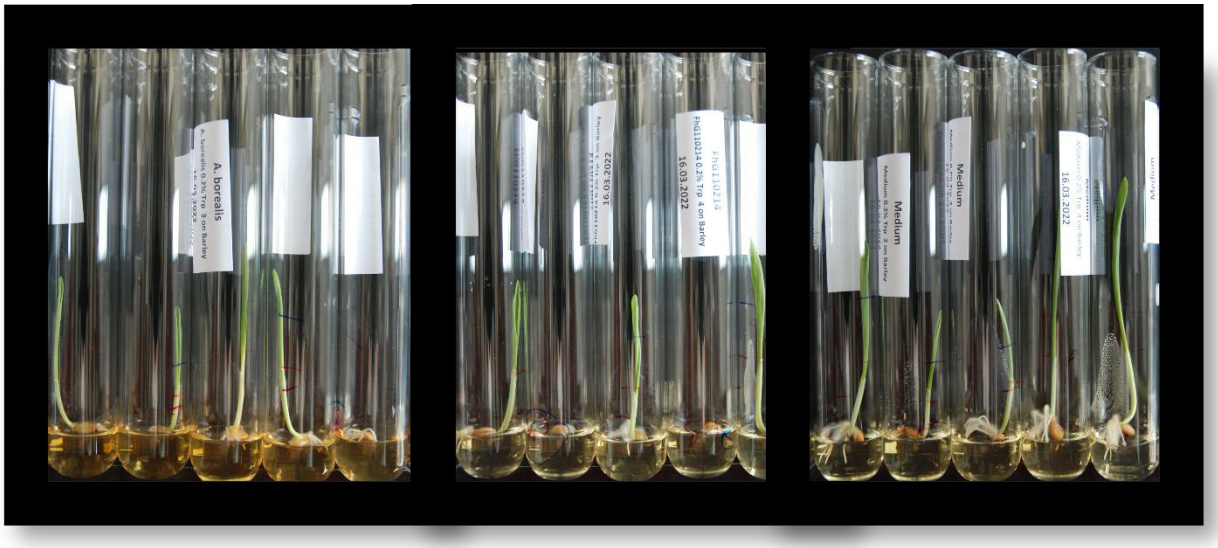

Figure S2-13: Experimental setup for plant-growth assays of barley seedlings. Five replicates of the same condition are shown. Barley seeds are disinfected and grown for 6 days on SH medium that has been pretreated with antibiotics. Extracts from Trp-supplemented cultures were added to the medium as described in the extended Materials and Methods. (A) Extract of *A. borealis* DSM23886 w (B) FHG110214 (C) medium only.

