University of South Bohemia České Budějovice

Faculty of Science

THE SECRETS ENCODED IN THE DRAFT GENOME OF *Lentzea* sp. STRAIN BCCO 10_0061 ISOLATED FROM RECLAIMED MINE HEAPS

Bachelor's thesis

Keller Moritz

Supervisor: Ana Catalina Lara Rodriguez, PhD

Co-supervisor: Erika Corretto, PhD

Institute of Soil Biology, Biology Centre CAS

České Budějovice

2022

Keller M., 2022: The secrets encoded in the draft genome of *Lentzea* sp. BCCO 10_0061 isolated from reclaimed mine heaps. BSc. Thesis, in English 46 pages, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation

Many Actinobacteria including *Lentzea* have been found to produce medically viable compounds like antibiotics and antitumor agents.

This thesis deals with genome sequencing of *Lentzea* sp. BCCO 10_0061, assembly, annotation and evaluation of the potential for secondary metabolite production, as well as phylogenetic classification and creation of a metabolic profile.

Declaration

I declare that I am the author of this qualification thesis and that in writing it I have used the sources and literature displayed in the list of used sources only.

Place, date.

Student's signature

List of Abbreviations

aa	amino acid(s)
ANI	average nucleotide identity
atpD	gene for ATP synthase subunit beta (protein)
BCCO	Biology Centre Collection of Organisms
BGC	biosynthetic gene cluster
BLAST	Basic Local Alignment Search Tool
bp	base pair(s)
CAS	Czech Academy of Science
CCSACB	Culture Collection of Soil Actinomycetes České Budějovice
CDS	coding DNA sequence
dDDH	digital DNA-DNA hybridisation
GC content	content of guanosine and cytosine in the DNA
gyrB	gene for DNA gyrase subunit B (protein)
iso-MGS	iso-migrastatin
MUSCLE	MUltiple Sequence Comparison by Log- Expectation (alignment algorithm)
NCBI	National Center for Biotechnology Information
NRPS	nonribosomal peptide synthetase
ORF	open reading frame
PKS	polyketide synthase
rpoB	gene for DNA-directed RNA polymerase subunit beta (protein)
rRNA	ribosomal ribonucleic acid
STAG	studijní agenda univerzity - university study agenda (online platform)
tRNA	transfer ribonucleic acid
UPGMB	variation of UPGMA (Unweighted Pair Group Method with Arithmetic mean)

1 Table of content

2 Abstract

With the rise of multidrug-resistant pathogens, humanity is in a constant race for new antibiotics and bioactive compounds. Many groups of Actinobacteria, have been found to produce interesting and unique antibiotic and antitumor agents. However, previous studies mainly focused on *Streptomyces* spp. In the present study, we present the genetic potential of the strain *Lentzea* sp. BCCO 10_0061 and investigate its metabolic capabilities using solely bioinformatic resources. For this purpose, a high-quality assembly of 23 contigs with a genus-typical size of 10.3 Mb and a GC content of 68.86% was obtained. Multi-locus phylogenetic analysis revealed a relation to *Lentzea albidocapillata.* AntiSMASH analysis to explore metabolic capabilities detected 30 possible biosynthetic gene clusters (BGC) with candidates for iso-migrastatin, an inhibitor of cancer cell migration, and nystatin A1, an antifungal agent, being the ones we chose to explore in depth. This study generated a nearly complete genome and high quality assembly (contig level) of *Lentzea* sp. BCCO 10_0061, that allowed insights into its metabolic capabilities and therefore cemented the basis for future studies on functional analysis of the proposed BGCs.

Keywords:

Actinobacteria, *Lentzea*, genome assembly, iso-migrastatin, nystatin A1, biosynthetic gene clusters

3 Introduction

New therapeutically interesting compounds are urgently needed to combat life threatening diseases like infections by antibiotic resistant pathogens or cancer (Rybak, 2004; Talbot et al., 2006; Boucher et al., 2009). Despite great technological advancements in pharmaceutical engineering from the last 50 years, it is still true that the most promising source of new drugs are natural products mainly synthesized by microorganisms such as Bacteria and Fungi (Kaitin, 2010; Lewis, 2017; Luepke et al., 2017). Pharmaceutical research history has shown that new important natural products can be found when new screening systems are improved by implementing the latest high quality biological knowledge into the existing discovering pipelines (Romesberg and Craney, 2016; Lewis, 2017).

How to choose a bacterium for pharmacological screening programs is a daunting prospect given the taxonomical diversity among Bacteria. However, members of the order Actinomycetales remain the richest source of natural products. Actinomycetales produce around 45% of all microbial bioactive secondary metabolites with 80% of them (around 7,600 compounds) being synthetized just by the genus *Streptomyces* (Valli et al., 2012). The application of genomic technologies has opened the possibility of testing new strains for the production of bioactive compounds and streamlining into the lab the most promising ones (Abdelmohsen et al., 2015; Lewis, 2017; Li et al., 2022).

The phylum Actinobacteria comprises organisms of agricultural, biotechnological, and ecological importance. They are present in high abundance in soils and are able to produce a wide variety of secondary metabolites (Nouioui et al., 2018) as well as to degrade several polymeric carbohydrates (Maiti and Mandal, 2022). They are gram positive, strictly aerobic and form abundant aerial hyphae (Yassin et al., 1995). Recently, compounds produced by *Lentzea* spp. have been suggested as medically important (Li et al., 2022), thus further increasing the interest in genetic data from this rare Actinomycete.

3.1 The genus Lentzea as source of bioactive secondary metabolites

Lentzea is a genus of the phylum Actinobacteria. The genome size varies between 8.64 to 10.81 Mb (NCBI – Genome, accessed 12/09/21) and is characterized by a G+C content that ranges between 68.6-79.6% (Fang et al., 2017). Its phenotypical characteristics include vegetative branched mycelia and aerial mycelia that fragments into rod-shaped spores. Meso-diaminopimelic acid form the cell-wall peptidoglycan; while galactose, ribose and mannose are the main sugars resulting from a whole-cell extraction. The major cellular fatty acids are iso-C14:0 and iso-C16:0. The phospholipids are diphosphatidylglycerol, phosphatidylethanolamine, hydroxyl-phosphatidyethanolamine, phosphatidylinositol, phosphotidylinositolmannosides. *Lentzea* belongs to the family *Pseudonocardinaceae*, and is closely related to *Saccharothrix*, *Kutzneria* and *Actinosynema* (Figure 1). The branching characteristic of this family consists in the presence of menaquinone MK-9(H4) in the fatty acid extracts (Yassin et al., 1995).

