



Technische Universität München

TUM School of Natural Sciences

# **Biosynthetic Gene Cluster Evaluation Genome Mining for Natural Product Formation**

*Syed Nadim Marc Ahmad*

Vollständiger Abdruck der von der TUM School of Natural Sciences der Technischen Universität München zur Erlangung eines

**Doktors der Naturwissenschaften (Dr. rer. nat.)**

genehmigten Dissertation.

Vorsitz: Prof. Dr. Tom Nilges

Prüfer\*innen der Dissertation:

1. Priv.-Doz. Dr. Norbert Mehlmer

2. Prof. Dr. Michael Groll

3. Prof. Dr. Imke Schmitt

Die Dissertation wurde am 23.10.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Natural Sciences am 05.12.2023 angenommen.



# Contents

I Abstract.....	III
II Zusammenfassung.....	IV
III Acknowledgements.....	VI
IV List of abbreviations.....	VII
1 Introduction.....	1
1.1 General Introduction: from Fungi to Lichen.....	1
1.2 Natural Products in Fungi and Lichen.....	3
1.3 Sequencing and Assembly.....	5
1.4 Genomics and Metagenomics in Lichen.....	8
1.5 Genome Mining for Biosynthetic Gene Clusters – Secondary Metabolites in Lichen.....	9
2 Material and Methods.....	15
2.1 Kits & Consumables.....	15
2.2 Lichen Sample Collection.....	16
2.3 High Molecular Weight DNA Extraction and Evaluation.....	17
2.4 Sequencing.....	18
2.5 Bioinformatic Analysis.....	19
3 Research.....	21
3.1 Summaries of included publications.....	21
Chapter I Biosynthetic potential of <i>Hypogymnia</i> holobionts - Insights into Secondary Metabolite Pathways.....	21
Chapter II Biosynthetic Gene Cluster Synteny - Orthologous Polyketide Synthases in <i>Hypogymnia physodes</i> , <i>Hypogymnia tubulosa</i> and <i>Parmelia sulcata</i> .....	22
3.2 Full length publications.....	23
Biosynthetic potential of <i>Hypogymnia</i> holobionts - Insights into Secondary Metabolite Pathways.....	23
Biosynthetic Gene Cluster Synteny - Orthologous Polyketide Synthases in <i>Hypogymnia physodes</i> , <i>Hypogymnia tubulosa</i> and <i>Parmelia sulcata</i> .....	41

<b>4 Discussion and Outlook.....</b>	<b>60</b>
<b>Lichen Natural Products in Drug Discovery .....</b>	<b>60</b>
<b>Genome Reconstruction of Lichen.....</b>	<b>62</b>
<b>Outlook - Regulation and Activation of Silent Biosynthetic Gene Clusters .....</b>	<b>64</b>
<b>Concluding Remarks .....</b>	<b>67</b>
<b>5 List of Publications .....</b>	<b>68</b>
<b>6. Reprint Permission.....</b>	<b>69</b>
<b>7. Figures &amp; Tables .....</b>	<b>70</b>
<b>8 Bibliography .....</b>	<b>71</b>
<b>9 Additional Full-Length Publications.....</b>	<b>84</b>
<b>9.1 Comparative Genome-Wide Analysis of Two <i>Caryopteris x Clandonensis</i> Cultivars: Insights on the Biosynthesis of Volatile Terpenoids .....</b>	<b>84</b>
<b>9.2 Differential RNA-Seq Analysis Predicts Genes Related to Terpene Tailoring in <i>Caryopteris x Clandonensis</i>.....</b>	<b>103</b>

# I Abstract

Lichens, a paradigm of mutualistic symbioses comprising photobionts (algae or cyanobacteria) and mycobionts (fungi), exhibit profound biosynthetic capabilities, giving rise to a myriad of structurally diverse secondary compounds. Exploiting the biotechnological potential inherent in lichens necessitates an in-depth understanding of their biosynthetic pathways and corresponding gene clusters. In the first part, an extensive study of biosynthetic gene clusters across all components of two lichen thalli from *Hypogymnia physodes* and *Hypogymnia tubulosa*, encompassing fungi, green algae, and bacteria is conducted. Leveraging two high-quality PacBio metagenomes, we identified a total of 460 biosynthetic gene clusters. Remarkably, mycobionts emerged as reservoirs of biosynthetic gene clusters, encompassing 73-114 clusters. Predominantly, Type I Polyketide Synthases (T1PKSs), Non-Ribosomal Peptide Synthetases (NRPSs), and terpene synthases prevailed among mycobiont biosynthetic genes.

Furthermore, an in-depth investigation focused on three prevalent lichen mycobionts was deployed: *H. physodes*, *H. tubulosa*, and *Parmelia sulcata*. Utilizing a high-quality PacBio metagenome of *P. sulcata*, we successfully extracted the mycobiont bin, uncovering 214 biosynthetic gene clusters. Notably, T1PKSs, NRPSs, and terpene synthases constituted the major gene clusters in this dataset. To investigate cluster synteny between this mycobiont bin and those from the previously described *Hypogymnia* species further bioinformatic processing was conducted. Through meticulous ketosynthase homology analysis, we unveiled nine highly conserved gene clusters shared among all three mycobiont species, encompassing both non-reducing and reducing polyketide synthases. Among these, four clusters exhibited putative involvement in the formation of orsellinic acid-derived lichen substances, including orcinol depsides, depsidones (e.g., lecanoric acid), methylated forms of orsellinic acid (e.g., atranorin), and melanins. This thesis represents the first comprehensive identification and analysis of biosynthetic gene clusters within complete lichen holobionts, shedding light on the unexplored biosynthetic potential residing within *Hypogymnia* species. Furthermore, syntenic comparison of these with *P. sulcata* render biosynthetic gene clusters related to natural product formation accessible for further evaluation. The insights gained from this research significantly contribute to the classification and dereplication of the extensive polyketide synthase diversity found in lichenized fungi. Moreover, the high-quality sequences of biosynthetic gene clusters obtained from these three prevalent lichen species establish a solid foundation for further exploration of biotechnological applications and deeper investigations into the natural product formation unique to lichens.

## II Zusammenfassung

Flechten, ein Paradigma mutualistischer Symbiosen aus Photobionten (Algen oder Cyanobakterien) und Mykobionten (Pilzen), zeigen ein ausgeprägtes biosynthetisches Potential, welches sich in der Vielzahl an strukturell unterschiedlichen Sekundärmetaboliten widerspiegelt. Um das biotechnologische Potenzial von Flechten nutzen zu können, ist ein tiefgehendes Verständnis ihrer biosynthetischen Synthesewege und der entsprechenden Gen-Clustern von entscheidender Bedeutung.

Im ersten Teil dieser Arbeit wird eine umfassende Studie der biosynthetischen Gen-Cluster in allen symbiotischen Organismen von Thalli aus *Hypogymnia physodes* und *Hypogymnia tubulosa* durchgeführt. Hierunter befinden sich einschließlich Pilze, Grünalgen und Bakterien. Unter Verwendung von zwei hochwertigen PacBio-Metagenomen wurden insgesamt 460 biosynthetische Gen-Cluster identifiziert. Bemerkenswerterweise erwiesen sich Mykobionten als Reservoirs biosynthetischer Gen-Cluster mit 73-114 Clustern. Vorherrschend waren dabei Typ-I-Polyketid-Synthasen (T1PKS), nicht-ribosomale Peptidsynthetasen (NRPS) und Terpensynthasen unter den biosynthetischen Genen der Mykobionten.

Des Weiteren wurde eine eingehende Untersuchung von drei weit verbreiteten Flechten-Mykobionten durchgeführt: *H. physodes*, *H. tubulosa* und *Parmelia sulcata*. Unter Verwendung eines hochwertigen PacBio-Metagenoms von *P. sulcata* wurde der Mykobionten-Bin extrahiert, der 214 biosynthetische Gen-Cluster umfasst. Hierbei stellen in diesem Datensatz ebenfalls T1PKS, NRPS und Terpensynthasen die Hauptengruppen dar. Um die Cluster-Syntenie zwischen diesem Mykobionten-Bin und denen der zuvor beschriebenen *Hypogymnia*-Arten weiter zu untersuchen, wurden weitere bioinformatische Analysen durchgeführt. Durch eine sorgfältige Analyse der Ketosynthase-Homologie wurden neun hochkonservierte Gen-Cluster identifiziert, die in allen drei Mykobionten-Arten vorkommen und sowohl nicht-reduzierende als auch reduzierende Polyketid-Synthasen umfassen. Vier dieser Gruppen weisen eine mögliche Beteiligung an der Bildung von Flechteninhaltsstoffen auf, die von Orsellinsäure abgeleitet sind, einschließlich Orcinol-Depside, Depsidonen (z. B. Lecanorsäure), methylierte Formen von Orsellinsäure (z. B. Atranorin) und Melanine. Diese Arbeit stellt die erste umfassende Identifizierung und Analyse biosynthetischer Gengruppen in vollständigen Flechten-Holobionten dar und wirft Licht auf das unerforschte biosynthetische Potenzial in *Hypogymnia*-Arten. Darüber hinaus ermöglicht der syntenische Vergleich mit *P. sulcata* eine weitere Untersuchung der mit der Naturstoffbildung in Verbindung stehenden biosynthetischen Gen-Cluster. Die gewonnenen Erkenntnisse tragen wesentlich zur Klassifizierung und Dereplikation der umfangreichen

Polyketid-Synthase-Diversität bei, die bei flechtenbildenden Pilzen gefunden wurde. Darüber hinaus legen die hochwertigen Sequenzen biosynthetischer Gen-Cluster von diesen drei verbreiteten Flechten-Arten eine solide Grundlage für die weitere Erforschung biotechnologischer Anwendungen und tiefgreifende Untersuchungen der einzigartigen Naturstoffbildung in Flechten.

# III Acknowledgements

First, I would like to thank my supervisors Norbert Mehlmer and Thomas Brück for granting me the opportunity to work under their supervision and complete my dissertation. Furthermore, I want to thank Imke Schmitt and Anjuli Calchera for their expertise and help, without you my dissertation would have taken far longer.

I would also like to deeply thank especially my dear friends Nathanael Arnold, Manfred Ritz, Selina Engelhart-Straub, Sophia Prem, Zora Rerop and Kevin Heieck for their friendship, cooperation, and collaboration. You were by my side all the time, offering me support, true companionship, and strength. Your wide-ranging knowledge, optimistic outlook, and eagerness to assist have fostered an inspiring research atmosphere, significantly enhancing my research advancement. Your perceptive input, thoughtful analysis, and valuable feedback have played a vital role in my research and honing my scientific abilities."

Finally, I would like to acknowledge the contributions of all the individuals including my family and friends who have helped me in any way during my PhD journey. Your assistance, advice, and encouragement have been essential in making my research work possible.



# IV List of abbreviations

**°C:** Degree Celsius

**A:** AMP-Binding site

**AT:** Acyltransferase

**ACP:** Acyl-Carrier Protein

**BGC:** Biosynthetic Gene Cluster

**BLAST:** Basic Local Alignment Search Tool

**bp:** Base pair

**BUSCO:** Benchmarking Universal Single-Copy Orthologs

**CCS:** Circular Consensus Sequence (HiFi-reads)

**cMT:** C-Methyltransferase

**CoA:** Coenzyme A

**CP:** Carrier Protein

**DH:** Dehydratase

**DNA:** Deoxyribonucleic Acid

**EI:** Electron Ionisation

**ER:** Enoylreductase

**FID:** Flame Ionisation Detector

**Gb:** Gigabases

**GC:** Gas chromatography

**gDNA:** Genomic DNA

**GO:** Gene Ontology

**HMW:** High Molecular Weight

**ITS:** Internal Transcribed Spacer

**kb:** Kilobases

**KS:** Ketosynthase

**KR:** Ketoreductase

**MeOH:** Methanol

**MS:** Mass Spectrometry

**ng:** Nano Gram

**NGS:** Next generation sequencing

**PacBio:** Pacific Bioscience

**NR-PKS:** Non-Reducing Polyketide Synthase

**NRPS:** Non-Ribosomal Peptide Synthetases

**PKS:** Polyketide Synthase

**pM:** Pico Molar

**PT:** Product Template

**QUAST:** Quality Assessment Tool

**RBH:** Reciprocal Best Hit

**RNA:** Ribonucleic Acid

**R-PKS:** Reducing Polyketide Synthase

**T1PKS:** Type I Polyketide Synthase

**T3PKS:** Type III Polyketide Synthase

**TD:** Terminal Domain

**SAT:** Starter unit:acyl-carrier protein transferases

**ZMW:** Zero-Mode Waveguide

***H. physodes:*** *Hypogymnia physodes*, HPH

***H. tubulosa:*** *Hypogymnia tubulosa*, HTU

***P. sulcata:*** *Parmelia sulcata*, PSU

# 1 Introduction

## 1.1 General Introduction: from Fungi to Lichen

Representing a large share of the genetic diversity on Earth, fungi have had a tremendous influence on the development of life through their own evolution and diversification [1]–[4]. In a re-evaluation on fungal diversity, Hawksworth & Lücking estimated numbers currently ranging between 2.2-3.8 million of fungi species, of which only 120,000 have been described [5]. Through their inherent ability to develop adaptation mechanisms, fungi have optimized their survival. Among these are coping mechanisms especially in regard to harsh abiotic stresses, processing of various substrates and the ability to form mutualistic unions between organisms of different superkingdoms (i.e. plants, animals or bacteria) providing benefits of a symbiotic relationship [6], [7]. Thus formed symbioses enabled many groups of organisms to adapt, grow and evolve ecologically, in particular lichens [8]–[10].

In the past, lichens were considered a mutualistic symbiosis consisting of a fungus and partners capable of photosynthesis, such as algae or cyanobacteria [11]. The mycobiont (fungus) determines the morphology of the lichen thallus, which is historically distinguished into three main groups: fruticose (hair-like, strap-shaped or shrubby, protruding from the substrate surface), crustose (crust-like, tight attachment to substrate) and foliose (leaf-like, flat and only attached in parts to the substrate) [12]. In their symbiosis, the hosted photobionts provide oxygen and carbon products through photosynthesis to be harvested by the fungus and are in return provided with water, mineral nutrients, protection from environmental conditions and herbivores [13]. Through their high adaptability, lichens are distributed in most ecosystems on Earth. These are subdivided into their occurrence: growing on plants or trees (epiphytic); populating on leaves (epiphyllic, mostly in sub-/tropics); growing on rock surfaces or embedded in the upper layer (epilithic) and being part of the soil composition (epigeic) [6], [14]. Regarding the composition of the lichen community, recent research has indicated that lichen individuals may contain not only the primary fungus responsible for lichen formation and the primary photosynthetic partner, but also additional fungi from other phyla, and other organisms [15]–[17]. Among these are Basidiomycota [15], or Ascomycota [18], as well as various types of bacteria and other algae [19], [20]. Today, lichens are often referred to as ecosystems [21] or holobionts [22].

Lichens are classified based on their mycobiont, of which the majority (98%) belong to Ascomycetes [23]. The photobiont is typically either a cyanobacterium (e.g. *Nostoc*) or an eukaryotic green algae (e.g. *Trebouxia*) [24]. Notably, of all Ascomycetes nearly half are

## Introduction

lichenized (approximately 20.000 species) [6], [24]–[26]. As new species are constantly described, higher numbers are expected.

Particularly of interest are lichens belonging to the *Parmeliaceae* family, as this is considered the most widespread family of lichen-forming fungi, comprising over 2726 species (about 10% of total lichen species) and 79 genera [27], [28]. In addition to the temperate, boreal, sub-alpine, and Mediterranean areas of the Northern Hemisphere, members of the family are extensively dispersed in the Southern Hemisphere (Australia, South America, and South Africa) [27], [29]. The allocated species can be subdivided mainly into one of five main clades, including the, alectorioid, cetrarioid, usneoid, hypogymnioid and parmelioid. This family is well studied regarding phylogeny and systematics, however evaluations on genome level are still scarce for most of the allocated species [27]. In this thesis two species from the clade hypogymnioid and one from parmelioid are evaluated in more detail, namely *Hypogymnia physodes*, *Hypogymnia tubulosa* and *Parmelia sulcata*.

*H. physodes* is one of the most ubiquitous and abundant macrolichens throughout Europe, and also occurs in North America, Africa, and Asia [30]. It is a polymorphic, foliose species, forming lip-shaped soralia located at the tips of the lobes, and up to 10 cm diameter in size. It is found on acidic substrata, including the bark of trees, and siliceous rocks. *H. tubulosa* is similar to *H. physodes*, but the lobes are more tubular, and soralia rounded, capitate, not lip-shaped. It has a similar distribution and grows on similar substrata as *H. physodes*, but is less frequent [31], [32]. Both species are associated with green algae of the genus *Trebouxia* and often grow together. The upper surface of *H. physodes* and *H. tubulosa* is bluish to whitish grey, due to the presence of atranorin in the cortex. The medullary layer contains various colorless depsides and depsidones, the lower surface is black, due to the presence of melanins.

*P. sulcata* is one of the most common species of lichen-forming fungi worldwide. It is widely distributed in temperate and cold regions of both hemispheres, and typically grows on the bark of trees. The lobed thallus has a light grey surface, usually with white ridges and soredia. *P. sulcata* belongs to a species complex, characterized by high genetic diversity [33]–[36].

## **1.2 Natural Products in Fungi and Lichen**

Fungi are known as a major source for a manifold of natural products, with the majority exhibiting biological activities [37], [38]. The synthesized metabolites are categorized into two groups: primary and secondary metabolites. Compounds belonging to the primary metabolites are crucial for growth, reproduction, and development of the organism, as well as metabolism. Examples for primary metabolites are high molecular weight polymers like lignin, cellulose or proteins, as well as sugars (e.g. glucose) or amino acids [39]–[41]. Secondary metabolites also often referred to as natural products, have an important influence on the adaptation of biological systems and the organism itself. Among these are environmental interactions like defence or warfare mechanisms, as well as roles in development [39], [42]. The distinction in unique secondary metabolite composition can also be utilized for their chemotaxonomic determination [41]. Some of these secondary metabolites are of medical or industrial importance, such as antibiotics (cyclosporin A) or hypocholesterolemic pharmaceuticals like lovastatin [40], [43].

The advent of the genomic era equips us with the tools to bypass the need for isolating and cultivating individual organisms to access the genes responsible for natural product biosynthesis [44]. Apart from extensively studied and easily accessible organisms, non-traditional sources are expected to provide a vast array of novel compounds [45]. For instance, symbiotic organisms, which have remained largely unexplored in a systematic manner, hold significant potential as a source of natural products [44]. Within the fungal kingdom, symbiotic lichenized fungi are particularly projected to possess a yet untapped reservoir of diverse secondary metabolites [39], [46].

Over the past few decades, the distinct secondary metabolite profile of these composite organisms has attracted growing interest [47]. Lichens have a significant impact on various community processes, largely due to their diverse and abundant secondary chemistry. These processes include e.g., decomposition, nutrient accumulation and rock weathering [48]. Over 1,000 secondary metabolites have been identified in lichens, many of which are exclusive to this symbiotic organism [49]–[52]. The majority of these characteristic lichen secondary metabolites originate from fungi and are typically deposited outside the cells, such as on the outer surface of medullary hyphae [49]. Lichen substances are often irregularly distributed within the thallus, with pigments typically concentrated in the upper cortex and colorless substances found in the medullary layer [49]. The concentration of lichen substances generally ranges from 0.1% to 5% of the dry weight of the lichen thallus, although in some cases, it can reach up to 30% [51], [53]. These secondary metabolites play diverse ecological roles, including light screening, chemical weathering, allelopathic effects, and

## Introduction

defence against herbivores [54]–[57]. In addition to benefiting the lichens themselves, these compounds exhibit various biological activities, some of which are of pharmaceutical interest. The activities of lichen compounds encompass antiviral, antibiotic, antitumor, allergenic, plant growth inhibitory, antiherbivore, and enzyme inhibitory effects [47], [50], [58], [59].

Lichen mycobionts are predominantly involved in the synthesis of a plethora of natural products [59]. Among the latter, the most prominent classes include depsides, depsidones, dibenzofurans and phenolic compounds, however also aliphatic acids, triterpenes, anthraquinones, secalonic acids, pulvinic acid derivatives, and xanthenes are observed [27], [49], [60]. These compounds exhibit bioactivities of antimicrobial [47], [61]–[63], antifungal [64], anti-inflammatory [65], antioxidant [66], [67] and antitumoral [63], [68] nature. Examples of medically relevant compounds include gyrophoric acid, atranorin and physodic acid [69]. Compounds like physodic acid, evernic acid, atranorin and usnic acid displayed an inhibitory effect on metabolic enzymes [58], [60]. To visualize the distribution of secondary metabolites in the lichen *Hypogymnia physodes*, *Hypogymnia tubulosa* and *Parmelia sulcata*,

Table 1 summarizes the respective natural products previously reported. Some of these exhibit pharmaceutically relevant activities [30], [70]–[72].

Table 1. Lichen products from *Hypogymnia tubulosa*, *Hypogymnia physodes* and *Parmelia sulcata*. Occurrence of respective metabolites is allocated to lichen region.

Lichen species	Natural product	Ref
<i>Hypogymnia tubulosa</i>	atranorin, chloroatranorin (cortical depsides); physodic acid, 3-hydroxyphysodic acid, 4-O-methyl physodic acid (medullary depsidones)	[71], [73]
<i>Hypogymnia physodes</i>	atranorin, chloroatranorin (cortical depsides), physodic acid, physodalic acid, 3-hydroxyphysodic acid, protocetraric acid, fumarprotocetraric acid (medullary depsidones)	[70], [73]–[75]
<i>Parmelia sulcata</i>	atranorin, chloroatranorin (cortical depsides), salazinic acid, consalazinic acid, lobaric acid (medullary depsidones), lecanoric acid (medullary depside)	[76]–[79]

Considering the exhibited manifold product spectra, lichen natural products are promising leads in drug discovery for the further development regarding pharmaceutical applications.

Despite the potential of lichen compounds, their widespread utilization is impeded by two primary factors. Firstly, the slow growth rate exhibited by many lichen species restricts their large-scale cultivation. Secondly, isolating and in vitro culturing of the symbiotic partners, in some cases, is challenging or even impossible [80]–[82]. Directly harvesting and isolating

compounds from lichen thalli is typically unsustainable and often unfeasible due to the substantial biomass requirement [83]. Experimental characterization of the molecular mechanisms responsible for the biosynthesis of characteristic lichen metabolites faces particular difficulties in aposymbiotic mycobiont cultivation and the slow growth rates observed in lichen-forming fungi [51], [84]. Additionally, abiotic factors such as microhabitat, elevation, hydration response, chemical signals, UV radiation, seasonality, and microclimatic fluctuations can influence the biosynthesis of natural products [51], [58].

Furthermore, the biosynthesis of substances may be influenced by biological factors, including interactions with microorganisms within the lichen thallus, predation by insects or snails, and the presence of competing organisms [52]. Moreover, mycobionts only produce significant amounts of natural products under specific inducing conditions, resulting in notable disparities between the secondary metabolite production in axenic cultures and natural environments [47], [58].

### 1.3 Sequencing and Assembly

#### 1.3.1 Sequencing

Two different sequencing methods were used in this thesis. To elaborate these, Figure 1 provides a graphical summary of different characteristics of each method. The input DNA fragment size mainly distinguishes the compared methods.

As reported in previous studies, the choice between these two methods highly depends on the desired output. To deepen the understanding, the advantages and disadvantages of each method are elaborated further. The underlying approach for NovaSeq 6000 (Illumina, San Diego, CA, USA) is the sequencing by synthesis, where nucleotides are fused to the extended DNA chain by a DNA polymerase [85], [86]. In addition, paired-end sequencing is feasible, enabling the sequencing high amounts of reads in high-quality and coverage [85]. As the sequenced fragments are 150 base pairs (bp) long, various limitations occur, regarding *de novo* assembly, detection of structural variants or transcript isoforms, as well as regions of high sequencing complexity [85]–[87]. However, most available bioinformatics tools are tailored to process short-read sequencing data [85], [88], [89]. The deployed long-read sequencing method on PacBio Sequel IIe (Pacific Bioscience, Menlo Park, CA, USA) overcomes the previously described limitations of short-read sequencing, as it allows for the accurate resolution of repetitive and complex genome regions such as segmental duplications [88], [90].

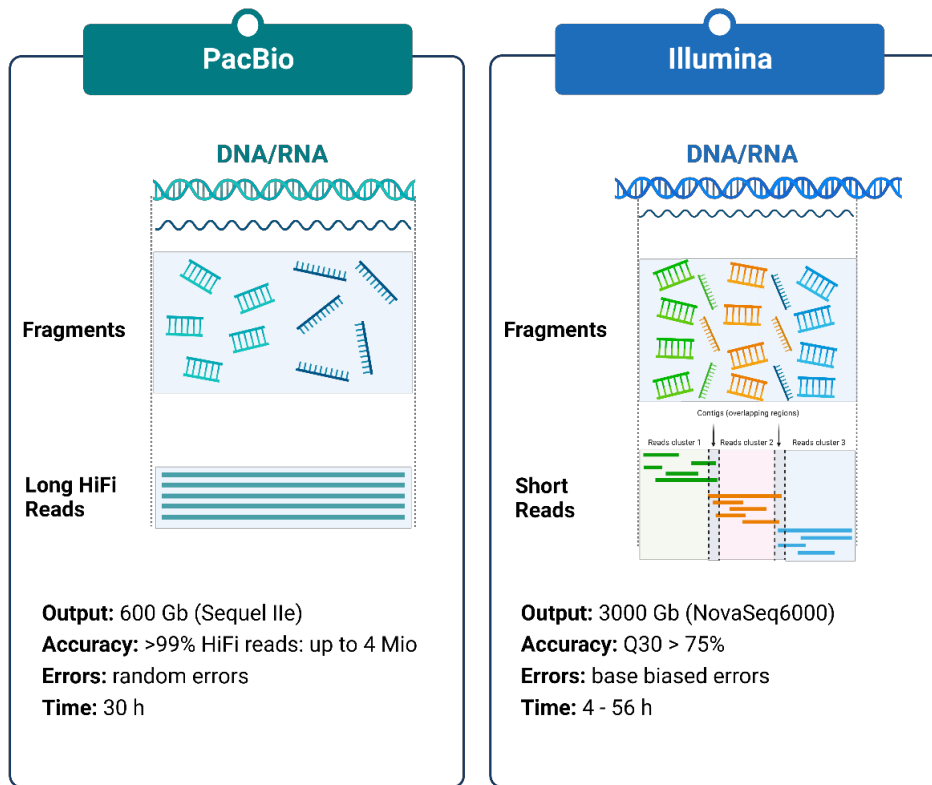


Figure 1. Overview of sequencing techniques deployed in this thesis. Here PacBio Sequel IIe long-read sequencing is compared to short-read sequencing on Illumina NovaSeq6000. In addition, metrics on output, accuracy, errors, and elapsing time are provided. Figure created with Biorender.

Due to the sequencing of up to 10 kilobase pairs (kb) continuous native DNA fragments, these complex regions can concisely be resolved [91], [92]. The foundation for this method lies in the SMRT cell harbouring zero-mode waveguides (ZMW), where ideally only one polymerase is immobilized to allow for the sequencing of a single template DNA fragment is sequenced. Here nucleotides labelled with fluorophores emitting unique spectra are utilized in the replication process. Upon binding of the nucleotides, a distinctive light pulse occurs, this transpires simultaneously across all ZMWs. These emissions are recorded and allocated, allowing for the interpretation as nucleotide sequences. For each complete strand replication, the term “pass” is introduced. Through the sample preparation process, the template DNA is circularized by adapter ligation, rendering it accessible for continuous replication by the polymerase, yielding multiple passes of the same template. Resulting Circular Consensus Sequences (CCS) reads (also declared as HiFi-reads) mitigate random errors in sequencing through correction by the other subreads, yielding accuracies auf up to 99.9% [85], [93]. Genome sequencing was conducted solely on PacBio, whereas RNA sequencing was performed on both platforms.

### 1.3.2 Assembly



By metagenomic sequencing of complete lichen thalli, the mycobiont exhibits the highest coverages in relation to other present species [94]. The presence of such a composition poses difficulties for genome assemblers that rely on solid  $k$ -mer analysis. High abundance of certain species can result in their disproportionate representation, while low-abundance species like the photobiont may face obstacles to form complete assemblies. To overcome these challenges, the long CCS reads obtained from PacBio Sequel IIe were utilized and assembled using metaFlye. This assembler is specifically tailored to handle read coverages with significant non-uniformity, making it particularly suitable for this scenario. To gain a deeper understanding of this assembler, the sequential processing steps are further elaborated. As common solid  $k$ -mer analysis only yields assemblies of the most prominent species, a different approach is deployed by metaFlye. In this case, a combination of local  $k$ -mer distribution and global  $k$ -mer counting is utilized. Additionally, an algorithm is employed to locate repeat edges in the respective metagenome assembly graph, rendering it robust to the previously elaborated strongly non-uniform read coverage distribution. This is particularly important for metagenomes harboring species with high genome similarities. These are distinguished as strain genomes, which often exhibit conserved sequence regions same as unique genome sections. A graphical summary how metaFlye functions is provided in Figure 2. Here the distinct genomes are numbered from one to four and their respective sequence path is depicted. This summary shows a subsection of an assembly graph generated by metaFlye, highlighting unique and repeat edges in black arrows and colored lines. MetaFlye allocates overlapping regions as repeat edges as seen in the colored blocks A, B and C. Following this, contiguous metagenomic assemblies are enabled by metaFlye [95].

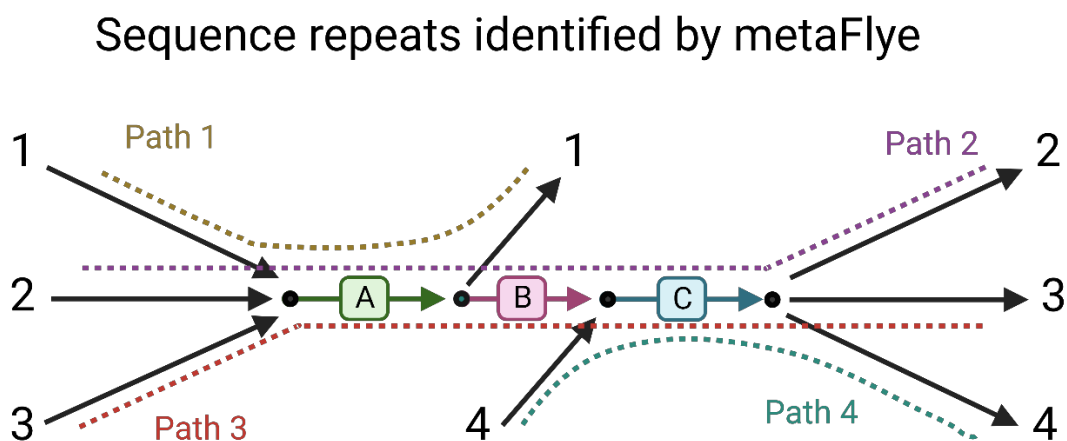


Figure 2. Graphical visualization of a metaFlye subgraph generated by four distinct genomes (1-4) in an assembly. Coloured lines depict repeat edges, whereas black arrows are unique edges. The edges A, B and C (coloured blocks) were deemed to be repetitive sequences by metaFlye, based on the distinct paths overlapping in the respective region. Figure adapted from Kolmogorov et. al [95] and created with Biorender.

## **1.4 Genomics and Metagenomics in Lichen**

The genomic sequences of obligatory symbionts can be acquired either by directly accessing individual symbionts or by extracting the relevant information from metagenomic samples comprising multiple species [96]. The conventional approach for obtaining genetic information involves isolating and culturing individual symbionts to gain pure, single-species DNA. However, in obligate symbioses, this presents a challenge due to the difficulty of cultivating each symbiont independently. With the advent of high-throughput sequencing technology, studying slow-growing symbiotic organisms has become feasible without the need to establish and grow axenic cultures of individual symbionts [81]. Utilizing a metagenomic shotgun sequencing approach, it is now possible to reconstruct high-quality genomes from samples containing multiple species. The genome quality was further enhanced by the introduction of long-read sequencing techniques in comparison to short-read sequencing [97]. Various methods can be employed to enrich DNA for targeted organisms, such as removing unwanted tissue or cells prior to or during DNA extraction, amplifying the targeted cell portion, or utilizing single-cell sequencing. Examples of these approaches include studies focusing the obligate fungal symbiont of aphids [98], or the primary fungal symbiont of lichens [99]. An alternative metagenomic approach involves initially sequencing of the entire symbiotic association as a metagenomic sample and subsequently assigning the genetic information to the symbiotic partners through bioinformatic analysis [96]. The approach of reconstructing genome sequences directly from metagenomic samples offers a means to bypass laborious cultivation or enrichment procedures, although it can present methodological challenges [100]–[102]. This method can lead to an uneven distribution of available genetic information within the symbiotic association, resulting in highly variable sequencing depth and genome coverage for individual partners, such as the host species or primary symbiont [81], [95], [103]. Assemblers, including those specifically designed for metagenomic samples, may exhibit sensitivity to low and uneven coverage, underscoring the critical role of assembler selection in ensuring the quality of genome reconstruction [95], [101], (please refer to the review on short and long-read assemblers from Xia et al [104]). Moreover, the potential generation of chimeric contigs, which are assembled from sequences originating from multiple genomes pose challenges to overcome during bioinformatic analysis [95]. Assembling repetitive DNA regions within a single genome, homologous regions of closely related strains, or conserved regions of different species can be particularly problematic, and the failure to resolve these regions may introduce rearrangement - and intergenomic assembly errors [105]. Additionally, post-assembly sequence binning procedures involve assigning contigs to taxonomic groups, typically based on similarity to sequenced genomes, as well as features such as GC content,

oligonucleotide frequency, and coverage depth [102]. Inaccurate assignments can lead to the formation of chimeric genome bins, which comprise contigs from multiple species. Furthermore, the success of similarity-based classification relies on the availability of high-quality, accurate, and contamination-free reference databases, particularly for whole-genome analysis [100]. There are more binning methods available such as *k*-mer based approaches which are described in the following reviews [104], [106]. However, in this thesis the taxonomic approach was deemed most suitable, based on sequence homology (see 2.5.2). Since 2011, publicly available genomes of lichen-forming fungi have been increasing [81], [96]. However, the majority of these genomes are derived from cultivated mycobionts [84], [107]–[109]. Given the challenges associated with axenic cultivation, metagenomic approaches have emerged as valuable tools for exploring the genomic characteristics of lichen-forming fungi, particularly for species that are difficult to isolate and grow in vitro [110]. These approaches are based e.g. on taxonomy, phylogeny or functional annotation [111]. Consequently, researchers have started employing metagenomics tools to investigate ecological, evolutionary, and biotechnological aspects of the lichen symbiosis. Notably, high-throughput metagenome or metatranscriptome sequencing directly from lichen thalli has been employed in various studies [112]. In summary, high-throughput metagenome sequencing plays a crucial role in studying lichen-forming fungi and photobionts that are difficult to cultivate, offering a valuable solution to overcome these cultivation challenges. However, to date, only a limited number of studies have utilized approaches to reconstruct genomes of lichen symbionts directly from whole metagenomic lichen thalli.

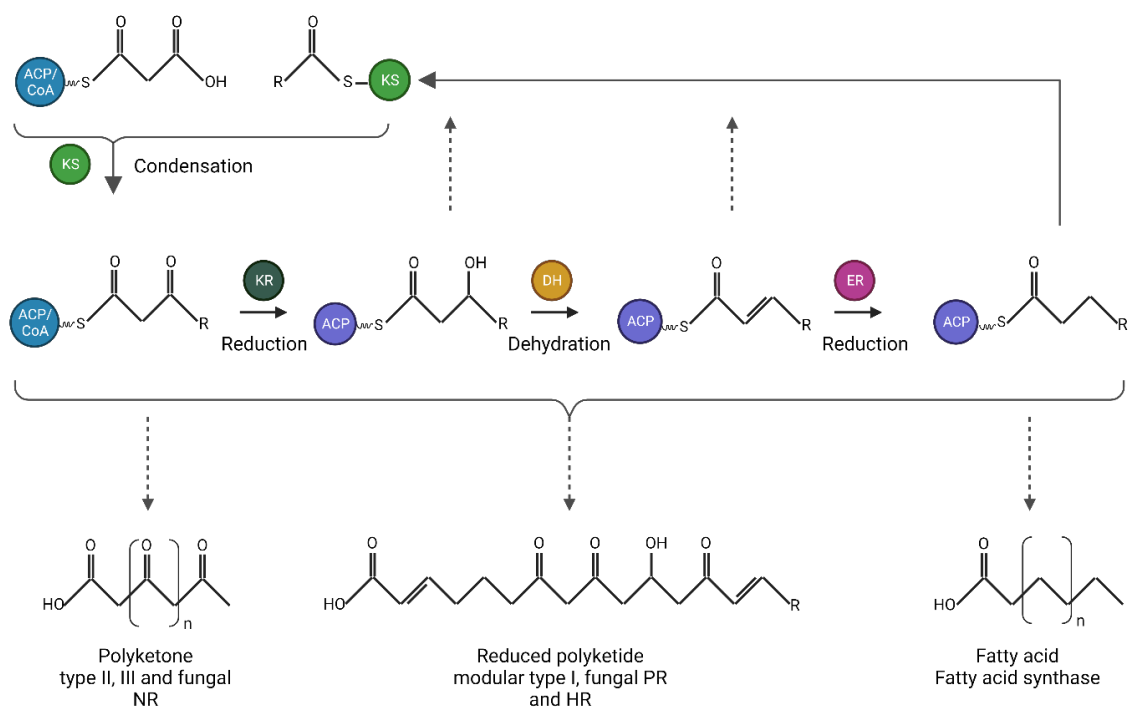
### **1.5 Genome Mining for Biosynthetic Gene Clusters – Secondary Metabolites in Lichen**

Despite the abundant array of chemical diversity observed in lichens, the utilization of biotechnological applications is currently constrained due to the limited availability of genomes and biosynthetic studies focused on lichen-forming fungi. In the last decade genome mining was introduced as the key technology to explore and exploit the natural product richness [113], [114]. Notwithstanding their chemical diversity, the production of all secondary metabolites is mediated by a limited number of shared biosynthetic pathways [115]. Lichen substances, known for their remarkable diversity, can be categorized into three primary groups based on their biosynthesis within the mycobiont. The majority of secondary metabolites found in lichens originate from the acetyl-polymalonyl pathway, which includes substance classes such as depsides, depsidones, dibenzofurans, anthraquinones, chromones, and xanthonones [49], [52], [116], [117]. Additionally, metabolite classes derived

## Introduction

from the shikimic acid pathway, such as terphenylquinones and pulvinic acid derivatives, as well as metabolite classes derived from the mevalonic acid pathway, such as steroids and carotenoids, are commonly encountered [49], [66]. Polyketide synthase (PKS) mega-enzymes utilizing the acetyl-polymanoyl pathway are responsible for synthesizing the most prevalent classes of lichen secondary metabolites [6], [58]. These enzymes connect monoaromatic subunits (such as orcinol,  $\beta$ -orcinol type or methylphloroacetophenone) through ester, ether, or carbon-carbon bonds to form phenolic compounds, dibenzofurans, depsides, and depsidones [52]. Fungal PKSs are comprised of multidomain systems that employ a set of active site domains repeatedly in multiple catalytic cycles to elongate polyketide products [39], [117]–[119]. PKSs are categorized based on the extent of reductive processing as partially or highly reducing polyketide synthases (R-PKS) and non-reducing polyketide synthases (NR-PKS) [120], with both types found in lichen-forming fungi. However, the most common lichen substance classes, such as depsides and depsidones, are derived from NR-PKSs [49], [116], [121]. Conserved structures in R-PKS from N- to C-termini consist of ketosynthase (KS); acyl transferase (AT); dehydratase (DH); C-methyl transferase (cMT); enoylreductase (ER); ketoreductase (KR); and acyl carrier protein (ACP) domains [117], [122]. In contrast, unique PKS domains comprising starter unit:acyl-carrier protein transferases (SAT) and a product template (PT) domain are observed in NR-PKS [120], [123]. SAT domains link the chain-initiating compound to the enzyme, while the PT domain is responsible for regulating the cyclization reactions that convert highly reactive, fully elongated intermediates into specific aromatic compounds. [120], [123]–[125]. NR-PKS domains are mostly organized in the following order SAT-KS-AT-PT-ACP-ACP-TE [126]. Other occurring units comprise AMP-binding sites (A), carrier proteins (CP) and the terminal domain (TD) [60], [127]. Figure 3 depicts a schematic reaction mechanism for polyketides and the involved domains in each step.