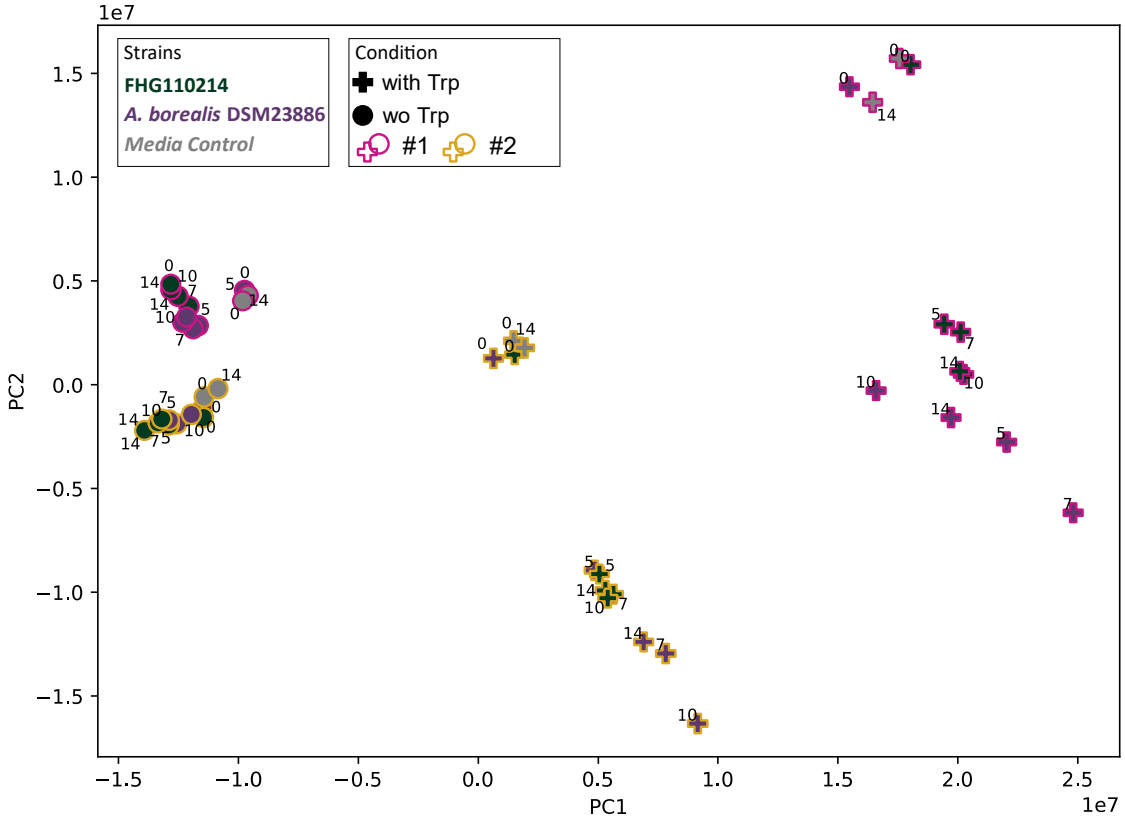

Figure S2-14: Principal component analysis of features annotated from crude extracts of *A. borealis* DSM 23886 and FHG110214 with and without supplementation of Trp. Biological replicates were normalised to an injection volume of 1 µL before PCA analysis. Border colours indicate biological replicates. Purple colours represent *A. borealis* DSM 23886, while green colours represent FHG110214. Cross symbols indicate that strains have been supplemented with 0.2% Tryptophan, while circles refer to the uninduced cultures. Media controls are highlighted in grey. All nodes are labelled with their sampling day after cultivation start at day 0. Clusters are grouped by their supplementation status (plus symbols vs. circles) rather than by strain (colours). Biological replicates are shifted but show the same patterns. Notably, extracts generated at the beginning of the cultivation period consistently cluster with the medium controls. The medium controls do not change over the cultivation course. Temporal shifts are most prominent for A. borealis DSM 23886 with Tryptophan induction, than for FHG110214 or *A. borealis* DSM 23886 without Trp supplementation.

Figure S2-15: Chemotype-barcoding upon Trp induction in strain FHG110214 and *A. borealis* DSM23886. Untargeted UHPLC-QTOF-HR-MS data were processed by dual-threshold bucketing; only buckets present in the 10,000-threshold set were retained. For the 48 samples, 2759 buckets were calculated. Grey barcode rows represent samples without Trp induction, serving as a reference. Induced samples are classified into response categories compared to the respective uninduced sample: extinguished (orange, indicating absence upon Trp), light increase (blue), light reduction (pink), strong increase (green), and strong reduction (red). Consistent with a global metabolic shift, feature counts increased with Trp (e.g., *A. borealis* DSM 23886: 165 → 684; FHG110214: 109 → 337).

6.3 HEL, A Multi-Host Synthetic Biology Platform for BGC Activation

6.3.1 Tables

Table 10:Strains used in this study

Species	Strain	Purpose
Escherichia coli	Epi300	High-copy, standard pCC2Fos
Escherichia coli	DH5alpha	Standard cloning
Escherichia coli	DB3.1	Standard cloning of *ccdB*-containing plasmids
Escherichia coli	Top10	Standard cloning
Escherichia coli	Rosetta	Expression of darobactin, evaluation of *trfA* cassette
Escherichia coli	Rosetta DarR	Darobactin resistant strain
Bacillus subtilis	SCK6	Heterologous host
Pseudomonas putida	KT2440	Heterologous host

Table 11: Vectors used in this study

Lab description/number	Manuscript descripton	Backbone	Resistance/E. coli	Origin
pENTR4	-	-	KanR	Thermo Fisher
puc19	-	-	AmpR	Lab standard
pCC2Fos		-	CmR	Lab standard
pGI10:	HEL-Fosmid, HELFos	pCC2Fos	*ccdB*, CmR	
pGI20	HELEntrEx	pENTR4	*ccdB*, KanR	
pGI7	HEL-Bridge A-H	puc19	AmpR	
pGI8	HEL-bridge B-H	puc19	AmpR	
pGI9	HEL-bridge C-H	puc19	AmpR	
pGI14	HEL-bridge D-H	puc19	AmpR	
pGI11	HEL-bridge E-H	puc19	AmpR	
pGI12	HEL-bridge F-H	puc19	AmpR	