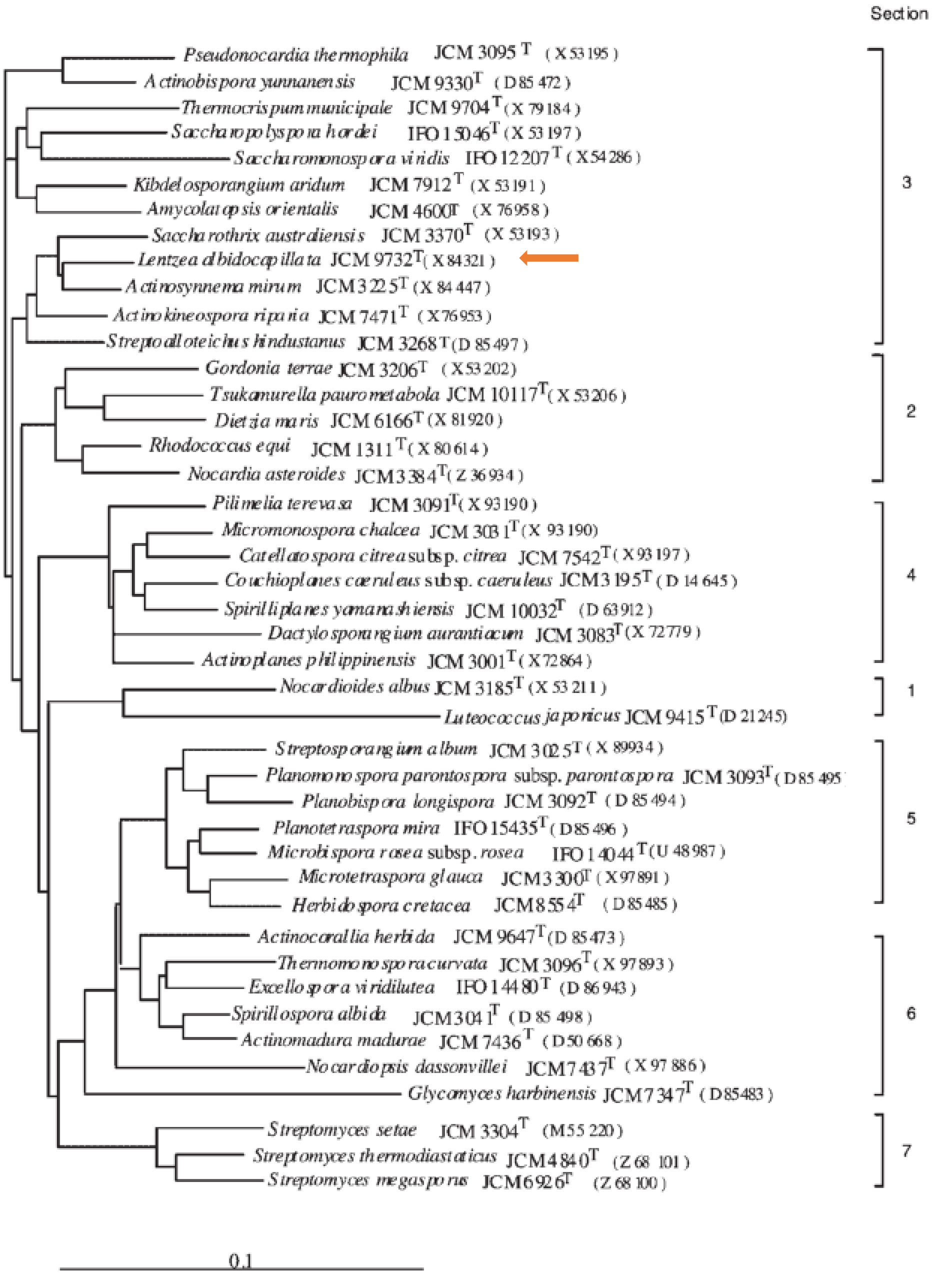

Section
Pseudonocardia thermophila JCM 3095T (X 53 195)
Actinobispora yunnanensis JCM 9330T (D 85 472)
Thermocrispum municipale JCM 9704T (X 79 184)
Saccharopolyspora hordei IFO 1 5046T (X 53 197)
Saccharomonospora viridis IFO 12 207T (X 54 286)
Kibdelosporangium aridum JCM 7912T (X 53 191)
Amycolatopsis orientalis JCM 4600T (X 76 958)
Saccharothrix australiensis JCM 3370T (X 53193)
Lentzea albidocapillata JCM 9732T (X 84 321)
Actinosynnema mirum JCM 3225T (X 84 447)
Actinokineospora riparia JCM 7471T (X 76 953)
Streptoalloteichus hindustanus JCM 3268T (D 85 497)
Gordonia terrae JCM 3206T (X 53 202)
Tsukamurella paurometabola JCM 10117T (X 53 206)
Dietzia maris JCM 6166T (X 81920)
Rhodococcus equi JCM 1311T (X 80 614)
Nocardia asteroides JCM 3384T (Z 36 934)
Pilimelia terevasa JCM 3091T (X 93 190)
Micromonospora chalcea JCM 3031T (X 93 190)
Catellatospora citrea subsp. citrea JCM 7542T (X 93 197)
Couchioplanes caeruleus subsp. caeruleus JCM 3195T (D 14 645)
Spirilliplanes yamanashiensis JCM 10032T (D 63 912)
Dactylosporangium aurantiacum JCM 3083T (X 72 779)
Actinoplanes philippinensis JCM 3001T (X 72 864)
Nocardioides albus JCM 3185T (X 53 211)
Luteococcus japonicus JCM 9415T (D 21245)
Streptosporangium album JCM 3025T (X 89934)
Planomonospora parontospora subsp. parontospora JCM 3093T (D 85 495
Planobispora longispora JCM 3092T (D 85 494)
Planotetraspora mira IFO 15435T (D 85 496)
Microbispora rosea subsp. rosea IFO 14044T (U 48 987)
Microtetraspora glauca JCM 3300T (X 97 891)
Herbidospora cretacea JCM 8554T (D 85 485)
Actinocorallia herbida JCM 9647T (D 85 473)
Thermomonospora curvata JCM 3096T (X 97 893)
Excellospora viridilutea IFO 14480T (D 86 943)
Spirillospora albida JCM 3041T (D 85 498)
Actinomadura madurae JCM 7436T (D 50 668)
Nocardiopsis dassonvillei JCM 7437T (X 97 886)
Glycomyces harbinensis JCM 7347T (D 85483)
Streptomyces setae JCM 3304T (M 55 220)
Streptomyces thermodiastaticus JCM 4840T (Z 68 101)
Streptomyces megasporus JCM 6926T (Z 68 100)
3
2
4
1
5
6
7
0.1