## Introduction



*Figure 3. Schematic reaction mechanism for the formation of polyketides and fatty acids, highlighting the potential for flux alteration towards polyketide synthesis. By modifying the reaction cycle, various polyketide derivatives can be generated with different levels of reduction. The respectively involved domains are depicted in coloured circles. Figure adapted from Weissman [128], created with Biorender.*

Another category of secondary metabolites includes non-ribosomal peptides produced by multidomain, multimodular enzymes known as non-ribosomal peptide synthetases (NRPSs). Operating independently of the ribosomal machinery, NRPSs assemble amino acids into peptides through amide bonds. Furthermore, the PKS and NRPS pathways can collaborate to generate hybrid polyketide synthase-non-ribosomal peptide synthetases (PKS-NRPSs), which produce structurally intricate hybrid molecules by fusing polyketides and amino acids via an amide bond [129]. Genes responsible for the production of distinct fungal secondary metabolites are frequently organized in a contiguous cluster [115]. Such gene clusters typically consist of a chemically defining core synthase or synthetase gene (e.g., PKS, NRPS, terpene synthase) that forms the backbone of the compound, along with genes encoding tailoring functions that can chemically modify the carbon backbone. In addition, these metabolic gene clusters often include transporter genes and pathway-specific regulatory genes, which play essential roles in the transport and regulation of the metabolic processes [39], [130]. The co-localization of genes involved in secondary metabolite biosynthesis within the genome has been effectively leveraged by advanced bioinformatic mining algorithms [131]. The review of Blin et al [132] summarizes various tools to assess BGCs, however in this study antiSMASH [133] was deployed to analyse metagenomic data on BGC level. Short-read sequencing methods can have limitations in accurately

## *Introduction*

representing isoforms among different species in a sample. This is due to the fact that short reads may not span across complex genomic or repetitive regions, as well as capture the full extent of genomic diversity in mixed microbial communities. As a result, it can be challenging to confidently assemble the complete genomes of all species present in the sample [134]–[136]. To overcome these limitations, long-read sequencing techniques are utilized instead. These technologies have the potential to generate more contiguous genome assemblies and improve our ability to accurately identify biosynthetic gene clusters (BGCs) and other functional genetic elements in complex microbial communities, resulting in high-quality data output [137]. Furthermore, nucleotide variances among symbionts can be detected by long reads, highlighting e.g. evolutionary adaptation of organisms through single nucleotide polymorphisms [138], [139]. Therefore, next-generation sequencing (NGS) has enabled the identification of potentially involved genes through deep genome sequencing. If multiple genes, that are involved in the formation of a specific secondary metabolite, are located in close proximity and not dispersed throughout the genome, they are referred to as a BGC [39], [115], [130], [140], [141]. A schematic BGC is provided in Figure 4. These advancements in genome sequencing techniques have significantly expanded the pool of available genomes that can be explored using these computational tools, thereby fostering natural product research. Consequently, genome mining has led to the identification of numerous BGCs [142], [143]. Recent studies have revealed that BGCs located within lichenized fungi are responsible for the production of various secondary metabolites, such as pigments, terpenes, and polyketides. [119], [144]–[147]. However, many of these identified gene clusters encode metabolites with unknown structures and functions, earning them the designation of orphan or cryptic gene clusters [39], [40], [133]. Additionally, the abundance of predicted biosynthetic gene clusters obtained through genome sequencing often does not correspond to the known expressed metabolic profile of the species, indicating the presence of a multitude of undiscovered chemical compounds [39], [130].

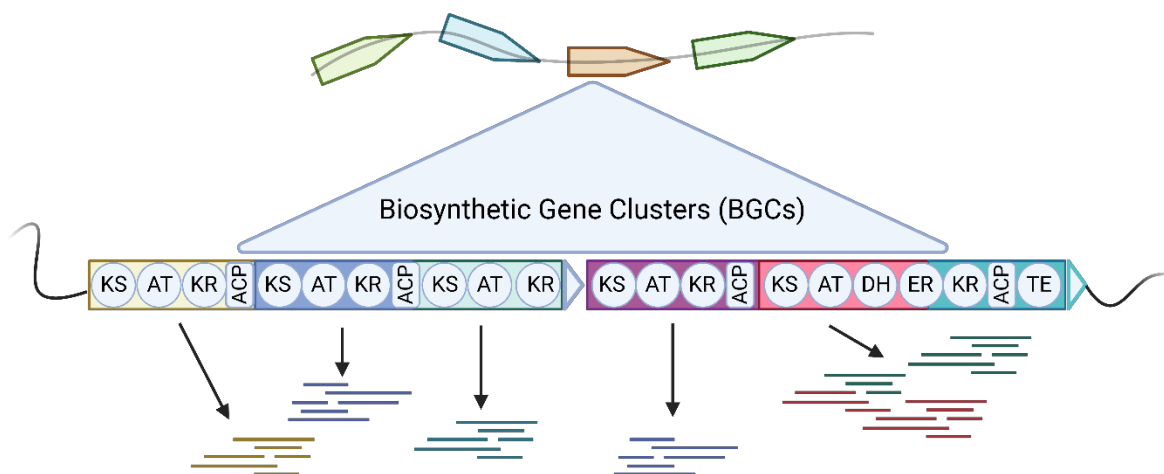


Figure 4. Schematic visualization of the genetic structure of a biosynthetic gene cluster (BGC). An example polyketide synthase (PKS) BGC is shown with the respective domains. Figure in adaption to Geers et. al [148] and created with Biorender.

To functionally annotate BGCs of composite organisms, sequencing of the entire metagenome is crucial to provide a comprehensive view of all BGCs present in the lichen symbiosis. This approach may facilitate uncovering interactions between the multiple species involved in the symbiosis at the level of BGCs [19]. Therefore, metagenomic tools were deployed to divide the obtained metagenomes taxonomically into respective bins. This renders symbiotic partners which are challenging to cultivate, accessible to genetic interrogation [149]. The genome of the mycobiont often harbors most of these BGCs [141], which are activated in response to environmental stimuli [150], [151]. In addition, various structurally similar compounds may be encoded by one BGC [152]–[154].

The functional characterization of fungal PKS gene clusters poses significant challenges. Unlike modular bacterial PKSs, fungal PKSs operate iteratively, concealing a cryptic programming within the PKS itself, which renders the polyketide structure unpredictable [115]. During each extension cycle, fungal PKSs have the ability to catalyse different sets of reactions by using the same catalytic domains, resulting in the generation of diverse and complex molecules from simple building blocks [155]. Our understanding of the intrinsic programming mechanisms that govern starter unit selection, chain length, and the degree of reduction, dehydration, and methylation in iterative PKSs is still limited [118], [156]. Additionally, the specificity and activity of tailoring enzymes, which play a crucial role in modifying the polyketide structure, remain mostly unknown, further complicating structure prediction [118]. To establish connections between genes and metabolite production

## Introduction

experimental characterization is imperative, such as through heterologous expression or gene knockout. However, these have been conducted on only a fraction of the numerous biosynthetic gene clusters predicted in sequenced genomes [157].

Due to their slow growth rates and the difficulties associated with establishing pure cultures for direct experimental genetic manipulations, the functional characterization of secondary metabolites in lichen-forming fungi is challenging [158]. Several potential assignments have been proposed using diverse methodologies. Various methods for proof of expression have been established to detect PKS genes in lichenized fungi, including the utilization of degenerate primers for amplifying PKS gene fragments from genomic DNA, the design of probes for library screening, and the application of complementary DNA (cDNA) templates generated from reverse-transcribed messenger RNA (mRNA) [159]. With the availability of transcriptomes and genomes of lichen-forming fungi, valuable insights into biosynthetic pathways can also be obtained directly from sequencing data of mycobiont cultures or metagenomic lichen thalli and comparing these based on sequence homology [109], [160]. Predicting the function of PKS genes relies on experimental evidence such as gene knockout experiments or heterologous expression [158], [161], [162]. In recent years, novel approaches have emerged to infer the functions of lichen-forming fungal secondary metabolites. These methods include assessing phylogenetic relatedness to characterized genes [163], [164], analysing transcriptional profiles [109], [165], and performing comparative homology mapping of entire gene clusters [118], [131], [160]. To achieve more accurate results, researchers often employ a combination of these approaches, leveraging the strengths of each method [109].

Further evaluation of identified PKS genes requires functional analysis which can be achieved by heterologous expression in adequate host organisms. For PKS genes various hosts can be considered, most commonly *Streptomyces* and *E. coli* are utilized, however *Saccharomyces cerevisiae* and *Aspergillus nidulans* are deployed as well [166].



## 2 Material and Methods

In the following paragraph an overview of the most important methods, procedures and tools used in this thesis is given. More concise information is enclosed in the respective materials and methods parts and the supplementary data sections of the corresponding publications included in this thesis.

### 2.1 Kits & Consumables

Kits utilized within the presented studies were obtained from the respective manufacturers. Table 2 summarizes deployed consumables.

*Table 2. Kits and Consumables deployed in this study.*

<b>Kit</b>	<b>Vendor</b>
Quick-DNA Fungal/Bacterial Miniprep Kit	Zymo Research, Freiburg, Germany
Genomic DNA Clean & Concentrator-10 Kit	Zymo Research, Freiburg, Germany
DNeasy PowerClean Clean up Kit	Qiagen, Venlo, The Netherlands
Qubit 2.0 Fluorometer	Thermo Scientific, Waltham, MA, USA
SMRTbell Express Template Prep Kit 2.0	Pacific Bioscience, Menlo Park, CA, USA
SMRTbell Enzyme Cleanup Kit	Pacific Bioscience, Menlo Park, CA, USA
AMPure PB Beads	Pacific Bioscience, Menlo Park, CA, USA
Barcoded Overhang Adapter Kit 8B	Pacific Bioscience, Menlo Park, CA, USA
Sequel II Binding Kit 2.0 + Internal Control	Pacific Bioscience, Menlo Park, CA, USA
SMRT Cell 8M tray	Pacific Bioscience, Menlo Park, CA, USA
Sequel II sequencing kit 2.0	Pacific Bioscience, Menlo Park, CA, USA

## 2.2 Lichen Sample Collection

To collect lichen species of the genus *Parmeliaceae* with adequate product spectra, the phenotypic characteristics were investigated. The chosen species comprise *Hypogymnia physodes* (HPH), *Hypogymnia tubulosa* (HTU) and *Parmelia sulcata* (PSU). To limit contamination of the collected samples to a minimum, handling with gloves was mandatory and only sterile devices were deployed to detach lichen material from the substrate.

*Parmeliaceae* typically grow on the bark of conifers. All samples were collected in Germany, Altenschneeberg (August 2022, latitude 49°26'14.3"N and longitude 12°32'50.9"E). A subsequent BLAST search on ITS sequences ensured correct sample collection and identified the obtained lichens as the aforementioned species.

### Gas chromatography - FID/MS analysis

Furthermore, gas chromatography with mass spectrometry (GC-MS) was utilized to investigate the secondary metabolite composition of the collected samples. Here, 500 mg of dry lichen biomass was subjected to maceration in 10 mL of MeOH for 24 hours at 300 rpm. The obtained extract was then investigated by using a Trace GC-MS Ultra system with DSQII (Thermo Scientific, Waltham, MA, USA). A volume of 1  $\mu$ L was injected via a TriPlus autosampler in split mode onto an SGE BPX5 column (30 m, I.D 0.25 mm, film 0.25  $\mu$ m). The injector temperature was set at 280 °C. Initial oven temperature was maintained at 50 °C for 2.5 minutes, followed by a temperature increase at a rate of 10 °C/min until reaching 320 °C, with a final hold step for 3 minutes. Helium functioned as the carrier gas with a flow rate of 0.8 mL/min and a split ratio of 8. Chromatograms and respective mass spectra were recorded using electron ionization (EI) at 70 eV. Detection of masses was monitored within the range of 50 m/z to 650 m/z in positive mode [167]. Same procedures were applied with GC-FID.

To identify obtained compounds, the resulting mass spectra were compared to the NIST/EPA/NIH MS library version 2.0. Among the identified compounds were secondary metabolites unique to lichen e.g., resorcylic acid derivatives, olivetolic acid, orsellinic acid and derivatives.

### **2.3 High Molecular Weight DNA Extraction and Evaluation**

Prior to DNA extraction obtained lichen thalli were examined by a binocular microscope. Here, any remaining wood, moss, and other lichen species were eliminated. In addition, possible contaminants were removed by dissecting visibly infected parts of the thalli. To prepare the lichen samples for sequencing, a previous high molecular weight genomic DNA (HMW gDNA) extraction is necessary. Therefore, the HMW gDNA was extracted by using the Quick-DNA Fungal/Bacterial Miniprep Kit (Zymo Research, Europe GmbH, Freiburg, Germany). Therefore, liquid nitrogen was deployed to ground dry thallus material from lichen samples. This was conducted by using the CryoMill (Retsch, Haan, Germany). Parameters for grinding were adjusted to sample characteristics e.g., thickness or sample volume. Subsequently, obtained homogenate was transferred to the Bashing Bead Buffer provided in the kit. Isolation of HMW gDNA was conducted according to manufacturer's instructions. With lichens harbouring high contents of pigments, polysaccharides and phenolic compounds, additional purification steps were necessary. These were performed by utilizing the Genomic DNA Clean and Concentrator-10 Kit (Zymo Research, Europe GmbH, Freiburg, Germany) and the DNeasy PowerClean Clean up Kit (Qiagen, Venlo, The Netherlands).

To ensure sufficient quality of the obtained HMW gDNAs, a consecutive assessment including analyses on Nanophotometer (Implen, Nanophotometer Pearl, Munich, Germany), Qubit 2.0 Fluorometer (Thermo Scientific, Waltham, MA, USA). Thresholds for this included a 260/280 absorbance ratio of 1.75-1.85 and a 260/230 absorbance ratio of 2.0-2.2. To enable successful long-read sequencing, the DNA fragment size needed to within 7 to 20 kilobases (kb). Therefore, TapeStation and FemtoPulse (Agilent Technologies, Santa Clara, CA, USA) analyses were conducted. Samples passing the quality control were subjected to further processing. SMRT bell libraries were constructed based on the instructions of the Low DNA Input Protocol of the SMRT bell Express Prep kit v2. (Pacific Bioscience, Menlo Park, CA, USA). Total DNA input for library generation was approximately 350-600 ng. Subsequent ligation with T-overhang SMRT bell adapters was performed at 20 °C for 1 hour, followed by two clean-up steps with AMPure PB beads. To ensure sufficient size and concentration of the final library, a consecutive assessment using TapeStation/FemtoPulse and Qubit Fluorometer 2.0 with Qubit dsDNA HS reagents Assay Kit was performed. Necessary amounts of components from the SMRTbell Kit were provided by SMRT-Link depending on input DNA content.

## **2.4 Sequencing**

### **2.4.1 Whole Genome Sequencing**

Prepared HMW gDNA libraries (SMRTbell libraries) of lichen samples were subjected to whole genome sequencing on a PacBio Sequel IIe device (Pacific Bioscience, Menlo Park, CA, USA). Pre-extension and adaptive loading (target of  $p1 + p2 = 0.95$ ) were set to two hours with an on-plate concentration of 90 pM. The movie time was set to 30 h [168]. After sequencing, the data were transferred to the SMRT-Link server (v11.0.0.146107) and processed in order to generate HiFi reads.

### **2.4.2 RNA Long Read IsoSeq and Illumina Short Read Sequencing**

In this study two approaches on RNA sequencing were performed, long read IsoSeq (Pacific Bioscience, Menlo Park, CA, USA) and short read Illumina (NovaSeq, Illumina, San Diego, CA, USA). The IsoSeq RNA sequencing method on PacBio Sequel IIe (Pacific Bioscience, Menlo Park, CA, USA) utilizes complete transcripts and enables isoform sequencing [85].

#### **PacBio IsoSeq**

To enhance the genome assembly quality, transcripts were sequenced to augment the depth and precision of the predicted gene models. RNA was extracted from frozen, unthawed lichen thalli using a CryoMill (Retsch, Haan, Germany) and a RNeasy Plant Mini Kit (Qiagen, Venlo, The Netherlands). To further clean the obtained RNA, a Turbo DNA free Kit (Invitrogen) was used. For long read RNA sequencing, an IsoSeq library preparation was performed on high quality RNA using the Sequel II Binding Kit 3.2 and SMRTbell prep kit 3.0. (Pacific Bioscience, Menlo Park, CA, USA). Transcriptome sequencing on Sequel IIe was performed at 24 h of movie time with an on-plate concentration of 80 pM.

#### **Illumina NovaSeq**

In case of short read RNA sequencing, samples were processed on a NovaSeq device, employing a paired-end run mode and  $2 \times 150$  bp read length. Therefore, total RNAs were extracted with TRI Reagent (Zymo Research, Europe GmbH, Freiburg, Germany) according to the manufacturer's instructions. Further purification was obtained by processing the samples with the RNA Clean and Concentrator-5 Kit (Zymo Research, Europe GmbH, Freiburg, Germany). This step was repeated until a 260/280 absorbance ratio of 1.9–2.1 and a 260/230 absorbance ratio of 1.8–2.2 were obtained. Solely RNAs with a RIN value  $>8.0$  (TapeStation) were considered for sequencing.

## 2.5 Bioinformatic Analysis

### 2.5.1 Metagenomic assembly

The assembly was conducted using metaFlye v2.9.1, circumventing the aforementioned challenges of non-uniform genome compositions in metagenomic samples. Furthermore, the assembled contigs were concurrently scaffolded using Flye, facilitating additional bioinformatic analyses [95].

### 2.5.2 Metagenomic Binning & Quality Assessment– DIAMOND & MEGAN

To distinguish between the obtained datasets, a taxonomic binning strategy was implemented. This involved conducting blastx using the DIAMOND v2.0.14 algorithm [169] on a custom database, as well as utilizing the MEGAN6 LR Community Edition v6.21.7 [170]. The DIAMOND database employed in the analysis encompassed protein sequences from diverse taxonomic groups, such as archaea, chlorophyta, cystobasidium fungi, bacteria, klebsormidophyceae, tremella, and viruses. To overcome challenges related to insertion and deletion errors in long read sequencing, the flags *--more-sensitive*, *--frameshift 15*, and *--rage-culling* were utilized in DIAMOND, allowing for a frame-shift-aware alignment mode [171]. The generated files underwent additional processing in MEGAN, where taxonomically classified sequences were associated with their respective bins [172]. Following that, contigs and scaffolds corresponding to the targeted nodes were extracted for subsequent analysis. To assess the comprehensiveness and reliability of the resulting bins for further investigation, multiple tools were employed, including BUSCO v5.3.2 (Benchmarking Universal Single-Copy Orthologs) [173], QUASt, v5.2.0 (Quality Assessment Tool) [174] and SeqKit v2.3.1 [175]. To verify the identities of the mycobiont, an analysis using ITSx v1.1.3 was conducted.

### 2.5.3 Gene Prediction

Gene prediction was carried out using AUGUSTUS v3.4.0/BRAKER v2.1.6, utilizing both the metagenomics data and corresponding transcriptomic data as hints [176]–[178]. AUGUSTUS uses various algorithms and statistical models to identify protein-coding genes within DNA sequences. Hereby considering factors such as coding sequences, introns, splice sites, and other structural features typically associated with genes. Furthermore, AUGUSTUS can be trained on species-specific data to improve its accuracy. This is especially useful when working with non-model organisms or species without well-annotated genomes [179], [180]. BRAKER is used gene prediction where no reference genome is available. It combines the predictions of two different gene prediction tools, GeneMark-ET, and AUGUSTUS, along with evidence from RNA-Seq data. GeneMark-ET is used initially to predict genes, and then AUGUSTUS further refines these predictions. RNA-Seq data, which provides information on gene expression, is incorporated to improve accuracy [176]–[178].

### 2.5.4 Functional annotation and visualization of biosynthetic gene clusters

This approach enabled the functional annotation of genes in the respective datasets. Additionally, the identification of biosynthetic gene clusters (BGCs) within the obtained bin was performed using antiSMASH v6.1.1 [133]. AntiSMASH primarily relies on Hidden Markov Models and BLAST to identify biosynthetic gene clusters and related sequences, using predictive algorithms to analyse gene boundaries, promoters, terminators, and chemical structure predictions [133]. The Gene Ontology (GO) terms were assigned using InterProScan (v5.59-91.0) for further characterization [181]. To visualize the distribution of BGCs in all three mycobionts an evaluation with BiG-SCAPE [182] and Cytoscape [183] was conducted, clustering the respective contigs by similarity networks. In order to gain insights into the polyketide synthesis in the mycobiont, PKS-related BGCs were investigated further.

### 2.5.5 Similarity Assessment of PKS genes

To investigate shared genes, at first a Reciprocal Best Hit (RBH) based on BLAST was performed on the respective primary mycobiont bins of PSU, *Hypogymnia physodes* (HPH), and *Hypogymnia tubulosa* (HTU) [184], [185] in all three combinations. The latter two were previously published [186]. Reciprocal best BLAST is a valuable technique in comparative genomics and phylogenetics identifying genes likely to have similar functions in different species. AntiSMASH, InterProScan, ITSx, SeqKit and RBH analyses were performed on the Galaxy servers [187]. Obtained pairs from RBH were compared for duplicates, highlighting highly conserved PKS-related genes in the investigated mycobiont genomes.

### 2.5.6 Phylogenetic Trees of identified PKS

Therefore, the sequences of KS regions were extracted and compared on a phylogenetic level using the Geneious Tree Builder (Geneious Prime® 2022.0.1) and IQ-TREE [188]. The tree in Geneious was built with the Neighbor-Joining method with the genetic distance model Tamura-Nei. With IQ-TREE a maximum likelihood tree was built using 1000 ultrafast bootstraps and the substitution model defined by ModelFinder [189] (in this case LG+F+G8). Visualization was conducted with interactive Tree of Life (iTOL) [190] and Inkscape.

### 2.5.7 Syntenic evaluation of orthologous PKS BGCs

Subsequently, an alignment of related gene clusters was performed on EasyFig (v2.2.5) [191] to investigate PKS synteny between HPH, HTU and PSU. Additionally, a progressiveMAUVE [192] alignment of the whole mycobiont bins was deployed to visualize synteny on the genome level in Geneious (Geneious Prime® 2022.0.1). These analyses are based on sequence homology, highlighting conserved regions in the investigated BGCs.

## 3 Research

### 3.1 Summaries of included publications

#### Chapter I Biosynthetic potential of *Hypogymnia* holobionts - Insights into Secondary Metabolite Pathways

Lichens represent mutualistic partnerships involving a photosynthetic organism (algae or cyanobacteria) and a fungal organism. These symbiotic entities are renowned for their production of distinctive secondary compounds. To exploit this biosynthetic capacity for biotechnological purposes, it is crucial to gain a deeper understanding of the biosynthetic pathways and associated gene clusters. This study aims to offer a comprehensive overview of the biosynthetic gene clusters present in all components of a lichen thallus, including fungi, green algae, and bacteria. By utilizing two high-quality PacBio metagenomes, we identified a total of 460 biosynthetic gene clusters. The fungal component of lichens exhibited 73–114 clusters, while other ascomycetes associated with lichens displayed 8–40 clusters. The green algae of the *Trebouxia* genus showcased 14–19 clusters, while lichen-associated bacteria possessed 101–105 clusters. Notably, the mycobionts predominantly contained T1PKSs, followed by NRPSs and terpenes. In contrast, the *Trebouxia* reads mainly contained clusters associated with terpenes, followed by NRPSs and T3PKSs. Other lichen-associated ascomycetes and bacteria contained diverse combinations of biosynthetic gene clusters. This study provides the initial identification of biosynthetic gene clusters in complete lichen holobionts, thereby rendering the unexplored biosynthetic potential in two *Hypogymnia* species accessible for further investigation.

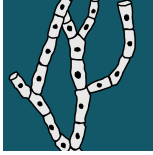
## **Chapter II Biosynthetic Gene Cluster Synteny - Orthologous Polyketide Synthases in *Hypogymnia physodes*, *Hypogymnia tubulosa* and *Parmelia sulcata***

Lichens form symbiotic partnerships involving a photosynthetic organism (algae or cyanobacteria) and a fungal organism, which collectively produce a diverse array of distinctive secondary compounds. To harness this biosynthetic potential for biotechnological advancements, it is crucial to gain a more comprehensive understanding of the biosynthetic pathways and corresponding gene clusters. In this study, we present a comparative analysis of the biosynthetic gene clusters from three lichen mycobionts: *Hypogymnia physodes*, *Hypogymnia tubulosa*, and *Parmelia sulcata*. Additionally, we provide a high-quality PacBio metagenome of *Parmelia sulcata*, from which we extracted the mycobiont bin containing 214 biosynthetic gene clusters. Most of these gene clusters were associated with T1PKSs, followed by NRPSs and terpenes. Our investigation primarily focused on biosynthetic gene clusters related to polyketide synthesis. Through ketosynthase homology, we identified nine highly syntenic clusters present in all three species. Among the four clusters associated with non-reducing PKSs, two are potentially involved in the synthesis of lichen substances derived from orsellinic acid (such as orcinol depsides and depsidones, e.g., lecanoric acid, physodic acid, lobaric acid), one is linked to compounds derived from methylated forms of orsellinic acid (beta orcinol depsides, e.g., atranorin), and one is associated with melanins. Furthermore, five clusters with orthologs in all three species are connected to reducing PKSs. Our study contributes to the classification and dereplication of the extensive polyketide synthase diversity found in lichenized fungi. The high-quality sequences of biosynthetic gene clusters obtained from these three common lichen species establish a foundation for further exploration of biotechnological applications and the molecular evolution of lichen substances.



**3.2 Full length publications**

Biosynthetic potential of *Hypogymnia* holobionts -  
Insights into Secondary Metabolite Pathways



Journal of  
*Fungi*

IMPACT  
FACTOR  
5.724

Indexed in:  
PubMed

Article

---

# Biosynthetic Potential of *Hypogymnia* Holobionts: Insights into Secondary Metabolite Pathways

---

Nadim Ahmad, Manfred Ritz, Anjuli Calchera, Jürgen Otte, Imke Schmitt, Thomas Brueck and  
Norbert Mehlmer

Special Issue

Advances in *Lichen* Species

Edited by


Prof. Dr. Juri Nascimbene and Dr. Gabriele Gheza



<https://doi.org/10.3390/jof9050546>

## Article

# Biosynthetic Potential of *Hypogymnia* Holobionts: Insights into Secondary Metabolite Pathways

Nadim Ahmad <sup>1,†</sup> , Manfred Ritz <sup>1,†</sup>, Anjuli Calchera <sup>2</sup> , Jürgen Otte <sup>2</sup> , Imke Schmitt <sup>2,3</sup>, Thomas Brueck <sup>1,\*</sup>   
and Norbert Mehlmer <sup>1,\*</sup> 

<sup>1</sup> Werner Siemens Chair of Synthetic Biotechnology, Department of Chemistry, Technical University of Munich (TUM), 85748 Garching, Germany

<sup>2</sup> Senckenberg Biodiversity and Climate Research Centre (SBIK-F), Senckenberganlage 25, 60325 Frankfurt am Main, Germany

<sup>3</sup> Institute of Ecology, Evolution and Diversity, Goethe University Frankfurt, Max-von-Laue-Straße 13, 60438 Frankfurt am Main, Germany

\* Correspondence: brueck@tum.de (T.B.); norbert.mehlmer@tum.de (N.M.)

† These authors contributed equally to this work.

**Abstract:** Lichens are symbiotic associations consisting of a photobiont (algae or cyanobacteria) and a mycobiont (fungus). They are known to produce a variety of unique secondary metabolites. To access this biosynthetic potential for biotechnological applications, deeper insights into the biosynthetic pathways and corresponding gene clusters are necessary. Here we provide a comprehensive view of the biosynthetic gene clusters of all organisms comprising a lichen thallus: fungi, green algae, and bacteria. We present two high-quality PacBio metagenomes, in which we identified a total of 460 biosynthetic gene clusters. Lichen mycobionts yielded 73–114 clusters, other lichen associated ascomycetes 8–40, green algae of the genus *Trebouxia* 14–19, and lichen-associated bacteria 101–105 clusters. The mycobionts contained mainly T1PKSs, followed by NRPSs, and terpenes; *Trebouxia* reads harbored mainly clusters linked to terpenes, followed by NRPSs and T3PKSs. Other lichen-associated ascomycetes and bacteria contained a mix of diverse biosynthetic gene clusters. In this study, we identified for the first time the biosynthetic gene clusters of entire lichen holobionts. The yet untapped biosynthetic potential of two species of the genus *Hypogymnia* is made accessible for further research.

**Keywords:** *Hypogymnia physodes*; *Hypogymnia tubulosa*; long read sequencing; polyketide synthesis; biosynthetic gene cluster; lichen; reference genome



check for  
updates

**Citation:** Ahmad, N.; Ritz, M.; Calchera, A.; Otte, J.; Schmitt, I.; Brueck, T.; Mehlmer, N. Biosynthetic Potential of *Hypogymnia* Holobionts: Insights into Secondary Metabolite Pathways. *J. Fungi* **2023**, *9*, 546. <https://doi.org/10.3390/jof9050546>

Academic Editors: Juri Nascimbene and Gabriele Gheza

Received: 7 April 2023

Revised: 21 April 2023

Accepted: 25 April 2023

Published: 9 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Lichens were historically regarded as a symbiotic organism comprising a fungus and photosynthetic partners (algae/cyanobacteria) [1]. However, more recent studies suggested that lichen individuals host not only the primary lichen-forming fungus (mycobiont) and primary photosynthetic partner (photobiont), but also additional fungi belonging to the Basidiomycota [2], or Ascomycota [3], as well as a diverse suite of bacteria and other algae [4,5]. Lichens have therefore been referred to as ecosystems [6], or holobionts [7]. Due to their unique secondary metabolite spectrum, these composite organisms have gained increasing attention over the past decades [8]. Lichen mycobionts synthesize a plethora of natural products [9]. The most prominent classes comprise depsides, depsidones, dibenzofurans, and phenolic compounds [10,11]. Confirmed bioactivities of these compounds encompass antimicrobial [8,12–14], antifungal [15], anti-inflammatory [16], antioxidant [17,18], and anticancer [14,19] functions. For the latter, medically relevant examples are physodic acid, gyrophoric acid, and atranorin [20]. An inhibitory effect on metabolic enzymes was also detected in evernic acid, physodic acid, usnic acid, and

atranorin [10,21]. Bacteria associated with lichens have also been shown to possess diverse compounds with pharmaceutically promising activities [22–24].

To exploit these compounds biotechnologically, it is crucial to elucidate their formation thoroughly. The above-mentioned compounds belong to the class of polyketides and are built by multi-enzyme polyketide synthases (PKS) via repetitive catalytic cycles of linking elongation units to a starter molecule [25,26]. In regard to fungal PKSs, a differentiation between reducing (R-PKS) and non-reducing PKS (NR-PKS) is made based on extent of reductive processing and domain composition [27]. For R-PKS, the conserved structure from N- to C-termini consists of keto synthase (KS), acyl transferase (AT), dehydratase (DH), C-methyl transferase (C-MeT), enoylreductase (ER), ketoreductase (KR), and acyl carrier protein (ACP) domains [27,28]. In comparison, NR-PKS contain starter unit:acyl-carrier protein transferases (SAT) and a product template (PT) domain, which are unique among the PKS families [26,29]. The former connects the chain initiating compound to the enzyme, whereas the latter regulates cyclization reactions of the highly reactive and completely elongated intermediate to specific aromatic compounds [26,29–31]. Organization of most NR-PKS domains was found to be SAT-KS-AT-PT-ACP-ACP-TE [32].

For further investigation of secondary metabolite production, the underlying mechanisms need to be elaborated. Through advancement in the fields of next generation sequencing (NGS), it became possible to locate potentially involved genes. When several genes participating in formation of a secondary metabolite are located in close vicinity rather than spread over the genome, the term biosynthetic gene cluster (BGC) is applied [33–37]. Recent studies have identified BGCs from lichenized fungi responsible for the biosynthesis of secondary metabolites, including those for pigments, terpenes, and polyketides [38–42].

In order to identify specific BGCs, it is necessary to have access to the molecular makeup of the symbionts. This can be achieved by simultaneously sequencing the complete metagenome, which provides a comprehensive view of the BGCs of all organisms involved in the lichen association. By doing so, potential interactions between multi-species symbioses could be detected at the level of BGCs [4]. In addition, genetic accessibility is enabled to symbiotic partners which are not or challenging to culture [43]. These clusters are often located on the genome of the fungal partner of the lichen [34] and are activated in response to environmental cues [44,45]. However, one BGC may encode for various structurally familiar compounds [46–48]. Sequencing a pool of microorganisms using short read sequencing methods can be challenging, as nucleotide variances among different species can be inadequately represented or lost during the metagenomics analysis. This is also true for complex regions, making it difficult to efficiently assemble the complete genomes of all species in the sample [49–51]. Long read sequencing techniques instead, are able to provide a more comprehensive view of the genome of different microorganisms within the metagenome sample, thus resulting in high quality data output [52]. Furthermore, the long reads can be used to detect nucleotide variances among symbionts [53]. For the first time, the BGCs of the complete holobiont-especially algae and bacteria- directly from lichen thalli were identified.

In this study, we present two new lichen metagenomes of the genus *Hypogymnia*, *Hypogymnia physodes* (HPH) and *H. tubulosa* (HTU). *H. physodes* is one of the most ubiquitous and abundant macrolichens throughout Europe, and also occurs in North America, Africa, and Asia. It is a foliose, polymorphic species, up to 10 cm diameter in size, and forms lip-shaped soralia at the tips of the lobes. It is found on acidic substrata, including the bark of trees, and siliceous rocks. *H. tubulosa* is similar to *H. physodes*, but the lobes are more tubular, and soralia rounded, capitate, not lip-shaped. It has a similar distribution and grows on similar substrata as *H. physodes*, but is less frequent [54,55]. Both species often grow together, and both are associated with green algae of the genus *Trebouxia*. Natural products found in *Hypogymnia* are atranorin, chloroatranorin; physodic acid, 3-hydroxyphysodic acid, physodalic acid; and occasionally protocetraric, fumarprotocetraric, and usnic acid, some of which are medicinally relevant [56–58]. Consequently, these

metagenomes provide an important resource for further research on the biosynthetic potential of the genus *Hypogymnia*.

## 2. Materials and Methods

### 2.1. Lichen Sample Collection

Samples were collected in Germany, Bavaria, Altenschneeberg (August 2022), from the bark of conifers, where they grew side by side. Precise locations of sequenced samples are listed in Table 1. To ensure correct lichen identity, a BLAST search on ITS sequences was performed. Lichen samples included in this study were identified as HPH and HTU.

**Table 1.** Geographic coordinates of the lichen sample collection.

Lichen Species	Latitude	Longitude
<i>Hypogymnia physodes</i> (HPH)	49°26'14.3" N	12°32'50.9" E
<i>Hypogymnia tubulosa</i> (HTU)	49°26'14.3" N	12°32'50.9" E

### 2.2. GC-MS Analysis of Lichen Compounds

In addition, parts of the collected samples were analyzed by GC-MS to investigate the secondary metabolite composition. A measure of 500 mg of dry lichen biomass was macerated in 10 mL methanol for 24 h at 300 rpm. The obtained extract was analyzed by a Trace GC-MS Ultra system with DSQII (Thermo Scientific, Waltham, MA, USA). A TriPlus autosampler was employed to inject 1 µL of sample volume in split mode onto a SGE BPX5 column (30 m, I.D 0.25 mm, film 0.25 µm); injector temperature was set to 280 °C. Initial oven temperature was kept at 50 °C for 2.5 min. The temperature was increased with a ramp rate of 10 °C/min to 320 °C with a final hold step for 3 min. The carrier gas in this study was helium with a flow rate of 0.8 mL/min and a split ratio of 8. The mass spectra and chromatograms were recorded at 70 eV (EI). Detection of masses took place between 50 *m/z* and 650 *m/z* in the positive mode [53]. Compounds were identified by spectral comparison with a NIST/EPA/NIH MS library version 2.0.

### 2.3. High Molecular Weight DNA (HMW gDNA) Extraction and Library Preparation

Prior to DNA extraction, the respective lichen thallus was investigated under a binocular microscope to remove moss, wood, and other lichens from the sample. In addition, parts of the thallus, which were visibly infected by parasitic fungi were removed to minimize contaminants in the sample.

HMW gDNA extraction was conducted as follows. Lichen HMW gDNA was extracted using the Quick-DNA Fungal/Bacterial Miniprep Kit (Zymo Research, Europe GmbH, Freiburg, Germany). Dry thallus material of *Hypogymnia* samples was ground to fine powder in liquid nitrogen. The homogenized material was transferred into Bashing Bead Buffer of the kit. Genomic HMW DNA was isolated according to the manufacturer's instructions. Due to the high content of polysaccharides, phenolic compounds and pigments further, purifications were necessary and carried out using the Genomic DNA Clean and Concentrator-10 Kit (Zymo Research, Europe GmbH, Freiburg, Germany) and the DNeasy PowerClean Clean up Kit (Qiagen, Venlo, The Netherlands). The quality of obtained HMW gDNAs was assessed with Nanophotometer (Implen, Munich, Germany), Qubit 2.0 Fluorometer (Thermo Scientific, Waltham, MA, USA) and TapeStation (Agilent Technologies, Santa Clara, CA, USA). SMRT bell libraries were constructed for samples passing the quality control (260/280 absorbance ratio of 1.75–1.85 and a 260/230 absorbance ratio of 2.0–2.2) according to the instructions for the Low DNA Input Protocol of the SMRT bell Express Prep kit v2. (Pacific Bioscience, Menlo Park, CA, USA).

Total input DNA (size: 10–18 kb) for the library generation was approximately 350–600 ng. The ligation with T-overhang SMRT bell adapters was performed at 20 °C for 1 h. After two clean up steps with AMPure PB beads, the size and concentration of the final

libraries were assessed using TapeStation and Qubit Fluorometer 2.0 with Qubit dsDNA HS reagents Assay Kit.

#### 2.4. Genome Sequencing

Whole genome sequencing of prepared lichen HMW gDNA libraries was conducted with a PacBio Sequel Iie device (Pacific Bioscience, Menlo Park, CA, USA). Pre-extension and adaptive loading (target of  $p1 + p2 = 0.95$ ) were set to two hours with an on-plate concentration of 90 pM. The movie time was set to 30 h [59].

#### 2.5. RNA Long Read Isoseq and Illumina Short Read Sequencing

In this study, long read IsoSeq (Pacific Bioscience, Menlo Park, CA, USA) and short read Illumina (NovaSeq, Illumina, San Diego, CA, USA) RNA sequencing were performed. To increase the quality of the genome assembly, transcripts were sequenced to add more depth and accuracy to the proposed gene models. For RNA extraction, frozen, unthawed lichen thalli were grinded using a CryoMill (Retsch, Haan, Germany) and a RNeasy Plant Mini Kit (Qiagen, Venlo, The Netherlands). To further clean the obtained RNA, a Turbo DNA free Kit (Invitrogen) was used. For long read RNA sequencing, an IsoSeq library preparation was performed on high quality RNA using the Sequel II Binding Kit 3.2 and SMRTbell prep kit 3.0. (Pacific Bioscience, Menlo Park, CA, USA). Transcriptome sequencing on Sequel Iie was performed at 24 h of movie time with an on-plate concentration of 80 pM. In case of short read RNA sequencing, samples were processed on a NovaSeq device, employing a paired-end run mode and  $2 \times 150$  bp read length. Therefore, total RNAs were extracted with TRI Reagent (Zymo Research, Europe GmbH, Freiburg, Germany) according to the manufacturer's instructions. Further purification was obtained by processing the samples with the RNA Clean and Concentrator-5 Kit (Zymo Research, Europe GmbH, Freiburg, Germany). This step was repeated until a 260/280 absorbance ratio of 1.9–2.1 and a 260/230 absorbance ratio of 1.8–2.2 were obtained. Only RNAs with a RIN value  $>8.0$  (TapeStation) were used for sequencing.

#### 2.6. Bioinformatic and Statistical Analysis

Metagenomic reads generated from whole lichen thalli predominantly contain fungal sequences from the mycobiont [60], which may pose problems to genome assemblers using solid  $k$ -mers. In this case, highly prevalent species will be over-represented and low-abundance species such as the photobiont will fail to assemble. To circumvent these issues, obtained Sequel Iie CCS long reads were assembled using metaFlye v2.9.1, as it is suitable to process read coverages of high non-uniformity. In addition, a simultaneous scaffolding of the assembled contigs was performed with flye, enabling further bioinformatic processing [61].