pGI13	HEL-bridge G-H	puc19	AmpR
pGI25	HELFos-GFP	pGI10	CmR, *ccdBR*
pGI33	HEL-Storage AB	puc19	*ccdB*, AmpR
pGI67	HEL-Storage CD	puc19	*ccdB*, AmpR
pGI68	HEL-Storage EF	puc19	*ccdB*, AmpR
pGI69	HEL-Storage GH	puc19	*ccdB*, AmpR
pGI70	HEL-Storage FG	puc19	*ccdB*, AmpR
pGI71	HEL-Storage DE	puc19	*ccdB*, AmpR
pGI60	HEL-Storage BC	puc19	*ccdB*, AmpR
pGI29	HELFosBSub	pCC2Fos	*CmR*, ccdBR
pGI31	HELFosBSub-GFP	pGI10	CmR
pGI34	HELFos-GFP_emptyCtrl	pGI25	CmR
pGI35	HELFos-GFP_weak	pGI25	CmR
pGI36	HELFos-GFP_medium	pGI25	CmR
pGI37	HELFos-GFP_strong	pGI25	CmR
pGI81	HELFos-GFP_pBAD	pGI25	CmR
pGI82	HELFos-GFP_Lux	pGI25	CmR
pGI83	HELFos-GFP_Lac	pGI25	CmR
PGI98	HELFos-daroKan	pGI105	KanR, CmR
PGI105	HELFos-daro	pCC2Fos	ccdB, CmR
PGI106	HEL-Fos-trfAGFP	pGI25	CmR, KanR
PGI117	HELFos-Bsub-GFP	pGI31	CmR
PGI118	HELFos-Bsub-daro	pGI105	CmR

Table 12:HEL Parts used in this study

Lab description/number	BB
pGI51	A-pBAD-B
pGI52	A-pLUX-B
pGI53	A-pLac-B
pGI54	A-Spacer500bps-B
pGI55	A-Spacer333bps-B
pGI57	A-Spacer2500bps-B
pGI58	A-Spacer1500bps-B
pGI61	B-J23100-C
pGI62	B-J23106-C
pGI63	B-J23117-C
pGI73	A-J23100-B
pGI74	A-J23106-B
pGI75	A-J23117-B
PGI88	B-oriT-C
PGI92	C-aac-D
PGI93	B-KanR-C
PGI94	D-pBAD-E
PGI95	D-T7lacI-E
PGI99	E-trfAcass-F
PGI100	F-pBAD-G
PGI101	F-pLAC-G
PGI102	F-LUX-G
PGI103	F-T7lacI-G
PGI113	F-Pml-XylS-G
PCR-based Parts	
HEL8	A-Empty control-H
HEL10	A-J23100-H
HEL13	A-J23106-H

HEL16	A-J23117-H
HEL22	A-amyE part1-B
HEL33	A-pBAD-B
HEL34	D-pBAD-E
HEL36	A-pLAC-B
HEL39	A-pLUX-B
HEL47	B-J23100-C
HEL48	B-J23106-C
HEL49	B-J23117-C
HEL50	A-Spacer2500-B
HEL51	A-Spacer1333-B
HEL52	A-Spacer550-B
HEL53	A-Spacer333-B
HEL11	C-pveg-H

6.3.2 Supplementary Figures

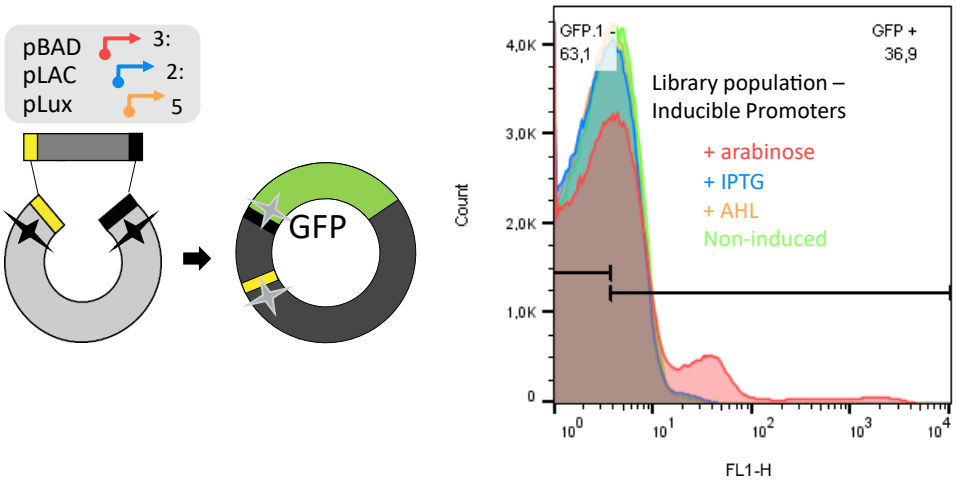

Figure S3 1: Cloning design for HELFos-GFP retrofitted with inducible promoter libraries (left). The obtained transformants were subsequently analysed using Flow cytometry (right). The x-axis represents the fluorescent signal, while the y-axis indicates the cell counts. Only the induction with IPTG and arabinose showed a fraction of cells indicating a GFP signal. Most of the cells stayed non-fluorescent.