Figure 1 Neighbor-joining tree based on 16S rRNA gene sequences of selected Actinomycetes. The nucleotide sequence accession numbers are in parentheses. Section 1: Micrococcus, Microbacterium, and reated genera. Section 2: Mycobacterium, Nocardia, and related genera. Section 3: Family Pseudonocardiaceae. Section 4: Family Micromonosporaceae. Section 5: Family Thermomonosporaceae. Section 6: Family Streptosporangiaceae. Section 7: Family Streptomycetaceae. (modified from Jarerat, Pranamuda and Tokiwa, 2002) The orange arrow indicates a representatiove of the genus Lentzea.

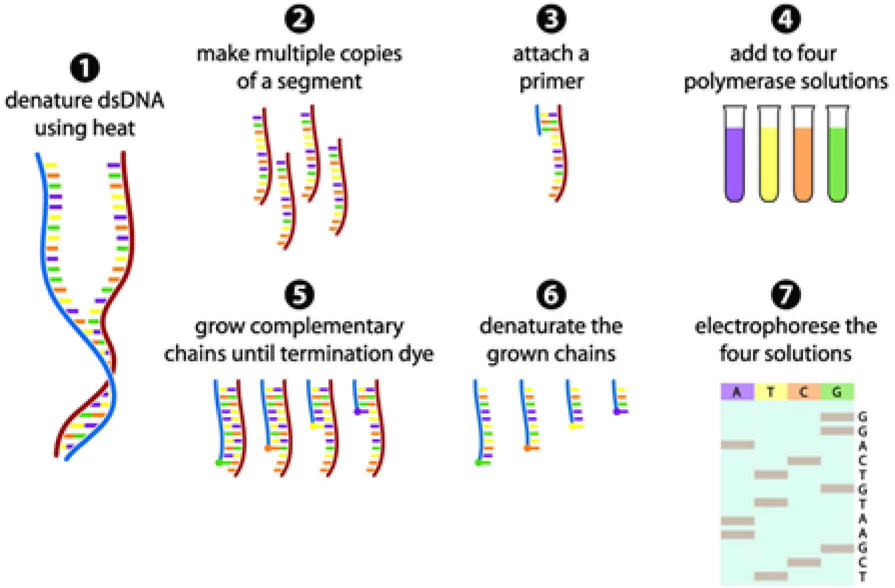
Lentzea strains have mostly been isolated from soil samples. Many originated from locations of relatively stable humidity and temperature such as mines or caves (Fang et al., 2017), but there are also some extremotolerant and extremophile species (Wichner et al., 2017).

Lentzea is a group for which little data is available, especially regarding its genetic characteristics. For example, there are currently (January 2021) only 17 fully available genomes of *Lentzea* spp. and 16 of them are on the contig or scaffold level, lacking essential elements like the 16S rRNA and some housekeeping genes (NCBI – Genome, accessed 12/09/21).

This thesis focused on a promising strain, *Lentzea* sp. BCCO 10_0061, which was isolated from a mining heap. It is known that these environments force the microorganisms to adopt surviving strategies such as the production of active metabolites that allow and facilitate their life and proliferation under extreme conditions (i.e. presence of acidic, oxidative and alkaline agents used for mining) (Petersen, 2016). BCCO 10_0061 is deposited in the Culture Collection of Soil Actinomycetes České Budějovice. The Actinomycetes collection was founded in 2007 at the Institute of Soil Biology, Biology Center CAS and "serves as a depository for environmental and clinical Actinomycetes with a focus on research, biotechnologies, medicine and other practical applications" (www.actinomycetes.cz, accessed on 15.10.22).

3.2 Genome sequencing

The first method of sequencing was introduced by Frederick Sanger in 1977 (Figure 2). It uses radioactively-labelled di-deoxynucleotide triphosphates, which were utilized for chain termination (Genomics and Schroeder, 2022).



❶
denature dsDNA using heat
❷
make multiple copies of a segment
❸
attach a primer
❹
add to four polymerase solutions
❺
grow complementary chains until termination dye
❻
denaturate the grown chains
❼
electrophorese the four solutions
A
T
C
G
G
G
A
C
T
G
T
A
A
G
C
T

Figure 2 Schematic representation of Sanger sequencing (from Genomics and Schroeder, 2022).

One of these terminating nucleotides (ddA, ddT, ddG and ddC) were each used in one reaction mixture with other deoxynucleotide triphosphates, a polymerase, a DNA primer and a template to generate DNA fragments of different lengths. Then electrophoreses of all four samples was performed to reconstruct the correct DNA sequence based on the succession of the resulting bands, which could be identified due to the nucleotide specific radioactive labels. Some improvements on this method were made by Leroy Hood and Michael Hunpiller in 1987 by using fluorescence-labelled ddNTPs, which allowed to perform the whole analysis in one vial and automate the detection (Genomics and Schroeder, 2022).

Next generation sequencing represented a major step forward, since it allows multiple reactions to take place simultaneously, therefore producing huge amounts of data in a short time. The first of these methods was pyrosequencing by Mostafa Ronaghi, Mathias Uhlen and Pal Nyren and is based on detection of luminescence from pyrophosphate, which is formed during the synthesis (Genomics and Schroeder, 2022).