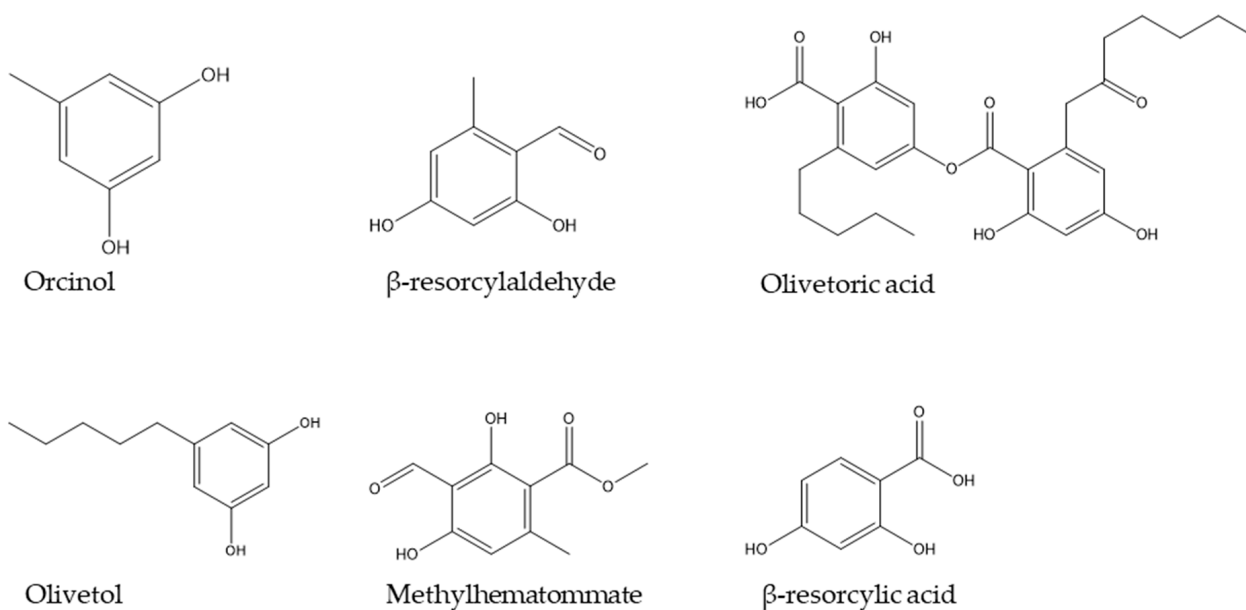
To distinguish obtained datasets, a taxonomical binning was conducted deploying blastx in the DIAMOND v2.0.14 algorithm [62] on a custom made database and MEGAN6 LR Community Edition v6.21.7 pipeline [63]. The DIAMOND database contained protein sequences of the following taxonomic groups: fungi, bacteria, archaea, viruses, chlorophyta, klebsormidiophyceae, tremella, and cystobasidium. For DIAMOND, the flags *–more-sensitive* *–frameshift 15* and *–rage-culling* were employed to account for insertion and deletion errors in long read sequencing with a frame-shift-aware alignment mode [64]. Resulting files were further processed in MEGAN by matching taxonomically assigned sequences to respective bins [65]. Consequently, contigs and scaffolds of desired nodes were extracted and subjected to consecutive analysis. To assess genome completeness and quality of resulting bins for further investigation, BUSCO, v5.3.2 (Benchmarking Universal Single-Copy Orthologs) [66], QUASt, v5.2.0 (Quality Assessment Tool) [67] and SeqKit v2.3.1 [68] were applied. For the validation of mycobiont and photobiont identities and distribution in investigated bins, an ITSx v1.1.3 analysis was performed [69]. Gene prediction with AUGUSTUS v3.4.0/BRAKER v2.1.6 is conducted, based on metagenomics data and respective transcriptomic data as hints [70–72]. This enables the functional annotation of genes in

the respective datasets. A consecutive annotation of BGCs present in the obtained bins is performed by antiSMASH v6.1.1 [73], whereas Cluster of Orthologous Groups (COG) and Gene Ontology (GO) terms are levied by the EggNOG Mapper (v2.1.5) [74]. AntiSMASH, EggNOG, ITSx, and SeqKit analyses were performed on the Galaxy servers [75].

### 3. Results and Discussion

#### 3.1. Volatile Compound Identification

GC-MS was used to identify volatile compounds in the macerated lichen samples. As shown in literature, lichen contain a plethora of secondary metabolites varying from depsides, depsidones, to dibenzofurans over to resorcylic-like compounds and anthraquinones [17]. In Figure 1, the polyketide-related compounds for both lichen samples are depicted. These are based on GC-MS data, which can be found in Table S1. Results showed similar compounds for both samples. These complex structures harbor aromatic ring elements as well as aliphatic elements and different chiral centers. In total chemical synthesis, complex biological molecules are not produced in high yields due to these centers and also to chemically reactive substituents as seen for the total synthesis of tetracycline [76]. Thus, a biological production system would be favorable. In nature, complex molecules are mostly secondary metabolites as part of multi-enzyme biosynthesis pathways [77]. The connected enzymes can be found on distinct clusters. Therefore, it is necessary to provide a comprehensive and high-quality genome to enable accessibility of the production-related genes.



**Figure 1.** Molecule structures of lichen compounds identified by GC-MS.

#### 3.2. Genome Sequencing and Quality Assessment

To present sequencing quality, the sequencing metrics of the two lichen samples HPH and HTU are listed in Table 2. Obtained metrics are comparable in each sample with only slight deviations. Due to mean HiFi Read Quality well above Q20, all samples were eligible for further bioinformatic analyses.

After taxonomic evaluation of both lichen samples, bins on family and genus level representing the different taxonomic groups of the investigated *Hypogymnia* holobionts were selected, in order to provide a sufficient insight on community composition. Table 3 shows the chosen bins for HPH and HTU divided into eukaryotic and bacterial origin. The identified fungi belonging to *Herpotrichiellaceae* and *Pleosporineae* are lichen-associated Ascomycota which possibly belong to the so-called “black fungi” (*Eurotiomycetes* and *Dothideomycetes*) [78].

**Table 2.** Metagenomic PacBio sequencing of *Hypogymnia physodes* (HPH) and *Hypogymnia tubulosa* (HTU).

Analysis Metrics	HPH	HTU
Total Bases (Gb)	678.17	656.10
HiFi Reads	3,584,910	3,461,456
HiFi Yield (Gb)	24.65	25.61
HiFi Read Length (mean, bp)	6877	7398
HiFi Read Quality (median)	Q42	Q43
HiFi Number of Passes (mean)	18	17

**Table 3.** Taxonomic binning of the lichen samples HPH and HTU.

Taxonomic Group	HPH	HTU
Fungi	<i>Parmeliaceae</i>	<i>Parmeliaceae</i>
	<i>Pleosporineae</i>	<i>Pleosporineae</i>
	<i>Herpotrichiellaceae</i>	-
Green Algae	<i>Trebouxia</i>	<i>Trebouxia</i>
Bacteria	<i>Lichenibacterium</i>	<i>Lichenibacterium</i>
	<i>Acetobacteraceae</i>	<i>Acetobacteraceae</i>
	<i>Granulicella</i>	<i>Granulicella</i>
	<i>Beijerinckiaceae</i>	<i>Beijerinckiaceae</i>
	-	<i>Verrucomicrobiota</i>

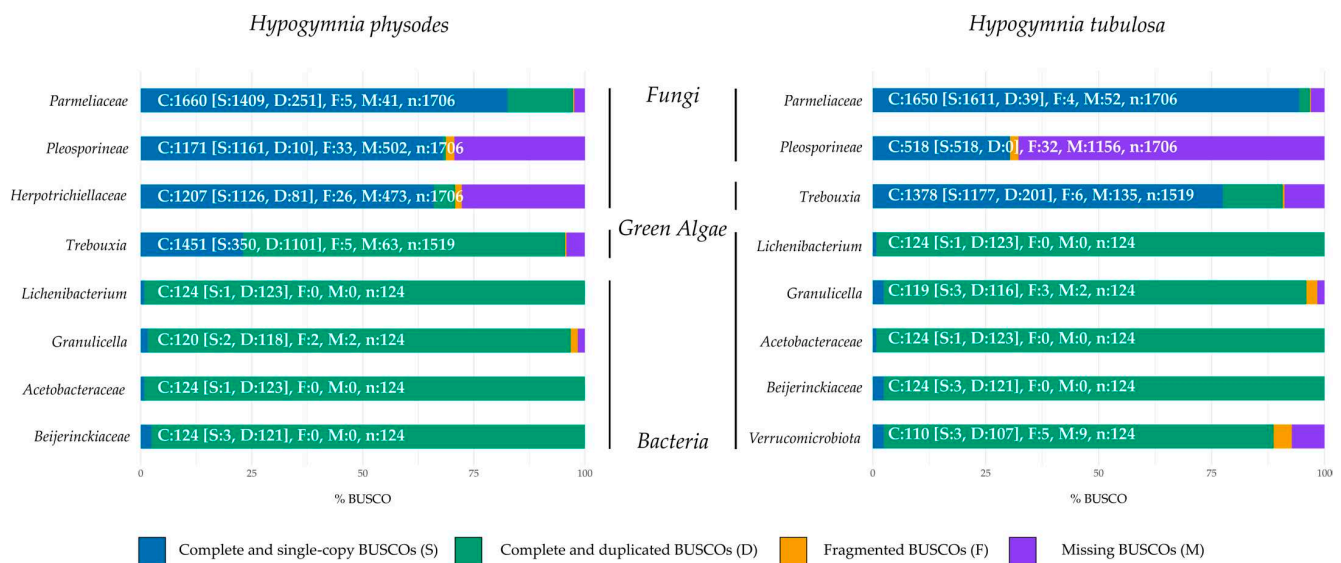
To evaluate genome completeness and reliability for further data processing, a BUSCO [66] assessment was performed on the respective bins of the lichen metagenome. In both samples, the bin “*Parmeliaceae*” included sequences of the mycobiont and the bin “*Trebouxia*” sequences of the photobiont. Unfortunately, the amount of lichen genomes available in databases is too low to provide a selection on lower taxonomic level. Bin compositions mainly deviated in regard to *Herpotrichiellaceae*, which was not prominent in HTU, whereas the bin of *Verrucomicrobiota* was not as prevalent in HPH as the other investigated bins.

Subsequently, the metagenomic bins were compared to the respective BUSCO gene sets of *Ascomycota*, *Chlorophyta* and *Bacteria*. These taxonomic groups were selected to provide a reliable insight in regard to genome representation of the lichen community. Obtained BUSCO results are shown in percentages to allow for an accurate comparison between the investigated bins. Figure 2 depicts the normalized and summarized results for each lichen sample by respective bin. All columns also include the absolute numbers to allow for more precise comparison between different bins.

In order to evaluate genome completeness of the mycobiont and other fungal bins, the orthologous gene sets of the phylum *Ascomycota* odb10 were chosen. In HPH, the fungal bins contained predominantly complete and single BUSCOs (66–82.6%) with the highest values in *Parmeliaceae*. Duplicated genes were observed in all fungal bins ranging from 0.6 to 14.7%, increasing from *Pleosporineae* over *Herpotrichiellaceae* to *Parmeliaceae*. These findings are based on fungal diversity in lichen, as the presence of different fungal genomes results in multiple detections of investigated gene sets by BUSCO [3]. In regard to missing BUSCO gene sets, approximately 29% were absent in *Herpotrichiellaceae* and *Pleosporineae*. As the primary mycobiont makes up most of the lichen biomass, this genome is predominately sequenced compared to other fungi included in the metagenome. Fragmentation on BUSCO gene sets ranged from 0.3 to 1.9%.

The photobiont (*Trebouxia*) was compared to the gene set of *Chlorophyta* odb10, which yielded mainly complete and duplicated BUSCOs (72.5%), with complete and single BUSCOs making up the majority of the remaining gene sets (23%). These findings are in line with the previously elaborated influence of biomass distribution in lichens on sequencing depth. The share of missing BUSCOs was about 4%, whereas 0.3% were fragmented.





**Figure 2.** BUSCO genome completeness assessment of the *Hypogymnia physodes* (HPH) and *Hypogymnia tubulosa* (HTU) holobiont. Lineage datasets utilized to assess completeness: *Ascomycota*, *Chlorophyta*, and *Bacteria*.

BUSCO evaluation of the bacterial fractions of the investigated lichen community against the gene set *Bacteria* yielded high duplication rates in every bin (95–99.2%). As the investigated bins encased more than one bacterial species, duplication rates were inevitable [79]. By investigating the genome completeness of the fungal bins present in HTU, the *Parmeliaceae* bin exhibits mainly complete and single BUSCOs with only slight duplication or fragmentation. In contrast, a high rate of missing BUSCOs was seen in the bin containing *Pleosporineae*. The node of this bin was smaller in comparison to the one in HPH (see Table S2). These findings may also be based on the share of primary mycobiont present in the lichen sample.

Comparable results are observed in the photobiont bin. Here the duplication rate is considerably lower than in the respective bin of HPH. These findings may suggest a uniform composition of algal partners in this lichen sample and a more diverse composition of algal strains in HPH. The bacterial fraction of HTU displayed comparable results to HPH regarding duplication rates, based on the same rationale. The assessment of gene set completeness showed a similar pattern to that of genome completeness across all taxonomic groups belonging to eukaryota (Supplementary Figure S1). For bacterial bins, more complete and single BUSCOs were observed by analysis of gene set completeness. However, the rate of missing and fragmented BUSCOs increased.

A second approach to evaluate the contiguity of the assembled metagenomes involves the utilization of the quality assessment tool for genome assemblies (QUAST). Thus, the different bins were assessed based on their genome size, number of contigs, and N50 values; a summary of the statistics for each bin are provided in Tables 4 and 5. Notably, the N50 values for the primary symbionts, *Parmeliaceae* sp. and *Trebouxia* sp., were around 1 Mb for both lichen samples. These findings, combined with the low L50 value for each, suggest with a high degree of confidence, that the genomes are assembled contiguously. A 1.4-fold difference in total contig length of *Parmeliaceae* is observed between HPH and HTU, which could be due to natural variations in genome size. However, both bins align with the average size of the assembled *Parmelia* spp. genome reported in the literature (45 Mb, NCBI BioSample: SAMN17391792). A 1.6-fold difference in genome size can be seen in the *Trebouxia* bins; here, it is worth mentioning that in HTU, the missing BUSCOs differ by a factor of two. This may cause variations in observed total contig length. Deviation from the 69.35 Mb genome size (NCBI BioSample: SAMD00066476 [80]) reported in the literature

may be explained by the nature of the *Trebouxia* bin, as it possibly encloses more than one algal partner in HPH [81].

**Table 4.** QUASt results of *Hypogymnia physodes* (HPH) from all investigated bins. From left to right, bins were abbreviated as follows: *Parmeliaceae*, *Pleosporineae*, *Herpotrichiellaceae*, *Trebouxia*, *Lichenibacterium*, *Acetobacteraceae*, *Granulicella*, and *Beijerinckiaceae*.

Analysis Metrics	Parm	Pleo	Herpo	Treb	Licheni	Aceto	Granu	Beij
# contigs	697	614	992	497	1215	1807	1551	977
Largest contig	2,288,512	226,469	240,890	2,824,615	1,201,869	5,239,257	4,442,654	1,606,592
Total length	58,447,069	23,701,369	26,591,620	129,019,835	41,345,960	91,111,576	63,381,977	27,344,988
N50	1,017,384	50,386	34,766	1,073,013	80,917	152,096	78,893	51,614
L50	20	156	238	37	75	94	149	115
# N's per 100 kbp	0.51	0	0	0.39	0	0.22	0	0.73
Average coverage	251.57	12.12	5.40	4.88	16.89	16.73	24.19	16.24
GC content	51.57	50.69	53.01	55.03	64.51	66.14	60.91	66.93

**Table 5.** Results from QUASt analysis in regard to selected bins from *Hypogymnia tubulosa* (HTU). From left to right, bins were abbreviated as follows: *Parmeliaceae*, *Pleosporineae*, *Trebouxia*, *Lichenibacterium*, *Acetobacteraceae*, *Granulicella*, *Beijerinckiaceae*, and *Verrucomicrobiota*.

Analysis Metrics	Parm	Pleo	Treb	Licheni	Aceto	Granu	Beij	Verruco
# contigs	102	385	726	1562	2647	602	685	708
Largest contig	3,149,693	100,164	4,890,443	1,108,507	2,374,220	4,366,399	1,094,751	4,601,209
Total length	39,875,705	10,235,779	76,898,651	56,755,182	107,354,954	35,634,928	20,508,045	33,036,627
N50	1,305,677	32,841	2,503,874	71,728	80,216	225,427	55,316	191,798
L50	11	108	12	149	276	19	69	27
# N's per 100 kbp	0	0	0.13	0.35	0.28	0.28	0	0.3
Average coverage	284.19	15.85	3.77	32.01	24.34	75.31	15.71	13.92
GC content	51.75	51.35	55.22	64.64	66.60	59.14	66.40	61.18

The bacterial bins in HPH and HTU exhibit a high number of contigs of up to 2647. However, it should be noted that the reported genome sizes for *Lichenibacterium*, *Acetobacteraceae*, and *Granulicella* are around 6 Mb (NCBI BioSample: SAMN09781801), 3 Mb (NCBI BioSample: SAMN29020633), and 6 Mb (NCBI BioSample: SAMN28407668, [82]), respectively. Usually, bacterial genomes processed by long read sequencers are expected to be one circular contig only. This suggests that multiple organisms belonging to these genera may be present in a single bin. Additionally, the high duplication rates observed in all bacterial bins by the above BUSCO data is consistent with the high number of contigs and contig lengths found. By investigating the average coverage in both lichen samples, a strong bias towards the mycobiont is observed, 251- and 284-fold coverage in HPH and HTU, respectively. These findings are in line with the aforementioned distribution of mycobiont biomass in lichen. The algal and other fungal partners in both samples are comparable in coverage. As the bacterial bins of HPH and HTU differed in size, it was also expected to be observed in regard to coverage.

Comparison of GC content yields similar results throughout all compared taxonomic groups in both lichen samples.

### 3.3. Gene Models and Functional Annotation

Annotation of gene models was conducted with AUGUSTUS, utilizing the presented metagenome assemblies and a long-read IsoSeq database functioning as hints [70–72]. Here transcriptomic data from each lichen sample was deployed as an individual training set to ensure correct gene predictions for all factions of the respective metagenome [83]. For both lichen samples, the predicted genes resembling putative proteins were summed up from all encased bins investigated in this study. This yielded a total gene count of 133,963 for HPH and 96,257 for HTU. In addition, statistics on average genes length, gene density,

and introns per gene were evaluated to further validate obtained gene predictions. Exact numbers for each bin are depicted in Table 6.

**Table 6.** Gene prediction statistics on the taxonomic groups of HPH and HTU. The bins are abbreviated in decreasing order as follows: *Parmeliaceae*, *Pleosporineae*, *Herpotrichiellaceae*, *Trebouxia*, *Lichenibacterium*, *Acetobacteraceae*, *Granulicella*, *Beijerinckiaceae*, and *Verrucomicrobiota*. Average was abbreviated as “Av.”.

Bin	HPH				HTU			
	Gene Count	Av. Gene Length	Gene Density (Genes/Mb)	Introns/Gene	Gene Count	Av. Gene Length	Gene Density (Genes/Mb)	Introns/Gene
Parm	18,804	1455.3	307.48	2.72	13,242	1440.9	329.77	2.06
Pleo	7318	1575.3	385.67	1.75	3159	1870.8	307.74	2.88
Herpo	9247	1589.5	274.67	1.79	-	-	-	-
Treb	19,421	1641.00	148.22	7.33	13,458	1576.5	172.46	7.18
Licheni	13,225	681.4	319.89	-	11,950	649.9	210.57	-
Granu	19,260	931.8	303.87	-	11,423	861.9	221.00	-
Aceto	34,871	783.1	382.73	-	23,724	740.7	320.39	-
Beij	11,817	763.3	431.85	-	8303	709.7	404.87	-
Verruco	-	-	-	-	10,998	782.0	332.90	-

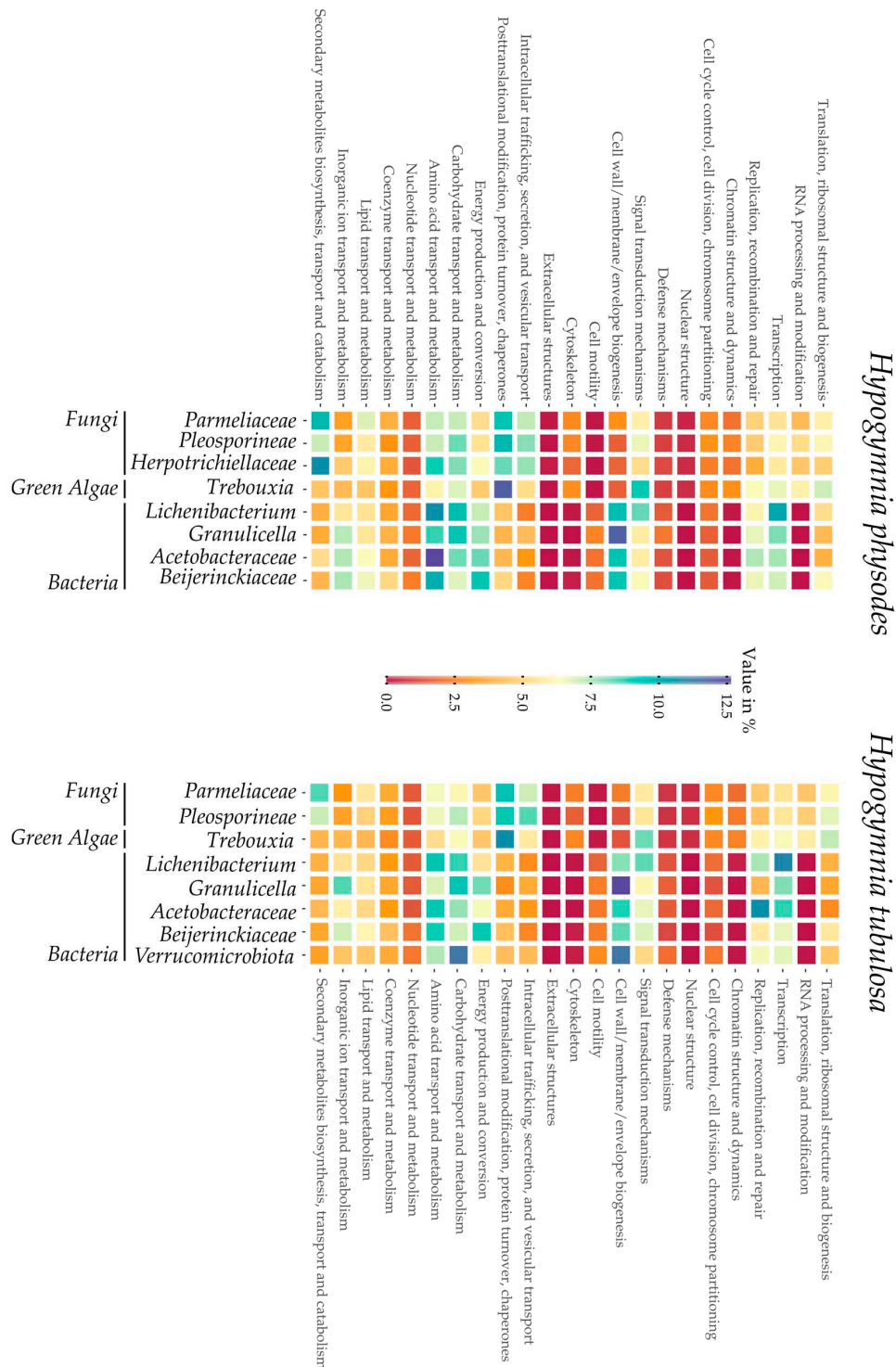
To further elaborate the metagenomes, a functional annotation was conducted based on the databases of Cluster of Orthologous Groups (COG) and Gene Ontology (GO) terms. Results were similar in both lichen samples and their respective bins. Annotation rates differed from ~85% in *Parmeliaceae* to ~94% in *Trebouxia* regarding the eukaryotic fraction, whereas approximately ~93% in the bacterial bins could be annotated. Interestingly, the bins of *Parmeliaceae* and *Trebouxia* harbored ~26% and ~20% of predicted genes with no assigned function, making further research on these organisms highly interesting.

Visualization of the COG distribution in the investigated lichen samples by each bin is depicted in Figure 3. From left to right, the bins are sorted by taxonomic group, beginning with the fungal fraction of the lichen metagenome over the green algae and ending with the bacterial bins. This overview provides insights into the respective functionalities of the predicted genes in each bin. Furthermore, strong differences between eukaryotes and prokaryotes are highlighted by a deep red or blue coloring in e.g., chromatin structure or cell wall biogenesis. On first sight, both heatmaps display a comparable coloring in the respective bins, thus highlighting the high similarity of the samples HPH and HTU. In regard to biosynthetic potential of the *Hypogymnia* holobiont, the row dedicated to “secondary metabolite biosynthesis” is elaborated further. Intriguingly, both *Parmeliaceae* bins possess high amounts of proteins predicted to be involved in secondary metabolite production. Other fungal bins equally displayed increased values compared to the bacterial fraction or the photobiont in the same category. These findings are coherent with the observation of enhanced secondary metabolite production in lichen mycobionts and fungi [17,84–87].

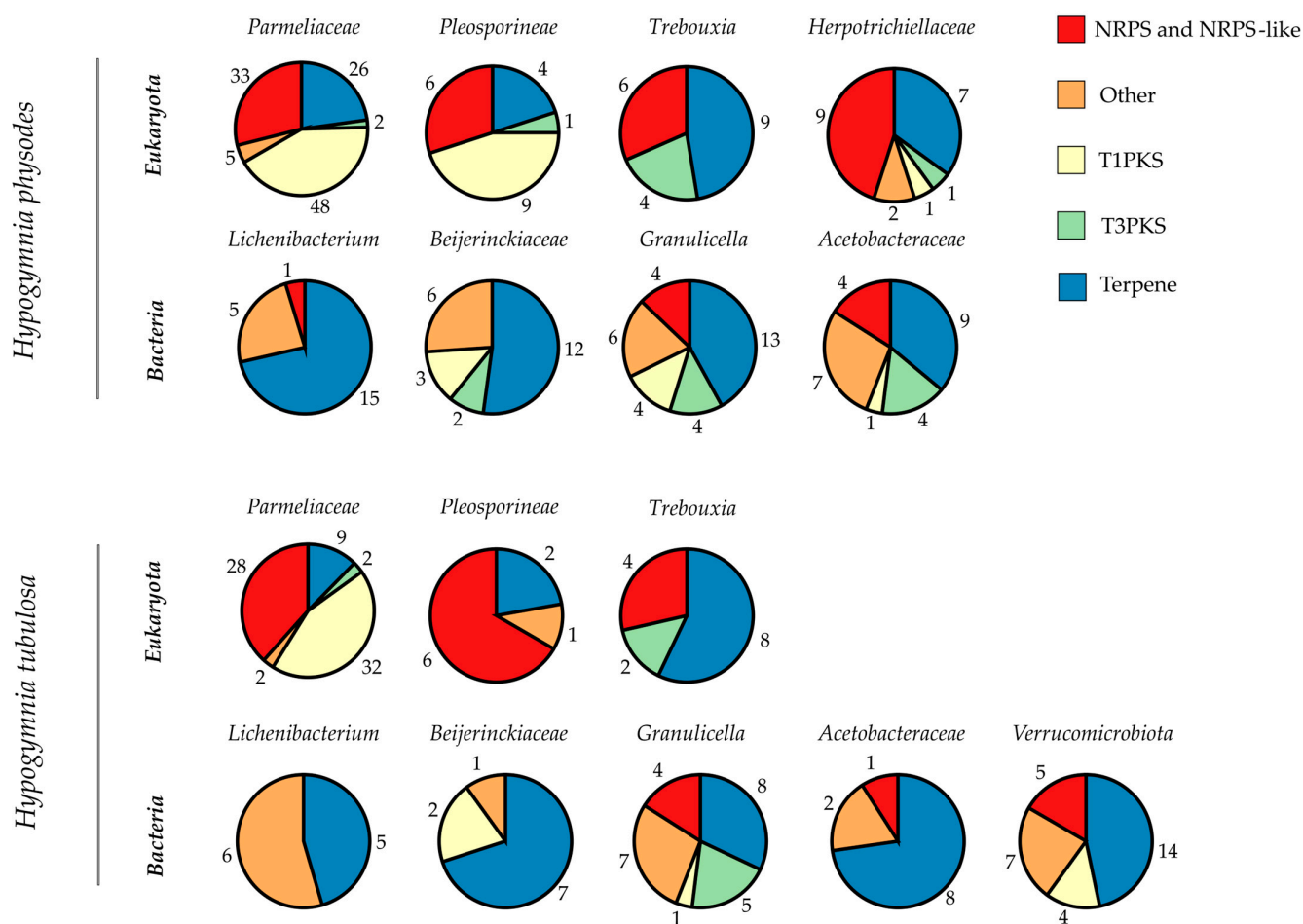
### 3.4. Biosynthetic Gene Clusters

Since the genus *Hypogymnia* is known to produce a manifold secondary metabolite spectrum, a deeper investigation into the biosynthetic pathways may yield intriguing results [8,88,89]. Therefore, the here-presented metagenomes were evaluated for biosynthetic gene clusters (BGCs) by antiSMASH 6.1.1 fungal/bacterial version [73], based on the previously conducted gene annotation. Furthermore, both *Trebouxia* bins were annotated by the fungal version of antiSMASH because plantSMASH yielded insufficient results as previously described [90]. With this in silico approach, the biosynthetic enzymes involved in the formation of unique lichen substances are rendered accessible for wet lab research. To visualize the biosynthetic potential of the genus *Hypogymnia*, each bin’s BGC composition was categorized by putative function (Figure 4). Bins are grouped according to lichen sample and subdivided by superkingdom. The pie charts were computed based on total

numbers of BGCs in the respective bin, including numbers representing the amount of BGCs of the various types. Main BGC categories identified by antiSMASH were nonribosomal peptide synthetases (NRPS and NRPS-like), type I and III polyketide synthases (T1PKS, T3PKS), and terpenes. All remaining types were grouped as “Other”. The complete list of all identified BGCs is given in Tables S3 and S4.



**Figure 3.** Heatmaps summarizing Cluster of Orthologous Groups (COG) from all bins of both lichen samples *Hypogymnia physodes* (HPH) and *Hypogymnia tubulosa* (HTU). Bins are grouped by taxonomic group in the following order: fungi, green algae, and bacteria.



**Figure 4.** Biosynthetic Gene Cluster (BGC) analysis of *Hypogymnia physodes* (HPH) and *Hypogymnia tubulosa* (HTU), results by antiSMASH 6.1.1. Each bin was analyzed separately in order to give an overview of BGCs in different lichen fractions for both samples. Bins are grouped by the corresponding lichen. The fractions of the pie charts are computed based on the total number of BGCs contained in that bin. Numbers next to each fraction reflect total numbers of the particular BGC type present. The different BGC types are depicted by the color code on the top right.

Most BGCs annotated throughout both lichen samples were related to terpene formation, followed by decreasing numbers of NRPS over T1PKS and ending with T3PKS. To allow for a direct comparison, the total numbers of BGCs per bin are listed in Table 7. For the first time, the photobiont and the bacterial fractions were investigated directly from a lichen thallus. This gives information on the BGC composition of the complete holobiont as found in nature.

**Table 7.** Total numbers of BGCs per bin divided by sample.

	Parm	Pleo	Herpo	Treb	Licheni	Beij	Granu	Aceto	Verruco
BGCs HPH	114	20	20	19	21	23	36	25	/-
BGCs HTU	73	8	/-	14	11	10	25	11	30

The prokaryotic fraction in HPH exhibited an average of 26 BGCs over all four bins. For HTU, an average of ~17 BGCs was reported. Some of the bacterial bins exhibit BGC totals in the ranges comparable to *Streptomyces* strains, which are known for their multitude of BGCs (25–70) [91]. However, as the bins contain more than one bacterial genome the

exact numbers of BGCs may differ between individual species. Main BGC class assigned by antiSMASH was related to terpene formation. Throughout all bacterial bins, only 26 BGCs were with assigned function. The influence of lichen-associated bacteria on the symbiosis still remains mostly unclear [92]. Identification of bacterial BGCs may yield insights on further functions in addition to the already known provision of vitamins and cofactors used for degradation of phenolic compounds [93]. As most of the BGCs remain without assigned function, these may be intriguing for further research.

A closer look into the photobiont bins yields a comparable distribution of BGCs in the respective categories, even if the BUSCO and QUASt results differed strongly (see Figure 2 and Tables 4 and 5). For the algal bins, 19 (HPH) and 14 (HTU) BGCs were annotated which exceeds the numbers reported in previous studies [90]. This might be due to the nature of the investigated bins, enclosing more than one species. Therefore, an ITSx analysis was performed to elucidate the bin composition of *Trebouxia* and *Parmeliaceae*. For the bins in HPH only single ITS sequences were found, which is also true for the mycobiont bin in HTU. In regard to the photobiont bin in HTU, no ITS sequence was extractable. However, the absence of ITS sequences does not eliminate the possibility that other parts of the respective genome are present in the respective bins. This also applies to the findings of the BUSCO analysis. Results on ITSx are listed in Table S5. As algae from the genus *Trebouxia* exhibit slow growth rates or are often not yet amenable for cultivation [43], this metagenomic approach renders unidentified BGCs and genes accessible for genome mining. Intriguingly, antiSMASH yielded no assignment to any of the annotated BGCs, making these clusters interesting for further research. As the main class of photobiont BGCs is related to terpene production a formation of relevant natural products may be possible.

In comparison, the *Parmeliaceae* bins harbor the highest number of BGCs (114 in HPH; 73 in HTU) in the respective lichen. The richness in BGCs is in line with findings in previous studies concerning HTU, as the described BGC content in lichen ranges between 27 and 80 [10]. In the case of HPH, the total number of BGCs exceeded these amounts. Compared to other organisms harboring high amounts of BGCs such as *Streptomyces* spp. (23–80) [94,95], *Cyanobacteria* spp. (1–42) [96,97] *Mixobacteria* spp. (30–46) [98,99], and *Nocardia* spp. (~36) [100,101], the investigated *Parmeliaceae* bins showed comparable or higher totals of BGCs. For both mycobiont bins, the majority of BGCs were putative PKS followed by terpene related clusters. In regard to annotation by antiSMASH with the MIBiG database and Pfam, only 27 (HPH) and 16 (HTU) BGCs were with assigned function, highlighting the yet still untapped potential for further investigation.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jof9050546/s1>, Figure S1: BUSCO gene set completeness; Table S1: GC-MS spectra for HPH and HTU; Table S2: Bin comparison of both lichen samples; Table S3: Summary of antiSMASH HPH results on BGCs; Table S4: Summary of antiSMASH HTU results on BGCs; Table S5: ITSx analysis results.

**Author Contributions:** Conceptualization, N.A., M.R., I.S., T.B. and N.M.; methodology, N.A., M.R. and A.C.; software, N.A., M.R. and N.M.; validation, N.A., M.R. and A.C.; formal analysis, N.A. and M.R.; investigation, N.A., M.R., A.C. and J.O.; resources, T.B.; data curation, N.A. and M.R.; writing—original draft preparation, N.A.; writing—review and editing, N.A., M.R., A.C., I.S., N.M. and T.B.; visualization, M.R.; supervision, N.M. and T.B.; project administration, N.M. and T.B.; funding acquisition, T.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the German Federal Ministry of Education and Research, grant number 031B0824A.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data available in a publicly accessible repository. The data presented in this study are openly available in the National Center for Biotechnology Information (NCBI). BioSample accession number: *Hypogymnia physodes* SAMN34074577, *Hypogymnia tubulosa* SAMN34074619.