Figure S3 2: *B. subtilis* SCK6 overnight cultures grown in Spizizen medium carrying different genomic integrations. Left: The wild type strain, middle: the strain carrying the Bacillus-adapted GFP cassette, right: the strain carrying the *Bacillus*-adapted darobactin cassette.

7 Overview of Manuscripts and Contributions

7.1.1 Genome Mining of Bacteroidota

Manuscripts

- ◊ Genome Mining on novel Bacteroidota strains: 4 manuscripts – 3 Shared first Authorship, 1 first authorship
- ◊ A genetically tractable branch of environmental *Pedobacter* from the phylum Bacteroidota represents a hotspot for natural product discovery – Shared Correspondence

Outcomes

⇒ Standardised, collection-scale workflow for assembled genomes covering the complete process from quality and taxonomy assignment to BGC prediction, similarity clustering, descriptive analysis, and visualisation.

⇒ Description of novel isolates from the Bacteroidota family, including Pedobacter strains from amphibian specimens, expanding reference diversity and showcasing the genomic workflow.

⇒ Prioritised *Pedobacter* clusters for metabolomics-guided isolation, leading to novel Cryopeptin derivatives and a new metabolite-BGC linkage in the genus.

7.1.2 Profiling of Novel Acidobacteriota

Manuscripts and preprints

- ◊ High-throughput cultivation for the selective isolation of Acidobacteria from termite nests as a potential new source of natural products – Co-authorship
- ◊ Exploring the Biosynthetic and Plant Growth-Promoting Potential of Novel Acidobacteriota Strains – First Authorship, (preprint)

Outcomes

⇒ Description of novel Acidobacteriota strains from termite nest and soil material, using genome sequencing, OSMAC cultivation, and antimicrobial screenings.

⇒ Global metabolome profiling under Tryptophan supplementation, including the identification and isolation of antifungal-associated features.

⇒ Metabolic mining for a set of phytohormones leading to the absolute quantification of phytohormones IAA and iP.

⇒ Establishment of a genomic workflow globally analysing the presence of PGPT traits in large sets of genomes based on the PLABASE database.

⇒ Experimentally evaluated whether PGPT-associated traits translate to plants.

7.1.3 HEL, A Multi-Host Synthetic Biology Platform for BGC

⇒ Design of a vector set able to allow retrofitting of a BGC containing formid with a) Expression modification cassettes (library-style) b) Chassis-specific compatibility cassettes (constant)

8 ACKNOWLEDGEMENTS

I would like to thank **Prof. Till Schäberle**, who created an open, good-humoured environment that supported genuine scientific independence. His curiosity—always inviting us to look beyond one's own horizon- set a tone of respectful freedom that I sincerely appreciate. I also thank **Prof. Sylvia Schnell** for reviewing my thesis and granting me access to her laboratory; the quiet evenings under the greenhouse lights remain treasured memories. Thanks to **Prof. Maren Ziegler** and **Prof. Martin Rühl** for taking the time to serve on the examining committee. For careful readings of the thesis, thank you, Ignazio, Marius, Maria and Lennart.

This work was supported by the Christiane Nüsslein-Volhard (CNV) Foundation and the Promotionsabschlusstipendium der Justus-Liebig-Universität Gießen. I truly appreciate both the financial support and the personal encouragement.

This thesis rests on everyday support.

Marius, day-to-day supervisor and a brilliant tornado, our combination may not have added to duration, but certainly to joy.

To my working group for support that went well beyond scientific input: I will long remember the conference trips, camping, and get-togethers. Former and current PhD colleagues, beyond our group, made being a PhD student far more enjoyable. I am grateful to the students I worked with—Jil, Mi, and Walter—for their curiosity and persistence.

Papa, Mausi, Natascha, Herbert, Mama, Cedric, Rahel, Luis and Lisa—thank you for your continuous love and support that doesn't need many words.

Philipp, thanks for accompanying and supporting me almost my entire life and for being the best father to our daughter; I'm grateful that we keep building the family life that works for all of us.

Lennart, thank you for your passion, both in work and beyond, along with your unconditional acceptance, and a well that holds endless empathy and kindness for me and for the world around us.

To the strong women in my professional and private life—Natascha, Sanja, Maria, Joana, Arzu, Desi and many others—who speak up when things don't work as they should, who organise space for discussion and for work, and who still find the energy to support others even when their own reserves are low: our collective power (and yes, sometimes anger) has been a source of motivation. I value it deeply.