The next generation sequencing platform used for this study was Illumina Miseq platform, which is based on fluorescently-labelled, reversibly-terminating nucleotides. The main steps consist of library preparation, cluster generation, sequencing, and data analysis (Illumina.com - An introduction to Next-Generation Sequencing Technology, accessed 01/09/21) (Figure 3).

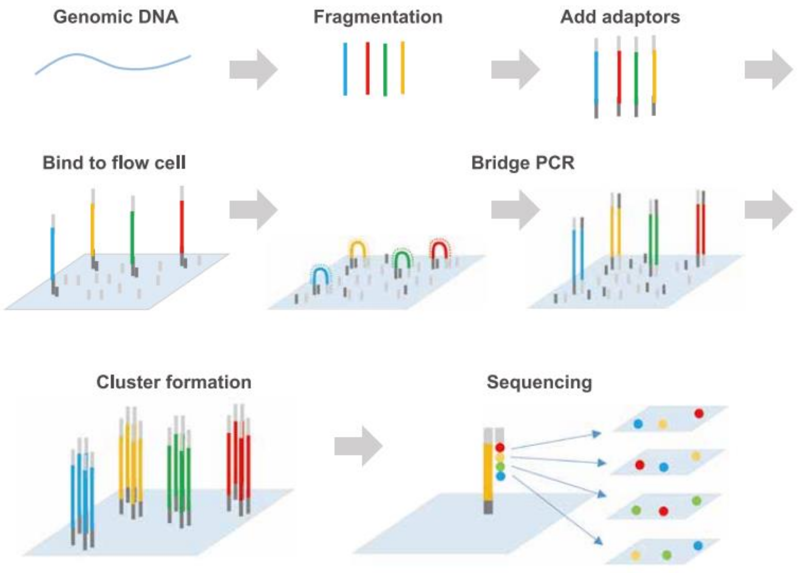


Figure 3 Illumina sequencing (taken from GenScript.com - Advancing genomics, medicine and health together – by semiconductor DNA synthesis technology, accessed 22/04/22)

After extraction, the DNA is fragmented, and adapters are ligated. These contain the sequencing primer binding sites, indices, and complementary regions for hybridisation with the flow cell oligos (Illumina.com - Next-generation sequencing for beginners, accessed 01/09/21). The library is then loaded into a flow cell and the adapters hybridize onto the flow cell oligos. The DNA fragments then bridge to adjacent oligos and are amplified forming clonal clusters. All reverse strands are then cleaved and removed. Then sequencing can begin using reversible-terminator, fluorescently labelled nucleotides, which are competing for incorporation (Illumina.com - An introduction to Next-Generation Sequencing Technology, accessed 01/09/21). After each incorporation the fluorophores are excited, the signal is recorded, and the terminating ends are cleaved. This is done simultaneously for all clusters on the flow cell, resulting in fast data acquisition. After the first reading cycle, another bridge amplification is done to generate the reverse strand which is similarly sequenced. Forward and reverse reads are paired based on the indices and contiguous reads can be generated (Illumina.com - Next-generation sequencing for beginners, accessed 01/09/21). This allows for longer and more accurate reads. The MiSeq platform produces approximately five million paired end reads with a length of 150 bp within a one-day experiment, which is sufficient data for the assembly of small genomes (Caporaso et al., 2012).

3.3 Genome assembly and annotation

Before starting with the genome assembly, the quality of the reads should be checked and if needed, the quality can be improved by trimming or deleting low-quality reads, by removing leftover sequencing adapters and accidental contamination with different sources such as human or mouse. These contamination reads can be identified by mapping all the reads to a reference genome (Riehl, B., Chivian, D., Canon, S. & Land, M., 2018). Afterwards, the cleaned and improved reads can be assembled into a complete chromosome or contigs. The term coverage describes the number of unique reads aligned to a certain part of the genome and can be calculated by aligning the reads to the assembly. Contigs with low mean coverage (< 20x) are usually considered unreliable (Brunstein, 2014) and are often the result of degraded input DNA, homologous regions, low complexity regions and high GC content (Thermo Fisher Scientific Inc., 2019). In general, the mean coverage for a good de-novo assembly of bacterial genomes should be around 50x (Illumina Inc., 2010).

KBase is an open-source platform, where data can be uploaded and analysed by multiple tools and compared to the built-in database of genomes and biochemistry. The platform also focuses on the possibility to share and discover research and allows multiple, remote parties to work together on one project (Arkin et al., 2018).

Within this platform, quality control of the reads can be realised by using tools to analyse raw sequence data from high throughput sequencing like FastQC (Babraham bioinformatics - FastQC A quality control tool for high throughput sequence data, accessed 12/06/22). The analysis returns multiple results such as some basic statistics like encoding, total sequences, number of sequences flagged as poor quality, sequence length and GC content as well as per base sequence quality, per sequence quality scores, per base sequence content, per sequence GC content, per base N content, sequence length distribution, sequence duplication levels, overrepresented sequences, and adapter content. Tools like this can be used to assess, if further read processing is necessary, before the assembly process can be started.

For assembly, SPAdes assembler was used. It was designed especially for single-cell sequencing and was thereby made to be robust against non-uniform read coverage and higher levels of sequencing errors. However, it was later proven to be just as valuable for multi-cell isolates (Prjibelski, accessed 12/04/22).