**Acknowledgments:** On the occasion of his retirement from the University of Prince Edward Island, Canada, we dedicate this manuscript to Russell G. Kerr, an exceptional teacher, influencer, entrepreneur, and visionary in the field of sustainable natural products research. Moreover, we would like to thank Martina Haack for enabling the GC measurements.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Plitt, C.C. A Short History of Lichenology. *Bryologist* **1919**, *22*, 77. [[CrossRef](#)]
2. Spribille, T.; Tuovinen, V.; Resl, P.; Vanderpool, D.; Wolinski, H.; Aime, M.C.; Schneider, K.; Stabenheiner, E.; Toome-Heller, M.; Thor, G.; et al. Basidiomycete yeasts in the cortex of ascomycete macrolichens. *Science* **2016**, *353*, 488. [[CrossRef](#)] [[PubMed](#)]
3. Muggia, L.; Grube, M. Fungal Diversity in Lichens: From Extremotolerance to Interactions with Algae. *Life* **2018**, *8*, 15. [[CrossRef](#)] [[PubMed](#)]
4. Aschenbrenner, I.A.; Cernava, T.; Berg, G.; Grube, M. Understanding Microbial Multi-Species Symbioses. *Front. Microbiol.* **2016**, *7*, 180. [[CrossRef](#)]
5. Smith, H.B.; Grande, F.D.; Muggia, L.; Keuler, R.; Divakar, P.K.; Grewe, F.; Schmitt, I.; Lumbsch, H.T.; Leavitt, S.D. Metagenomic data reveal diverse fungal and algal communities associated with the lichen symbiosis. *Symbiosis* **2020**, *82*, 133–147. [[CrossRef](#)]
6. Hawksworth, D.L.; Grube, M. Lichens redefined as complex ecosystems. *New Phytol.* **2020**, *227*, 1281. [[CrossRef](#)]
7. Rolshausen, G.; Grande, F.D.; Otte, J.; Schmitt, I. Lichen holobionts show compositional structure along elevation. *Mol. Ecol.* **2022**, *00*, 1–12. [[CrossRef](#)]
8. Ranković, B.; Kosanić, M. Lichens as a Potential Source of Bioactive Secondary Metabolites. In *Lichen Secondary Metabolites*; Springer: Cham, Switzerland, 2019; pp. 1–29. [[CrossRef](#)]
9. Huneck, S.; Yoshimura, I. Identification of Lichen Substances. In *Identification of Lichen Substances*; Springer: Berlin/Heidelberg, Germany, 1996; pp. 11–123. [[CrossRef](#)]
10. Calchera, A.; Grande, F.D.; Bode, H.B.; Schmitt, I. Biosynthetic Gene Content of the ‘Perfume Lichens’ *Evernia prunastri* and *Pseudevernia furfuracea*. *Molecules* **2019**, *24*, 203. [[CrossRef](#)]
11. Elix, J.A.; Stocker-Wörgötter, E. Biochemistry and secondary metabolites. In *Lichen Biology*; Cambridge University Press: Cambridge, UK, 2008; pp. 104–133.
12. Ristić, S.; Ranković, B.; Kosanić, M.; Stanojković, T.; Stamenković, S.; Vasiljević, P.; Manojlović, I.; Manojlović, N. Phytochemical study and antioxidant, antimicrobial and anticancer activities of *Melanelia subaurifera* and *Melanelia fuliginosa* lichens. *J. Food Sci. Technol.* **2016**, *53*, 2804–2816. [[CrossRef](#)]
13. Sisodia, R.; Geol, M.; Verma, S.; Rani, A.; Dureja, P. Antibacterial and antioxidant activity of lichen species *Ramalina roesleri*. *Nat. Prod. Res.* **2013**, *27*, 2235–2239. [[CrossRef](#)]
14. Kosanić, M.; Manojlović, N.; Janković, S.; Stanojković, T.; Ranković, B. *Evernia prunastri* and *Pseudoevernia furfuraceae* lichens and their major metabolites as antioxidant, antimicrobial and anticancer agents. *Food Chem. Toxicol.* **2013**, *53*, 112–118. [[CrossRef](#)]
15. Karabulut, G.; Ozturk, S. Antifungal activity of *Evernia prunastri*, *Parmelia sulcata*, *Pseudevernia furfuracea* var. *Furfuracea*. *Pak. J. Bot.* **2015**, *47*, 1575–1579.
16. Joshi, T.; Sharma, P.; Joshi, T.; Chandra, S. In silico screening of anti-inflammatory compounds from Lichen by targeting cyclooxygenase-2. *J. Biomol. Struct. Dyn.* **2019**, *38*, 3544–3562. [[CrossRef](#)]
17. Goga, M.; Elečko, J.; Marcinčinová, M.; Ručová, D.; Bačkorová, M.; Bačkor, M. Lichen Metabolites: An Overview of Some Secondary Metabolites and Their Biological Potential. In *Co-Evolution of Secondary Metabolites*; Reference Series in Phytochemistry; Springer: Cham, Switzerland, 2020; pp. 175–209. [[CrossRef](#)]
18. Kosanić, M.; Ranković, B. Studies on Antioxidant Properties of Lichen Secondary Metabolites. In *Lichen Secondary Metabolites*; Springer: Cham, Switzerland, 2019; pp. 129–153. [[CrossRef](#)]
19. Solárová, Z.; Liskova, A.; Samec, M.; Kubatka, P.; Büsselberg, D.; Solár, P. Anticancer Potential of Lichens’ Secondary Metabolites. *Biomolecules* **2020**, *10*, 87. [[CrossRef](#)]
20. Cardile, V.; Graziano, A.C.E.; Avola, R.; Piovano, M.; Russo, A. Potential anticancer activity of lichen secondary metabolite physodic acid. *Chem. Biol. Interact.* **2017**, *263*, 36–45. [[CrossRef](#)]
21. Boustie, J.; Grube, M. Lichens—A promising source of bioactive secondary metabolites. *Plant Genet. Resour.* **2005**, *3*, 273–287. [[CrossRef](#)]
22. Parrot, D.; Legrave, N.; Delmail, D.; Grube, M.; Suzuki, M.; Tomasi, S. Review—Lichen-Associated Bacteria as a Hot Spot of Chemodiversity: Focus on Uncialamycin, a Promising Compound for Future Medicinal Applications. *Plant. Med.* **2016**, *82*, 1143–1152. [[CrossRef](#)]
23. Suzuki, M.T.; Parrot, D.; Berg, G.; Grube, M.; Tomasi, S. Lichens as natural sources of biotechnologically relevant bacteria. *Appl. Microbiol. Biotechnol.* **2016**, *100*, 583–595. [[CrossRef](#)]

24. Hei, Y.; Zhang, H.; Tan, N.; Zhou, Y.; Wei, X.; Hu, C.; Liu, Y.; Wang, L.; Qi, J.; Gao, J.M. Antimicrobial activity and biosynthetic potential of cultivable actinomycetes associated with Lichen symbiosis from Qinghai-Tibet Plateau. *Microbiol. Res.* **2021**, *244*, 126652. [[CrossRef](#)]
25. Gokulan, K.; Khare, S.; Cerniglia, C. METABOLIC PATHWAYS| Production of Secondary Metabolites of Bacteria. In *Encyclopedia of Food Microbiology*, 2nd ed.; Academic Press: Cambridge, MA, USA, 2014; pp. 561–569. [[CrossRef](#)]
26. Crawford, M.J.; Townsend, C.A. New insights into the formation of fungal aromatic polyketides. *Nat. Rev. Microbiol.* **2010**, *8*, 879–889. [[CrossRef](#)]
27. Cox, R.J. Curiouser and curiouser: Progress in understanding the programming of iterative highly-reducing polyketide synthases. *Nat. Prod. Rep.* **2023**, *40*, 9–27. [[CrossRef](#)] [[PubMed](#)]
28. Dutta, S.; Whicher, J.R.; Hansen, D.A.; Hale, W.A.; Chemler, J.A.; Congdon, G.R.; Narayan, A.R.H.; Håkansson, K.; Sherman, D.H.; Smith, J.L.; et al. Structure of a modular polyketide synthase. *Nature* **2014**, *510*, 512–517. [[CrossRef](#)] [[PubMed](#)]
29. Huitt-Roehl, C.R.; Hill, E.A.; Adams, M.M.; Vagstad, A.L.; Li, J.W.; Townsend, C.A. Starter Unit Flexibility for Engineered Product Synthesis by the Nonreducing Polyketide Synthase PksA. *ACS Chem. Biol.* **2015**, *10*, 1443–1449. [[CrossRef](#)] [[PubMed](#)]
30. Crawford, J.M.; Korman, T.P.; Labonte, J.W.; Vagstad, A.L.; Hill, E.A.; Kamari-Bidkorpeh, O.; Tsai, S.-C.; Townsend, C.A. Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization. *Nature* **2009**, *461*, 1139–1143. [[CrossRef](#)]
31. Li, Y.; Xu, W.; Tang, Y. Classification, prediction, and verification of the regioselectivity of fungal polyketide synthase product template domains. *J. Biol. Chem.* **2010**, *285*, 22764–22773. [[CrossRef](#)]
32. Wang, Y.; Kim, J.A.; Cheong, Y.H.; Koh, Y.J.; Hur, J.S. Isolation and characterization of a non-reducing polyketide synthase gene from the lichen-forming fungus *Usnea longissimi*. *Mycol. Prog.* **2012**, *11*, 75–83. [[CrossRef](#)]
33. Medema, M.H.; Kottmann, R.; Yilmaz, P.; Cummings, M.; Biggins, J.B.; Blin, K.; de Bruijn, I.; Chooi, Y.-H.; Claesen, J.; Coates, R.C.; et al. Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol.* **2015**, *11*, 625–631. [[CrossRef](#)]
34. Pizarro, D.; Divakar, P.K.; Grewe, F.; Crespo, A.; Grande, F.D.; Lumbsch, H.T. Genome-Wide Analysis of Biosynthetic Gene Cluster Reveals Correlated Gene Loss with Absence of Usnic Acid in Lichen-Forming Fungi. *Genome Biol. Evol.* **2020**, *12*, 1858–1868. [[CrossRef](#)]
35. Keller, N.P.; Turner, G.; Bennett, J.W. Fungal secondary metabolism—From biochemistry to genomics. *Nat. Rev. Microbiol.* **2005**, *3*, 937–947. [[CrossRef](#)]
36. Keller, N.P. Fungal secondary metabolism: Regulation, function and drug discovery. *Nat. Rev. Microbiol.* **2018**, *17*, 167–180. [[CrossRef](#)]
37. Rokas, A.; Wisecaver, J.H.; Lind, A.L. The birth, evolution and death of metabolic gene clusters in fungi. *Nat. Rev. Microbiol.* **2018**, *16*, 731–744. [[CrossRef](#)]
38. Singh, G.; Calchera, A.; Schulz, M.; Drechsler, M.; Bode, H.B.; Schmitt, I.; Grande, F.D. Climate-specific biosynthetic gene clusters in populations of a lichen-forming fungus. *Environ. Microbiol.* **2021**, *23*, 4260–4275. [[CrossRef](#)]
39. Singh, G. Linking Lichen Metabolites to Genes: Emerging Concepts and Lessons from Molecular Biology and Metagenomics. *J. Fungi* **2023**, *9*, 160. [[CrossRef](#)]
40. Llewellyn, T.; Nowell, R.W.; Aptroot, A.; Temina, M.; Prescott, T.A.K.; Barraclough, T.G.; Gaya, E. Metagenomics Shines Light on the Evolution of ‘Sunscreen’ Pigment Metabolism in the Teloschistales (Lichen-Forming Ascomycota). *Genome Biol. Evol.* **2023**, *15*, evad002. [[CrossRef](#)]
41. Kim, W.; Liu, R.; Woo, S.; Bin Kang, K.; Park, H.; Yu, Y.H.; Ha, H.-H.; Oh, S.-Y.; Yang, J.H.; Kim, H.; et al. Linking a gene cluster to atranorin, a major cortical substance of lichens, through genetic dereplication and heterologous expression. *MBio* **2021**, *12*, e0111121. [[CrossRef](#)]
42. Kealey, J.T.; Craig, J.P.; Barr, P.J. Identification of a lichen depside polyketide synthase gene by heterologous expression in *Saccharomyces cerevisiae*. *Metab. Eng. Commun.* **2021**, *13*, e001172. [[CrossRef](#)]
43. Muggia, L.; Zellnig, G.; Rabensteiner, J.; Grube, M. Morphological and phylogenetic study of algal partners associated with the lichen-forming fungus *Tephromela atra* from the Mediterranean region. *Symbiosis* **2010**, *51*, 149–160. [[CrossRef](#)]
44. Chiang, Y.M.; Chang, S.L.; Oakley, B.R.; Wang, C.C.C. Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms. *Curr. Opin. Chem. Biol.* **2011**, *15*, 137–143. [[CrossRef](#)]
45. Zheng, W.; Wang, X.; Zhou, H.; Zhang, Y.; Li, A.; Bian, X. Establishment of recombineering genome editing system in *Paraburkholderia megapolitana* empowers activation of silent biosynthetic gene clusters. *Microb. Biotechnol.* **2020**, *13*, 397–405. [[CrossRef](#)]
46. Singh, G.; Calchera, A.; Merges, D.; Valim, H.; Otte, J.; Schmitt, I.; Grande, F.D. A Candidate Gene Cluster for the Bioactive Natural Product Gyrophoric Acid in Lichen-Forming Fungi. *Microbiol. Spectr.* **2022**, *10*, e0010922. [[CrossRef](#)]
47. Wasil, Z.; Pahirulzaman, K.A.K.; Butts, C.; Simpson, T.J.; Lazarus, C.M.; Cox, R.J. One pathway, many compounds: Heterologous expression of a fungal biosynthetic pathway reveals its intrinsic potential for diversity. *Chem. Sci.* **2013**, *4*, 3845–3856. [[CrossRef](#)]
48. Martinet, L.; Naômé, A.; Deflandre, B.; Maciejewska, M.; Tellatin, D.; Tenconi, E.; Smargiasso, N.; de Pauw, E.; van Wezel, G.P.; Rigali, S. A single biosynthetic gene cluster is responsible for the production of bagremycin antibiotics and feroverdin iron chelators. *MBio* **2019**, *10*, e01230-19. [[CrossRef](#)] [[PubMed](#)]
49. Tsai, Y.C.; Conlan, S.; Deming, C.; Segre, J.A.; Kong, H.H.; Korfach, J.; Oh, J.; Program, N.C.S. Resolving the complexity of human skin metagenomes using single-molecule sequencing. *MBio* **2016**, *7*, e01948-15. [[CrossRef](#)] [[PubMed](#)]
50. Cuscó, A.; Pérez, D.; Viñes, J.; Fàbregas, N.; Francino, O. Long-read metagenomics retrieves complete single-contig bacterial genomes from canine feces. *BMC Genom.* **2021**, *22*, 330. [[CrossRef](#)]



51. Bickhart, D.M.; Kolmogorov, M.; Tseng, E.; Portik, D.M.; Korobeynikov, A.; Tolstoganov, I.; Uritskiy, G.; Liachko, I.; Sullivan, S.T.; Shin, S.B.; et al. Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities. *Nat. Biotechnol.* **2022**, *40*, 711–719. [[CrossRef](#)]
52. Xie, H.; Yang, C.; Sun, Y.; Igarashi, Y.; Jin, T.; Luo, F. PacBio Long Reads Improve Metagenomic Assemblies, Gene Catalogs, and Genome Binning. *Front. Genet.* **2020**, *11*, 1077. [[CrossRef](#)]
53. Chen, L.; Na Zhao, N.; Cao, J.; Liu, X.; Xu, J.; Ma, Y.; Yu, Y.; Zhang, X.; Zhang, W.; Guan, X.; et al. Short- and long-read metagenomics expand individualized structural variations in gut microbiomes. *Nat. Commun.* **2022**, *13*, 3175. [[CrossRef](#)]
54. Wirth, V.; Hauck, M.; Schultz, M. *Die Flechten Deutschlands*; Ulmer: Stuttgart, Germany, 2013; pp. 1–672.
55. Purvis, O.W.; Coppins, B.J.; Hawksworth, D.L.; James, P.W.; Moore, D.M. *The Lichen Flora of Great Britain and Ireland*; Cambridge University Press: Cambridge, UK, 1992; p. 710. Available online: <https://research-scotland.ac.uk/handle/20.500.12594/6198> (accessed on 29 March 2023).
56. Ranković, B.; Kosanić, M.; Manojlović, N.; Rančić, A.; Stanojković, T. Chemical composition of *Hypogymnia physodes* lichen and biological activities of some its major metabolites. *Med. Chem. Res.* **2014**, *23*, 408–416. [[CrossRef](#)]
57. Stojanović, G.; Zlatanović, I.; Zrnzević, I.; Stanković, M.; Jovanović, V.S.; Zlatković, B. *Hypogymnia tubulosa* extracts: Chemical profile and biological activities. *Nat. Prod. Res.* **2017**, *32*, 2735–2739. [[CrossRef](#)]
58. Yilmaz, M.; Tay, T.; Kivanç, M.; Türk, H.; Türk, A.Ö. The antimicrobial activity of extracts of the lichen *Hypogymnia tubulosa* and its 3-hydroxyphysodic acid constituent. *Z. Naturforsch. Sect. C J. Biosci.* **2005**, *60*, 35–38. [[CrossRef](#)]
59. Ritz, M.; Ahmad, N.; Brueck, T.; Mehlmer, N. Comparative Genome-Wide Analysis of Two *Caryopteris x Clandonensis* Cultivars: Insights on the Biosynthesis of Volatile Terpenoids. *Plants* **2023**, *12*, 632. [[CrossRef](#)]
60. Tzovaras, B.G.; Segers, F.H.I.D.; Bicker, A.; Dal Grande, F.; Otte, J.; Anvar, S.Y.; Hankeln, T.; Schmitt, I.; Ebersberger, I. What Is in *Umbilicaria pustulata*? A Metagenomic Approach to Reconstruct the Holo-Genome of a Lichen. *Genome Biol. Evol.* **2020**, *12*, 309–324. [[CrossRef](#)]
61. Kolmogorov, M.; Bickhart, D.M.; Behsaz, B.; Gurevich, A.; Rayko, M.; Shin, S.B.; Kuhn, K.; Yuan, J.; Polevikov, E.; Smith, T.P.L.; et al. metaFlye: Scalable long-read metagenome assembly using repeat graphs. *Nat. Methods* **2020**, *17*, 1103–1110. [[CrossRef](#)]
62. Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using diamond. *Nat. Methods* **2015**, *12*, 59–60. [[CrossRef](#)]
63. Huson, D.H.; Albrecht, B.; Bağcı, C.; Bessarab, I.; Górska, A.; Jolic, D.; Williams, R.B.H. MEGAN-LR: New algorithms allow accurate binning and easy interactive exploration of metagenomic long reads and contigs. *Biol. Direct* **2018**, *13*, 6. [[CrossRef](#)]
64. Bağcı, C.; Patz, S.; Huson, D.H. DIAMOND+MEGAN: Fast and Easy Taxonomic and Functional Analysis of Short and Long Microbiome Sequences. *Curr. Protoc.* **2021**, *1*, e59. [[CrossRef](#)]
65. Bağcı, C.; Beier, S.; Górska, A.; Huson, D.H. Introduction to the analysis of environmental sequences: Metagenomics with MEGAN. *Methods Mol. Biol.* **2019**, *1910*, 591–604. [[CrossRef](#)]
66. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)]
67. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [[CrossRef](#)]
68. Shen, W.; Le, S.; Li, Y.; Hu, F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS ONE* **2016**, *11*, e0163962. [[CrossRef](#)]
69. Bengtsson-Palme, J.; Ryberg, M.; Hartmann, M.; Branco, S.; Wang, Z.; Godhe, A.; De Wit, P.; Sánchez-García, M.; Ebersberger, I.; De Sousa, F.; et al. Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods Ecol. Evol.* **2013**, *4*, 914–919. [[CrossRef](#)]
70. Hoff, K.J.; Lomsadze, A.; Borodovsky, M.; Stanke, M. Whole-genome annotation with BRAKER. *Methods Mol. Biol.* **2019**, *1962*, 65–95. [[CrossRef](#)] [[PubMed](#)]
71. Hoff, K.J.; Lange, S.; Lomsadze, A.; Borodovsky, M.; Stanke, M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **2016**, *32*, 767–769. [[CrossRef](#)] [[PubMed](#)]
72. Brůna, T.; Hoff, K.J.; Lomsadze, A.; Stanke, M.; Borodovsky, M. BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinf.* **2021**, *3*, lqaa108. [[CrossRef](#)]
73. Blin, K.; Shaw, S.; Kloosterman, A.M.; Charlop-Powers, Z.; van Wezel, G.P.; Medema, M.H.; Weber, T. antiSMASH 6.0: Improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **2021**, *49*, W29–W35. [[CrossRef](#)]
74. Huerta-Cepas, J.; Forslund, K.; Coelho, L.P.; Szklarczyk, D.; Jensen, L.J.; von Mering, C.; Bork, P. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **2017**, *34*, 2115–2122. [[CrossRef](#)]
75. Galaxy | Europe. Available online: <https://usegalaxy.eu/> (accessed on 17 March 2023).
76. Zakeri, B.; Wright, G.D. Chemical biology of tetracycline antibiotics. *Biochem. Cell Biol.* **2008**, *86*, 124–136. [[CrossRef](#)]
77. Staunton, J.; Weissman, K.J. Polyketide biosynthesis: A millennium review. *Nat. Prod. Rep.* **2001**, *18*, 380–416. [[CrossRef](#)]
78. Muggia, L.; Fleischhacker, A.; Kopun, T.; Grube, M. Extremotolerant fungi from alpine rock lichens and their phylogenetic relationships. *Fungal Divers.* **2016**, *76*, 119–142. [[CrossRef](#)]
79. Manni, M.; Berkeley, M.R.; Seppey, M.; Zdobnov, E.M. BUSCO: Assessing Genomic Data Quality and Beyond. *Curr. Protoc.* **2021**, *1*, e323. [[CrossRef](#)]

80. Kono, M.; Tanabe, H.; Ohmura, Y.; Satta, Y.; Terai, Y. Physical contact and carbon transfer between a lichen-forming *Trebouxia* alga and a novel Alphaproteobacterium. *Microbiol.* **2017**, *163*, 678–691. [[CrossRef](#)]
81. Blaha, J.; Baloch, E.; Grube, M. High photobiont diversity associated with the euryoecious lichen-forming ascomycete *Lecanora rupicola* (Lecanoraceae, Ascomycota). *Biol. J. Linn. Soc.* **2006**, *88*, 283–293. [[CrossRef](#)]
82. Wang, L.; Liu, W.; Liang, J.; Zhao, L.; Li, Q.; Zhou, C.; Cen, H.; Weng, Q.; Zhang, G. Mining of novel secondary metabolite biosynthetic gene clusters from acid mine drainage. *Sci. Data* **2022**, *9*, 760. [[CrossRef](#)]
83. Ismail, W.M.; Ye, Y.; Tang, H. Gene finding in metatranscriptomic sequences. *BMC Bioinf.* **2014**, *15*, S8. [[CrossRef](#)]
84. Ola, A.R.B.; Thomy, D.; Lai, D.; Brötz-Oesterhelt, H.; Proksch, P. Inducing secondary metabolite production by the endophytic fungus *Fusarium tricinctum* through coculture with *Bacillus subtilis*. *J. Nat. Prod.* **2013**, *76*, 2094–2099. [[CrossRef](#)]
85. Shwab, E.K.; Keller, N.P. Regulation of secondary metabolite production in filamentous ascomycetes. *Mycol. Res.* **2008**, *112*, 225–230. [[CrossRef](#)]
86. Devi, R.; Kaur, T.; Guleria, G.; Rana, K.L.; Kour, D.; Yadav, N.; Yadav, A.N.; Saxena, A.K. Fungal secondary metabolites and their biotechnological applications for human health. In *New and Future Developments in Microbial Biotechnology and Bioengineering*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 147–161. [[CrossRef](#)]
87. Bills, G.F.; Gloer, J.B. Biologically Active Secondary Metabolites from the Fungi. In *The Fungal Kingdom*; Wiley Online Library: Hoboken, NJ, USA, 2017; pp. 1087–1119. [[CrossRef](#)]
88. Studzińska-Sroka, E.; Zarabska-Bożjewicz, D. *Hypogymnia physodes*—A lichen with interesting medicinal potential and ecological properties. *J. Herb. Med.* **2019**, *17–18*, 100287. [[CrossRef](#)]
89. Elečko, J.; Vilková, M.; Frenák, R.; Routray, D.; Ručová, D.; Bačkor, M.; Goga, M. A Comparative Study of Isolated Secondary Metabolites from Lichens and Their Antioxidative Properties. *Plants* **2022**, *11*, 1077. [[CrossRef](#)]
90. O'Neill, E. Mining Natural Product Biosynthesis in Eukaryotic Algae. *Mar. Drugs* **2020**, *18*, 90. [[CrossRef](#)]
91. Belknap, K.; Park, C.J.; Barth, B.M.; Andam, C.P. Genome mining of biosynthetic and chemotherapeutic gene clusters in *Streptomyces* bacteria. *Sci. Rep.* **2020**, *10*, 2003. [[CrossRef](#)]
92. Bates, S.T.; Cropsey, G.W.G.; Caporaso, J.G.; Knight, R.; Fierer, N. Bacterial communities associated with the lichen symbiosis. *Appl. Environ. Microbiol.* **2011**, *77*, 1309–1314. [[CrossRef](#)] [[PubMed](#)]
93. Cernava, T.; Erlacher, A.; Aschenbrenner, I.A.; Krug, L.; Lassek, C.; Riedel, K.; Grube, M.; Berg, G. Deciphering functional diversification within the lichen microbiota by meta-omics. *Microbiome* **2017**, *5*, 82. [[CrossRef](#)] [[PubMed](#)]
94. Liu, Z.; Zhao, Y.; Huang, C.; Luo, Y. Recent Advances in Silent Gene Cluster Activation in *Streptomyces*. *Front. Bioeng. Biotechnol.* **2021**, *9*, 88. [[CrossRef](#)] [[PubMed](#)]
95. Caicedo-Montoya, C.; Manzo-Ruiz, M.; Rios-Esteva, R. Pan-Genome of the Genus *Streptomyces* and Prioritization of Biosynthetic Gene Clusters with Potential to Produce Antibiotic Compounds. *Front. Microbiol.* **2021**, *12*, 2786. [[CrossRef](#)]
96. Popin, R.V.; Alvarenga, D.O.; Castelo-Branco, R.; Fewer, D.P.; Sivonen, K. Mining of Cyanobacterial Genomes Indicates Natural Product Biosynthetic Gene Clusters Located in Conjugative Plasmids. *Front. Microbiol.* **2021**, *12*, 3353. [[CrossRef](#)]
97. Dittmann, E.; Gugger, M.; Sivonen, K.; Fewer, D.P. Natural Product Biosynthetic Diversity and Comparative Genomics of the Cyanobacteria. *Trends Microbiol.* **2015**, *23*, 642–652. [[CrossRef](#)]
98. Gregory, K.; Salvador, L.A.; Akbar, S.; Adaikpoh, B.I.; Stevens, D.C. Survey of Biosynthetic Gene Clusters from Sequenced Myxobacteria Reveals Unexplored Biosynthetic Potential. *Microorganisms* **2019**, *7*, 181. [[CrossRef](#)]
99. Moghaddam, J.A.; Crüsemann, M.; Alanjary, M.; Harms, H.; Dávila-Céspedes, A.; Blom, J.; Poehlein, A.; Ziemert, N.; König, G.M.; Schäberle, T.F. Analysis of the Genome and Metabolome of Marine Myxobacteria Reveals High Potential for Biosynthesis of Novel Specialized Metabolites. *Sci. Rep.* **2018**, *8*, 16600. [[CrossRef](#)]
100. Männle, D.; McKinnie, S.M.K.; Mantri, S.S.; Steinke, K.; Lu, Z.; Moore, B.S.; Ziemert, N.; Kaysser, L. Comparative Genomics and Metabolomics in the Genus *Nocardia*. *mSystems* **2020**, *5*, e00092-20. [[CrossRef](#)]
101. Doroghazi, J.R.; Metcalf, W.W. Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. *BMC Genom.* **2013**, *14*, 611. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Biosynthetic Gene Cluster Synteny -  
Orthologous Polyketide Synthases in  
*Hypogymnia physodes*, *Hypogymnia  
tubulosa* and *Parmelia sulcata*

# Biosynthetic gene cluster synteny: Orthologous polyketide synthases in *Hypogymnia physodes*, *Hypogymnia tubulosa*, and *Parmelia sulcata*

Nadim Ahmad<sup>1</sup> | Manfred Ritz<sup>1</sup> | Anjuli Calchera<sup>2</sup> | Jürgen Otte<sup>2</sup> |  
Imke Schmitt<sup>2,3</sup> | Thomas Brueck<sup>1</sup>  | Norbert Mehlmer<sup>1</sup> 

<sup>1</sup>Department of Chemistry, Werner Siemens Chair of Synthetic Biotechnology, TUM School of Natural Sciences, Technical University of Munich (TUM), Garching, Germany

<sup>2</sup>Senckenberg Biodiversity and Climate Research Centre (SBIK-F), Frankfurt am Main, Germany

<sup>3</sup>Institute of Ecology, Evolution and Diversity, Goethe University Frankfurt, Frankfurt am Main, Germany

## Correspondence

Thomas Brueck and Norbert Mehlmer, Department of Chemistry, Werner Siemens Chair of Synthetic Biotechnology, TUM School of Natural Sciences, Technical University of Munich (TUM), 85748 Garching, Germany. Email: [brueck@tum.de](mailto:brueck@tum.de) and [norbert.mehlmer@tum.de](mailto:norbert.mehlmer@tum.de)

## Funding information

Federal Ministry of Education and Research, Grant/Award Number: 031B0824A

## Abstract

Lichens are symbiotic associations consisting of a photobiont (algae or cyanobacteria) and a mycobiont (fungus), which together generate a variety of unique secondary metabolites. To access this biosynthetic potential for biotechnological applications, deeper insights into the biosynthetic pathways and corresponding gene clusters are necessary. Here, we provide a comparative view of the biosynthetic gene clusters of three lichen mycobionts derived from *Hypogymnia physodes*, *Hypogymnia tubulosa*, and *Parmelia sulcata*. In addition, we present a high-quality PacBio metagenome of *Parmelia sulcata*, from which we extracted the mycobiont bin containing 214 biosynthetic gene clusters. Most biosynthetic gene clusters in these genomes were associated with T1PKSs, followed by NRPSs and terpenes. This study focused on biosynthetic gene clusters related to polyketide synthesis. Based on ketosynthase homology, we identified nine highly syntenic clusters present in all three species. Among the four clusters belonging to nonreducing PKSs, two are putatively linked to lichen substances derived from orsellinic acid (orcinol depsides and depsidones, e.g., lecanoric acid, physodic acid, lobaric acid), one to compounds derived from methylated forms of orsellinic acid (beta orcinol depsides, e.g., atranorin), and one to melanins. Five clusters with orthologs in all three species are linked to reducing PKSs. Our study contributes to sorting and dereplicating the vast PKS diversity found in lichenized fungi. High-quality sequences of biosynthetic gene clusters of these three common species provide a foundation for further exploration into biotechnological applications and the molecular evolution of lichen substances.

## KEYWORDS

biosynthetic gene cluster, long-read sequencing, *Parmelia sulcata*, Parmeliaceae, phylogeny, polyketide synthesis

Nadim Ahmad and Manfred Ritz contributed equally to this work

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *MicrobiologyOpen* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

In the past, lichens were considered a mutualistic symbiosis consisting of a fungus and partners capable of photosynthesis, such as algae or cyanobacteria (Plitt, 1919). Recent research has indicated that lichen individuals may contain not only the primary fungus responsible for lichen formation (mycobiont) and the primary photosynthetic partner (photobiont) but also additional fungi from other phyla and other organisms. Among these are Basidiomycota (Spribille et al., 2016) or Ascomycota (Muggia & Grube, 2018), as well as various types of bacteria and other algae (Aschenbrenner et al., 2016; Smith et al., 2020). Today, lichens are often referred to as ecosystems (Hawksworth & Grube, 2020) or holobionts (Rolshausen et al., 2022). Over the past few decades, the distinct secondary metabolite profile of these composite organisms has attracted growing interest (Ranković & Kosanić, 2019). A plethora of natural products are synthesized by lichen mycobionts (Huneck & Yoshimura, 1996). Among the latter, the most prominent classes include depsides, depsidones, dibenzofurans, and phenolic compounds (Elix & Stocker-Wörgötter, 2008; Calchera et al., 2019). These compounds exhibit bioactivities of antimicrobial (Kosanić et al., 2013; Ranković & Kosanić, 2019; Ristić et al., 2016; Sisodia et al., 2013), antifungal (Karabulut and Ozturk, 2015), anti-inflammatory (Joshi et al., 2019), antioxidant (Goga et al., 2020; Kosanić & Ranković, 2019), and antitumoral (Kosanić et al., 2013; Solárová et al., 2020) nature. Examples of medically relevant compounds include gyrophoric acid, atranorin, and physodic acid (Cardile et al., 2017). Compounds like physodic acid, evernic acid, atranorin, and usnic acid displayed an inhibitory effect on metabolic enzymes (Boustie & Grube, 2005; Calchera et al., 2019).

To utilize these compounds biotechnologically, it is essential to have a thorough understanding of their biosynthesis. The compounds mentioned above are classified as polyketides and are created through multiple catalytic cycles in which elongation units are linked to a starter molecule by mega-enzyme polyketide synthases (PKS) (Crawford & Townsend, 2010). In the context of fungal PKS, reducing PKS (R-PKS) and nonreducing PKS (NR-PKS) are distinguished based on their domain composition and the extent of reductive processing involved in their catalytic cycles. R-PKSs contain additional domains responsible for reducing the intermediate polyketide chain during synthesis, while NR-PKSs lack these domains and thus do not perform any reduction steps (Cox, 2023). Conserved structures in R-PKS from N- to C-termini consist of ketosynthase (KS); acyl transferase (AT); dehydratase (DH); C-methyl transferase (cMT); enoylreductase (ER); ketoreductase (KR); and acyl carrier protein (ACP) domains (Cox, 2023; Dutta et al., 2014). In contrast, unique PKS domains comprising starter unit:acyl-carrier protein transferases (SAT) and a product template (PT) domain are observed in NR-PKS (Crawford & Townsend, 2010; Huitt-Roehl et al., 2015). SAT domains link the chain-initiating compound to the enzyme, while the PT domain is responsible for regulating the cyclization reactions that convert highly reactive, fully elongated intermediates into specific aromatic compounds (Crawford & Townsend, 2010; Crawford et al.,

2009; Huitt-Roehl et al., 2015; Li et al., 2010). NR-PKS domains are mostly organized in the following order SAT-KS-AT-PT-ACP-ACP-TE (Wang et al., 2021). Other occurring units comprise AMP-binding sites (A), carrier proteins (CP), and the terminal domain (TD); these are however dedicated to hybrid PKS-NRPS (Meier & Burkart, 2011; Boettger & Hertweck, 2013; Calchera et al., 2019).

Next-generation sequencing (NGS) has enabled the identification of potentially involved genes in natural product formation through deep genome sequencing. If multiple genes that are involved in the formation of a specific secondary metabolite are coregulated and located in proximity (not dispersed throughout the genome), they are referred to as a biosynthetic gene cluster (BGC) (Keller, 2019; Keller et al., 2005; Medema et al., 2015; Pizarro et al., 2020; Rokas et al., 2018). Recent studies have revealed that BGCs located within lichenized fungi are responsible for the production of various secondary metabolites, such as pigments, terpenes, and polyketides (Kealey et al., 2021; Kim et al., 2021; Llewellyn et al., 2023; Singh, 2023; Singh et al., 2021b).

To identify particular BGCs, it is essential to gain access to the genomic composition of the symbiotic partners. This can be accomplished by sequencing the entire meta-genome, which provides a comprehensive view of all BGCs present in the organisms involved in the lichen symbiosis. This approach may facilitate uncovering interactions between the multiple species involved in the symbiosis at the level of BGCs (Aschenbrenner et al., 2016). Therefore, metagenomic tools are deployed to divide the obtained metagenomes taxonomically into respective bins. This renders symbiotic partners, which are challenging to cultivate, accessible to genetic interrogation (Muggia et al., 2010). The genome of the mycobiont often harbors most of these BGCs (Pizarro et al., 2020), which are activated in response to environmental stimuli (Chiang et al., 2011; Zheng et al., 2020). In addition, various structurally similar compounds may be encoded by one BGC (Martinet et al., 2019; Singh et al., 2022; Wasil et al., 2013).

Short-read sequencing methods can have limitations in accurately representing isoforms among different species in a sample. This is because short reads may not span across complex genomic or repetitive regions, as well as capture the full extent of genomic diversity in mixed microbial communities. As a result, it can be challenging to confidently assemble the complete genomes of all species present in the sample (Bickhart et al., 2022; Cuscó et al., 2021; Tsai et al., 2016). To overcome these limitations, long-read sequencing techniques are utilized instead. These technologies have the potential to generate more contiguous genome assemblies and improve our ability to accurately identify BGCs and other functional genetic elements in complex microbial communities, resulting in high-quality data output (Xie et al., 2020). Furthermore, nucleotide variances among symbionts can be detected by long reads (Chen et al., 2022).

In this study, we present a new lichen metagenome of *Parmelia sulcata* (PSU) (BioSample SAMN35345252) and the BGCs of the mycobiont. *P. sulcata* is one of the most common species of lichen-forming fungi worldwide. It is widely distributed in temperate and

cold regions of both hemispheres and typically grows on the bark of trees. The lobed thallus has a light gray surface, usually with white ridges and soredia. *P. sulcata* belongs to a species complex, characterized by high genetic diversity (Crespo et al., 1997, 1999; Feuerer & Thell, 2002; Molina et al., 2011).

One of the aims of the present study was to better understand the diversity and distribution of BGCs among lichenized fungi, categorize and group PKSs, and identify gene clusters linked to known compounds. We selected three species with overlapping natural product profiles. The upper surface of *P. sulcata*, *H. physodes*, and *H. tubulosa* is bluish to whitish gray, due to the presence of atranorin in the cortex. The medullary layer contains various colorless depsides and depsidones (Table 1), and the lower surface is black due to the presence of melanins. We hypothesize that some of the BGCs found in the three species are highly similar (orthologous) because they are linked to the same or a structurally similar compound, as has been shown in other, non-lichenized, fungi (Theobald et al., 2018).

The identified BGCs of *P. sulcata* were compared for orthologs with previously published metagenomes from *H. physodes* and *H. tubulosa* (Ahmad et al., 2023). Table 1 summarizes the natural products previously reported in the compared lichen. Some of these exhibit medicinally relevant activities (Ranković et al., 2014; Stojanović et al., 2018; Yilmaz et al., 2005). Consequently, this syntenic comparison provides further insights into relevant natural product formation rendering the biosynthetic potential of the three lichen mycobionts from *P. sulcata*, *H. physodes*, and *H. tubulosa* accessible for biotechnological exploitation.

## 2 | MATERIAL AND METHODS

### 2.1 | Lichen sample collection

Samples were collected in Germany, Altenschneeberg (August 2022), from the bark of conifers. Precise locations of sequenced samples are latitude 49°26'14.3"N and longitude 12°32'50.9"E. To ensure correct lichen identity, a BLAST search on ITS sequences (see Supporting Information: Table S1: [10.5281/zenodo.8205254](https://doi.org/10.5281/zenodo.8205254)) was performed. The lichen sample included in this study was identified as PSU.

### 2.2 | GC-MS analysis of lichen compounds

Furthermore, a part of the collected samples were subjected to GC-MS analysis to investigate the composition of secondary metabolites. This method was chosen as it enables the identification of volatile compounds in lichen, such as orsellinic acid derivatives. Therefore, approximately 500 mg of dry lichen biomass was macerated in 10 mL of methanol for 24 h at 300 rpm. The resulting extract was then analyzed using a Trace GC-MS Ultra system with DSQII (Thermo Scientific). An autosampler TriPlus was utilized to inject a sample volume of 1 µL in split mode onto an SGE BPX5 column (30 m, ID 0.25 mm, film 0.25 µm). The injector temperature was set at 280°C. The initial oven temperature was maintained at 50°C for 2.5 min, followed by a temperature increase at a rate of 10°C/min until reaching 320°C, with a final hold step for 3 min. Helium was used as the carrier gas with a flow rate of 0.8 mL/min and a split ratio of 8. Mass spectra and chromatograms were recorded using electron ionization at 70 eV. Masses were detected within the range of 50 m/z to 650 m/z in positive mode (Ringel et al., 2020). Identification of compounds was achieved by comparing their spectra with the NIST/EPA/NIH MS library version 2.0. Identified compounds in Supporting Information: Table S2 depict the top 10 hits found in this sample (Supporting Information: Table S2: [10.5281/zenodo.8205254](https://doi.org/10.5281/zenodo.8205254)). Some of these compounds however depict silylated molecules that do not belong to this sample, as no silylation was conducted. These are residuals from previous experiments, which remained on the liner or the column.

### 2.3 | High molecular weight DNA (HMW gDNA) extraction and library preparation

Before DNA extraction, the lichen thallus was examined under a binocular microscope to eliminate any moss, wood, and other lichens present in the sample. Additionally, visibly infected parts of the thallus were removed to minimize potential contaminants.

HMW gDNA extraction was conducted as follows. The Quick-DNA Fungal/Bacterial Miniprep Kit (Zymo Research, Europe GmbH) was used to extract lichen HMW genomic DNA. Dry thallus material from PSU samples was ground into a fine powder using liquid nitrogen. The homogenized material was then transferred to the Bashing Bead Buffer provided in the kit. Genomic HMW DNA was isolated according to the

**TABLE 1** Lichen substances found in *Parmelia sulcata*, *Hypogymnia physodes*, and *Hypogymnia tubulosa*. Location in thallus and substance class are given in parentheses.

Lichen species	Natural product	Reference
<i>Parmelia sulcata</i>	Atranorin, chloroatranorin (cortical depsides), salazinic acid, consalazinic acid, lobaric acid (medullary depsidones), lecanoric acid (medullary depside)	Brodo et al. (2003), Candan et al. (2007), Duarte (2022), Galloway (2007)
<i>Hypogymnia physodes</i>	Atranorin, chloroatranorin (cortical depsides), physodic acid, physodalic acid, 3-hydroxyphysodic acid, protocetraric acid, fumarprotocetraric acid (medullary depsidones)	Molnár and Farkas (2011), Purvis (1992), Ranković et al. (2014), Solhaug et al. (2009)
<i>Hypogymnia tubulosa</i>	Atranorin, chloroatranorin (cortical depsides), physodic acid, 3-hydroxyphysodic acid, 4-O-methyl physodic acid (medullary depsidones)	Purvis (1992), Stojanović et al. (2018)

manufacturer's instructions. Due to the high content of polysaccharides, phenolic compounds, and pigments, additional purification steps were necessary. These purifications were performed using the Genomic DNA Clean and Concentrator-10 Kit (Zymo Research, Europe GmbH) and the DNeasy PowerClean Clean-up Kit (Qiagen). The quality of the obtained HMW genomic DNAs was assessed using a Nanophotometer (Implen, Nanophotometer Pearl), Qubit 2.0 Fluorometer (Thermo Scientific), and TapeStation (Agilent Technologies).

SMRT bell libraries were constructed for the samples that passed the quality control, which included a 260/280 absorbance ratio of 1.75–1.85 and a 260/230 absorbance ratio of 2.0–2.2. The library was prepared following the instructions for the Low DNA Input Protocol of the SMRT bell Express Prep kit v2 (Pacific Bioscience). Total input DNA with a size range of 10–18 kb for library generation was approximately 350–600 ng. Ligation with T-overhang SMRT bell adapters was performed at 20°C for 1 h. Following two cleanup steps with AMPure PB beads, the size and concentration of the final library were assessed using TapeStation and Qubit Fluorometer 2.0 with Qubit dsDNA HS reagents Assay Kit.

## 2.4 | Genome and RNA Illumina short-read sequencing

The obtained library was subjected to whole-genome sequencing on a PacBio Sequel IIe device (Pacific Bioscience). Pre-extension and adaptive loading (target of  $p1 + p2 = 0.95$ ) were set to 2 h with an on-plate concentration of 90 pM. The movie time was set to 30 h (Ritz et al., 2023).

Additionally, RNA sequencing was conducted using short-read Illumina technology (NovaSeq, Illumina). This transcriptomic data were generated to provide additional depth and accuracy to the predicted gene models. For RNA extraction, frozen lichen thalli were ground using a CryoMill (Retsch), and the RNeasy Plant Mini Kit (Qiagen) was used. To further purify the extracted RNA, the Turbo DNA-free Kit (Invitrogen) was employed. For short-read RNA sequencing, the samples underwent processing on a NovaSeq instrument using a paired-end run mode and a read length of  $2 \times 150$  bp. To begin, total RNAs were extracted using TRI Reagent (Zymo Research, Europe GmbH) following the manufacturer's instructions. Subsequently, the samples underwent further purification using the RNA Clean and Concentrator-5 Kit (Zymo Research, Europe GmbH). This purification step was repeated until it had a 260/280 absorbance ratio between 1.9 and 2.1, as well as a 260/230 absorbance ratio between 1.8 and 2.2. Only RNAs with a RIN value greater than 8.0 (TapeStation) were considered suitable for sequencing.

## 2.5 | Bioinformatic and statistical analysis

Metagenomic reads derived from entire lichen thalli primarily consist of fungal sequences originating from the mycobiont (Greshake

Tzovaras et al., 2020). This composition presents challenges for genome assemblers that rely on solid  $k$ -mers. The abundance of certain species can lead to their over-representation, while low-abundance species like the photobiont may fail to assemble. To address these challenges, the obtained long CCS reads from PacBio Sequel IIe were therefore assembled using metaFlye v2.9.1. This assembler is specifically designed to manage read coverages with high nonuniformity, making it well-suited for this case. Additionally, the assembled contigs were simultaneously scaffolded using flye, enabling further bioinformatic processing (Kolmogorov et al., 2020).