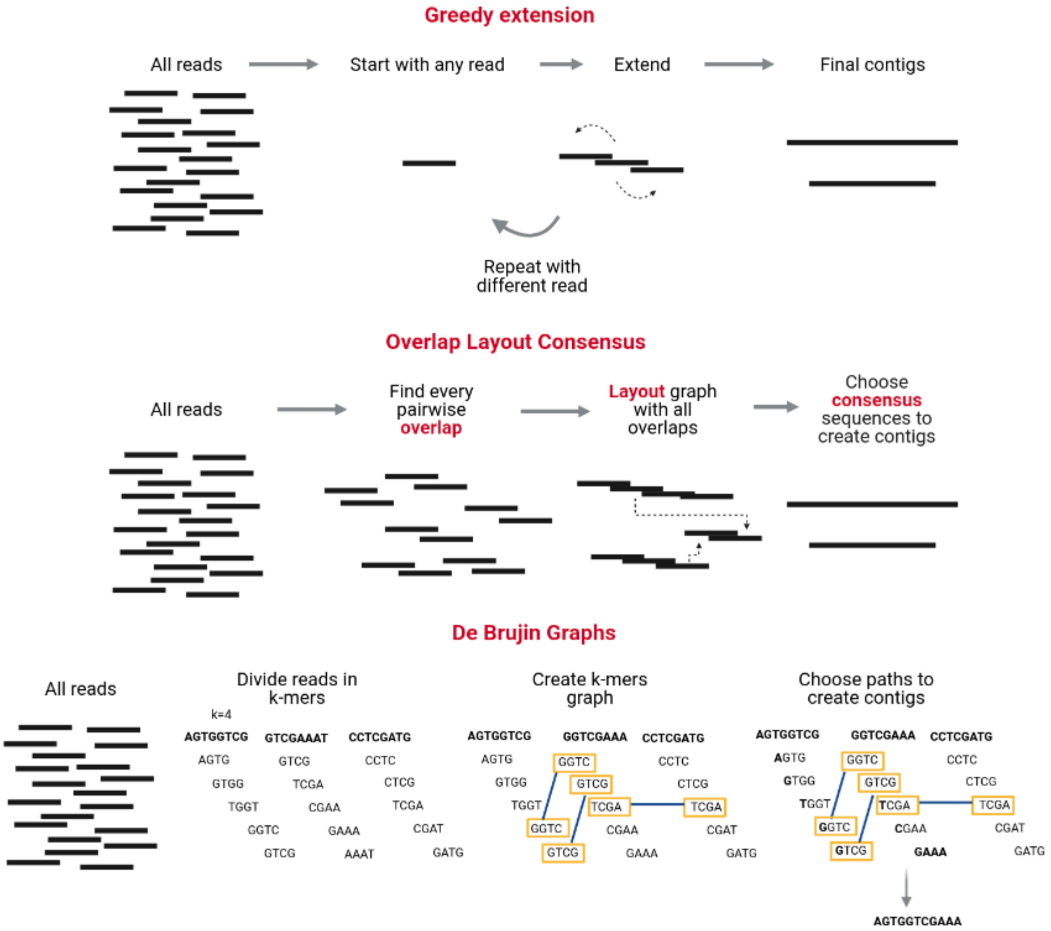


Figure 4 Genome assembly method schematic (taken from Github.io – Metagenome assembly, accessed 16/04/22)

SPAdes assembly is based on the “de Bruijn graph”-approach (Figure 4). Reads are divided into all possible k-mers (k typically is around 55 in modern NGS) and then aligned based on k-mer identity (Compeau, Pevzner and Tesler, 2011). This generates a de Bruijn graph, which is then simplified, yielding the assembly graph. Then mate-pair reads are aligned and a universal repeat resolving- and scaffolding algorithm is applied, which returns the contigs (Prjibelski, accessed 12/04/22).

Completeness and contamination of the assembly is controlled by CheckM, which uses a broad range of lineage-specific, collocated marker genes to assess of genomes by determining their presence (also multiple times) and absence (Parks et al., 2015). For Actinomycetales, a set of 368 markers are checked (Parks et al., 2015).

Genes can be predicted and assigned a function through annotation. Numerous software options are available. However, they work based on similar principles: identifying regions that do not code for protein, identifying the open reading frames (ORF) and then assigning a function by comparison. In this study, we used two different programs RAST and Prokka and compared their annotations of the available *Lentzea* genomes. RAST has a database with groups of proteins having similar function called FIGfams. These groups are curated by experts or determined computationally based on the sequence identity and function. Genes can then be assigned a function based on similarity and relationships (Aziz et al., 2008). First, tRNA and rRNA genes are identified, then all putative protein-encoding regions are determined and labelled as putative genes. A small set of nearly universal genes (for example the tRNA synthetases) are pinpointed along the genome and used to find the closest relatives. To save computational power, first all remaining putative genes are checked against FIGfams occurring in these relatives, then against all FIGfams and the remaining putative proteins are blasted against a non-redundant protein database. Finally, an initial metabolic reconstruction is made by grouping genes into functional subsystems. In contrast, Prokka starts by predicting the coding sequences and then hierarchically compares them to databases of greater size and lower trustworthiness and finally, assigns functions based on e-values with a threshold of 10^{-6} (Seemann, 2014).

3.4 Prediction of secondary metabolite gene clusters

AntiSMASH is used to predict biosynthetic gene clusters and possible structures of the produced secondary metabolites. The pipeline translates all protein-coding genes and searches them with profile Hidden Markov Models based on multiple sequence alignment. Specific models for each BGC group exist (Medema et al., 2011). After the identification of the group, the cluster is compared to a database of already known clusters and a similarity score is assigned (Blin et al., 2021). Based on this score, a compound with similar structure and function may be produced by the organism. To confirm the chemical structure and biological function of the selected gene cluster, the compound needs to be produced in the lab, isolated and analysed with mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR) techniques to elucidate the structure. The biological functions like antifungal or antibacterial properties can be tested firstly with *in vitro* tests such as the plate diffusion technique (Berkow, Lockhart and Ostrosky-Zeichner, 2020) or more advanced microfluidic cell culture systems to test for antimetastatic properties (Kitaeva et al., 2020).

4 Work aim

Lentzea is a poorly researched and rare genus of *Actinomycetes*, which are well known producers of many medically relevant compounds such as antibacterial agents (erythromycin, streptomycin, and kanamycin); antifungal agents (nystatin); immunosuppressants (rapamycin); antiparasitic, anti-lymphatic filariasis, and anti-onchocerciasis agents (ivermectin) (Quinn et al., 2020).

The goal of this thesis is to generate a whole-genome assembly of *Lentzea* sp. BCCO 10_0061, provide a genetic description and compare housekeeping genes (16S rRNA, *atpD*, *rpoB* and *gyrB*) of selected strains to assess the strain's phylogenetic placement. Furthermore, analysis of putative biosynthetic gene clusters is performed to identify compounds of medical interest.

5 Material and Methods

The following passage contains classified information, and it is contained only in the full version of the thesis that is deposited at the Faculty of Science of the USB. Content is omitted due to future publication.

6 Results

The following passage contains classified information, and it is contained only in the full version of the thesis that is deposited at the Faculty of Science of the USB. Content is omitted due to future publication.

7 Discussion

The following passage contains classified information, and it is contained only in the full version of the thesis that is deposited at the Faculty of Science of the USB. Content is omitted due to future publication.

8 Conclusion

This study aimed to increase the information available on genetic potential of *Lentzea* spp. and investigated their metabolic capabilities. In this context, the complete genome of *Lentzea* sp. BCCO 10_0061 was generated and thereby adds to the existing, limited pool of data. Computational annotations were performed, and the organism was classified within the *Lentzea* genus with the closest similar complete genome available being *Lentzea albidocapillata* (GCF- 900176525.1). A total of 30 biosynthetic gene clusters were identified and the likelihood of production of iso-migrastatin and nystatin A1 or similar products was assessed. The presented results open the opportunity for future research. Possible follow up experiments could include transcriptomic analyses of BCCO 10_0061 to identify the conditions under which the identified BGCs are active. Subsequently, isolation and structural analysis of the secondary metabolites using MS and different NMR techniques as well as antifungal tests (i.e. inhibition of *Candida*, for nystatin A1) or antimetastatic effects (for iso-MGS) of the products predicted in this study.

9 Reference List

Abdelmohsen, U. R. et al. (2015) ‘Elicitation of secondary metabolism in actinomycetes’, Biotechnology advances, 33(6 Pt 1), pp. 798–811. doi: 10.1016/j.biotechadv.2015.06.003.

Advancing genomics, medicine and health together – by semiconductor DNA synthesis technology (no date) Genscript.com. Available at: https://www.genscript.com/advancing-genomics-medicine-and-health-together-by-semiconductor-dna-synthesis-technology-summary.html (Accessed: 22 April 2022).

An introduction to Next-Generation Sequencing Technology (no date) Illumina.com. Available at: https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf (Accessed: 1 September 2021).

Arkin, A. P. et al. (2018) ‘KBase: The United States department of energy systems biology knowledgebase’, Nature biotechnology, 36(7), pp. 566–569. doi: 10.1038/nbt.4163.

Aziz, R. K. et al. (2008) ‘The RAST Server: rapid annotations using subsystems technology’, BMC genomics, 9(1), p. 75. doi: 10.1186/1471-2164-9-75.

Babraham bioinformatics - FastQC A quality control tool for high throughput sequence data (no date) Babraham.ac.uk. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Accessed: 12 June 2022).

Berkow, E. L., Lockhart, S. R. and Ostrosky-Zeichner, L. (2020) ‘Antifungal susceptibility testing: Current approaches’, Clinical microbiology reviews, 33(3). doi: 10.1128/CMR.00069-19.

Biomatters Ltd. (2020) Geneoius Prime 2020.1.2. Available at: http://www.geneious.com/.

Blin, K. et al. (2021) ‘antiSMASH 6.0: improving cluster detection and comparison capabilities’, Nucleic acids research, 49(W1), pp. W29–W35. doi: 10.1093/nar/gkab335.

Boucher, H. W. et al. (2009) ‘Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America’, Clinical infectious diseases: an official publication of the Infectious Diseases Society of America, 48(1), pp. 1–12. doi: 10.1086/595011.

Brautaset, T. et al. (2000) ‘Biosynthesis of the polyene antifungal antibiotic nystatin in Streptomyces noursei ATCC 11455: analysis of the gene cluster and deduction of the biosynthetic pathway’, Chemistry & biology, 7(6), pp. 395–403. doi: 10.1016/s1074-5521(00)00120-4.

Brettin, T. et al. (2015) ‘RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes’, Scientific reports, (5(article 8365)), pp. 1–5. doi: 10.1038/srep08365.

Brunstein, J. (2014) In-depth coverage: some useful NGS terms, Mlo-online.com. Available at: https://www.mlo-online.com/home/article/13007639/indepth-coverage-some-useful-ngs-terms (Accessed: 10 November 2022).

Caporaso, J. G. et al. (2012) ‘Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms’, The ISME journal, 6(8), pp. 1621–1624. doi: 10.1038/ismej.2012.8.

Collection of Actinomycetes (no date) Bcco.cz. Available at: https://actinomycetes.bcco.cz/ (Accessed: 15 October 2022).

Compeau, P. E. C., Pevzner, P. A. and Tesler, G. (2011) ‘How to apply de Bruijn graphs to genome assembly’, Nature biotechnology, 29(11), pp. 987–991. doi: 10.1038/nbt.2023.

Copp, J. N. and Neilan, B. A. (2006) ‘The phosphopantetheinyl transferase superfamily: phylogenetic analysis and functional implications in cyanobacteria’, Applied and environmental microbiology, 72(4), pp. 2298–2305. doi: 10.1128/AEM.72.4.2298-2305.2006.

Cude, W. N. et al. (2012) ‘Production of the antimicrobial secondary metabolite indigoidine contributes to competitive surface colonization by the marine roseobacter Phaeobacter sp. strain Y4I’, Applied and environmental microbiology, 78(14), pp. 4771–4780. doi: 10.1128/AEM.00297-12.

Edgar, R. C. (2004) ‘MUSCLE: multiple sequence alignment with high accuracy and high throughput’, Nucleic acids research, 32(5), pp. 1792–1797. doi: 10.1093/nar/gkh340.

Edler, D. et al. (2021) ‘raxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML’, Methods in ecology and evolution, 12(2), pp. 373–377. doi: 10.1111/2041-210x.13512.

Fang, B.-Z. et al. (2017) ‘Lentzea cavernae sp. nov., an actinobacterium isolated from a karst cave sample, and emended description of the genus Lentzea’, International journal of systematic and evolutionary microbiology, 67(7), pp. 2357–2362. doi: 10.1099/ijsem.0.001958.

Genomics, F. L. and Schroeder, K. (2022) A history of sequencing, Front Line Genomics. Available at: https://frontlinegenomics.com/a-history-of-sequencing/ (Accessed: 10 November 2022).

Gerber, N. N. and Lechevalier, H. A. (1965) ‘Geosmin, an earthy-smelling substance isolated from Actinomycetes’, Applied microbiology, 13(6), pp. 935–938. doi: 10.1128/am.13.6.935-938.1965.