To differentiate the acquired data sets, a taxonomic binning approach was employed. This involved performing blastx using the DIAMOND v2.0.14 algorithm (Buchfink et al., 2015) on a custom-made database, as well as utilizing the MEGAN6 LR Community Edition v6.21.7 (Huson et al., 2018). The DIAMOND database used in the analysis encompassed protein sequences from various taxonomic groups, including fungi, bacteria, archaea, viruses, chlorophyta, klebsormidophyceae, tremella, and cystobasidium. To circumvent obstacles like insertion and deletion errors in long-read sequencing, the flags `--more-sensitive --frameshift 15` and `--rage-culling` were employed in DIAMOND to allow for a frame-shift-aware alignment mode (Bağcı et al., 2021). The resulting files were subjected to further processing in MEGAN, where taxonomically assigned sequences were matched to their respective bins (Bağcı et al., 2019). Subsequently, contigs and scaffolds corresponding to the desired nodes were extracted for subsequent analysis. To evaluate the completeness and quality of the resulting bins for further investigation, several tools were utilized, including BUSCO v5.3.2 (Benchmarking Universal Single-Copy Orthologs) (Simão et al., 2015), QUAST, v5.2.0 (Quality Assessment Tool) (Gurevich et al., 2013), and SeqKit v2.3.1 (Shen et al., 2016). To validate the identities of the mycobiont, an ITSx v1.1.3 analysis was performed (Bengtsson-Palme et al., 2013). Gene prediction was carried out using AUGUSTUS v3.4.0/BRAKER v2.1.6, leveraging both the metagenomics data and corresponding transcriptomic data as hints (Brûna et al., 2021; Hoff et al., 2016, 2019). This facilitated the functional annotation of genes in the respective data sets. Furthermore, the identification of BGCs present in the obtained bin was performed using antiSMASH v6.1.1 (Blin et al., 2021). The Gene Ontology (GO) terms were levied by InterProScan (v5.59-91.0) (Jones et al., 2014).

To visualize the distribution of BGCs in all three mycobionts, an evaluation with BiG-SCAPE (Navarro-Muñoz et al., 2019) and Cytoscape (Shannon et al., 2003) was conducted, clustering the respective contigs by shared name. To gain insights into the polyketide synthesis in the mycobiont, PKS-related BGCs were investigated further. At first, a Reciprocal Best Hit (RBH) based on BLAST was performed on the respective primary mycobiont bins of PSU, *Hypogymnia physodes* (HPH), and *Hypogymnia tubulosa* (HTU) (Camacho et al., 2009; Cock et al., 2015) in all three combinations. The latter two were previously published (Ahmad et al., 2023). AntiSMASH, InterProScan, ITSx, SeqKit, and RBH analyses were performed on the Galaxy servers (Afgan et al., 2022). Obtained pairs from RBH were compared for duplicates, highlighting highly

conserved PKS-related genes in the investigated mycobiont genomes. Therefore, the sequences of KS regions were extracted and compared on a phylogenetic level to already published data sets (Gerasimova et al., 2022; Mosunova et al., 2022; Singh et al., 2022) and sequences from the MIBiG database (Terlouw et al., 2023). Therefore, a midpoint rooted maximum likelihood phylogenetic tree (IQ-TREE v2.1.2; Nguyen et al., 2015), levied with 1000 ultrafast bootstraps (Minh et al., 2013) and the elaborated substitution model (ModelFinder; Kalyaanamoorthy et al., 2017) LG + F + G8 was built based on an alignment with MAFFT v7.508 (Katoh & Standley, 2013). The included sequences were grouped in the vicinity of the KS sequences from HPH, HTU, and PSU to highlight the validity of the obtained results. It yielded the following metrics, based on Input data of 76 sequences with 3742 amino-acid sites, exhibiting 917 constant and invariant sites, 2142 parsimony informative sites, and 3009 distinct site patterns. This tree was visualized with iTOL v5 (Letunic & Bork, 2021) and Inkscape. The according fasta and alignment file is provided in Supporting Information: Tables S3 and S4: [10.5281/zenodo.8205254](https://doi.org/10.5281/zenodo.8205254). Subsequently, an alignment of related gene clusters was performed on EasyFig (v2.2.5) (Sullivan et al., 2011) to investigate PKS synteny between HPH, HTU, and PSU. Additionally, a progressiveMAUVE (Darling et al., 2004) alignment of the whole mycobiont bins was deployed to visualize synteny on the genome level in Geneious (Geneious Prime® 2022.0.1).

### 3 | RESULTS AND DISCUSSION

#### 3.1 | Genome sequencing and quality assessment

The same lichen-related substances were observed in the GC-MS analysis (see Supporting Information: Figure S1 and Table S2: [10.5281/zenodo.8205254](https://doi.org/10.5281/zenodo.8205254)) of PSU as in the previously described species HPH and HTU (Ahmad et al., 2023). Therefore, this lichen was deemed to harbor familiar BGCs. To present sequencing quality, the metrics of PSU are listed in Table 2. As the mean HiFi Read Quality exhibited a value well above Q20, further bioinformatic analyses were conducted.

After taxonomic analysis, the mycobiont (Parmeliaceae) bin was further investigated. Evaluating the contiguity of the assembled metagenome involved the utilization of the quality assessment tool for genome assemblies (QUAST). The metagenome was assessed based on the genome size, number of contigs, and N50 values. A summary of the statistics is provided in the upper part of Table 3. Notably, the N50 value was beneath the recommended threshold of 1 Mb from PacBio, indicating poor contiguity. To address this issue, ITS analysis was performed yielding two ITS sequences, which were both identified as PSU in the NCBI BLAST search against the nr database using megablast. This may be due to the intertwined growth of two closely related subspecies of PSU or the lack of data to align the respective ITS sequences to. Subsequent binning clustered all the related contigs together, which may have impacted the QUAST results regarding, for example, N50. Nevertheless, the low L50 value

**TABLE 2** Metagenomic PacBio sequencing of *Parmelia sulcata* (PSU).

Analysis metrics	PSU
Total bases (Gb)	683.00
HiFi reads	3,799,596
HiFi yield (Gb)	26.01
HiFi read length (mean, bp)	6845
HiFi read quality (median)	Q43
HiFi number of passes (mean)	18

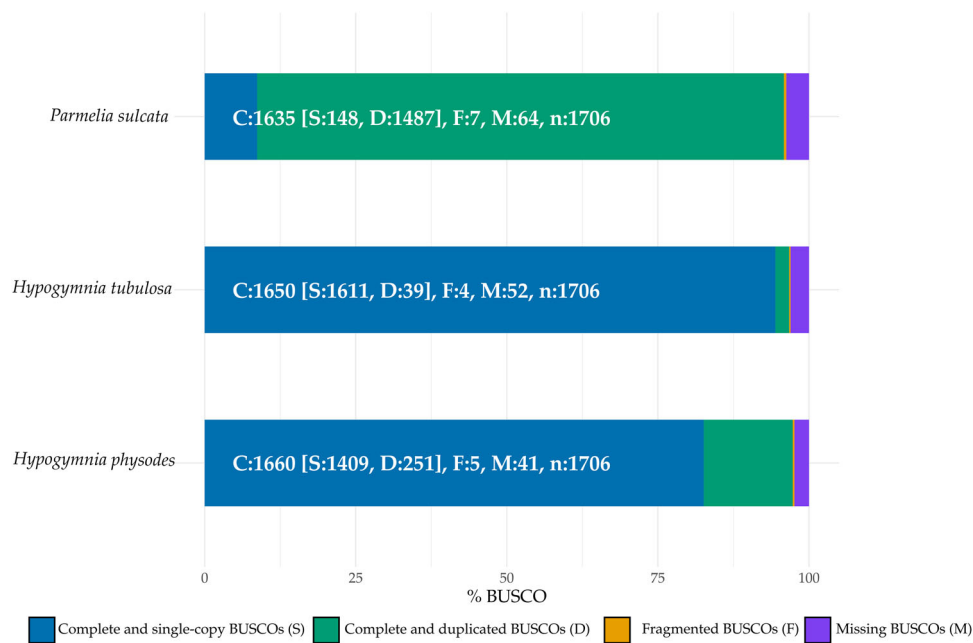
**TABLE 3** QUAST and gene prediction results of Parmeliaceae bin, including the mycobiont sequences.

	Analysis metrics	Parmeliaceae	
QUAST	Number of contigs	328	
	Largest contig	2,191,686	
	Total length	82,929,871	
	N50	698,771	
	L50	41	
	Number of Ns per 100 kbp	0	
	Average coverage	107.9	
	GC content	47.62	
	Gene prediction	Gene count	24,437
		Average gene length	1547.1
Gene density (genes/Mb)		294.67	
Introns/gene		2.29	

suggests sufficient contiguity of the obtained genome. The possible presence of two closely related lichen species in the mycobiont bin is further supported by a comparison of the obtained genome size to that of *Parmelia* spp. reported in the literature (45 Mb, NCBI BioSample: SAMN17391792). Average coverage of 107× was deemed to be highly sufficient for further processing. Comparison of GC content with the genome from literature yields equivalent results. For further assessment, a gene prediction was conducted with Augustus-BRAKER based on short-read RNA sequencing data. The resulting statistics are depicted in the lower part of Table 3. The observed gene count was 24,437, which exceeded the median value of 11,000 (Stajich, 2017), further hinting at the presence of two mycobionts in the investigated bin.

Genome completeness and reliability for further data processing were evaluated with a BUSCO (Simão et al., 2015) assessment. The results were normalized and summarized in Figure 1. Ascomycota odb10 was utilized as the orthologous gene set. Obtained BUSCO results are depicted in percentages to allow for an accurate comparison between the investigated bins. Additionally, absolute





**FIGURE 1** BUSCO genome completeness assessment of the mycobiont subsets of *Parmelia sulcata*, *Hypogymnia tubulosa*, and *Hypogymnia physodes*; the latter two were previously described (Ahmad et al., 2023). The subset checked for completeness was Ascomycota with the orthologous gene set *ascomycota\_odb10*.

numbers were included in all columns to allow for more precise comparison between different bins.

At first sight, PSU exhibits a high number of complete and duplicated BUSCOs supporting the statement of two closely related PSU species. The remaining bins harbored more complete and single BUSCOs. The fragmentation rate was low throughout all investigated mycobionts. The sum of missing BUSCOs was comparable in the three bins.

### 3.2 | BGC annotation and phylogeny

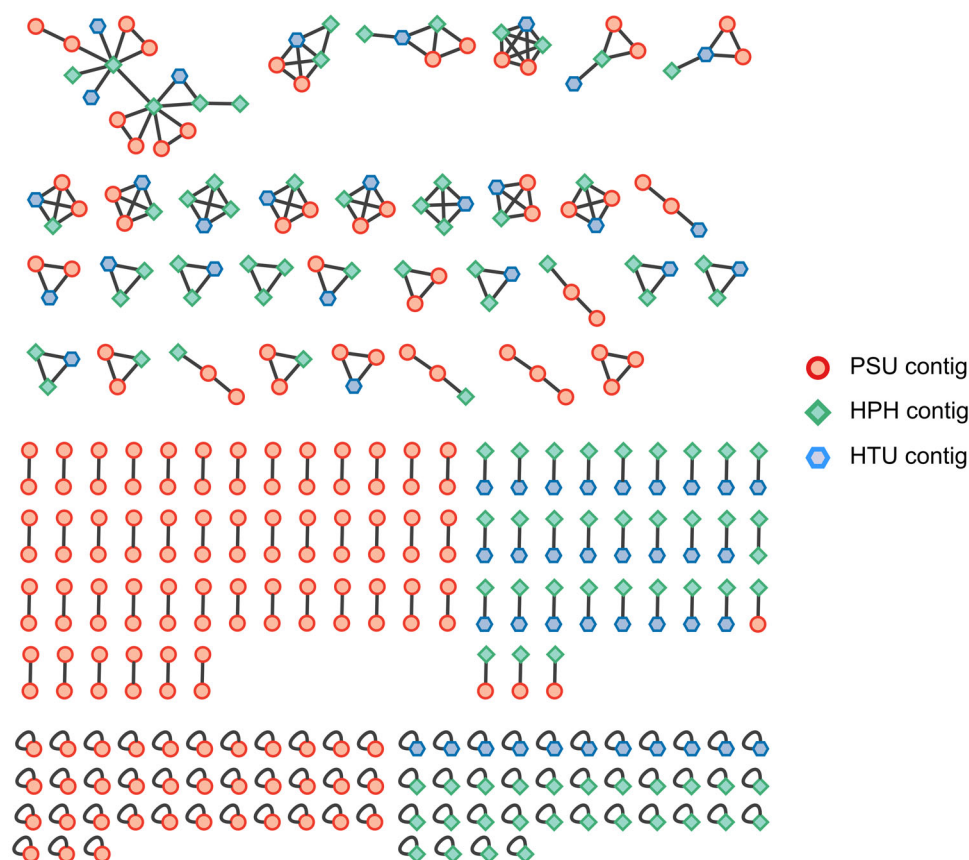
The genus *Parmelia* produces a plethora of secondary metabolites, rendering further evaluation of inherent pathways highly interesting (Candan et al., 2007; Elečko et al., 2022; Gandhi et al., 2022; Ranković & Kosanić, 2019; Ranković et al., 2007). Therefore, the obtained metagenome of *P. sulcata* was investigated for BGCs by antiSMASH 6.1.1 fungal/bacterial version (Blin et al., 2021), leveraging the prior conducted gene annotation. This yielded the following amount of BGCs for the three mycobiont bins: HTU 73, HPH 114, and PSU 214. These findings are coherent with the observation of enhanced secondary metabolite production in lichen mycobionts and fungi (Bills & Gloer, 2017; Devi et al., 2020; Goga et al., 2020; Ola et al., 2013; Shwab & Keller, 2008). The number of BGCs present in HPH and HTU from previous studies (Ahmad et al., 2023) was in line with the literature, as the described BGC content in lichen ranges between 27 and 80 (Calchera et al., 2019). Regarding PSU, the total number of BGCs exceeded others potentially due to the bin containing two closely related mycobionts. As *P. sulcata* is a widely

distributed lichen species and known for its high variability, the possibility of collecting two closely related, intricately growing species is high (Brodo et al., 2003). These are referred to as “cryptic” and are described as different species with similar morphology (Bickford et al., 2007; Brodo et al., 2003; Hawksworth & Rossman, 2007; Molina et al., 2011). The phenomenon of cryptic species is often observed in *P. sulcata* (Molina et al., 2011). This renders the physical isolation of only one specimen from an environmental sample challenging or even not feasible. The same applies to separation on a metagenomic level, as most currently available genomes were assembled de novo and the overall amount is still limited, making binning on a species level challenging. However, if the amount of BGCs was divided by two, more comparable results may be obtained. By comparison to different organisms, which also exhibit a high richness in BGCs like *Nocardia* spp. (~36) (Doroghazi & Metcalf, 2013; Männle et al., 2020), *Myxobacteria* spp. (30–46) (Amiri Moghaddam et al., 2018; Gregory et al., 2019), *Streptomyces* spp. (23–80) (Caicedo-Montoya et al., 2021; Liu et al., 2021), and *Cyanobacteria* spp. (1–42) (Dittmann et al., 2015; Popin et al., 2021), the presented mycobiont genome shows comparable or higher totals of BGCs when the nature of the deployed bin is considered. For the new genome of PSU, antiSMASH assigned 10 out of 214 BGCs with a specific function (100% similarity to MIBiG clusters), leaving the majority of the BGCs with unknown or uncharacterized functions. Among the annotated “most similar cluster” were, for example, 6-hydroxymellein, 1,3,6,8-tetrahydroxynaphthalene, and clavarinic acid. Please refer to Supporting Information: Table S5 for further information: [10.5281/zenodo.8205254](https://doi.org/10.5281/zenodo.8205254). A visual representation of all BGCs from the three investigated mycobionts is depicted in

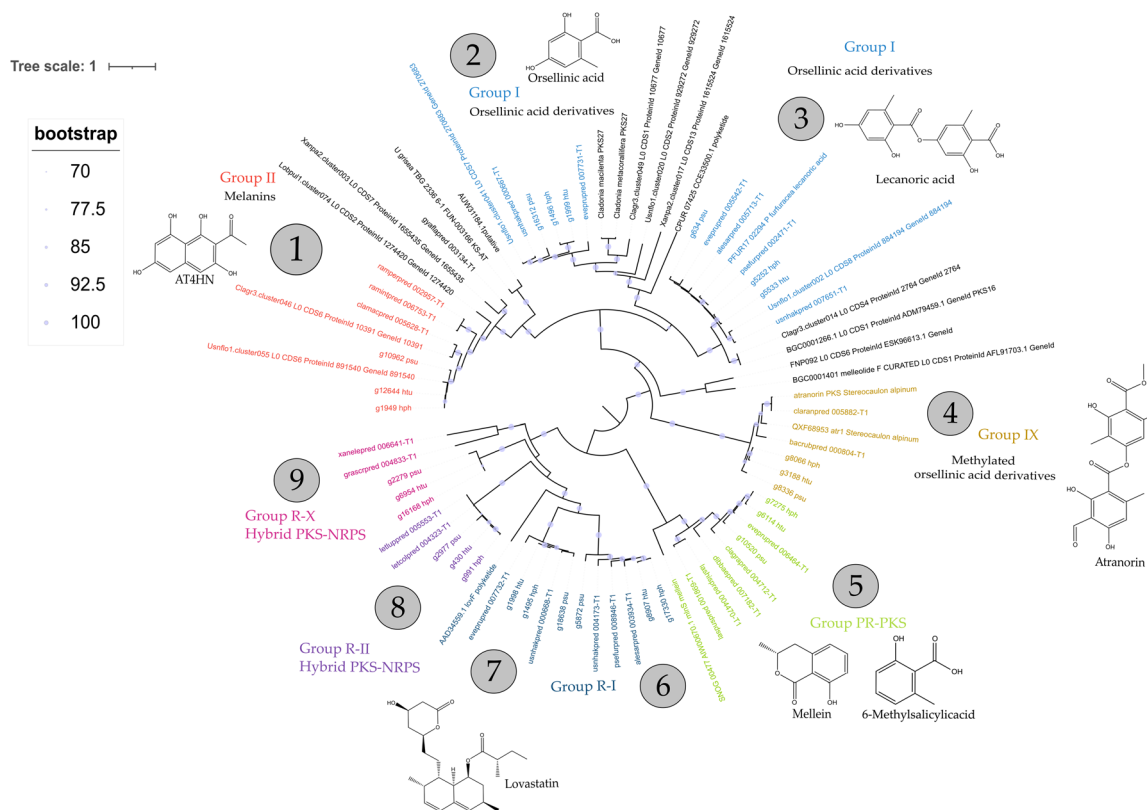
Figure 2. The analysis with BiG-SCAPE yielded a similarity network which was visualized with Cytoscape. Investigated BGCs were color-coded based on their respective mycobiont. This visualization highlights the connection of various BGCs between the investigated mycobionts. It comes to attention that only one cluster contains solely one BGC belonging to PSU; all others contain two BGCs from PSU if this mycobiont is involved. By focusing on the middle segment of Figure 2, it becomes evident that PSU exhibits a high amount of paired BGCs (45) and also triples (two, located at the lower part of the upper segment), which is not observed in the other two mycobionts. Due to the high duplicates but also the presence of singletons grouped in the lower part of the figure, the assumption on two closely related *P. sulcata* species is fortified, as singletons exhibit no sufficient sequence similarity to other BGCs present in the investigated data set. These unique or highly divergent BGCs do not share common biosynthetic gene homology with other clusters in the data set. This renders them highly interesting for further investigation as these may provide insights into the potential production of novel or uncharacterized secondary metabolites (Sánchez-Navarro et al., 2022). Since the most abundant group of BGCs is related to polyketide formation, these were further investigated.

Subsequently, BGCs related to PKS formation were subjected to RBH with BLAST (Camacho et al., 2009), where all three mycobiont data sets were compared and paired. The resulting orthologous pairs yielded highly conserved regions accessible for further analysis (Supporting Information: Table S6: 10.5281/zenodo.8205254).

As the ketosynthase (KS) region of PKS represents their most conserved domain (Amnuaykanjanasin et al., 2009), a phylogenetic tree was constructed based on these sequences (Figure 3). Depicted clades were numbered from one to nine and additionally included gene names tagged with the lichen mycobiont they are derived from. In every clade, three KS regions belonging to the different mycobionts aligned with reference sequences. These findings strongly highlight the degree of similarity between the compared PKS throughout all examined samples. In addition, both *Hypogymnia* species clade more closely than the PSU sample, which reflects the phylogenetic relationships (Divakar et al., 2015). Furthermore, elaborated KS regions were compared phylogenetically to previous studies to elucidate putative function and natural product formation (Gerasimova et al., 2022; Mosunova et al., 2022; Singh et al., 2022). This tree depicts the KS regions of the lichen compared in this study with reference sequences from previous studies (Gerasimova



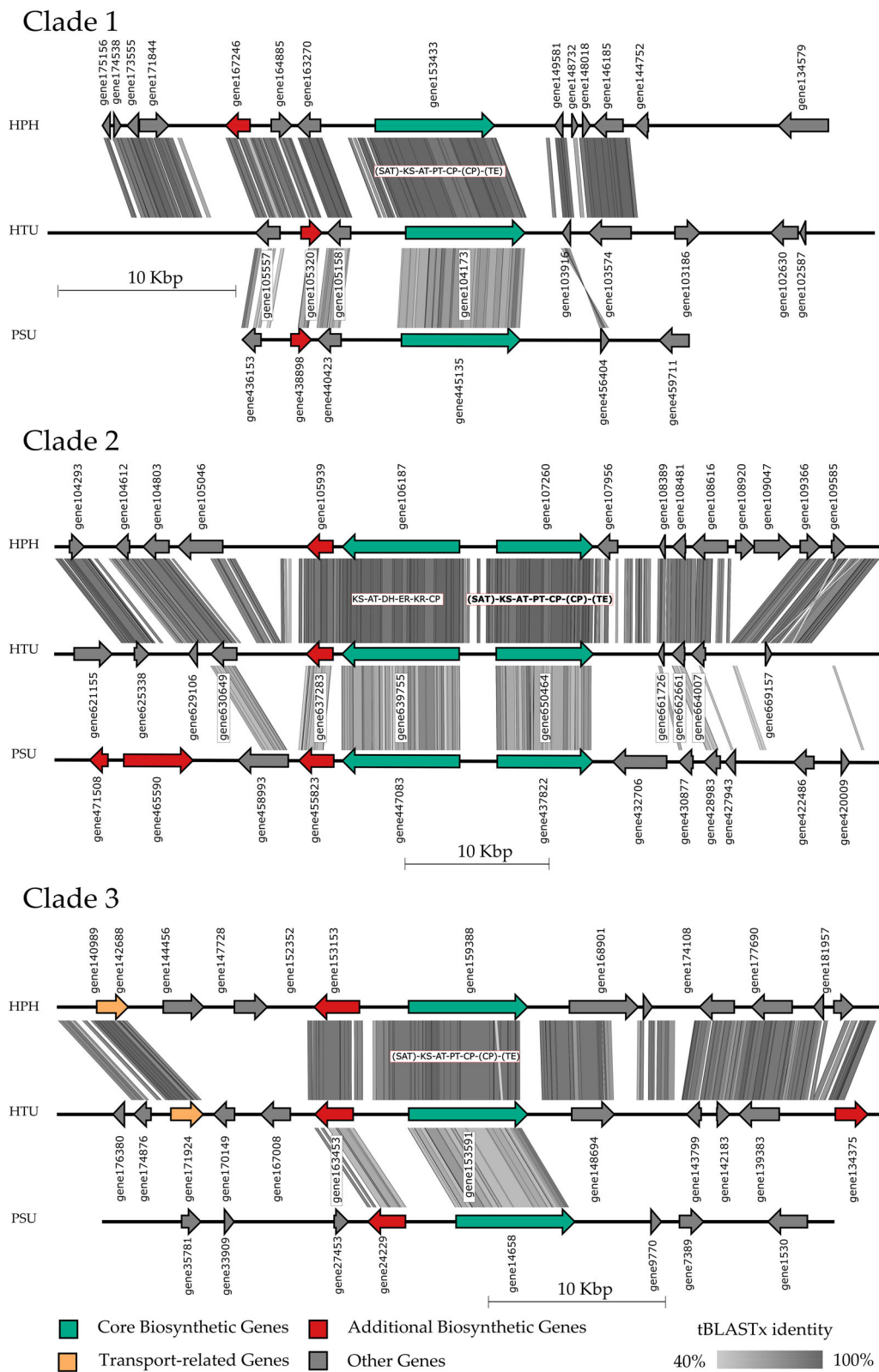
**FIGURE 2** BiG-SCAPE analysis and visualization of biosynthetic gene clusters (BGCs). Here, BGCs of all three lichen mycobionts were analyzed in BiG-SCAPE and clustered in Cytoscape. The upper segment represents the similarity network between the contigs of the investigated three lichen mycobionts. In the left middle segment, contig pairs of *Parmelia sulcata* (PSU) are grouped, whereas on the right side, the pairs of *Hypogymnia physodes* and PSU are located. The singletons of each mycobiont are depicted in the lower section. These represent contigs with no significant sequence similarity to others in the pool.



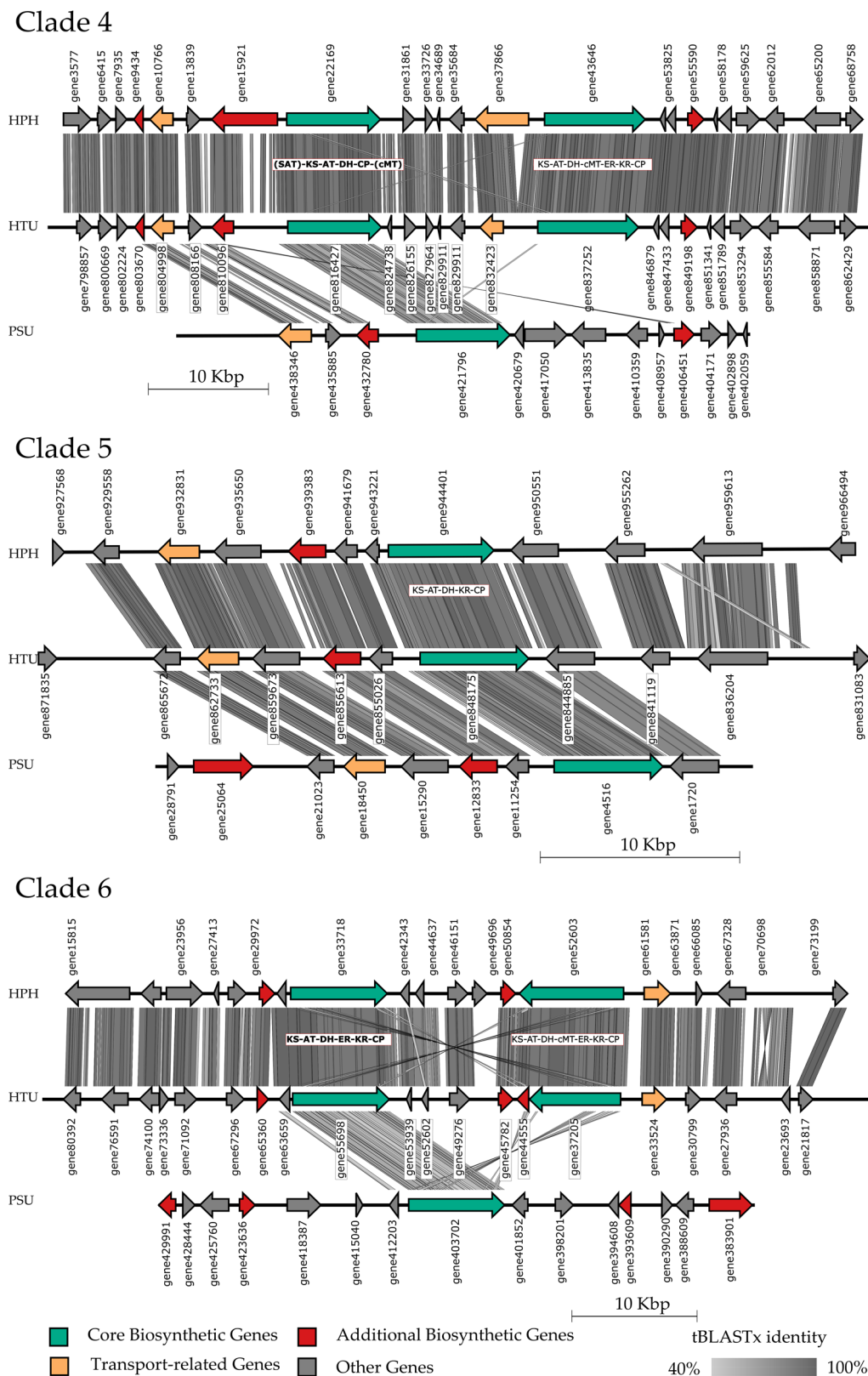
**FIGURE 3** Phylogenetic tree of orthologous KS regions compared to reference sequences. Phylogenetic relationships of nine biosynthetic gene clusters (Clades 1–9), which had an ortholog in each of the three studied species. The phylogeny is based on the ketosynthase domain of the PKS. Gene names were tagged with the abbreviation of the respective lichen mycobiont. Obtained KS sequences were compared on a phylogenetic level with the data sets from previous studies, including sequences from the MIBiG database, and the groups were named accordingly (Gerasimova et al., 2022; Mosunova et al., 2022; Singh et al., 2022). All clades were allocated to the respective group with the corresponding putatively produced chemical compound. AT4HN, 2-acetyl-1,3,6,8-tetrahydroxynaphthalene.

et al., 2022; Mosunova et al., 2022; Singh et al., 2022) and the MIBiG database. The upper part of Figure 3 depicts PKS belonging to the group of nonreducing PKS (NR-PKS), comprising Clades 1–4. The other part of the figure shows Clades 6–9 representing reducing PKS (R-PKS), whereas Clade 5 contains a partially reducing PKS (PR-PKS). The position of the latter is in line with the nature of the described reductive functions, being in the middle of the two types. Clade names are based on previous PKS phylogenies (Gerasimova et al., 2022; Mosunova et al., 2022; Singh et al., 2022). Clade 1 is putatively linked to naphthalene-like compounds, which is also in line with annotations by antiSMASH. This BGC may be associated with melanin biosynthesis in the three studied lichenized fungi. All of the species are characterized by a black (melanized) lower surface of the thallus (Elvebakk, 2011). Clades 2 and 3 belong to Group I and are linked to orsellinic acid and its derivatives, such as the di-depside lecanoric acid (Schroeckh et al., 2009) and the tri-depside gyrophoric acid (Singh et al., 2022). We observed orsellinic acid in the GC-MS analysis. Orcinol-type depsidones may also be linked to these clusters (Singh et al., 2021a). Interestingly, the BGCs in Clade 2 contain a second, reducing PKS (Figure 3). Alternatively, a FAS comprising HexA and HexB subunits could be responsible for the synthesis of the acyl chains, as seen in the biosynthesis of norsolorinic acid in the

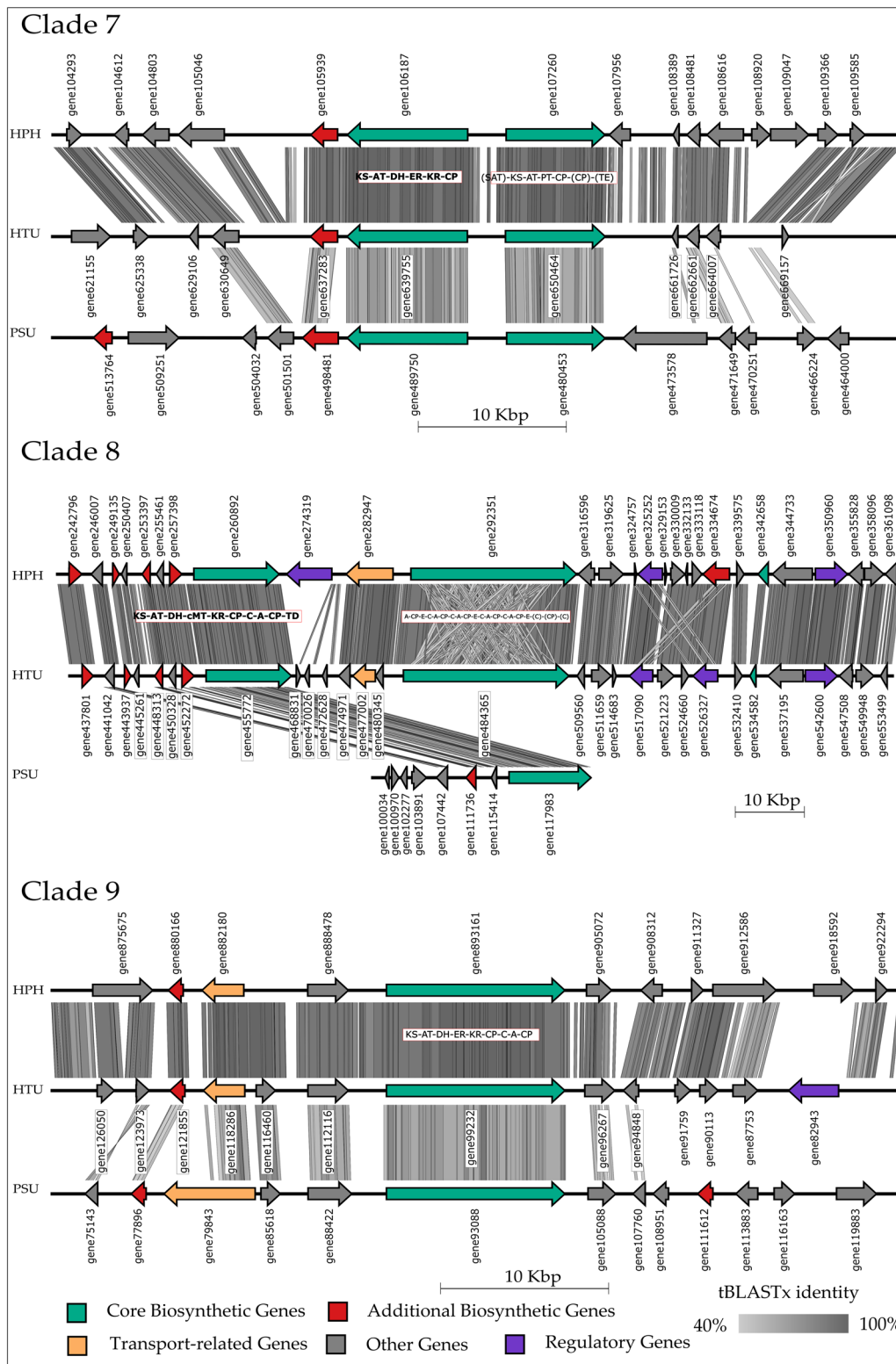
aflatoxin pathway (Brown et al., 1996; Watanabe & Townsend, 2002). Homologs of HexA and HexB genes have been found in the lichenized fungi *Pseudevernia furfuracea* and *Cladonia grayi* (Singh et al., 2021a). Clade 4 contains a PKS with a cMet domain and is linked to methylated orsellinic acid and its derivatives, for example, the beta orcinol-type depside atranorin. This BGC has been functionally characterized and was found in the genomes of several atranorin-producing lichens (Kim et al., 2021). Regarding Clade 5 (PR-PKS), a possible product is mellein or 6-methyl salicylic acid (6-MSA), resembling other polyketides produced by lichen. For the R-PKS, Clades 6, 7, and 8 were assigned to R-I and R-II, allegedly producing lovastatin-like compounds. Interestingly, Clade 9 was putatively assigned to the novel R-PKS group X being located at the very edge of this group, estimating product formation challenging. It is to be mentioned that biosynthetic core genes of Clades 8 and 9 were annotated as hybrid PKS-NRPS. To further validate the obtained results, the whole PKS sequences of HPH, HTU, and PSU allocated to the respective clades were compared with some of the reference sequences grouping in the vicinity of these via blastp and another phylogenetic tree (Supporting Information: Figure S4: 10.5281/zenodo.8205254). The phylogenetic tree only slightly differs from Figure 3 in the position of Clade 8 and the similarity of Clade 8 to



**FIGURE 4** Synteny plot of Clades 1–3 of the orthologous PKS genes. Three out of nine highly syntenic biosynthetic gene clusters found in the three studied lichen-forming fungi. Clade 1 is linked to melanin biosynthesis, and Clades 2 and 3 are linked to the synthesis of orsellinic acid and related orcinol-type depsides and depsidones. Phylogenetic relationships of the nine clusters (based on the ketosynthase domain of the T1PKS core gene) and *putatively* associated compounds are presented in Figure 3. ACP, acyl carrier protein; AT, acyltransferase; CP, carrier protein; DH, dehydratase; ER, enoylreductase; KR, ketoreductase; KS, ketosynthase; PT, product template; SAT, starter unit:acyl-carrier protein transferases; TE, thioesterase. Domains in brackets were annotated via antiSMASH as outside or incomplete modules.



**FIGURE 5** Synteny plot of clades 4–6 of the orthologous PKS genes. Three out of nine highly syntenic biosynthetic gene clusters found in the three studied lichen-forming fungi. Clade 4 is linked to the synthesis of methylated *orsellinic acid* and related beta-ornicol-type depsides, such as atranorin. Clade 5 is linked to mullein or 6-MSA biosynthesis, whereas Clade 6 putatively produces lovastatin-like compounds. Phylogenetic relationships of the nine clusters (based on the ketosynthase domain of the T1PKS core gene) and putatively associated compounds are presented in Figure 3. AT, acyltransferase; cMT, C-methyltransferase; CP, carrier protein; DH, dehydratase; ER, enoylreductase; KR, ketoreductase; KS, ketosynthase; SAT, starter unit; acyl-carrier protein transferases. Domains in brackets were annotated via antiSMASH as outside or incomplete modules.



**FIGURE 6** Synteny plot of clades 7–9 of the orthologous PKS genes. Three out of nine highly syntenic biosynthetic gene clusters found in the three studied lichen-forming fungi. Clades 7 and 8 are linked to the putative synthesis of lovastatin-like compounds. Clades 8 and 9 are annotated as hybrid PKS-NRPS BGCs. A, AMP-binding; AT, acyltransferase; C, condensation; cMT, C-methyltransferase; CP, carrier protein; DH, dehydratase; ER, enoylreductase; KR, ketoreductase; KS, ketosynthase; TD, terminal domain. Domains in brackets were annotated via antiSMASH as outside or incomplete modules.

Clades 6 and 7. Deviations in the tree are due to differences in sequence homology resulting in a light disposition of the clades. The blast results (Supporting Information: Table S7: [10.5281/zenodo.8205254](https://doi.org/10.5281/zenodo.8205254)) yielded percent identities of an average of 65%, excluding alignments with the lovastatin PKS lovF in Clades 6 to 8 (average 32%) and the reference sequences in Clade 9 (average 40%). This also explains the position of the identified genes of HPH, HTU, and PSU in Clade 9 in relation to the reference sequences. With the nevertheless high homology in KS domains, an investigation on the synteny level may permit a deeper insight into the homology of PKS-related BGCs.

### 3.3 | Synteny of investigated mycobiont BGCs

As the distribution of presented BGCs in Figure 2 slightly differs, an investigation of synteny on the genomic level may yield more insights into conserved regions throughout the evaluated mycobionts. In Supporting Information: Figures S2 and S3 ([10.5281/zenodo.8205254](https://doi.org/10.5281/zenodo.8205254)), the bin alignments of HPH/HTU and HTU/PSU with progressiveMAUVE (Darling et al., 2004) were pictured. The red bars in the alignments indicate regions of large-scale rearrangements or inversions, which were predominantly observed between HPH and HTU. By comparing the number of lines connecting the synteny blocks of the respective lichen mycobiont pair, the close relation of HPH and HTU becomes more prominent while a lower amount connects the blocks of HTU and PSU. As HPH and HTU share a high synteny, only a comparison of HTU and PSU was attached as an additional comparison with HPH was considered redundant.

To gain deeper insights into the conserved PKS of the investigated mycobionts, a synteny evaluation of the BGCs derived from the phylogenetic tree (Figure 3) was conducted. The resulting synteny plots were constructed using EasyFig and comprise each of three of the phylogenetic clades. All plots include gene names with correlating sequences, which can be found in the Supporting Information of the respective mycobiont (please refer to Supporting Information: Table S8 and Folder S1: [10.5281/zenodo.8205254](https://doi.org/10.5281/zenodo.8205254)). Additionally, the domain composition of the homologous core gene was included, highlighting the phylogenetically allocated core gene in bold font for BGCs with more core genes. A high cluster homology can be observed between HPH and HTU throughout all computed synteny plots. In Figure 4, the core biosynthetic genes are conserved in all three clades, which also applies to additional biosynthetic genes. The cluster homology between HTU and PSU is mostly confined to the annotated genes. However, several HPH coding regions in Clade 1, for example, exhibit high homology to noncoding regions in HTU. These may be artifacts from gene prediction, indicating genes missed by annotation (Calchera et al., 2019). For Clade 1 and Clade 2, an inversion is observed while comparing HTU to PSU. The domain composition of orthologous core genes in Clade 1, Clade 3, and the first core gene (bold) of Clade 2 indicates that the PKS is nonreducing. On the other hand, the remaining Clade 2 core gene suggests reducing the activity of the corresponding PKS.