Gurevich, A. et al. (2013) ‘QUAST: quality assessment tool for genome assemblies’, Bioinformatics (Oxford, England), 29(8), pp. 1072–1075. doi: 10.1093/bioinformatics/btt086.

Hussain, A. et al. (2017) ‘Novel bioactive molecules from Lentzea violacea strain AS 08 using one strain-many compounds (OSMAC) approach’, Bioorganic & medicinal chemistry letters, 27(11), pp. 2579–2582. doi: 10.1016/j.bmcl.2017.03.075.

Illumina Inc. (2010) ‘De Novo Assembly Using Illumina Reads’. Available at: https://www.illumina.com/Documents/products/technotes/technote_denovo_assembly_ecoli.pdf (Accessed: 10 November 2022).

Jarerat, A., Pranamuda, H. and Tokiwa, Y. (2002) ‘Poly(L-lactide)-degrading activity in various Actinomycetes’, Macromolecular bioscience, 2(9), pp. 420–428. doi: 10.1002/mabi.200290001.

Kaitin, K. I. (2010) ‘Deconstructing the drug development process: the new face of innovation’, Clinical pharmacology and therapeutics, 87(3), pp. 356–361. doi: 10.1038/clpt.2009.293.

Kim, B.-G. et al. (2009) ‘Identification of functionally clustered nystatin-like biosynthetic genes in a rare actinomycetes, Pseudonocardia autotrophica’, Journal of industrial microbiology & biotechnology, 36(11), pp. 1425–1434. doi: 10.1007/s10295-009-0629-5.

Kitaeva, K. V. et al. (2020) ‘Cell culture based in vitro test systems for anticancer drug screening’, Frontiers in bioengineering and biotechnology, 8, p. 322. doi: 10.3389/fbioe.2020.00322.

Krügel, H. et al. (1999) ‘Functional analysis of genes from Streptomyces griseus involved in the synthesis of isorenieratene, a carotenoid with aromatic end groups, revealed a novel type of carotenoid desaturase’, Biochimica et biophysica acta. Molecular and cell biology of lipids, 1439(1), pp. 57–64. doi: 10.1016/s1388-1981(99)00075-x.

Langmead, B. and Salzberg, S. L. (2012) ‘Fast gapped-read alignment with Bowtie 2’, Nature methods, 9(4), pp. 357–359. doi: 10.1038/nmeth.1923.

Lewis, K. (2017) ‘New approaches to antimicrobial discovery’, Biochemical pharmacology, 134, pp. 87–98. doi: 10.1016/j.bcp.2016.11.002.

Li, C. et al. (2022) ‘Discovery of unusual dimeric piperazyl cyclopeptides encoded by a Lentzea flaviverrucosa DSM 44664 biosynthetic supercluster’, Proceedings of the National Academy of Sciences of the United States of America, 119(17). doi: 10.1073/pnas.2117941119.

Lim, S.-K. et al. (2009) ‘iso-Migrastatin, migrastatin, and dorrigocin production in Streptomyces platensis NRRL 18993 is governed by a single biosynthetic machinery featuring an acyltransferase-less type I polyketide synthase’, The journal of biological chemistry, 284(43), pp. 29746–29756. doi: 10.1074/jbc.M109.046805.

Luepke, K. H. et al. (2017) ‘Past, present, and future of antibacterial economics: Increasing bacterial resistance, limited antibiotic pipeline, and societal implications’, Pharmacotherapy, 37(1), pp. 71–84. doi: 10.1002/phar.1868.

Maiti, P. K. and Mandal, S. (2022) ‘Comprehensive genome analysis of Lentzea reveals repertoire of polymer-degrading enzymes and bioactive compounds with clinical relevance’, Scientific reports, 12(1), p. 8409. doi: 10.1038/s41598-022-12427-7.

Matthies, C., Erhard, H.-P. and Drake, H. L. (1997) ‘Effects of pH on the comparative culturability of fungi and bacteria from acidic and less acidic forest soils’, Journal of basic microbiology, 37(5), pp. 335–343. doi: 10.1002/jobm.3620370506.

Medema, M. H. et al. (2011) ‘antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences’, Nucleic acids research, 39(Web Server issue), pp. W339-46. doi: 10.1093/nar/gkr466.

Mikheenko, A. et al. (2016) ‘Icarus: visualizer for de novo assembly evaluation’, Bioinformatics (Oxford, England), 32(21), pp. 3321–3323. doi: 10.1093/bioinformatics/btw379.

Morgulis, A. et al. (2008) ‘Database indexing for production MegaBLAST searches’, Bioinformatics (Oxford, England), 24(16), pp. 1757–1764. doi: 10.1093/bioinformatics/btn322.

Nouioui, I. et al. (2018) ‘Genome-based taxonomic classification of the phylum Actinobacteria’, Frontiers in microbiology, 9, p. 2007. doi: 10.3389/fmicb.2018.02007.

Nurk, S. et al. (2013) ‘Assembling genomes and mini-metagenomes from highly chimeric reads’, in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 158–170.

O’Leary, N. A. et al. (2016) ‘Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation’, Nucleic acids research, 44(D1), pp. D733-45. doi: 10.1093/nar/gkv1189.

Overbeek, R. et al. (2014) ‘The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST)’, Nucleic acids research, 42(Database issue), pp. D206-14. doi: 10.1093/nar/gkt1226.

Parks, D. H. et al. (2015) ‘CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes’, Genome research, 25(7), pp. 1043–1055. doi: 10.1101/gr.186072.114.

Pei, S. et al. (2021) ‘Complete genome sequence of Streptomyces Sp. HSG2 from rhizosphere soil of mangrove in Qingmei Gang, Sanya’, Research Square. doi: 10.21203/rs.3.rs-298540/v1.

Petersen, J. (2016) ‘Heap leaching as a key technology for recovery of values from low-grade ores – A brief overview’, Hydrometallurgy, 165, pp. 206–212. doi: 10.1016/j.hydromet.2015.09.001

Ping, M. et al. (2021) ‘Proposal of Lentzea deserti (Okoro et al. 2010) Nouioui et al. 2018 as a later heterotypic synonym of Lentzea atacamensis (Okoro et al. 2010) Nouioui et al. 2018 and an emended description of Lentzea atacamensis’, PloS one, 16(2), p. e0246533. doi: 10.1371/journal.pone.0246533.

Pommerehne, K. et al. (2019) ‘The antitumor antibiotic rebeccamycin-challenges and advanced approaches in production processes’, Applied microbiology and biotechnology, 103(9), pp. 3627–3636. doi: 10.1007/s00253-019-09741-y.

Prjibelski, A. (no date) ‘New Frontiers of Genome Assembly with SPAdes 3.1’. Available at: https://www.open-bio.org/bosc2014/BOSC2014-Genome-03-SPAdes-Prjibelski.pdf (Accessed: 12 April 2022).

Quinn, G. A. et al. (2020) ‘Streptomyces from traditional medicine: sources of new innovations in antibiotic discovery’, Journal of medical microbiology, 69(8), pp. 1040–1048. doi: 10.1099/jmm.0.001232.

Reverchon, S. et al. (2002) ‘Characterization of indigoidine biosynthetic genes in Erwinia chrysanthemi and role of this blue pigment in pathogenicity’, Journal of bacteriology, 184(3), pp. 654–665. doi: 10.1128/JB.184.3.654-665.2002.

Riehl, B., Chivian, D., Canon, S. & Land, M. (2018) BBTools. Available at: https://github.com/kbaseapps/BBTools/tree/cd0e4223a1091df6db817855955223207f1d50a1 /ui/narrative/methods/RQCFilter.

Romesberg, F. E. and Craney, A. (2016) ‘Discovery of novel antibacterials’, Bioorganic & medicinal chemistry, 24(24), pp. 6225–6226. doi: 10.1016/j.bmc.2016.11.046.

Rybak, M. J. (2004) ‘Resistance to antimicrobial agents: an update’, Pharmacotherapy, 24(12 Pt 2), pp. 203S–15S. doi: 10.1592/phco.24.18.203S.52236.

Schrey, S. D. et al. (2012) ‘Production of fungal and bacterial growth modulating secondary metabolites is widespread among mycorrhiza-associated streptomycetes’, BMC microbiology, 12(1), p. 164. doi: 10.1186/1471-2180-12-164.

Seemann, T. (2014) ‘Prokka: rapid prokaryotic genome annotation’, Bioinformatics (Oxford, England), 30(14), pp. 2068–2069. doi: 10.1093/bioinformatics/btu153.

Shan, D. et al. (2005) ‘Synthetic analogues of migrastatin that inhibit mammary tumor metastasis in mice’, Proceedings of the National Academy of Sciences of the United States of America, 102(10), pp. 3772–3776. doi: 10.1073/pnas.0500658102.

Stamatakis, A. (2014) ‘RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies’, Bioinformatics (Oxford, England), 30(9), pp. 1312–1313. doi: 10.1093/bioinformatics/btu033.

Talbot, G. H. et al. (2006) ‘Bad bugs need drugs: an update on the development pipeline from the Antimicrobial Availability Task Force of the Infectious Diseases Society of America’, Clinical infectious diseases: an official publication of the Infectious Diseases Society of America, 42(5), pp. 657–668. doi: 10.1086/499819.

Thermo Fisher Scientific Inc. (2019) ‘The importance of coverage: advantages of amplicon-based approaches in next-generation sequencing’, pp. 1–4. Available at: https://www.thermofisher.com/at/en/home/life-science/sequencing/sequencing-learning-center/next-generation-sequencing-information/ngs-basics/importance-coverage-throughput.html (Accessed: 10 November 2022).

Tian, W. and Skolnick, J. (2003) ‘How well is enzyme function conserved as a function of pairwise sequence identity?’, Journal of molecular biology, 333(4), pp. 863–882. doi: 10.1016/j.jmb.2003.08.057.

Valli, S. et al. (2012) ‘Antimicrobial potential of Actinomycetes species isolated from marine environment’, Asian pacific journal of tropical biomedicine, 2(6), pp. 469–473. doi: 10.1016/S2221-1691(12)60078-1.

Völler, G. H. et al. (2012) ‘Characterization of new class III lantibiotics--erythreapeptin, avermipeptin and griseopeptin from Saccharopolyspora erythraea, Streptomyces avermitilis and Streptomyces griseus demonstrates stepwise N-terminal leader processing’, Chembiochem: a European journal of chemical biology, 13(8), pp. 1174–1183. doi: 10.1002/cbic.201200118.

Volokhan, O. et al. (2005) ‘An unexpected role for the putative 4’-phosphopantetheinyl transferase-encoding gene nysF in the regulation of nystatin biosynthesis in Streptomyces noursei ATCC 11455’, FEMS microbiology letters, 249(1), pp. 57–64. doi: 10.1016/j.femsle.2005.05.052.

Weisman, C. M., Murray, A. W. and Eddy, S. R. (2022) ‘Mixing genome annotation methods in a comparative analysis inflates the apparent number of lineage-specific genes’, Current biology: CB, 32(12), pp. 2632-2639.e2. doi: 10.1016/j.cub.2022.04.085.

Wichner, D. et al. (2017) ‘Isolation and anti-HIV-1 integrase activity of lentzeosides A-F from extremotolerant lentzea sp. H45, a strain isolated from a high-altitude Atacama Desert soil’, The Journal of antibiotics, 70(4), pp. 448–453. doi: 10.1038/ja.2016.78.

Williams, J. C. et al. (2019) ‘Synthesis of the siderophore coelichelin and its utility as a probe in the study of bacterial metal sensing and response’, Organic letters, 21(3), pp. 679–682. doi: 10.1021/acs.orglett.8b03857.

Woo, E. J. et al. (2002) ‘Migrastatin and a new compound, isomigrastatin, from Streptomyces platensis’, The Journal of antibiotics, 55(2), pp. 141–146. doi: 10.7164/antibiotics.55.141.

Yassin, A. F. et al. (1995) ‘Lentzea gen. nov., a new genus of the order Actinomycetales’, International journal of systematic bacteriology, 45(2), pp. 357–363. doi: 10.1099/00207713-45-2-357.

10 Appendix

The following passage contains classified information, and it is contained only in the full version of the thesis that is deposited at the Faculty of Science of the USB. Content is omitted due to future publication.