Regarding cluster homology in Figure 5, similar results are observed as in Figure 4, with HPH and HTU depicting high sequence homology. PSU exhibits the highest cluster homology in Clade 5 when compared to the remaining two clades. In addition to the core genes, accessory genes show high homology to HPH and HTU. The PKSs in the investigated clades represent partially reducing activities in Clade 5 and reducing PKS in Clade 6, whereas Clade 4 exhibits a nonreducing PKS based on the sequence of domains. The distribution and composition of genes (core and accessory) differ between PSU and the remaining mycobionts, in particular Clades 4 and 6.

The last group of clades is visualized in Figure 6. Notably, Clades 2 and 7 include the same genes from HPH and HTU but different genes from PSU. Homology between HPH and HTU is comparable to the previously described figures while PSU differs. The BGC of PSU in Clade 8 is small when compared to those of HPH and HTU; however, the orthologous core genes express high homologies, including some accessory genes. Clade 9 exhibits high similarities in gene composition throughout all compared. In Clades 8 and 9, only regulatory genes were annotated throughout all investigated clades. The absence of regulatory genes in the other BGCs may have several reasons. Possibly these genes were incompletely annotated or artifacts in gene prediction persisted. Other potential causes involve a distributed regulation from external regulatory genes (Sun et al., 2022) or an alternative, non-gene-based regulatory mechanism (Wang et al., 2021). Another conceivable reason may be a novel or yet uncharacterized regulatory gene (Keller, 2019).

By combining phylogenetic analyses with subsequent synteny evaluation of the KS regions, we were able to demonstrate that gene cluster similarity strongly correlates with the underlying KS topology (Calchera et al., 2019; Jenke-Kodama et al., 2005; Kroken et al., 2003; Ziemert & Jensen, 2012). Compared BGCs pose an intriguing opportunity for further investigation of polyketide synthesis. As these derive from lichen with comparable secondary metabolite production according to GC-MS, the underlying genes for these natural products should be highly conserved, rendering them accessible for experimental exploitation.

## 4 | CONCLUSIONS

In this study, we provide a new genome of PSU, which was extracted from a metagenomic sample. Additionally, inherent PKS genes were compared with sequences from *Hypogymnia physodes* and *Hypogymnia tubulosa* from a previous study (Ahmad et al., 2023). Thus, orthologous PKS genes were evaluated on a phylogenetic level with reference sequences from Gerasimova et al., 2022; Mosunova et al., 2022; Singh et al., 2022; and the MIBiG database. The obtained phylogenetic tree provides information about the allocated putatively produced compounds of each clade. While most of the investigated clades show high sequence homology when compared to the respective reference sequences (Gerasimova et al., 2022; Mosunova et al., 2022; Singh et al., 2022), Clades 8 and 9 exhibit low homologies and need further investigation. A syntenic evaluation of

the nine orthologous BGC triplets from the three lichen samples compared in this study highlights the similarity of the biosynthetic core genes and renders them ready for wet lab experiments. This genome mining approach identified the sequences involved in the putative formation of various polyketides, which need to be further investigated by expression in suitable organisms in wet-lab experiments. As lichen secondary metabolites are still yet an untapped pool of compounds with pharmaceutical relevance, this study gives access to a new high-quality genome ready for genome mining.

Based on these findings, further experiments can be conducted to shed light on the biosynthetic machinery of lichen PKS and their intriguing product spectra. The obtained results can be used to verify the predicted function and help to dereplicate PKS for future studies.

## AUTHOR CONTRIBUTIONS

**Nadim Ahmad:** Conceptualization (equal); data curation (lead); formal analysis (equal); investigation (equal); methodology (equal); software (equal); validation (equal); visualization (equal); writing—original draft (lead); writing—review & editing (equal). **Manfred Ritz:** Conceptualization (equal); formal analysis (equal); investigation (equal); methodology (equal); software (equal); validation (equal); visualization (equal); writing—review & editing (equal). **Anjuli Calchera:** Investigation (equal); methodology (equal); validation (equal); writing—review & editing (equal). **Jürgen Otte:** Investigation (equal). **Imke Schmitt:** Conceptualization (equal); writing—review & editing (equal). **Thomas Brueck:** Funding acquisition (lead); resources (lead); supervision (equal); writing—review & editing (equal). **Norbert Mehlmer:** Conceptualization (equal); project administration (lead); supervision (equal); writing—review & editing (equal).

## ACKNOWLEDGMENTS

This research was funded by the German Federal Ministry of Education and Research, grant number 031B0824A. We thank Nathanael Arnold for proofreading the manuscript and Martina Haack for enabling the GC measurements. The Galaxy server that was used for some calculations is in part funded by Collaborative Research Centre 992!—Medical Epigenetics (DFG grant SFB 992/1 2012) and German Federal Ministry of Education and Research (BMBF grants 031 A538A/A538C RBC, 031L0101B/031L0101C de.NBI-epi, 031L0106 de.STAIR (de.NBI)). Open Access funding enabled and organized by Projekt DEAL.

## CONFLICT OF INTEREST STATEMENT

None declared.

## DATA AVAILABILITY STATEMENT

The sequence data presented in this study are openly available in the National Center for Biotechnology Information (NCBI) under BioSample accession numbers: *Parmelia sulcata* SAMN35345252, *Hypogymnia physodes* SAMN34074577, and *Hypogymnia tubulosa* SAMN34074619. Supporting Information: Figures S1–S4 and Tables S1–S8 are available in the Zenodo repository: [10.5281/zenodo.8205254](https://doi.org/10.5281/zenodo.8205254)

## ETHICS STATEMENT

None required.

## ORCID

Thomas Brueck  <http://orcid.org/0000-0002-2113-6957>

Norbert Mehlmer  <http://orcid.org/0000-0002-6854-4341>

## REFERENCES

- Afgan, E., Nekrutenko, A., Grünig, B. A., Blankenberg, D., Goecks, J., Schatz, M. C., Ostrovsky, A. E., Mahmoud, A., Lonie, A. J., Syme, A., Fouilloux, A., Bretaudeau, A., Nekrutenko, A., Kumar, A., Eschenlauer, A. C., DeSanto, A. D., Guerler, A., Serrano-Solano, B., Batut, B., ... Briggs, P. J. (2022). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Research*, 50, W345–W351. <https://doi.org/10.1093/NAR/GKAC247>
- Ahmad, N., Ritz, M., Calchera, A., Otte, J., Schmitt, I., Brueck, T., & Mehlmer, N. (2023). Biosynthetic potential of hypogymnia holobionts: Insights into secondary metabolite pathways. *Journal of Fungi*, 9, 546. <https://doi.org/10.3390/JOF9050546>
- Amiri Moghaddam, J., Crüsemann, M., Alanjary, M., Harms, H., Dávila-Céspedes, A., Blom, J., Poehlein, A., Ziemert, N., König, G. M., & Schäberle, T. F. (2018). Analysis of the genome and metabolome of marine myxobacteria reveals high potential for biosynthesis of novel specialized metabolites. *Scientific Reports*, 8(8), 16600. <https://doi.org/10.1038/s41598-018-34954-y>
- Amnuaykanjanasin, A., Phonghanpot, S., Sengpanich, N., Cheevadhanarak, S., & Tanticharoen, M. (2009). Insect-specific polyketide synthases (PKSs), potential PKS-nonribosomal peptide synthetase hybrids, and novel PKS clades in tropical fungi. *Applied and Environmental Microbiology*, 75, 3721–3732. [https://doi.org/10.1128/AEM.02744-08/SUPPL\\_FILE/FILE\\_S1.DOC](https://doi.org/10.1128/AEM.02744-08/SUPPL_FILE/FILE_S1.DOC)
- Aschenbrenner, I. A., Cernava, T., Berg, G., & Grube, M. (2016). Understanding microbial multi-species symbioses. *Frontiers in Microbiology*, 7. <https://doi.org/10.3389/FMICB.2016.00180>
- Bağcı, C., Beier, S., Górska, A., & Huson, D. H. (2019). Introduction to the analysis of environmental sequences: Metagenomics with MEGAN. In *Methods in Molecular Biology* (pp. 591–604). Humana Press Inc. [https://doi.org/10.1007/978-1-4939-9074-0\\_19/FIGURES/10](https://doi.org/10.1007/978-1-4939-9074-0_19/FIGURES/10)
- Bağcı, C., Patz, S., & Huson, D. H. (2021). DIAMOND+MEGAN: Fast and easy taxonomic and functional analysis of short and long microbiome sequences. *Current Protocols*, 1, e59. <https://doi.org/10.1002/CPZ1.59>
- Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., De Wit, P., Sánchez-García, M., Ebersberger, I., de Sousa, F., Amend, A. S., Jumpponen, A., Unterseher, M., Kristiansson, E., Abarenkov, K., Bertrand, Y. J. K., Sanli, K., Eriksson, K. M., Vik, U., ... Nilsson, R. H. (2013). Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in Ecology and Evolution*, 4, 914–919. <https://doi.org/10.1111/2041-210X.12073>
- Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., Ingram, K. K., & Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, 22, 148–155. <https://doi.org/10.1016/j.tree.2006.11.004>
- Bickhart, D. M., Kolmogorov, M., Tseng, E., Portik, D. M., Korobeynikov, A., Tolstoganov, I., Uritskiy, G., Liachko, I., Sullivan, S. T., Shin, S. B., Zorea, A., Andreu, V. P., Panke-Buisse, K., Medema, M. H., Mizrahi, I., Pevzner, P. A., & Smith, T. P. L. (2022). Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities. *Nature Biotechnology*, 40, 711–719. <https://doi.org/10.1038/S41587-021-01130-Z>



- Bills, G. F., & Gloer, J. B. (2017). Biologically active secondary metabolites from the fungi. In J. Heitman, B. J. Howlett, P. W. Crous, E. H. Stukenbrock, T. Y. Stukenbrock, & A. R. Gow (Eds.), *The fungal kingdom* (pp. 1087–1119). ASM Press. <https://doi.org/10.1128/9781555819583.CH54>
- Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., van Wezel, G. P., Medema, M. H., & Weber, T. (2021). antiSMASH 6.0: Improving cluster detection and comparison capabilities. *Nucleic Acids Research*, 49, W29–W35. <https://doi.org/10.1093/NAR/GKAB335>
- Boustie, J., & Grube, M. (2005). Lichens—a promising source of bioactive secondary metabolites. *Plant Genetic Resources*, 3, 273–287. <https://doi.org/10.1079/PGR200572>
- Brodo, I. M., Duran Sharnoff, S., & Sharnoff, S. (2003). Lichens of North America. *International Microbiology*, 62(6), 149–150. <https://doi.org/10.1007/S10123-003-0124-1>
- Brown, D. W., Adams, T. H., & Keller, N. P. (1996). Aspergillus has distinct fatty acid synthases for primary and secondary metabolism. *Proceedings of the National Academy of Sciences*, 93, 14873–14877. <https://doi.org/10.1073/PNAS.93.25.14873>
- Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M., & Borodovsky, M. (2021). BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, 3, 1–11. <https://doi.org/10.1093/NARGAB/LQAA108>
- Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using diamond. *Nature Methods*, 12, 59–60. <https://doi.org/10.1038/nmeth.3176>
- Caicedo-Montoya, C., Manzo-Ruiz, M., & Ríos-Esteva, R. (2021). Pan-genome of the genus *Streptomyces* and prioritization of biosynthetic gene clusters with potential to produce antibiotic compounds. *Frontiers in Microbiology*, 12, 2786. <https://doi.org/10.3389/FMICB.2021.677558/BIBTEX>
- Calchera, A., Dal Grande, F., Bode, H. B., & Schmitt, I. (2019). Biosynthetic gene content of the 'perfume lichens' *Evernia prunastri* and *Pseudevernia furfuracea*. *Molecules*, 24(1):203. <https://doi.org/10.3390/molecules24010203>
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10, 421. <https://doi.org/10.1186/1471-2105-10-421/FIGURES/4>
- Candan, M., Yilmaz, M., Tay, T., Erdem, M., & Türk, A. O. (2007). Antimicrobial activity of extracts of the lichen *Parmelia sulcata* and its salazinic acid constituent. *Zeitschrift für Naturforschung C Journal of Biosciences*, 62, 619–621. <https://doi.org/10.1515/ZNC-2007-7-827/MACHINEREADABLECITATION/RIS>
- Cardile, V., Graziano, A. C. E., Avola, R., Piovano, M., & Russo, A. (2017). Potential anticancer activity of lichen secondary metabolite physodic acid. *Chemico-Biological Interactions*, 263, 36–45. <https://doi.org/10.1016/J.CBI.2016.12.007>
- Chen, L., Zhao, N., Cao, J., Liu, X., Xu, J., Ma, Y., Yu, Y., Zhang, X., Zhang, W., Guan, X., Yu, X., Liu, Z., Fan, Y., Wang, Y., Liang, F., Wang, D., Zhao, L., Song, M., & Wang, J. (2022). Short- and long-read metagenomes expand individualized structural variations in gut microbiomes. *Nature Communications*, 13, 1–12. <https://doi.org/10.1038/s41467-022-30857-9>
- Chiang, Y. M., Chang, S. L., Oakley, B. R., & Wang, C. C. C. (2011). Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms. *Current Opinion in Chemical Biology*, 15, 137–143. <https://doi.org/10.1016/J.CBPA.2010.10.011>
- Cox, R. J. (2023). Curiouser and curiouser: Progress in understanding the programming of iterative highly-reducing polyketide synthases. *Natural Product Reports*, 40, 9–27. <https://doi.org/10.1039/D2NP00007E>
- Cock, P. J. A., Chilton, J. M., Grüning, B., Johnson, J. E., & Soranzo, N. (2015). NCBI BLAST+ integrated into Galaxy. *GigaScience*, 4, 39. <https://doi.org/10.1186/S13742-015-0080-7/2707769>
- Crawford, J. M., & Townsend, C. A. (2010). New insights into the formation of fungal aromatic polyketides. *Nature Reviews Microbiology*, 8, 879–889. <https://doi.org/10.1038/nrmicro2465>
- Crawford, J. M., Korman, T. P., Labonte, J. W., Vagstad, A. L., Hill, E. A., Kamari-Bidkorpheh, O., Tsai, S.-C., & Townsend, C. A. (2009). Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization. *Nature*, 461, 1139–1143. <https://doi.org/10.1038/nature08475>
- Crespo, A., Bridge, P. D., & Hawksworth, D. L. (1997). Amplification of fungal rDNA-ITS regions from non-fertile specimens of the lichen-forming genus *Parmelia*. *Lichenol*, 29, 275–282. <https://doi.org/10.1006/LICH.1996.0071>
- Crespo, A., Bridge, P. D., Hawksworth, D. L., Grube, M., & Cubero, O. F. (1999). Comparison of rRNA genotype frequencies of *Parmelia sulcata* from long established and recolonizing sites following sulphur dioxide amelioration. *Plant Systematics and Evolution*, 217, 177–183. <https://doi.org/10.1007/BF00984363>
- Cuscó, A., Pérez, D., Viñes, J., Fàbregas, N., & Francino, O. (2021). Long-read metagenomics retrieves complete single-contig bacterial genomes from canine feces. *BMC Genomics*, 22, 1–15. <https://doi.org/10.1186/S12864-021-07607-0>
- Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, 14, 1394–1403. <https://doi.org/10.1101/GR.2289704>
- Devi, R., Kaur, T., Guleria, G., Lata Rana, K., Kour, D., Yadav, N., & Saxena, A. K. (2020). Fungal secondary metabolites and their biotechnological applications for human health. In A. A. Rastegari, A. N. Yadav, & N. Yadav (Eds.), *New and future developments in microbial biotechnology and bioengineering* (pp. 147–161). Elsevier. <https://doi.org/10.1016/B978-0-12-820528-0.00010-7>
- Dittmann, E., Gugger, M., Sivonen, K., & Fewer, D. P. (2015). Natural product biosynthetic diversity and comparative genomics of the Cyanobacteria. *Trends in Microbiology*, 23, 642–652. <https://doi.org/10.1016/J.TIM.2015.07.008>
- Divakar, P. K., Crespo, A., Wedin, M., Leavitt, S. D., Hawksworth, D. L., Myllys, L., McCune, B., Randlane, T., Bjerke, J. W., Ohmura, Y., Schmitt, I., Boluda, C. G., Alors, D., Roca-Valiente, B., Del-Prado, R., Ruibal, C., Buaruang, K., Núñez-Zapata, J., Amo de Paz, G., ... Lumbsch, H. T. (2015). Evolution of complex symbiotic relationships in a morphologically derived family of lichen-forming fungi. *New Phytologist*, 208, 1217–1226. <https://doi.org/10.1111/NPH.13553>
- 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/FIGURES/6>
- Duarte, N. (2022). Special issue on plant and marine-derived natural product research in drug discovery: Strengths and perspective. *Pharmaceuticals*, 15, 1249. <https://doi.org/10.3390/PH15101249>
- Elečko, J., Vilková, M., Frenák, R., Routray, D., Ručová, D., Bačkor, M., & Goga, M. (2022). A comparative study of isolated secondary metabolites from lichens and their antioxidative properties. *Plants*, 11, 1077. <https://doi.org/10.3390/PLANTS11081077/S1>
- Elix, J. A., & Stocker-Wörgötter, E. (2008). Biochemistry and secondary metabolites. *Lichen Biology*, 104–133.
- Elvebakk, A. (2011). A review of the genus *Hypogymnia* (Parmeliaceae) in Chile. *The Bryologist*, 114, 379–388. <https://doi.org/10.1639/0007-2745-1142379>
- Feuerer, T., & Thell, A. (2002). *Parmelia ernstiae* a new macrolichen from Germany. *Mitteilungen des Instituts für Allg Bot Hamburg*. 30–32, 49–60.

- Galloway, D. J. (2007). *Flora of New Zealand: lichens: Including lichen-forming and lichenicolous fun.* Manaaki Whenua Press.
- Gandhi, A. D., Umamahesh, K., Sathiyaraj, S., Suriyakala, G., Velmurugan, R., Al Farraj, D. A., Gawwad, M. R. A., Murugan, K., Babujanarthanam, R., & Saranya, R. (2022). Isolation of bioactive compounds from lichen *Parmelia sulcata* and evaluation of antimicrobial property. *Journal of Infection and Public Health*, 15, 491–497. <https://doi.org/10.1016/J.JIPH.2021.10.014>
- Gerasimova, J. V., Beck, A., Werth, S., & Resl, P. (2022). High diversity of type I polyketide genes in *Bacidia rubella* as revealed by the comparative analysis of 23 lichen genomes. *Journal of Fungi*, 8, 449. <https://doi.org/10.3390/JOF8050449>
- Goga, M., Elečko, J., Marcinčinová, M., Ručová, D., Bačkorová, M., & Backor, M. (2020). Lichen metabolites: An overview of some secondary metabolites and their biological potential. In J.-M. Mérillon, & K. G. Ramawat (Eds.), *Co-evolution of secondary metabolites*. Reference Series in Phytochemistry (pp. 175–209). Springer Cham. [https://doi.org/10.1007/978-3-319-96397-6\\_57/FIGURES/37](https://doi.org/10.1007/978-3-319-96397-6_57/FIGURES/37)
- Gregory, K., Salvador, L. A., Akbar, S., Adaikpoh, B. I., & Stevens, D. C. (2019). Survey of biosynthetic gene clusters from sequenced myxobacteria reveals unexplored biosynthetic potential. *Microorganisms*, 7, 181. <https://doi.org/10.3390/MICROORGANISMS7060181>
- Greshake Tzovaras, B., Segers, F. H. I. D., Bicker, A., Dal Grande, F., Otte, J., Anvar, S. Y., Hankeln, T., Schmitt, I., & Ebersberger, I. (2020). What is in *Umbilicaria pustulata*? A metagenomic approach to reconstruct the holo-genome of a lichen. *Genome Biology and Evolution*, 12, 309–324. <https://doi.org/10.1093/GBE/EVAA049>
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, 29, 1072–1075. <https://doi.org/10.1093/BIOINFORMATICS/BTT086>
- Hawksworth, D. L., & Grube, M. (2020). Lichens redefined as complex ecosystems. *New Phytol*, 227, 1281. <https://doi.org/10.1111/NPH.16630>
- Hawksworth, D. L., & Rossman, A. Y. (2007). Where are all the undescribed fungi? *Phytopathology*, 87, 888–891. <https://doi.org/10.1094/PHYTO1997879888>
- Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., & Stanke, M. (2016). BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*, 32, 767–769. <https://doi.org/10.1093/BIOINFORMATICS/BTV661>
- Hoff, K. J., Lomsadze, A., Borodovsky, M., & Stanke, M. (2019). Whole-genome annotation with BRAKER. *Methods in Molecular Biology*, 1962, 65–95. [https://doi.org/10.1007/978-1-4939-9173-0\\_5/FIGURES/5](https://doi.org/10.1007/978-1-4939-9173-0_5/FIGURES/5)
- Huneck, S., & Yoshimura, I. (1996). Identification of Lichen Substances. *Identify Lichen Subst*, 11–123. [https://doi.org/10.1007/978-3-642-85243-5\\_2](https://doi.org/10.1007/978-3-642-85243-5_2)
- Huson, D. H., Albrecht, B., Bağcı, C., Bessarab, I., Górska, A., Jolic, D., & Williams, R. B. H. (2018). MEGAN-LR: New algorithms allow accurate binning and easy interactive exploration of metagenomic long reads and contigs. *Biology Direct*, 13, 6. <https://doi.org/10.1186/S13062-018-0208-7/FIGURES/8>
- Huitt-Roehl, C. R., Hill, E. A., Adams, M. M., Vagstad, A. L., Li, J. W., & Townsend, C. A. (2015). Starter unit flexibility for engineered product synthesis by the nonreducing polyketide synthase PksA. *ACS Chemical Biology*, 10, 1443–1449. <https://doi.org/10.1021/acschembio.5b00005>
- Jenke-Kodama, H., Sandmann, A., Müller, R., & Dittmann, E. (2005). Evolutionary implications of bacterial polyketide synthases. *Molecular Biology and Evolution*, 22, 2027–2039. <https://doi.org/10.1093/MOLBEV/MSI193>
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, 30, 1236–1240. <https://doi.org/10.1093/BIOINFORMATICS/BTU031>
- Joshi, T., Sharma, P., Joshi, T., & Chandra, S. (2019). In silico screening of anti-inflammatory compounds from Lichen by targeting cyclooxygenase-2. *Journal of Biomolecular Structure and Dynamics*, 38(12), 3544–3562. <https://doi.org/10.1080/07391102.2019.1664328>
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermini, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(14), 587–589. <https://doi.org/10.1038/nmeth.4285>
- Karabulut, G., & Ozturk, S. (2015). Antifungal activity of evernia prunastri, parmelia sulcata, pseudovernia furfuracea var. furfuracea. *Pakistan Journal of Botany*, 47, 1575–1579.
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30, 772–780. <https://doi.org/10.1093/MOLBEV/MST010>
- Kealey, J. T., Craig, J. P., & Barr, P. J. (2021). Identification of a lichen depside polyketide synthase gene by heterologous expression in *Saccharomyces cerevisiae*. *Metabolic Engineering Communications*, 13, e00172. <https://doi.org/10.1016/J.MEC.2021.E00172>
- Keller, N. P. (2019). Fungal secondary metabolism: Regulation, function and drug discovery. *Nature Reviews Microbiology*, 17(3), 167–180. <https://doi.org/10.1038/s41579-018-0121-1>
- Keller, N. P., Turner, G., & Bennett, J. W. (2005). Fungal secondary metabolism — from biochemistry to genomics. *Nature Reviews Microbiology*, 3, 937–947. <https://doi.org/10.1038/nrmicro1286>
- Kim, W., Liu, R., Woo, S., Kang, K. B., Park, H., Yu, Y. H., Ha, H. H., Oh, S. Y., Yang, J. H., Kim, H., Yun, S. H., & Hur, J. S. (2021). Linking a gene cluster to atranorin, a major cortical substance of lichens, through genetic dereplication and heterologous expression. *mBio*, 12, e0111121. [https://doi.org/10.1128/MBIO.01111-21/SUPPL\\_FILE/MBIO.01111-21-ST002.DOCX](https://doi.org/10.1128/MBIO.01111-21/SUPPL_FILE/MBIO.01111-21-ST002.DOCX)
- Kolmogorov, M., Bickhart, D. M., Behasz, B., Gurevich, A., Rayko, M., Shin, S. B., Kuhn, K., Yuan, J., Pevnikov, E., Smith, T. P. L., & Pevzner, P. A. (2020). metaFlye: Scalable long-read metagenome assembly using repeat graphs. *Nature Methods*, 17, 1103–1110. <https://doi.org/10.1038/s41592-020-00971-x>
- Kosanić, M., Manojlović, N., Janković, S., Stanojković, T., & Ranković, B. (2013). Evernia prunastri and Pseudovernia furfuracea lichens and their major metabolites as antioxidant, antimicrobial and anticancer agents. *Food and Chemical Toxicology*, 53, 112–118. <https://doi.org/10.1016/J.FCT.2012.11.034>
- Kosanić, M., & Ranković, B. (2019). Studies on antioxidant properties of lichen secondary metabolites. *Lichen Secondary Metabolism*, 129–153. [https://doi.org/10.1007/978-3-030-16814-8\\_4](https://doi.org/10.1007/978-3-030-16814-8_4)
- Kroken, S., Glass, N. L., Taylor, J. W., Yoder, O. C., & Turgeon, B. G. (2003). Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proceedings of the National Academy of Sciences*, 100, 15670–15675. [https://doi.org/10.1073/PNAS.2532165100/SUPPL\\_FILE/2165TABLE2.RTF](https://doi.org/10.1073/PNAS.2532165100/SUPPL_FILE/2165TABLE2.RTF)
- Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, 49, W293–W296. <https://doi.org/10.1093/NAR/GKAB301>
- Li, Y., Xu, W., & Tang, Y. (2010). Classification, prediction, and verification of the regioselectivity of fungal polyketide synthase product template domains. *Journal of Biological Chemistry*, 285, 22764–22773. <https://doi.org/10.1074/JBC.M110.128504>
- Liu, Z., Zhao, Y., Huang, C., & Luo, Y. (2021). Recent advances in silent gene cluster activation in *Streptomyces*. *Frontiers in Bioengineering and Biotechnology*, 9, 88. <https://doi.org/10.3389/FBIOE.2021.632230/BIBTEX>

- Llewellyn, T., Nowell, R. W., Aptroot, A., Temina, M., Prescott, T. A. K., Barraclough, T. G., & Gaya, E. (2023). Metagenomics shines light on the evolution of 'sunscreen' pigment metabolism in the teloschistales (Lichen-Forming Ascomycota). *Genome Biology and Evolution*. <https://doi.org/10.1093/GBE/EVAD002>
- Männle, D., McKinnie, S. M. K., Mantri, S. S., Steinke, K., Lu, Z., Moore, B. S., Ziemert, N., & Kayser, L. (2020). Comparative genomics and metabolomics in the genus *Nocardia*. *mSystems*, 5, e00125-20. <https://doi.org/10.1128/MSYSTEMS.00125-20>
- Martinet, L., Naomé, A., Deflandre, B., Maciejewska, M., Tellatin, D., Tenconi, E., Smargiasso, N., de Pauw, E., van Wezel, G. P., & Rigali, S. (2019). A single biosynthetic gene cluster is responsible for the production of bagremycin antibiotics and ferroverdin iron chelators. *MBio*, 10. <https://doi.org/10.1128/mBio.01230-19>
- Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, B. J., Blin, K., Bruijn, D. I., Chooi, Y. H., Claesen, J., Coates, C. R., Cruz-Morales, P., Duddela, S., Düsterhus, S., Edwards, J. D., Fewer, P. D., Garg, M., Geiger, C., Gomez-Escribano, P. J., Greule, A., Hadjithomas, M., ... Glöckner Oliver, F. (2015). Minimum information about a biosynthetic gene cluster. *Nature Chemical Biology*, 11, 625–631. <https://doi.org/10.1038/nchembio.1890>
- Meier, J. L., & Burkart, M. D. (2011). Proteomic analysis of polyketide and nonribosomal peptide biosynthesis. *Current Opinion in Chemical Biology*, 15, 48–56. <https://doi.org/10.1016/j.cbpa.2010.10.021>
- Minh, B. Q., Nguyen, M. A. T., & Von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, 30, 1188–1195. <https://doi.org/10.1093/MOLBEV/MST024>
- Molina, M. C., Divakar, P. K., Millanes, A. M., Sánchez, E., Del-Prado, R., Hawksworth, D. L., & Crespo, A. (2011). *Parmelia sulcata* (Ascomycota: Parmeliaceae), a sympatric monophyletic species complex. *The Lichenologist*, 43, 585–601. <https://doi.org/10.1017/S0024282911000521>
- Molnár, K., & Farkas, E. (2011). Depsides and depsidones in populations of the lichen *Hypogymnia physodes* and its genetic diversity. *Annales Botanici Fennici*, 48, 473–482. <https://doi.org/10.5735/0850480605>
- Mosunova, O. V., Navarro-Muñoz, J. C., Haksar, D., van Neer, J., Hoeksma, J., den Hertog, J., & Collemare, J. (2022). Evolution-Informed discovery of the naphthalenone biosynthetic pathway in fungi. *mBio*, 13, e0022322. [https://doi.org/10.1128/MBIO.00223-22-SUPPL\\_FILE/MBIO.00223-22-S0010.XLSX](https://doi.org/10.1128/MBIO.00223-22/SUPPL_FILE/MBIO.00223-22-S0010.XLSX)
- Muggia, L., & Grube, M. (2018). Fungal diversity in lichens: From extremotolerance to interactions with algae. *Life*, 8(2), 15. <https://doi.org/10.3390/LIFE8020015>
- Muggia, L., Zellnig, G., Rabensteiner, J., & Grube, M. (2010). Morphological and phylogenetic study of algal partners associated with the lichen-forming fungus *Tephromela atra* from the Mediterranean region. *Symbiosis*, 51, 149–160. <https://doi.org/10.1007/s13199-010-0060-8>
- Navarro-Muñoz, J. C., Selem-Mojica, N., Mallowney, 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. (2019). A computational framework to explore large-scale biosynthetic diversity. *Nature Chemical Biology*, 16, 60–68. <https://doi.org/10.1038/s41589-019-0400-9>
- Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32, 268–274. <https://doi.org/10.1093/MOLBEV/MSU300>
- Ola, A. R. B., Thomy, D., Lai, D., Brötz-Oesterhelt, H., & Proksch, P. (2013). Inducing secondary metabolite production by the endophytic fungus *Fusarium tricinctum* through coculture with *Bacillus subtilis*. *Journal of Natural Products*, 76, 2094–2099. [https://doi.org/10.1021/NP400589H/SUPPL\\_FILE/NP400589H\\_SI\\_001.PDF](https://doi.org/10.1021/NP400589H/SUPPL_FILE/NP400589H_SI_001.PDF)
- Pizarro, D., Divakar, P. K., Grewe, F., Crespo, A., Dal Grande, F., & Thorsten Lumbsch, H. (2020). Genome-wide analysis of biosynthetic gene cluster reveals correlated gene loss with absence of usnic acid in lichen-forming fungi. *Genome Biology and Evolution*, 12, 1858–1868. <https://doi.org/10.1093/GBE/EVAA189>
- Plitt, C. C. (1919). A short history of lichenology. *Bryologist*, 22, 77. <https://doi.org/10.2307/3238526>
- Popin, R. V., Alvarenga, D. O., Castelo-Branco, R., Fewer, D. P., & Sivonen, K. (2021). Mining of cyanobacterial genomes indicates natural product biosynthetic gene clusters located in conjugative plasmids. *Frontiers in Microbiology*, 12, 3353. <https://doi.org/10.3389/FMICB.2021.684565/BIBTEX>
- Purvis, O. W., Coppins, B. J., Hawksworth, D., & James, P. W. (1992). *The lichen flora of Great Britain and Ireland*. 710. Natural History Museum.
- Ranković, B., & Kosanić, M. (2019). Lichens as a potential source of bioactive secondary metabolites. In B. Ranković (Ed.), *Lichen secondary metabolites* (pp. 1–29). Springer International Publishing. [https://doi.org/10.1007/978-3-030-16814-8\\_1](https://doi.org/10.1007/978-3-030-16814-8_1)
- Ranković, B., Kosanić, M., Manojlović, N., Rančić, A., & Stanojković, T. (2014). Chemical composition of *Hypogymnia physodes* lichen and biological activities of some its major metabolites. *Medicinal Chemistry Research*, 23, 408–416. <https://doi.org/10.1007/s00044-013-0644-y>
- Ranković, B., Mišić, M., & Sukdolak, S. (2007). Evaluation of antimicrobial activity of the lichens *Lasallia pustulata*, *Parmelia sulcata*, *Umbilicaria crustulosa*, and *Umbilicaria cylindrica*. *Microbiology*, 76, 723–727. <https://doi.org/10.1134/S0026261707060112/METRICS>
- Ringel, M., Reinbold, M., Hirte, M., Haack, M., Huber, C., Eisenreich, W., Masri, M. A., Schenk, G., Guddat, L. W., Loll, B., Kerr, R., Garbe, D., & Brück, T. (2020). Towards a sustainable generation of pseudopterosin-type bioactives. *Green Chemistry*, 22, 6033–6046. <https://doi.org/10.1039/D0GC01697G>
- Ristić, S., Ranković, B., Kosanić, M., Stanojković, T., Stamenković, S., Vasiljević, P., Manojlović, I., & Manojlović, N. (2016). Phytochemical study and antioxidant, antimicrobial and anticancer activities of *Melanelia subaurifera* and *Melanelia fuliginosa* lichens. *Journal of Food Science and Technology*, 53, 2804–2816. <https://doi.org/10.1007/s13197-016-2255-3>
- Ritz, M., Ahmad, N., Brueck, T., & Mehler, N. (2023). Comparative Genome-Wide analysis of two caryopteris x clandonensis cultivars: Insights on the biosynthesis of volatile terpenoids. *Plants*, 12, 632. <https://doi.org/10.3390/PLANTS12030632/S1>
- Rokas, A., Wisecaver, J. H., & Lind, A. L. (2018). The birth, evolution and death of metabolic gene clusters in fungi. *Nature Reviews Microbiology*, 16, 731–744. <https://doi.org/10.1038/s41579-018-0075-3>
- Rolshausen, G., Dal Grande, F., Otte, J., & Schmitt, I. (2022). Lichen holobionts show compositional structure along elevation. *Molecular Ecology*, 1–12. <https://doi.org/10.1111/MEC.16471>
- Sánchez-Navarro, R., Nuhamunada, M., Mohite, O. S., Wasmund, K., Albertsen, M., Gram, L., Nielsen, P. H., Weber, T., & Singleton, C. M. (2022). Long-read metagenome-assembled genomes improve identification of novel complete biosynthetic gene clusters in a complex microbial activated sludge ecosystem. *mSystems*, 7, e0063222. <https://doi.org/10.1128/MSYSTEMS.00632-22>
- Schroeckh, V., Scherlach, K., Nützmann, H. W., Shelest, E., Schmidt-Heck, W., Schuemann, J., Martin, K., Hertweck, C., & Brakhage, A. A. (2009). Intimate bacterial-fungal interaction triggers biosynthesis of archetypal polyketides in *Aspergillus nidulans*. *Proceedings of the National Academy of Sciences*, 106, 14558–14563. [https://doi.org/10.1073/PNAS.0901870106/SUPPL\\_FILE/0901870106SI.PDF](https://doi.org/10.1073/PNAS.0901870106/SUPPL_FILE/0901870106SI.PDF)
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A

- software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13, 2498–2504. <https://doi.org/10.1101/GR.1239303>
- Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE*, 11, e0163962. <https://doi.org/10.1371/JOURNAL.PONE.0163962>
- Shwab, E. K., & Keller, N. P. (2008). Regulation of secondary metabolite production in filamentous ascomycetes. *Mycological Research*, 112, 225–230. <https://doi.org/10.1016/J.MYCRES.2007.08.021>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31, 3210–3212. <https://doi.org/10.1093/BIOINFORMATICS/BTV351>
- Singh, G. (2023). Linking Lichen metabolites to genes: Emerging concepts and lessons from molecular biology and metagenomics. *Journal of Fungi*, 9, 160. <https://doi.org/10.3390/JOF9020160>
- Singh, G., Armaleo, D., Dal Grande, F., & Schmitt, I. (2021a). Depside and depsidone synthesis in lichenized fungi comes into focus through a genome-wide comparison of the olivetoric acid and physodic acid chemotypes of pseudovernia furfuracea. *Biomolecules*, 11, 1445. <https://doi.org/10.3390/BIOM11101445/S1>
- Singh, G., Calchera, A., Merges, D., Valim, H., Otte, J., Schmitt, I., & Dal Grande, F. (2022). A candidate gene cluster for the bioactive natural product gyrophoric acid in lichen-forming fungi. *Microbiology Spectrum*, 10, e0010922. [https://doi.org/10.1128/SPECTRUM.00109-22/SUPPL\\_FILE/REVIEWER-COMMENTS.PDF](https://doi.org/10.1128/SPECTRUM.00109-22/SUPPL_FILE/REVIEWER-COMMENTS.PDF)
- Singh, G., Calchera, A., Schulz, M., Drechsler, M., Bode, H. B., Schmitt, I., & Grande, F. D. (2021b). Climate-specific biosynthetic gene clusters in populations of a lichen-forming fungus. *Environmental Microbiology*, 23(8), 4260–4275. <https://doi.org/10.1111/1462-2920.15605>
- Smith, H. B., Dal Grande, F., Muggia, L., Keuler, R., Divakar, P. K., Grewe, F., Schmitt, I., Thorsten Lumbsch, H., & Leavitt, S. D. (2020). Metagenomic data reveal diverse fungal and algal communities associated with the lichen symbiosis. *Symbiosis* 82, 133–147. <https://doi.org/10.1007/s13199-020-00699-4>
- Sisodia, R., Geol, M., Verma, S., Rani, A., & Dureja, P. (2013). Antibacterial and antioxidant activity of lichen species *Ramalina roesleri*. *Natural Product Research*, 27, 2235–2239. <https://doi.org/10.1080/14786419.2013.811410>
- Solárová, Z., Liskova, A., Samec, M., & Kubatka, P. (2020). Anticancer potential of lichens' secondary metabolites. *Biomolecules*, 10(1), 87. <https://doi.org/10.3390/BIOM10010087>
- Solhaug, K. A., Lind, M., Nybakken, L., & Gauslaa, Y. (2009). Possible functional roles of cortical depsides and medullary depsidones in the foliose lichen *Hypogymnia physodes*. *Flora - Morphology, Distribution, Functional Ecology of Plants*, 204, 40–48. <https://doi.org/10.1016/J.FLORA.2007.12.002>
- Spribile, T., Tuovinen, V., Resl, P., VandEerpool, D., Wolinski, H., Aime, M. C., Schneider, K., Stabentheiner, E., Toome-Heller, M., Thor, G., Mayrhofer, H., Johannesson, H., & Mccutcheon, J. P. (2016). Basidiomycete yeasts in the cortex of ascomycete macrolichens. *Science*, 353, 488. <https://doi.org/10.1126/SCIENCE.AAF8287>
- Stajich, J. E. (2017). Fungal genomes and insights into the evolution of the kingdom. *Microbiology Spectrum*, 5. <https://doi.org/10.1128/MICROBIOLSPEC.FUNK-0055-2016>
- Stojanović, G., Zlatanović, I., Zrnzević, I., Stanković, M., Stankov Jovanović, V., & Zlatković, B. (2018). *hypogymnia tubulosa* extracts: Chemical profile and biological activities. *Natural Product Research*, 32, 2735–2739. <https://doi.org/10.1080/1478641920171375926>
- Sullivan, M. J., Petty, N. K., & Beatson, S. A. (2011). Easyfig: A genome comparison visualizer. *Bioinformatics*, 27, 1009–1010. <https://doi.org/10.1093/BIOINFORMATICS/BTR039>
- Sun, Y., Chen, B., Li, X., Yin, Y., & Wang, C. (2022). Orchestrated biosynthesis of the secondary metabolite cocktails enables the producing fungus to combat diverse bacteria. *mBio*, 13, e0180022. [https://doi.org/10.1128/MBIO.01800-22/SUPPL\\_FILE/MBIO.01800-22-S0010.XLSX](https://doi.org/10.1128/MBIO.01800-22/SUPPL_FILE/MBIO.01800-22-S0010.XLSX)
- 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., ... 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>
- Theobald, S., Vesth, T. C., Rendsvig, J. K., Nielsen, K. F., Riley, R., de Abreu, L. M., Salamov, A., Frisvad, J. C., Larsen, T. O., Mikael Rørdam Andersen, M. R., & Hoof, J. B. (2018). Uncovering secondary metabolite evolution and biosynthesis using gene cluster networks and genetic dereplication. *Scientific Reports*, 8(18), 1–12. <https://doi.org/10.1038/s41598-018-36561-3>
- Tsai, Y.-C., Conlan, S., Deming, C., Segre, J. A., Kong, H. H., Korch, J., Ohl, J., & NISC Comparative Sequencing Program. (2016). Resolving the complexity of human skin metagenomes using single-molecule sequencing. *mBio*, 7. <https://doi.org/10.1128/mBio.01948-15>
- Wang, W., Drott, M., Greco, C., Luciano-Rosario, D., Wang, P., & Keller, N. P. (2021). Transcription factor repurposing offers insights into the evolution of biosynthetic gene cluster regulation. *mBio*, 12, e0139921. <https://doi.org/10.1128/MBIO.01399-21>
- Wasil, Z., Pahirulzaman, K. A. K., Butts, C., Simpson, T. J., Lazarus, C. M., & Cox, R. J. (2013). One pathway, many compounds: Heterologous expression of a fungal biosynthetic pathway reveals its intrinsic potential for diversity. *Chemical Science*, 4, 3845–3856. <https://doi.org/10.1039/C3SC51785C>
- Watanabe, C. M. H., & Townsend, C. A. (2002). Initial characterization of a type I fatty acid synthase and polyketide synthase multienzyme complex NorS in the biosynthesis of aflatoxin B1. *Chemistry & Biology*, 9, 981–988. [https://doi.org/10.1016/S1074-5521\(02\)00213-2](https://doi.org/10.1016/S1074-5521(02)00213-2)
- Xie, H., Yang, C., Sun, Y., Igarashi, Y., Jin, T., & Luo, F. (2020). PacBio long reads improve metagenomic assemblies, gene catalogs, and genome binning. *Frontiers in Genetics*, 11, 1077. <https://doi.org/10.3389/fgene.2020.516269>
- Zheng, W., Wang, X., Zhou, H., Zhang, Y., Li, A., & Bian, X. (2020). Establishment of recombinering genome editing system in *Paraburkholderia megapolitana* empowers activation of silent biosynthetic gene clusters. *Microbial Biotechnology*, 13, 397–405. <https://doi.org/10.1111/1751-7915.13535>
- Ziemert, N., & Jensen, P. R. (2012). Phylogenetic approaches to natural product structure prediction. *Methods in Enzymology*, 517, 161–182. <https://doi.org/10.1016/B978-0-12-404634-4.00008-5>

**How to cite this article:** Ahmad, N., Ritz, M., Calchera, A., Otte, J., Schmitt, I., Brueck, T., & Mehlmer, N. (2023). Biosynthetic gene cluster synteny: Orthologous polyketide synthases in *Hypogymnia physodes*, *Hypogymnia tubulosa*, and *Parmelia sulcata*. *MicrobiologyOpen*, 12, e1386. <https://doi.org/10.1002/mbo3.1386>

## 4 Discussion and Outlook

### Lichen Natural Products in Drug Discovery

In a pharmaceutical context, natural products are chemical compounds that arise from metabolic processes within living organisms. These products can be broadly categorized as either primary or secondary metabolites, although the latter term is commonly used to refer to natural products [39]–[41]. Primary metabolites play essential roles in the fundamental growth and development of an organism [66], whereas secondary metabolites are not directly involved in its basic survival or reproductive processes [193], [194]. Instead, secondary metabolites are often produced in response to environmental stresses, such as defence against predators, protection against harmful ultraviolet radiation, sequestration of limited mineral ions, or attraction of insects for pollination. Additionally, secondary metabolites can arise as byproducts of primary metabolism without a clearly defined purpose or function [39], [42], [195].

Natural products provide numerous advantages. Firstly, they exhibit unparalleled structural and chemical diversity, often characterized by a higher number of chiral centres, entropically favourable rigid structures with greater abundance of ring structures compared to synthetic compounds. The latter tend to be flat in general and lack stereocentres. Additionally, natural products occupy a significantly larger molecular space in terms of chemical properties compared to compounds generated through combinatorial chemistry [196]. Interestingly, many natural products defy the conventional Lipinski's rule of five for drug-like properties, yet still possess potent oral activity as drugs [197]. Secondly, the majority of secondary metabolites have evolved to fulfil specific biological functions. Each bond and stereochemical arrangement within their structures is meticulously crafted by the organism to facilitate optimal binding to their intended targets and trigger a desired biological response. Moreover, due to the adaptable nature of these compounds, they may also exhibit activity against alternative biological targets, as they possess the essential pharmacophore required for interaction. Lastly, secondary metabolites are shaped by evolutionary processes to suit the lichens needs in response to changing climates and ecological conditions. This implies the existence of novel molecular scaffolds that can effectively address the challenges posed by evolving conditions [198].

Lichen cells harbour a plethora of diverse natural metabolites and bioactive compounds, which have garnered heightened interest owing to their potential applications in various industries, including pharmaceuticals, biotechnology, medicine, cosmetics, and other related fields [199]. The predominant categories consist of depsides, depsidones, dibenzofurans,

and phenolic compounds. Additionally, there are observations of aliphatic acids, triterpenes, anthraquinones, secalonic acids, pulvinic acid derivatives, and xanthenes [27], [49], [60]. Numerous investigations have provided evidence for the capacity of distinct lichen species to synthesize diverse natural compounds exhibiting various physicochemical and biological properties [200]. In this regard, the investigated lichens *Hypogymnia physodes* (HPH), *Hypogymnia tubulosa* (HTU) and *Parmelia sulcata* (PSU) harbour compounds e.g. atranorin, physodic acid and lecanoric acid which have proven biological activities [70], [71], [78]. For atranorin and lecanoric acid orthologous BGCs were identified in Ahmad et al [201] throughout all evaluated lichen metagenomes. Other putatively produced compounds are orsellinic acid derivatives, lovastatin-like structures, melanins and mellein, for which also orthologous BGCs were assigned.

Regarding drug discovery, the natural product spectrum of lichens remains an intriguing pool of mostly yet untapped compounds. However, many promising drug candidates have been derived from and influenced by natural sources [193], [194], [202]–[204]. In general, a mere 24.6% of all recently approved drugs (1981–2019) have been entirely synthesized [205]. Natural products serve as a valuable reservoir of compounds and form the foundation of the majority of the existing drugs [44], [45]. Moreover, natural products provide a broader coverage of the biologically relevant chemical space compared to synthetic compounds [45].

The findings in the included publications by Ahmad et al entitled “Biosynthetic potential of *Hypogymnia* holobionts - Insights into Secondary Metabolite Pathways” [186] and “Biosynthetic Gene Cluster Synteny - Orthologous Polyketide Synthases in *Hypogymnia physodes*, *Hypogymnia tubulosa* and *Parmelia sulcata*” [201] highlight, that the majority of annotated BGCs of the mycobionts were not or only poorly assigned with any function by antiSMASH. Table 3 summarizes these results by their annotation score into three categories. All BGCs with similarities above 90% are deemed to exhibit the inferred function. In this comparison it becomes apparent that only a small fraction of these lichen BGCs is truly connectable to already known clusters, rendering the vast majority cryptic. The poorly annotated clusters exhibit partial similarities to reference sequences, however no certain function may be derived from these, even if an assigned function is allocated.

Table 3: BGCs present in investigated lichen samples. Percentage shows annotation score based on similarity by antiSMASH. Annotation values above 90% are deemed to be valid. After each lichen sample the total of present BGCs is mentioned to allow for direct comparison.

Lichen species	> 90%	< 90%	0%
HTU (n=73)	5	11	57
HPH (n=114)	3	24	87
PSU (n=214)	10	46	158

## *Discussion and Outlook*

With the rise of multidrug-resistant pathogens and the growing resistance to conventional antibiotics, there is a pressing demand for the exploration of new compounds that exhibit efficacy against these challenges [206], [207]. Other fields of application are e.g. agriculture [208], food industry [209] and cosmetics [210]. Due to their inherent complexity, natural products and their derivatives often demonstrate enhanced effectiveness in combating the elevating antibiotic resistance [211], [212]. Given that less than 10% of the Earth's biodiversity has undergone investigation for biological activity, natural sources continue to hold immense promise for the discovery of new lead compounds [213]–[216]. In the future, the emergence of novel natural product frameworks suitable for pharmaceutical development is anticipated to arise from genomic sources [205]. As mentioned before, lichens do harbour compounds with e.g., antimicrobial or antifungal activity, rendering these as intriguing targets for experimental investigation. In regard to flavour compounds, several BGCs related to terpene production were identified by Ahmad et al [186], which may be of interest. Subsequently, heterologous expression of promising candidates will give evidence on product formation and further enzyme engineering approaches may yield new compounds with yet unknown function.

In the presented studies by Ahmad et al [186], [201] various BGCs were made accessible by genome mining, not exclusively from the mycobiont but also from the remaining lichen community, allowing for the investigation of new compounds yet uncharacterized in their function. Among these are the orthologous PKS related BGCs with various inferred functions across the three investigated lichen samples. In light of the BigPharm Project, olivetolic acid (OA) was the compound of interest produced by all three lichens. Derived from this, a closer look into the identified PKS clades pinpoints the putative BGCs to produce OA as clades 2 and 3. Here, orsellinic acid derivatives are inferred as putatively produced compounds. Subsequently, PKS genes or entire BGCs from these clades should be utilized in heterologous expressions to investigate the respective product spectra.

### **Genome Reconstruction of Lichen**

The rich and biologically active secondary metabolites found in lichens offer promising prospects for the discovery of new natural product leads, particularly in the face of increasing drug resistance [58], [204], [211], [217]–[219]. However, the utilization of lichens in biotechnological applications is currently limited due to three main challenges: a) the intricate symbiotic nature of lichens makes experimental strategies difficult or unfeasible, b) the slow growth rates and c) the scarcity of genetic information pertaining to the mycobiont responsible for synthesizing the desired lichen compounds [52], [81], [96]. The advent of

high-throughput sequencing of metagenomic samples presents a promising approach for reconstructing genomes of symbionts that are challenging to cultivate or cannot be cultivated at all [149], [160]. However, metagenomic approaches pose significant analytical challenges, primarily revolving around two key aspects: a) assembly strategy, which involves finding a suitable approach to handle extreme coverage biases or uneven genome distributions, and b) taxonomic binning, which necessitates the application of an appropriate assignment method to extract the relevant genetic information of interest accurately and without contamination from other co-occurring organisms in the DNA mixture [95], [101], [102]. Through long-read sequencing approaches the previously described limitations of short-read sequencing can be overcome, as it allows for the accurate resolution of repetitive and complex genome regions such as segmental duplications [88], [90]. Also the sequencing of entire lichen thalli is feasible, without the need for axenic cultivation of separated symbionts [112]. Obtained HiFi reads were consecutively assembled in this thesis by metaFlye, addressing the aforementioned challenges of assembling metagenomes from a lichen community. With the resulting high-quality metagenomes improved binning confidence may be achieved, as more contiguous DNA fragments were sequenced, circumventing loss of information crucial to the taxonomic binning approach [137]. Subsequent functional annotation was achieved by deploying antiSMASH, rendering predicted BGCs accessible for further investigation. In both presented studies by Ahmad et.al entitled “Biosynthetic potential of *Hypogymnia* holobionts - Insights into Secondary Metabolite Pathways” [186] and “Biosynthetic Gene Cluster Synteny - Orthologous Polyketide Synthases in *Hypogymnia physodes*, *Hypogymnia tubulosa* and *Parmelia sulcata*” [201] the genomes of the complete holobionts and in more detail the mycobionts are elaborated. All genomes were reconstructed from metagenomic sequenced lichen thalli using bioinformatic tools for assembly of highly non-uniform metagenomes and taxonomic binning. Obtained genomes were assessed thoroughly and deemed valid, allowing for genome mining of the harboured BGCs. The most prominent class of BGCs were those related to polyketide formation, which were further elaborated on phylogenetic and syntenic level in Ahmad et al entitled “Biosynthetic Gene Cluster Synteny - Orthologous Polyketide Synthases in *Hypogymnia physodes*, *Hypogymnia tubulosa* and *Parmelia sulcata*” [201].



### Outlook - Regulation and Activation of Silent Biosynthetic Gene Clusters

The primary objective of this thesis is to gain a comprehensive understanding of the biosynthetic gene repertoire within lichen and especially their mycobiont, with the prospective of establishing connections between genes and secondary metabolites. This research aims to lay the groundwork for future biotechnological approaches that leverage the potential of lichen secondary metabolism. Additionally, exploring the biosynthetic genes present in lichenized fungi can provide valuable insights into the evolution of these crucial gene families. The existence of a substantial number of cryptic or silent gene clusters, in comparison to the known repertoire of metabolites, suggests the existence of numerous undiscovered compounds that have yet to be identified [160]. These unknown metabolites may play vital roles in maintaining the lichen symbiosis or facilitating communication between symbiotic partners. They may also be specifically relevant during certain life stages, such as the search for compatible symbiont partners in the environment or during thallus formation [52]. However, accessing and comprehending the functions of these lichen-derived metabolites remains a challenge at present.

To decipher the secondary metabolite formation in lichen, the underlying regulatory mechanisms need to be understood. Especially if the identified BGCs of this thesis shall be used for heterologous expression experiments in the future. In the following possible targets and hints are provided to allow for more focused approaches in investigating the product spectra of lichen BGCs. In Ahmad et al “Biosynthetic Gene Cluster Synteny - Orthologous Polyketide Synthases in *Hypogymnia physodes*, *Hypogymnia tubulosa* and *Parmelia sulcata*” [201] various BGCs were identified as orthologous throughout all investigated lichens. Additionally, the BGCs were assigned with putatively produced compounds, which in case of e.g., orsellinic acid or olivetolic acid were observed in the GC analysis.

The regulation of secondary metabolism is well-established in certain organisms, particularly filamentous fungi belonging to the Ascomycota phylum [220]. Within BGCs, the genes are co-regulated, meaning, that they are either simultaneously activated or repressed. Many BGCs contain a specific transcription factor gene that is crucial for the expression of the entire cluster. Common types of transcription factors found in BGCs include zinc binuclear proteins and Cys2His2 zinc finger proteins [220]. However, alternative types of transcription factors, such as bANK transcription factors characterized by a bZIP basic region and ankyrin repeats, may also be present in some cases [221]–[225].

Aside from pathway-specific regulators, some BGCs are globally regulated in response to environmental cues. Global regulators of secondary metabolism can upregulate certain clusters, while downregulating others. Examples of such regulators include AreA, which

responds to nitrogen sources, PacC [226], which modulates BGC expression in response to pH fluctuations, and CreA, which responds to different carbon sources [227]. Furthermore, the regulation of secondary metabolism is interconnected with fungal development. For instance, filamentous fungi synthesize mycotoxins at the commencement of sporulation, which exhibit toxic effects on vertebrates. In certain cases, secondary metabolites serve as pigments, that protect spores [228]. By assessing the antiSMASH results, it becomes apparent, that there are indeed BGCs present related to e.g., melanin or toxin production (please refer to Ahmad et al [201]). The deletion or disruption of key regulatory genes leads to the simultaneous suppression of secondary metabolite production and fungal development. Examples of this include the impairment of asexual sporulation and mycotoxin production [229], as well as conidiation and melanin expression [230]. In ascomycetes, certain global regulators that modulate secondary metabolism through chromatin-remodelling mechanisms have been characterized. One notable regulator is LaeA, initially discovered in *Aspergillus nidulans* but subsequently identified in other *Aspergillus* species as well. Deletion of LaeA leads to a reduction in secondary metabolite production through transcriptional repression, while its overexpression enhances the biosynthesis of specific natural products like penicillin and lovastatin [231]. The coordination of secondary metabolism with fungal development is exemplified by the VelB/VelA/LaeA complex, also known as the velvet complex. This complex plays a crucial role in integrating environmental cues, particularly light signals, to regulate secondary metabolism in fungi [227], [232]–[235]. In Ahmad et al [201] BGCs putatively related to production of lovastatin-like compounds were identified (clades 6-8), making these findings intriguing for future wet-lab experiments.

Engaging the activation of silent BGCs, various methods have been developed, such as OSMAC (one strain, many compounds) and co-cultivation. This is especially of interest in regard to experiments with separated lichen mycobionts for whole cell biocatalysis.

In order to enhance the production of a broader spectrum of natural products from microorganisms, various culturing conditions are typically employed. It has been observed that a single strain possesses the ability to synthesize a diverse repertoire of structurally distinct compounds under specific growth conditions, although the complete synthesis of all these compounds simultaneously is hindered by the energetic and metabolic trade-offs [194]. Remarkably, minute alterations in cultivation parameters, such as the composition of the culture medium, pH, temperature, salinity, aeration, and shape of the culture vessel, have the potential to entirely modify, induce, or optimize the physiological behaviour of a microorganism, thereby exerting a significant impact on the biosynthesis of these metabolites [236]. OSMAC has opened up new avenues for industrial-scale production of desired compounds and was firstly applied by Fuchser and Zeeck in 1997 [237].

## Discussion and Outlook

The utilization of the OSMAC approach in the exploration of lichen natural products has garnered considerable attention. Lichenologists worldwide have shown increasing interest in studying not only the physiology but also the biosynthesis of metabolites [238]. Consequently, they have sought ways to culture the mycelium from lichen thalli independently, without the presence of algal or cyanobacterial partners, in the hope of producing lichen compounds. The development of the tissue culture method by Yoshimura et al., enabling the separation and individual cultivation of the lichen-forming fungus from the lichen thallus, has significantly advanced the realization of this goal [239].

However, there exists an inverse relationship between the growth rate of the mycobiont and the yield of metabolites [240]. Culturing conditions, such as the type and availability of carbon and nitrogen sources, exert a notable influence on the growth and metabolite production of the mycobiont [241]–[243]. For instance, *Usnea ghattensis* cultures exhibited high growth rates and the production of usnic acid, when simple sugars (e.g., glucose and sucrose) or polyethylene glycol were used as sole carbon sources. In contrast, nitrogen sources such as amino acids (glycine, asparagine, alanine, or vitamins), particularly glycine, support mycobiont growth but impaired efficient usnic acid production [242]. Another example for the influence of changes in culture conditions was observed with a mycobiont isolated from a *Parmotrema reticulatum* thallus. Here robust colonies were developed when grown on solid LB medium, however atranorin, the primary cortical lichen depside, was only detected after 5 to 10 months of growth. On the other hand, colonies grown on MEYE and MY10 media with a gradual desiccation treatment did not synthesize any lichen polyketides but rather produced primary triacylglyceride and fatty acid as the major metabolites [244]. The BGC for the putative production of atranorin was found throughout all investigated lichen datasets, so this finding may be applicable to separated mycobionts from HPH, HTU and PSU gained from future experiments. However, the main issue of the OSMAC approach is still the lack of stimuli or cues sent by the symbiotic partners of the mycobiont, inducing secondary metabolite expression.

Under standard laboratory conditions, microorganisms are typically grown in axenic cultures, which can result in the suppression of BGCs [245]. However, co-cultivation approaches can be employed to mimic interactions, such as quorum sensing and siderophore-mediated communication, that naturally occur between microorganisms [245]–[250]. This co-cultivation strategy has the potential to induce both quantitative and qualitative changes in secondary metabolite production, including the generation of novel natural products that may remain undetectable under axenic laboratory conditions. Thus, co-cultivation serves as a means to mimic microbial competition, which is a common phenomenon in natural environments [245], [247].

## Concluding Remarks

Through this thesis, the comprehension of the genetic makeup associated with secondary metabolism in lichen and in detail their mycobionts has significantly expanded, paving the way for novel research avenues to establish connections between genes and secondary metabolites. Furthermore, it aimed to unlock the untapped potential of these organisms for various biotechnological applications. In the first part of this study, an extensive assessment of the accuracy and reliability of metagenomically assembled genomes in lichen was conducted, thereby validating the effectiveness of this approach in accessing the biosynthetic capabilities within complex symbiotic systems that are challenging to cultivate. Herby, the BGC composition of entire holobionts from *H. physodes* and *H. tubulosa* were presented for the first time.

In the second part, the focus lay on investigating the specific biosynthetic gene content of the lichen-forming fungi from the metabolite rich lichens *Hypogymnia physodes*, *Hypogymnia tubulosa* and *Parmelia sulcata*. The syntenic comparison of highly orthologous BGCs from the polyketide family can be utilized for heterologous expression experiments, to investigate polyketide product spectra.

In summary, the feasibility of metagenomic reconstruction to obtain genome sequences from complex mixed-species samples, thereby eliminating the laborious step of axenic cultivation was demonstrated. This approach is particularly valuable for investigating unculturable organisms or intricate ecological communities. By employing whole-genome mining, comparative mapping, and phylogenetic analyses, a comprehensive understanding of the metabolic potential of a species can be achieved. Additionally, the newly established reference genomes of *H. physodes*, *H. tubulosa* and *P. sulcata* serve as invaluable resources for various applications, including genome mining, comparative (phylo)genomics, the development of high-resolution genetic markers, and population genetics. Specifically, these findings will significantly contribute to advancing our knowledge of lichen biosynthetic pathways and provide valuable tools for the development of heterologous expression systems. Ultimately, these advancements will enable harnessing the vast repertoire of secondary metabolites found in lichens for various biotechnological applications. The identified BGCs provide direct routes for the production of compounds like e.g., olivetolic acid or atranorin in expression experiments.

## 5 List of Publications

### **Biosynthetic potential of *Hypogymnia* holobionts - Insights into Secondary Metabolite Pathways**

Nadim Ahmad; Manfred Ritz; Anjuli Calchera; Jürgen Otte; Imke Schmitt; Thomas Brueck; Norbert Mehlmer

*Journal of Fungi* 2023, 9(5), 546

<https://doi.org/10.3390/jof9050546>

### **Biosynthetic Gene Cluster Synteny - Orthologous Polyketide Synthases in *Hypogymnia physodes*, *Hypogymnia tubulosa* and *Parmelia sulcata***

Nadim Ahmad; Manfred Ritz; Anjuli Calchera; Jürgen Otte; Imke Schmitt; Thomas Brueck; Norbert Mehlmer

*MicrobiologyOpen* 2023, 12 (5)

DOI: <https://doi.org/10.1002/mbo3.1386>

### **Comparative Genome-Wide Analysis of Two *Caryopteris x Clandonensis* Cultivars: Insights on the Biosynthesis of Volatile Terpenoids**

Nadim Ahmad, Manfred Ritz, Thomas Brück and Norbert Mehlmer

*Plants* 2023, 12(3), 632

<https://doi.org/10.3390/plants12030632>

### **Differential RNA-Seq Analysis Predicts Genes Related to Terpene Tailoring in *Caryopteris x Clandonensis***

Nadim Ahmad, Manfred Ritz, Thomas Brück and Norbert Mehlmer

*Plants* 2023, 12(12), 2305

<https://doi.org/10.3390/plants12122305>

## 6. Reprint Permission

### **Biosynthetic potential of *Hypogymnia* holobionts - Insights into Secondary Metabolite Pathways**

Published in MDPI, *Journal of Fungi* 2023, 9(5), 546(<https://doi.org/10.3390/jof9050546>)

The article is published and available under the Creative Commons Attribution License (CC BY)

### **Biosynthetic Gene Cluster Synteny - Orthologous Polyketide Synthases in *Hypogymnia physodes*, *Hypogymnia tubulosa* and *Parmelia sulcata*.**

Published in Wiley, *MicrobiologyOpen* 2023

The article is published and available under the Creative Commons Attribution License (CC BY)

#### **Permissions**

No special permission is required to reuse all or part of articles, including figures and tables. For articles published under an open access Creative Common CC BY license, any part of the article may be reused without permission provided that the original article is clearly cited. Reuse of an article does not imply endorsement by the authors.

## 7. Figures & Tables

Figure 1. Overview of sequencing techniques deployed in this thesis. Here PacBio Sequel IIe long-read sequencing is compared to short-read sequencing on Illumina NovaSeq6000. In addition, metrics on output, accuracy, errors, and elapsing time are provided. Figure created with Biorender. .... 6

Figure 2. Graphical visualization of a metaFlye subgraph generated by four distinct genomes (1-4) in an assembly. Coloured lines depict repeat edges, whereas black arrows are unique edges. The edges A, B and C (coloured blocks) were deemed to be repetitive sequences by metaFlye, based on the distinct paths overlapping in the respective region. Figure adapted from Kolmogorov et. al [95] and created with Biorender. .... 7

Figure 3. Schematic reaction mechanism for the formation of polyketides and fatty acids, highlighting the potential for flux alteration towards polyketide synthesis. By modifying the reaction cycle, various polyketide derivatives can be generated with different levels of reduction. The respectively involved domains are depicted in coloured circles. Figure adapted from Weissman [128], created with Biorender.....11

Figure 2. Schematic visualization of the genetic structure of a biosynthetic gene cluster (BGC). An example polyketide synthase (PKS) BGC is shown with the respective domains. Figure in adaption to Geers et. al [148] and created with Biorender.....13

Table 1. Lichen products from *Hypogymnia tubulosa*, *Hypogymnia physodes* and *Parmelia sulcata*. Occurrence of respective metabolites is allocated to lichen region. .... 4

Table 2. Kits and Consumables deployed in this study. ....15

Table 3: BGCs present in investigated lichen samples. Percentage shows annotation score based on similarity by antiSMASH. Annotation values above 90% are deemed to be valid. After each lichen sample the total of present BGCs is mentioned to allow for direct comparison.....61

## 8 Bibliography

- [1] D. L. Hawksworth, "The fungal dimension of biodiversity: magnitude, significance, and conservation," *Mycol. Res.*, vol. 95, no. 6, pp. 641–655, Jun. 1991, doi: 10.1016/S0953-7562(09)80810-1.
- [2] D. S. Hibbett *et al.*, "A higher-level phylogenetic classification of the Fungi," *Mycol. Res.*, vol. 111, no. 5, pp. 509–547, May 2007, doi: 10.1016/J.MYCRES.2007.03.004.
- [3] G. M. Mueller *et al.*, "Global diversity and distribution of macrofungi," *Biodivers. Conserv.*, vol. 16, no. 1, pp. 37–48, Jan. 2007, doi: 10.1007/s10531-006-9108-8.
- [4] H. E. O'Brien, J. L. Parrent, J. A. Jackson, J.-M. Moncalvo, and R. Vilgalys, "Fungal Community Analysis by Large-Scale Sequencing of Environmental Samples," *Appl. Environ. Microbiol.*, vol. 71, no. 9, pp. 5544–5550, Sep. 2005, doi: 10.1128/AEM.71.9.5544-5550.2005.
- [5] D. L. Hawksworth and R. Lücking, "Fungal Diversity Revisited: 2.2 to 3.8 Million Species," *Microbiol. Spectr.*, vol. 5, no. 4, Aug. 2017, doi: 10.1128/microbiolspec.FUNK-0052-2016.
- [6] T. H. Nash, "Lichen Biology," *Lichen Biol. Second Ed.*, pp. 1–486, Jan. 2008, doi: 10.1017/CBO9780511790478.
- [7] S. Onofri *et al.*, "Evolution and adaptation of fungi at boundaries of life," *Adv. Sp. Res.*, vol. 40, no. 11, pp. 1657–1664, Jan. 2007, doi: 10.1016/J.ASR.2007.06.004.
- [8] A. E. Douglas, "Symbiotic interactions.," *J. Mar. Biol. Assoc. United Kingdom*, vol. 74, no. 3, pp. 743–743, Aug. 1994, doi: 10.1017/S0025315400047810.
- [9] G. M. Gadd, "Fungi and Their Role in the Biosphere," *Encycl. Ecol. Five-Volume Set*, pp. 1709–1717, Jan. 2008, doi: 10.1016/B978-008045405-4.00734-5.
- [10] P. K. Divakar *et al.*, "Evolution of complex symbiotic relationships in a morphologically derived family of lichen-forming fungi," *New Phytol.*, vol. 208, no. 4, pp. 1217–1226, Dec. 2015, doi: 10.1111/NPH.13553.
- [11] C. C. Plitt, "A Short History of Lichenology," *Bryologist*, vol. 22, no. 6, p. 77, Nov. 1919, doi: 10.2307/3238526.
- [12] B. Büdel and C. Scheidegger, "Thallus morphology and anatomy," *Lichen Biol. Second Ed.*, pp. 40–68, Jan. 2008, doi: 10.1017/CBO9780511790478.005.
- [13] M.-C. ten Veldhuis, G. Ananyev, and G. C. Dismukes, "Symbiosis extended: exchange of photosynthetic O<sub>2</sub> and fungal-respired CO<sub>2</sub> mutually power metabolism of lichen symbionts," *Photosynth. Res.*, vol. 143, no. 3, pp. 287–299, Mar. 2020, doi: 10.1007/s11120-019-00702-0.
- [14] R. Lücking, Organization for Flora Neotropica., and New York Botanical Garden., "Foliicolous lichenized fungi," p. 866, 2008, Accessed: Jul. 07, 2023. [Online]. Available: <https://www.nhbs.com/flora-neotropica-volume-103-foliicolous-lichenized-fungi-book>
- [15] T. Spribille *et al.*, "Basidiomycete yeasts in the cortex of ascomycete macrolichens," *Science*, vol. 353, no. 6298, p. 488, Jul. 2016, doi: 10.1126/SCIENCE.AAF8287.
- [16] T. Cernava, H. Müller, I. A. Aschenbrenner, M. Grube, and G. Berg, "Analyzing the antagonistic potential of the lichen microbiome against pathogens by bridging metagenomic with culture studies," *Front. Microbiol.*, vol. 6, no. JUN, 2015, doi: 10.3389/FMICB.2015.00620.
- [17] K. Mark, L. Laanisto, C. G. Bueno, Ü. Niinemets, C. Keller, and C. Scheidegger, "Contrasting co-occurrence patterns of photobiont and cystobasidiomycete yeast associated with common epiphytic lichen species," *New Phytol.*, vol. 227, no. 5, pp. 1362–1375, Sep. 2020, doi: 10.1111/NPH.16475.
- [18] L. Muggia and M. Grube, "Fungal Diversity in Lichens: From Extremotolerance to Interactions with Algae," *Life 2018, Vol. 8, Page 15*, vol. 8, no. 2, p. 15, May 2018, doi: 10.3390/LIFE8020015.
- [19] I. A. Aschenbrenner, T. Cernava, G. Berg, and M. Grube, "Understanding Microbial Multi-Species Symbioses," *Front. Microbiol.*, vol. 7, no. FEB, Feb. 2016, doi: 10.3389/FMICB.2016.00180.
- [20] H. B. Smith *et al.*, "Metagenomic data reveal diverse fungal and algal communities associated with the lichen symbiosis," *Symbiosis*, vol. 82, no. 1–2, pp. 133–147, Nov. 2020, doi: 10.1007/s13199-020-00699-



## Bibliography

- 4.
- [21] D. L. Hawksworth and M. Grube, "Lichens redefined as complex ecosystems," *New Phytol.*, vol. 227, no. 5, p. 1281, Sep. 2020, doi: 10.1111/NPH.16630.
- [22] G. Rolshausen, F. Dal Grande, J. Otte, and I. Schmitt, "Lichen holobionts show compositional structure along elevation," *Mol. Ecol.*, vol. 00, pp. 1–12, 2022, doi: 10.1111/MEC.16471.
- [23] R. Honegger, "Metabolic Interactions at the Mycobiont-Photobiont Interface in Lichens," *Plant Relationships*, pp. 209–221, 1997, doi: 10.1007/978-3-662-10370-8\_12.
- [24] R. Lücking, B. P. Hodkinson, and S. D. Leavitt, "The 2016 classification of lichenized fungi in the Ascomycota and Basidiomycota – Approaching one thousand genera," *The Bryologist*, 2016. <https://www.jstor.org/stable/44250015> (accessed Jul. 10, 2023).
- [25] P. M. Kirk, P. F. Cannon, J. Stalpers, and D. W. Minter, "Dictionary of the Fungi 10th ed.-CABI (2008)," *CABI Publ. Gt. Britain.*, p. 708, 2008.
- [26] M. Grube and M. Wedin, "Lichenized Fungi and the Evolution of Symbiotic Organization," *Microbiol. Spectr.*, vol. 4, no. 6, Dec. 2016, doi: 10.1128/microbiolspec.FUNK-0011-2016.
- [27] M. P. Gómez-Serranillos, C. Fernández-Moriano, E. González-Burgos, P. K. Divakar, and A. Crespo, "Parmeliaceae family: phytochemistry, pharmacological potential and phylogenetic features," *RSC Adv.*, vol. 4, no. 103, pp. 59017–59047, Nov. 2014, doi: 10.1039/C4RA09104C.
- [28] A. Thell *et al.*, "A review of the lichen family Parmeliaceae - history, phylogeny and current taxonomy," *Nord. J. Bot.*, vol. 30, no. 6, pp. 641–664, Dec. 2012, doi: 10.1111/J.1756-1051.2012.00008.X.
- [29] A. Elvebakk, "A review of the genus Hypogymnia (Parmeliaceae) in Chile," <https://doi.org/10.1639/0007-2745-114.2.379>, vol. 114, no. 2, pp. 379–388, Jun. 2011, doi: 10.1639/0007-2745-114.2.379.
- [30] E. Studzińska-Sroka and D. Zarabska-Bożewicz, "Hypogymnia physodes – A lichen with interesting medicinal potential and ecological properties," *J. Herb. Med.*, vol. 17–18, p. 100287, Sep. 2019, doi: 10.1016/J.HERMED.2019.100287.
- [31] V. Wirth *et al.*, "Die Flechten Deutschlands," pp. 1–672, 2013.
- [32] W. (Ole W. Purvis, "The Lichen Flora of Great Britain and Ireland," p. 710, 1992, Accessed: Mar. 29, 2023. [Online]. Available: <https://research-scotland.ac.uk/handle/20.500.12594/6198>
- [33] A. Crespo, P. D. Bridge, and D. L. Hawksworth, "AMPLIFICATION OF FUNGAL rDNA-ITS REGIONS FROM NON-FERTILE SPECIMENS OF THE LICHEN-FORMING GENUS PARMELIA," *Lichenol.*, vol. 29, no. 3, pp. 275–282, May 1997, doi: 10.1006/LICH.1996.0071.
- [34] A. Crespo, P. D. Bridge, D. L. Hawksworth, M. Grube, and O. F. Cubero, "Comparison of rRNA genotype frequencies of *Parmelia sulcata* from long established and recolonizing sites following sulphur dioxide amelioration," *Plant Syst. Evol.*, vol. 217, no. 3–4, pp. 177–183, 1999, doi: 10.1007/BF00984363.
- [35] T. Feuerer and A. Thell, "Feuerer, T. & Thell, A. (2002) *Parmelia ernstiae* a new macrolichen from Germany," *Mitteilungen des Instituts für Allg. Bot. Hamburg.*, pp. 30-32 : 49-60, 2002.
- [36] M. del C. Molina *et al.*, "*Parmelia sulcata* (Ascomycota: Parmeliaceae), a sympatric monophyletic species complex," *Lichenol.*, vol. 43, no. 6, pp. 585–601, Nov. 2011, doi: 10.1017/S0024282911000521.
- [37] B. T. Pfannenstiel and N. P. Keller, "On top of biosynthetic gene clusters: How epigenetic machinery influences secondary metabolism in fungi," *Biotechnol. Adv.*, vol. 37, no. 6, p. 107345, Nov. 2019, doi: 10.1016/J.BIOTECHADV.2019.02.001.
- [38] E. Skellam, "Strategies for Engineering Natural Product Biosynthesis in Fungi," *Trends Biotechnol.*, vol. 37, no. 4, pp. 416–427, Apr. 2019, doi: 10.1016/J.TIBTECH.2018.09.003.
- [39] N. P. Keller, "Fungal secondary metabolism: regulation, function and drug discovery," *Nat. Rev. Microbiol.* 2019 173, vol. 17, no. 3, pp. 167–180, Dec. 2019, doi: 10.1038/s41579-018-0121-1.
- [40] A. A. Brakhage, "Regulation of fungal secondary metabolism," *Nat. Rev. Microbiol.* 2013, vol. 11, no. 1, pp. 21–32, Nov. 2013, doi: 10.1038/nrmicro2916.
- [41] J. R. Hanson, "The classes of natural product and their isolation," *Natural Products*, 2003.
- [42] N. J. Mouncey, H. Otani, D. Udvary, and Y. Yoshikuni, "New voyages to explore the natural product galaxy," *J. Ind. Microbiol. Biotechnol.*, vol. 46, no. 3–4, pp. 273–279, Mar. 2019, doi: 10.1007/S10295-018-02122-W.

## Bibliography

- [43] A. L. Demain and E. Martens, "Production of valuable compounds by molds and yeasts," *J. Antibiot.* 2017 704, vol. 70, no. 4, pp. 347–360, Oct. 2016, doi: 10.1038/ja.2016.121.
- [44] G. D. Wright, "Unlocking the potential of natural products in drug discovery," *Microb. Biotechnol.*, vol. 12, no. 1, pp. 55–57, Jan. 2019, doi: 10.1111/1751-7915.13351.
- [45] A. L. Harvey, R. Edrada-Ebel, and R. J. Quinn, "The re-emergence of natural products for drug discovery in the genomics era," *Nat. Rev. Drug Discov.* 2015 142, vol. 14, no. 2, pp. 111–129, Jan. 2015, doi: 10.1038/nrd4510.
- [46] G. F. Bills and J. B. Gloer, "Biologically Active Secondary Metabolites from the Fungi," *Microbiol. Spectr.*, vol. 4, no. 6, Dec. 2016, doi: 10.1128/microbiolspec.FUNK-0009-2016.
- [47] B. Ranković and M. Kosanić, "Lichens as a Potential Source of Bioactive Secondary Metabolites," *Lichen Second. Metab.*, pp. 1–29, 2019, doi: 10.1007/978-3-030-16814-8\_1.
- [48] J. Asplund and D. A. Wardle, "How lichens impact on terrestrial community and ecosystem properties," *Biol. Rev.*, vol. 92, no. 3, pp. 1720–1738, Aug. 2017, doi: 10.1111/BRV.12305.
- [49] J. A. Elix and E. Stocker-Wörgötter, "Biochemistry and secondary metabolites.," *Lichen Biol.*, pp. 104–133, 2008.
- [50] K. Molnár and E. Farkas, "Current results on biological activities of lichen secondary metabolites: a review," *Z. Naturforsch. C.*, vol. 65, no. 3–4, pp. 157–173, 2010, doi: 10.1515/ZNC-2010-3-401.
- [51] E. Stocker-Wörgötter, "Metabolic diversity of lichen-forming ascomycetous fungi: culturing, polyketide and shikimate metabolite production, and PKS genes," *Nat. Prod. Rep.*, vol. 25, no. 1, pp. 188–200, Feb. 2008, doi: 10.1039/B606983P.
- [52] M. J. Calcott, D. F. Ackerley, A. Knight, R. A. Keyzers, and J. G. Owen, "Secondary metabolism in the lichen symbiosis," *Chem. Soc. Rev.*, vol. 47, no. 5, pp. 1730–1760, Mar. 2018, doi: 10.1039/C7CS00431A.
- [53] S. Huneck, "Nature of Lichen Substances," *The Lichens*, pp. 495–522, Jan. 1973, doi: 10.1016/B978-0-12-044950-7.50020-9.
- [54] X. Chen *et al.*, "The Organisms on Rock Cultural Heritages: Growth and Weathering," *Geoheritage 2021* 133, vol. 13, no. 3, pp. 1–17, Jun. 2021, doi: 10.1007/S12371-021-00588-2.
- [55] K. Veres, M. Sinigla, K. Szabó, N. Varga, and E. Farkas, "The long-term effect of removing the UV-protectant usnic acid from the thalli of the lichen *Cladonia foliacea*," *Mycol. Prog.*, vol. 21, no. 9, p. 83, Sep. 2022, doi: 10.1007/S11557-022-01831-Y.
- [56] R. Kalra, X. A. Conlan, and M. Goel, "Lichen allelopathy: a new hope for limiting chemical herbicide and pesticide use," <https://doi.org/10.1080/09583157.2021.1901071>, vol. 31, no. 8, pp. 773–796, 2021, doi: 10.1080/09583157.2021.1901071.
- [57] J. Asplund *et al.*, "Contrasting responses of plant and lichen carbon-based secondary compounds across an elevational gradient," *Funct. Ecol.*, vol. 35, no. 2, pp. 330–341, Feb. 2021, doi: 10.1111/1365-2435.13712.
- [58] J. Boustie and M. Grube, "Lichens—a promising source of bioactive secondary metabolites," *Plant Genet. Resour.*, vol. 3, no. 2, pp. 273–287, Aug. 2005, doi: 10.1079/PGR200572.
- [59] S. Huneck and I. Yoshimura, "Identification of Lichen Substances," *Identif. Lichen Subst.*, pp. 11–123, 1996, doi: 10.1007/978-3-642-85243-5\_2.
- [60] A. Calchera, F. Dal Grande, H. B. Bode, and I. Schmitt, "Biosynthetic Gene Content of the 'Perfume Lichens' *Evernia prunastri* and *Pseudevernia furfuracea*," *Molecules*, vol. 24, no. 1, Jan. 2019, doi: 10.3390/molecules24010203.
- [61] S. Ristić *et al.*, "Phytochemical study and antioxidant, antimicrobial and anticancer activities of *Melanelia subaurifera* and *Melanelia fuliginosa* lichens," *J. Food Sci. Technol.*, vol. 53, no. 6, pp. 2804–2816, Jun. 2016, doi: 10.1007/s13197-016-2255-3.
- [62] R. Sisodia, M. Geol, S. Verma, A. Rani, and P. Dureja, "Antibacterial and antioxidant activity of lichen species *Ramalina roesleri*," *Nat. Prod. Res.*, vol. 27, no. 23, pp. 2235–2239, Dec. 2013, doi: 10.1080/14786419.2013.811410.
- [63] M. Kosanić, N. Manojlović, S. Janković, T. Stanojković, and B. Ranković, "*Evernia prunastri* and *Pseudevernia furfuracea* lichens and their major metabolites as antioxidant, antimicrobial and

## Bibliography

- anticancer agents," *Food Chem. Toxicol.*, vol. 53, pp. 112–118, Mar. 2013, doi: 10.1016/J.FCT.2012.11.034.
- [64] G. Karabulut and S. Ozturk, "ANTIFUNGAL ACTIVITY OF EVERNIA PRUNASTRI, PARMELIA SULCATA, PSEUDEVERNIA FURFURACEA VAR. FURFURACEA," *Pak. J. Bot.*, vol. 47, no. 4, pp. 1575–1579, 2015.
- [65] T. Joshi, P. Sharma, T. Joshi, and S. Chandra, "In silico screening of anti-inflammatory compounds from Lichen by targeting cyclooxygenase-2," <https://doi.org/10.1080/07391102.2019.1664328>, vol. 38, no. 12, pp. 3544–3562, Aug. 2019, doi: 10.1080/07391102.2019.1664328.
- [66] M. Goga, J. Elečko, M. Marcinčinová, D. Ručová, M. Bačkorová, and M. Bačkor, "Lichen Metabolites: An Overview of Some Secondary Metabolites and Their Biological Potential," in *Reference Series in Phytochemistry*, Springer Science and Business Media B.V., 2020, pp. 175–209. doi: 10.1007/978-3-319-96397-6\_57.
- [67] M. Kosanić and B. Ranković, "Studies on Antioxidant Properties of Lichen Secondary Metabolites," *Lichen Second. Metab.*, pp. 129–153, 2019, doi: 10.1007/978-3-030-16814-8\_4.
- [68] Z. Solárová, A. Liskova, M. Samec, P. Kubatka, D. Büsselberg, and P. Solár, "Anticancer Potential of Lichens' Secondary Metabolites," *Biomol. 2020, Vol. 10, Page 87*, vol. 10, no. 1, p. 87, Jan. 2020, doi: 10.3390/BIOM10010087.
- [69] V. Cardile, A. C. E. Graziano, R. Avola, M. Piovano, and A. Russo, "Potential anticancer activity of lichen secondary metabolite physodic acid," *Chem. Biol. Interact.*, vol. 263, pp. 36–45, Feb. 2017, doi: 10.1016/J.CBI.2016.12.007.
- [70] B. Ranković, M. Kosanić, N. Manojlović, A. Rančić, and T. Stanojković, "Chemical composition of Hypogymnia physodes lichen and biological activities of some its major metabolites," *Med. Chem. Res.*, vol. 23, no. 1, pp. 408–416, Jan. 2014, doi: 10.1007/s00044-013-0644-y.
- [71] G. Stojanović, I. Zlatanović, I. Zrnzević, M. Stanković, V. Stankov Jovanović, and B. Zlatković, "Hypogymnia tubulosa extracts: chemical profile and biological activities," *Nat. Prod. Res.*, vol. 32, no. 22, pp. 2735–2739, Nov. 2018, doi: 10.1080/14786419.2017.1375926.
- [72] M. Yılmaz, T. Tay, M. Kivanç, H. Türk, and A. Ö. Türk, "The Antimicrobial Activity of Extracts of the Lichen Hypogymnia tubulosa and Its 3-Hydroxyphysodic Acid Constituent," *Zeitschrift für Naturforsch. C*, vol. 60, no. 1–2, pp. 35–38, Feb. 2005, doi: 10.1515/znc-2005-1-207.
- [73] W. (Ole W. Purvis, "The Lichen Flora of Great Britain and Ireland," p. 710, 1992, Accessed: May 24, 2023. [Online]. Available: <https://research-scotland.ac.uk/handle/20.500.12594/6198>
- [74] K. A. Solhaug, M. Lind, L. Nybakken, and Y. Gauslaa, "Possible functional roles of cortical depsides and medullary depsidones in the foliose lichen Hypogymnia physodes," *Flora - Morphol. Distrib. Funct. Ecol. Plants*, vol. 204, no. 1, pp. 40–48, Jan. 2009, doi: 10.1016/J.FLORA.2007.12.002.
- [75] K. Molnár and E. Farkas, "Depsides and Depsidones in Populations of the Lichen Hypogymnia physodes and Its Genetic Diversity," <https://doi.org/10.5735/085.048.0605>, vol. 48, no. 6, pp. 473–482, Dec. 2011, doi: 10.5735/085.048.0605.
- [76] N. Duarte, "Special Issue on Plant and Marine-Derived Natural Product Research in Drug Discovery: Strengths and Perspective," *Pharm. 2022, Vol. 15, Page 1249*, vol. 15, no. 10, p. 1249, Oct. 2022, doi: 10.3390/PH15101249.
- [77] I. M. Brodo, S. Duran Sharnoff, and S. Sharnoff, "Lichens of North America," *Int. Microbiol. 2003 62*, vol. 6, no. 2, pp. 149–150, Jun. 2003, doi: 10.1007/S10123-003-0124-1.
- [78] M. Candan, M. Yılmaz, T. Tay, M. Erdem, and A. Ö. Türk, "Antimicrobial Activity of Extracts of the Lichen Parmelia sulcata and its Salazinic Acid Constituent," *Zeitschrift für Naturforsch. C*, vol. 62, no. 7–8, pp. 619–621, Aug. 2007, doi: 10.1515/znc-2007-7-827.
- [79] D. J. Galloway, "Flora of New Zealand: Lichens: Including Lichen-Forming and Lichenicolous Fun," *Flora New Zeal. Lichens Incl. Lichen-Forming Lichenicolous Fun*, p. 35, 2007.
- [80] P. D. Crittenden, J. C. David, D. L. Hawksworth, and F. S. Campbell, "Attempted isolation and success in the culturing of a broad spectrum of lichen-forming and lichenicolous fungi," *New Phytol.*, vol. 130, no. 2, pp. 267–297, Jun. 1995, doi: 10.1111/J.1469-8137.1995.TB03048.X.
- [81] M. Grube, G. Berg, Ó. S. Andrésson, O. Vilhelmsson, P. S. Dyer, and V. P. W. Miao, "Lichen Genomics," *Ecol. Genomics Fungi*, pp. 191–212, Sep. 2013, doi: 10.1002/9781118735893.CH9.

## Bibliography

- [82] T. R. McDonald, E. Gaya, and F. Lutzoni, "Twenty-five cultures of lichenizing fungi available for experimental studies on symbiotic systems," *Symbiosis*, vol. 59, no. 3, pp. 165–171, Mar. 2013, doi: 10.1007/s13199-013-0228-0.
- [83] V. Miao, M. F. Coëffet-LeGal, D. Brown, S. Sinnemann, G. Donaldson, and J. Davies, "Genetic approaches to harvesting lichen products," *Trends Biotechnol.*, vol. 19, no. 9, pp. 349–355, Sep. 2001, doi: 10.1016/S0167-7799(01)01700-0.
- [84] M. Abdel-Hameed, R. L. Bertrand, M. D. Piercey-Normore, and J. L. Sorensen, "Putative identification of the usnic acid biosynthetic gene cluster by de novo whole-genome sequencing of a lichen-forming fungus," *Fungal Biol.*, vol. 120, no. 3, pp. 306–316, Mar. 2016, doi: 10.1016/J.FUNBIO.2015.10.009.
- [85] T. Hu, N. Chitnis, D. Monos, and A. Dinh, "Next-generation sequencing technologies: An overview," *Hum. Immunol.*, vol. 82, no. 11, pp. 801–811, Nov. 2021, doi: 10.1016/J.HUMIMM.2021.02.012.
- [86] K. R. Kumar, M. J. Cowley, and R. L. Davis, "Next-Generation Sequencing and Emerging Technologies," *Semin. Thromb. Hemost.*, vol. 45, no. 07, pp. 661–673, Oct. 2019, doi: 10.1055/s-0039-1688446.
- [87] I. Abnizova, R. te Boekhorst, and Y. L. Orlov, "Computational Errors and Biases in Short Read Next Generation Sequencing," *J. Proteomics Bioinform.*, vol. 10, no. 1, 2017, doi: 10.4172/JPB.1000420.
- [88] R. Pereira, J. Oliveira, and M. Sousa, "Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics," *J. Clin. Med. 2020, Vol. 9, Page 132*, vol. 9, no. 1, p. 132, Jan. 2020, doi: 10.3390/JCM9010132.
- [89] A. Magi, M. Benelli, A. Gozzini, F. Girolami, F. Torricelli, and M. L. Brandi, "Bioinformatics for Next Generation Sequencing Data," *Genes 2010, Vol. 1, Pages 294-307*, vol. 1, no. 2, pp. 294–307, Sep. 2010, doi: 10.3390/GENES1020294.
- [90] S. Goodwin, J. Gurtowski, S. Ethe-Sayers, P. Deshpande, M. C. Schatz, and W. R. McCombie, "Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome," *Genome Res.*, vol. 25, no. 11, pp. 1750–1756, Nov. 2015, doi: 10.1101/GR.191395.115.
- [91] G. A. Logsdon, M. R. Vollger, and E. E. Eichler, "Long-read human genome sequencing and its applications," *Nat. Rev. Genet.*, vol. 21, no. 10, pp. 597–614, Oct. 2020, doi: 10.1038/s41576-020-0236-x.
- [92] L. Tedersoo, M. Albertsen, S. Anslan, and B. Callahan, "Perspectives and Benefits of High-Throughput Long-Read Sequencing in Microbial Ecology," *Appl. Environ. Microbiol.*, vol. 87, no. 17, pp. 1–19, Aug. 2021, doi: 10.1128/AEM.00626-21.
- [93] J. Eid *et al.*, "Real-Time DNA Sequencing from Single Polymerase Molecules," *Science (80-. )*, vol. 323, no. 5910, pp. 133–138, Jan. 2009, doi: 10.1126/science.1162986.
- [94] B. G. Tzovaras *et al.*, "What Is in *Umbilicaria pustulata*? A Metagenomic Approach to Reconstruct the Holo-Genome of a Lichen," *Genome Biol. Evol.*, vol. 12, no. 4, pp. 309–324, Apr. 2020, doi: 10.1093/GBE/EVAA049.
- [95] M. Kolmogorov *et al.*, "metaFlye: scalable long-read metagenome assembly using repeat graphs," *Nat. Methods*, vol. 17, no. 11, pp. 1103–1110, Nov. 2020, doi: 10.1038/s41592-020-00971-x.
- [96] M. Grube and M. Wedin, "Lichenized Fungi and the Evolution of Symbiotic Organization," *Microbiol. Spectr.*, vol. 4, no. 6, Dec. 2016, doi: 10.1128/microbiolspec.FUNK-0011-2016.
- [97] Z. F. Zhang, L. R. Liu, Y. P. Pan, J. Pan, and M. Li, "Long-read assembled metagenomic approaches improve our understanding on metabolic potentials of microbial community in mangrove sediments," *Microbiome 2023 111*, vol. 11, no. 1, pp. 1–20, Aug. 2023, doi: 10.1186/S40168-023-01630-X.
- [98] K. J. Vogel and N. A. Moran, "Functional and Evolutionary Analysis of the Genome of an Obligate Fungal Symbiont," *Genome Biol. Evol.*, vol. 5, no. 5, pp. 891–904, May 2013, doi: 10.1093/GBE/EVT054.
- [99] T. R. McDonald, O. Mueller, F. S. Dietrich, and F. Lutzoni, "High-throughput genome sequencing of lichenizing fungi to assess gene loss in the ammonium transporter/ammonia permease gene family," *BMC Genomics*, vol. 14, no. 1, p. 225, Apr. 2013, doi: 10.1186/1471-2164-14-225.
- [100] F. P. Breitwieser, J. Lu, and S. L. Salzberg, "A review of methods and databases for metagenomic classification and assembly," *Brief. Bioinform.*, vol. 20, no. 4, pp. 1125–1136, Jul. 2019, doi: 10.1093/bib/bbx120.
- [101] B. Greshake, S. Zehr, F. Dal Grande, A. Meiser, I. Schmitt, and I. Ebersberger, "Potential and pitfalls of eukaryotic metagenome skimming: a test case for lichens," *Mol. Ecol. Resour.*, vol. 16, no. 2, pp. 511–523, Mar. 2016, doi: 10.1111/1755-0998.12463.

## Bibliography

- [102] S. G. Tringe and E. M. Rubin, "Metagenomics: DNA sequencing of environmental samples," *Nat. Rev. Genet.* 2005 611, vol. 6, no. 11, pp. 805–814, Oct. 2005, doi: 10.1038/nrg1709.
- [103] J. C. Wooley, A. Godzik, and I. Friedberg, "A Primer on Metagenomics," *PLOS Comput. Biol.*, vol. 6, no. 2, p. e1000667, 2010, doi: 10.1371/JOURNAL.PCBI.1000667.
- [104] Y. Xia *et al.*, "Strategies and tools in illumina and nanopore-integrated metagenomic analysis of microbiome data," *iMeta*, vol. 2, no. 1, p. e72, Feb. 2023, doi: 10.1002/IMT2.72.
- [105] N. Sangwan, F. Xia, and J. A. Gilbert, "Recovering complete and draft population genomes from metagenome datasets," *Microbiome*, vol. 4, no. 1, p. 8, Dec. 2016, doi: 10.1186/s40168-016-0154-5.
- [106] C. Yang *et al.*, "A review of computational tools for generating metagenome-assembled genomes from metagenomic sequencing data," *Comput. Struct. Biotechnol. J.*, vol. 19, pp. 6301–6314, Jan. 2021, doi: 10.1016/j.csbj.2021.11.028.
- [107] D. Armaleo *et al.*, "The lichen symbiosis re-viewed through the genomes of *Cladonia grayi* and its algal partner *Asterochloris glomerata*," *BMC Genomics*, vol. 20, no. 1, p. 605, Dec. 2019, doi: 10.1186/s12864-019-5629-x.
- [108] E. E. Armstrong, S. Prost, D. Ertz, M. Westberg, A. Frisch, and M. Bendiksby, "Draft Genome Sequence and Annotation of the Lichen-Forming Fungus *Arthonia radiata*," *Genome Announc.*, vol. 6, no. 14, Apr. 2018, doi: 10.1128/genomeA.00281-18.
- [109] Y. Wang, X. Yuan, L. Chen, X. Wang, and C. Li, "Draft Genome Sequence of the Lichen-Forming Fungus *Ramalina intermedia* Strain YAF0013," *Genome Announc.*, vol. 6, no. 23, Jun. 2018, doi: 10.1128/GENOMEA.00478-18.
- [110] G. Singh, "Linking Lichen Metabolites to Genes: Emerging Concepts and Lessons from Molecular Biology and Metagenomics," *J. Fungi* 2023, Vol. 9, Page 160, vol. 9, no. 2, p. 160, Jan. 2023, doi: 10.3390/JOF9020160.
- [111] N. Taş *et al.*, "Metagenomic tools in microbial ecology research This review comes from a themed issue on Environmental biotechnology," *Curr. Opin. Biotechnol.*, vol. 67, pp. 184–191, 2021, doi: 10.1016/j.copbio.2021.01.019.
- [112] S. Biradar, R. Ramya, and A. Sankaranarayanan, "Mycobionts interactions in lichen," in *Microbial Symbionts*, Elsevier, 2023, pp. 215–233. doi: 10.1016/B978-0-323-99334-0.00040-2.
- [113] L. Albarano, R. Esposito, N. Ruocco, and M. Costantini, "Genome Mining as New Challenge in Natural Products Discovery," *Mar. Drugs* 2020, Vol. 18, Page 199, vol. 18, no. 4, p. 199, Apr. 2020, doi: 10.3390/MD18040199.
- [114] M. T. Robey, L. K. Caesar, M. T. Drott, N. P. Keller, and N. L. Kelleher, "An interpreted atlas of biosynthetic gene clusters from 1,000 fungal genomes," *Proc. Natl. Acad. Sci.*, vol. 118, no. 19, p. e2020230118, May 2021, doi: 10.1073/pnas.2020230118.
- [115] N. P. Keller, G. Turner, and J. W. Bennett, "Fungal secondary metabolism — from biochemistry to genomics," *Nat. Rev. Microbiol.* 2005 312, vol. 3, no. 12, pp. 937–947, Dec. 2005, doi: 10.1038/nrmicro1286.
- [116] G. Singh, D. Armaleo, F. Dal Grande, and I. Schmitt, "Depside and depsidone synthesis in lichenized fungi comes into focus through a genome-wide comparison of the olivetoric acid and physodic acid chemotypes of *pseudevernia furfuracea*," *Biomolecules*, vol. 11, no. 10, p. 1445, Oct. 2021, doi: 10.3390/BIOM11101445/S1.
- [117] R. J. Cox, "Curiouser and curiouser: progress in understanding the programming of iterative highly-reducing polyketide synthases," *Nat. Prod. Rep.*, vol. 40, no. 1, pp. 9–27, Jan. 2023, doi: 10.1039/D2NP00007E.
- [118] R. J. Cox, E. Skellam, and K. Williams, "Biosynthesis of Fungal Polyketides," *Physiol. Genet.*, pp. 385–412, 2018, doi: 10.1007/978-3-319-71740-1\_13.
- [119] W. Kim *et al.*, "Linking a Gene Cluster to Atranorin, a Major Cortical Substance of Lichens, through Genetic Dereplication and Heterologous Expression," *MBio*, vol. 12, no. 3, Jun. 2021, doi: 10.1128/mBio.01111-21.
- [120] J. M. Crawford and C. A. Townsend, "New insights into the formation of fungal aromatic polyketides," *Nat. Rev. Microbiol.* 2010 812, vol. 8, no. 12, pp. 879–889, Nov. 2010, doi: 10.1038/nrmicro2465.
- [121] L. Chen, X. Wei, and Y. Matsuda, "Depside Bond Formation by the Starter-Unit Acyltransferase Domain

## Bibliography

- of a Fungal Polyketide Synthase," *J. Am. Chem. Soc.*, vol. 144, no. 42, pp. 19225–19230, Oct. 2022, doi: 10.1021/jacs.2c08585.
- [122] S. Dutta *et al.*, "Structure of a modular polyketide synthase," *Nat. 2014 5107506*, vol. 510, no. 7506, pp. 512–517, Jun. 2014, doi: 10.1038/nature13423.
- [123] C. R. Huitt-Roehl, E. A. Hill, M. M. Adams, A. L. Vagstad, J. W. Li, and C. A. Townsend, "Starter Unit Flexibility for Engineered Product Synthesis by the Nonreducing Polyketide Synthase PksA," *ACS Chem. Biol.*, vol. 10, no. 6, pp. 1443–1449, Jun. 2015, doi: 10.1021/acscchembio.5b00005.
- [124] J. M. Crawford *et al.*, "Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization," *Nat. 2009 4617267*, vol. 461, no. 7267, pp. 1139–1143, Oct. 2009, doi: 10.1038/nature08475.
- [125] L. Y. I. II, X. W. I. I, T. Y., and I. I., "Classification, prediction, and verification of the regioselectivity of fungal polyketide synthase product template domains," *J. Biol. Chem.*, vol. 285, no. 30, pp. 22764–22773, Jul. 2010, doi: 10.1074/JBC.M110.128504.
- [126] Y. Wang, J. A. Kim, Y. H. Cheong, Y. J. Koh, and J. S. Hur, "Isolation and characterization of a non-reducing polyketide synthase gene from the lichen-forming fungus *Usnea longissima*," *Mycol. Prog.*, vol. 11, no. 1, pp. 75–83, Feb. 2012, doi: 10.1007/S11557-010-0730-1.
- [127] J. L. Meier and M. D. Burkart, "Proteomic analysis of polyketide and nonribosomal peptide biosynthesis," *Curr. Opin. Chem. Biol.*, vol. 15, no. 1, pp. 48–56, Feb. 2011, doi: 10.1016/j.cbpa.2010.10.021.
- [128] K. J. Weissman, "Chapter 1 Introduction to Polyketide Biosynthesis," in *Methods in Enzymology*, vol. 459, no. B, Academic Press, 2009, pp. 3–16. doi: 10.1016/S0076-6879(09)04601-1.
- [129] D. Boettger and C. Hertweck, "Molecular Diversity Sculpted by Fungal PKS-NRPS Hybrids," *ChemBioChem*, vol. 14, no. 1, pp. 28–42, Jan. 2013, doi: 10.1002/CBIC.201200624.
- [130] A. Rokas, J. H. Wisecaver, and A. L. Lind, "The birth, evolution and death of metabolic gene clusters in fungi," *Nat. Rev. Microbiol. 2018 1612*, vol. 16, no. 12, pp. 731–744, Sep. 2018, doi: 10.1038/s41579-018-0075-3.
- [131] J. C. Nielsen *et al.*, "Global analysis of biosynthetic gene clusters reveals vast potential of secondary metabolite production in *Penicillium* species," *Nat. Microbiol. 2017 26*, vol. 2, no. 6, pp. 1–9, Apr. 2017, doi: 10.1038/nmicrobiol.2017.44.
- [132] K. Blin, H. U. Kim, M. H. Medema, and T. Weber, "Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters," *Brief. Bioinform.*, vol. 20, no. 4, pp. 1103–1113, Jul. 2019, doi: 10.1093/BIB/BBX146.
- [133] K. Blin *et al.*, "antiSMASH 6.0: improving cluster detection and comparison capabilities," *Nucleic Acids Res.*, vol. 49, no. W1, pp. W29–W35, Jul. 2021, doi: 10.1093/NAR/GKAB335.
- [134] Y.-C. Tsai *et al.*, "Resolving the Complexity of Human Skin Metagenomes Using Single-Molecule Sequencing," *MBio*, vol. 7, no. 1, Mar. 2016, doi: 10.1128/mBio.01948-15.
- [135] A. Cuscó, D. Pérez, J. Viñes, N. Fàbregas, and O. Francino, "Long-read metagenomics retrieves complete single-contig bacterial genomes from canine feces," *BMC Genomics*, vol. 22, no. 1, pp. 1–15, Dec. 2021, doi: 10.1186/S12864-021-07607-0.
- [136] D. M. Bickhart *et al.*, "Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities," *Nat. Biotechnol.*, vol. 40, no. 5, pp. 711–719, May 2022, doi: 10.1038/S41587-021-01130-Z.
- [137] H. Xie, C. Yang, Y. Sun, Y. Igarashi, T. Jin, and F. Luo, "PacBio Long Reads Improve Metagenomic Assemblies, Gene Catalogs, and Genome Binning," *Front. Genet.*, vol. 11, p. 1077, Sep. 2020, doi: 10.3389/fgene.2020.516269.
- [138] L. Chen *et al.*, "Short- and long-read metagenomics expand individualized structural variations in gut microbiomes," *Nat. Commun. 2022 131*, vol. 13, no. 1, pp. 1–12, Jun. 2022, doi: 10.1038/s41467-022-30857-9.
- [139] M. Oliveira and L. Azevedo, "Molecular Markers: An Overview of Data Published for Fungi over the Last Ten Years," *J. Fungi*, vol. 8, no. 8, p. 803, Aug. 2022, doi: 10.3390/JOF8080803/S1.
- [140] M. H. Medema *et al.*, "Minimum Information about a Biosynthetic Gene cluster," *Nat. Chem. Biol. 2015 119*, vol. 11, no. 9, pp. 625–631, Aug. 2015, doi: 10.1038/nchembio.1890.
- [141] D. Pizarro, P. K. Divakar, F. Grewe, A. Crespo, F. Dal Grande, and H. T. Lumbsch, "Genome-Wide

## Bibliography

- Analysis of Biosynthetic Gene Cluster Reveals Correlated Gene Loss with Absence of Usnic Acid in Lichen-Forming Fungi,” *Genome Biol. Evol.*, vol. 12, no. 10, pp. 1858–1868, Oct. 2020, doi: 10.1093/gbe/evaa189.
- [142] J. R. Doroghazi *et al.*, “A roadmap for natural product discovery based on large-scale genomics and metabolomics,” *Nat. Chem. Biol.* 2014 1011, vol. 10, no. 11, pp. 963–968, Sep. 2014, doi: 10.1038/nchembio.1659.
- [143] H. Zhao and M. H. Medema, “Standardization for natural product synthetic biology,” *Nat. Prod. Rep.*, vol. 33, no. 8, pp. 920–924, Jul. 2016, doi: 10.1039/C6NP00030D.
- [144] G. Singh *et al.*, “Climate-specific biosynthetic gene clusters in populations of a lichen-forming fungus,” *Environ. Microbiol.*, vol. 23, no. 8, pp. 4260–4275, Aug. 2021, doi: 10.1111/1462-2920.15605.
- [145] G. Singh, “Linking Lichen Metabolites to Genes: Emerging Concepts and Lessons from Molecular Biology and Metagenomics,” *J. Fungi* 2023, Vol. 9, Page 160, vol. 9, no. 2, p. 160, Jan. 2023, doi: 10.3390/JOF9020160.
- [146] T. Llewellyn *et al.*, “Metagenomics Shines Light on the Evolution of ‘Sunscreen’ Pigment Metabolism in the Teloschistales (Lichen-Forming Ascomycota),” *Genome Biol. Evol.*, Jan. 2023, doi: 10.1093/GBE/EVAD002.
- [147] J. T. Kealey, J. P. Craig, and P. J. Barr, “Identification of a lichen depside polyketide synthase gene by heterologous expression in *Saccharomyces cerevisiae*,” *Metab. Eng. Commun.*, vol. 13, p. e00172, Dec. 2021, doi: 10.1016/J.MEC.2021.E00172.
- [148] A. U. Geers, Y. Buijs, M. L. Strube, L. Gram, and M. Bentzon-Tilia, “The natural product biosynthesis potential of the microbiomes of Earth – Bioprospecting for novel anti-microbial agents in the meta-omics era,” *Comput. Struct. Biotechnol. J.*, vol. 20, pp. 343–352, Jan. 2022, doi: 10.1016/J.CSBJ.2021.12.024.
- [149] L. Muggia, G. Zellnig, J. Rabensteiner, and M. Grube, “Morphological and phylogenetic study of algal partners associated with the lichen-forming fungus *Tephromela atra* from the Mediterranean region,” *Symbiosis*, vol. 51, no. 2, pp. 149–160, Sep. 2010, doi: 10.1007/s13199-010-0060-8.
- [150] Y. M. Chiang, S. L. Chang, B. R. Oakley, and C. C. C. Wang, “Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms,” *Curr. Opin. Chem. Biol.*, vol. 15, no. 1, pp. 137–143, Feb. 2011, doi: 10.1016/J.CBPA.2010.10.011.
- [151] W. Zheng, X. Wang, H. Zhou, Y. Zhang, A. Li, and X. Bian, “Establishment of recombineering genome editing system in *Paraburkholderia megapolitana* empowers activation of silent biosynthetic gene clusters,” *Microb. Biotechnol.*, vol. 13, no. 2, pp. 397–405, Mar. 2020, doi: 10.1111/1751-7915.13535.
- [152] G. Singh *et al.*, “A Candidate Gene Cluster for the Bioactive Natural Product Gyrophoric Acid in Lichen-Forming Fungi,” *Microbiol. Spectr.*, vol. 10, no. 4, Aug. 2022, doi: 10.1128/spectrum.00109-22.
- [153] Z. Wasil, K. A. K. Pahirulzaman, C. Butts, T. J. Simpson, C. M. Lazarus, and R. J. Cox, “One pathway, many compounds: heterologous expression of a fungal biosynthetic pathway reveals its intrinsic potential for diversity,” *Chem. Sci.*, vol. 4, no. 10, pp. 3845–3856, Aug. 2013, doi: 10.1039/C3SC51785C.
- [154] L. Martinet *et al.*, “A Single Biosynthetic Gene Cluster Is Responsible for the Production of Bagremycin Antibiotics and Ferroverdin Iron Chelators,” *MBio*, vol. 10, no. 4, Aug. 2019, doi: 10.1128/mBio.01230-19.
- [155] C. Hertweck, “The Biosynthetic Logic of Polyketide Diversity,” *Angew. Chemie Int. Ed.*, vol. 48, no. 26, pp. 4688–4716, Jun. 2009, doi: 10.1002/ANIE.200806121.
- [156] J. Schumann and C. Hertweck, “Advances in cloning, functional analysis and heterologous expression of fungal polyketide synthase genes,” *J. Biotechnol.*, vol. 124, no. 4, pp. 690–703, Aug. 2006, doi: 10.1016/J.JBIOTEC.2006.03.046.
- [157] B. R. Terlouw *et al.*, “MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters,” *Nucleic Acids Res.*, vol. 51, no. D1, pp. D603–D610, Jan. 2023, doi: 10.1093/NAR/GKAC1049.
- [158] A. N. Gagunashvili, S. P. Davidsson, Z. O. Jónsson, and Ó. S. Andrésón, “Cloning and heterologous transcription of a polyketide synthase gene from the lichen *Solorina crocea*,” *Mycol. Res.*, vol. 113, no. 3, pp. 354–363, Mar. 2009, doi: 10.1016/J.MYCRES.2008.11.011.
- [159] R. L. Bertrand and J. L. Sorensen, “A comprehensive catalogue of polyketide synthase gene clusters in lichenizing fungi,” *J. Ind. Microbiol. Biotechnol.*, vol. 45, no. 12, pp. 1067–1081, Dec. 2018, doi: 10.1007/S10295-018-2080-Y.

## Bibliography

- [160] R. L. Bertrand, M. Abdel-Hameed, and J. L. Sorensen, "Lichen Biosynthetic Gene Clusters Part II: Homology Mapping Suggests a Functional Diversity," *J. Nat. Prod.*, vol. 81, no. 4, pp. 732–748, Apr. 2018, doi: 10.1021/acs.jnatprod.7b00770.
- [161] Y. H. Chooi *et al.*, "Cloning and sequence characterization of a non-reducing polyketide synthase gene from the lichen *Xanthoparmelia semiviridis*," *Mycol. Res.*, vol. 112, no. 2, pp. 147–161, Feb. 2008, doi: 10.1016/J.MYCRES.2007.08.022.
- [162] C. Y. Chiang, M. Ohashi, and Y. Tang, "Deciphering chemical logic of fungal natural product biosynthesis through heterologous expression and genome mining," *Nat. Prod. Rep.*, vol. 40, no. 1, pp. 89–127, Jan. 2023, doi: 10.1039/D2NP00050D.
- [163] Y. Wang, J. Wang, Y. H. Cheong, and J. S. Hur, "Three New Non-reducing Polyketide Synthase Genes from the Lichen-Forming Fungus *Usnea longissima*," <https://doi.org/10.5941/MYCO.2014.42.1.34>, vol. 42, no. 1, pp. 34–40, 2014, doi: 10.5941/MYCO.2014.42.1.34.
- [164] N. Ziemert and P. R. Jensen, "Phylogenetic Approaches to Natural Product Structure Prediction," *Methods Enzymol.*, vol. 517, pp. 161–182, Jan. 2012, doi: 10.1016/B978-0-12-404634-4.00008-5.
- [165] D. Armaleo, X. Sun, and C. Culbertson, "Insights from the first putative biosynthetic gene cluster for a lichen depside and depsidone," <https://doi.org/10.3852/10-335>, vol. 103, no. 4, pp. 741–754, Jul. 2011, doi: 10.3852/10-335.
- [166] Y. Gao, Y. Zhao, X. He, Z. Deng, and M. Jiang, "Challenges of functional expression of complex polyketide biosynthetic gene clusters," *Curr. Opin. Biotechnol.*, vol. 69, pp. 103–111, Jun. 2021, doi: 10.1016/J.COPBIO.2020.12.007.
- [167] M. Ringel *et al.*, "Towards a sustainable generation of pseudopterosin-type bioactives," *Green Chem.*, vol. 22, no. 18, pp. 6033–6046, Sep. 2020, doi: 10.1039/D0GC01697G.
- [168] M. Ritz, N. Ahmad, T. Brueck, and N. Mehlmer, "Comparative Genome-Wide Analysis of Two *Caryopteris x Clandonensis* Cultivars: Insights on the Biosynthesis of Volatile Terpenoids," *Plants*, vol. 12, no. 3, p. 632, Feb. 2023, doi: 10.3390/PLANTS12030632/S1.
- [169] B. Buchfink, C. Xie, and D. H. Huson, "fast and sensitive protein alignment using diamond," vol. 12, no. 1, 2015, doi: 10.1038/nmeth.3176.
- [170] D. H. Huson *et al.*, "MEGAN-LR: New algorithms allow accurate binning and easy interactive exploration of metagenomic long reads and contigs," *Biol. Direct*, vol. 13, no. 1, pp. 1–17, Apr. 2018, doi: 10.1186/S13062-018-0208-7.
- [171] C. Bağcı, S. Patz, and D. H. Huson, "DIAMOND+MEGAN: Fast and Easy Taxonomic and Functional Analysis of Short and Long Microbiome Sequences," *Curr. Protoc.*, vol. 1, no. 3, p. e59, Mar. 2021, doi: 10.1002/CPZ1.59.
- [172] C. Bağcı, S. Beier, A. Górska, and D. H. Huson, "Introduction to the Analysis of Environmental Sequences: Metagenomics with MEGAN," in *Methods in Molecular Biology*, vol. 1910, Humana Press Inc., 2019, pp. 591–604. doi: 10.1007/978-1-4939-9074-0\_19.
- [173] F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov, "BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs," *Bioinformatics*, vol. 31, no. 19, pp. 3210–3212, Oct. 2015, doi: 10.1093/BIOINFORMATICS/BTV351.
- [174] A. Gurevich, V. Saveliev, N. Vyahhi, and G. Tesler, "QUAST: quality assessment tool for genome assemblies," *Bioinformatics*, vol. 29, no. 8, pp. 1072–1075, Apr. 2013, doi: 10.1093/bioinformatics/btt086.
- [175] W. Shen, S. Le, Y. Li, and F. Hu, "SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation," *PLoS One*, vol. 11, no. 10, p. e0163962, Oct. 2016, doi: 10.1371/JOURNAL.PONE.0163962.
- [176] K. J. Hoff, A. Lomsadze, M. Borodovsky, and M. Stanke, "Whole-Genome Annotation with BRAKER," in *Methods in Molecular Biology*, vol. 1962, Humana Press Inc., 2019, pp. 65–95. doi: 10.1007/978-1-4939-9173-0\_5.
- [177] K. J. Hoff, S. Lange, A. Lomsadze, M. Borodovsky, and M. Stanke, "BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS," *Bioinformatics*, vol. 32, no. 5, pp. 767–769, Mar. 2016, doi: 10.1093/BIOINFORMATICS/BTV661.
- [178] T. Brůna, K. J. Hoff, A. Lomsadze, M. Stanke, and M. Borodovsky, "BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database," *NAR Genomics Bioinforma.*, vol. 3, no. 1, pp. 1–11, Jan. 2021, doi: 10.1093/nargab/lqaa108.



## Bibliography

- [179] O. Keller, M. Kollmar, M. Stanke, and S. Waack, "A novel hybrid gene prediction method employing protein multiple sequence alignments," *Bioinformatics*, vol. 27, no. 6, pp. 757–763, Mar. 2011, doi: 10.1093/BIOINFORMATICS/BTR010.
- [180] T. Alioto, "Gene prediction," *Methods Mol. Biol.*, vol. 855, pp. 175–201, 2012, doi: 10.1007/978-1-61779-582-4\_6.
- [181] P. Jones *et al.*, "InterProScan 5: genome-scale protein function classification," *Bioinformatics*, vol. 30, no. 9, pp. 1236–1240, May 2014, doi: 10.1093/BIOINFORMATICS/BTU031.
- [182] J. C. Navarro-Muñoz *et al.*, "A computational framework to explore large-scale biosynthetic diversity," *Nat. Chem. Biol.* 2019 161, vol. 16, no. 1, pp. 60–68, Nov. 2019, doi: 10.1038/s41589-019-0400-9.
- [183] P. Shannon *et al.*, "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome Res.*, vol. 13, no. 11, pp. 2498–2504, Nov. 2003, doi: 10.1101/GR.1239303.
- [184] C. Camacho *et al.*, "BLAST+: architecture and applications," *BMC Bioinformatics*, vol. 10, no. 1, p. 421, Dec. 2009, doi: 10.1186/1471-2105-10-421.
- [185] P. J. A. Cock, J. M. Chilton, B. Grüning, J. E. Johnson, and N. Soranzo, "NCBI BLAST+ integrated into Galaxy," *Gigascience*, vol. 4, no. 1, Dec. 2015, doi: 10.1186/S13742-015-0080-7/2707769.
- [186] N. Ahmad *et al.*, "Biosynthetic Potential of Hypogymnia Holobionts: Insights into Secondary Metabolite Pathways," *J. Fungi* 2023, Vol. 9, Page 546, vol. 9, no. 5, p. 546, May 2023, doi: 10.3390/JOF9050546.
- [187] "Galaxy | Europe." <https://usegalaxy.eu/> (accessed Mar. 17, 2023).
- [188] L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, and B. Q. Minh, "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies," *Mol. Biol. Evol.*, vol. 32, no. 1, pp. 268–274, Jan. 2015, doi: 10.1093/molbev/msu300.
- [189] S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. Von Haeseler, and L. S. Jermin, "ModelFinder: fast model selection for accurate phylogenetic estimates," *Nat. Methods* 2017 146, vol. 14, no. 6, pp. 587–589, May 2017, doi: 10.1038/nmeth.4285.
- [190] I. Letunic and P. Bork, "Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation," *Nucleic Acids Res.*, vol. 49, no. W1, pp. W293–W296, Jul. 2021, doi: 10.1093/NAR/GKAB301.
- [191] M. J. Sullivan, N. K. Petty, and S. A. Beatson, "Easyfig: a genome comparison visualizer," *Bioinformatics*, vol. 27, no. 7, pp. 1009–1010, Apr. 2011, doi: 10.1093/BIOINFORMATICS/BTR039.
- [192] A. C. E. Darling, B. Mau, F. R. Blattner, and N. T. Perna, "Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements," *Genome Res.*, vol. 14, no. 7, pp. 1394–1403, Jul. 2004, doi: 10.1101/GR.2289704.
- [193] L. Muggia, I. Schmitt, and M. Grube, "Lichens as treasure chests of natural products," *SIM News*, vol. 59, no. 3, pp. 85–97, 2009.
- [194] M. Ren, S. Jiang, Y. Wang, X. Pan, F. Pan, and X. Wei, "Discovery and excavation of lichen bioactive natural products," *Front. Microbiol.*, vol. 14, p. 1177123, Apr. 2023, doi: 10.3389/fmicb.2023.1177123.
- [195] A. H. Schweiger, G. M. Ullmann, N. M. Nürk, D. Triebel, R. Schobert, and G. Rambold, "Chemical properties of key metabolites determine the global distribution of lichens," *Ecol. Lett.*, vol. 25, no. 2, pp. 416–426, Feb. 2022, doi: 10.1111/ELE.13930.
- [196] M. Feher and J. M. Schmidt, "Property Distributions: Differences between Drugs, Natural Products, and Molecules from Combinatorial Chemistry," *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 1, pp. 218–227, Jan. 2003, doi: 10.1021/CI0200467.
- [197] L. Z. Benet, C. M. Hosey, O. Ursu, and T. I. Oprea, "BDDCS, the Rule of 5 and drugability," *Adv. Drug Deliv. Rev.*, vol. 101, pp. 89–98, Jun. 2016, doi: 10.1016/J.ADDR.2016.05.007.
- [198] N. D. Yuliana, A. Khatib, Y. H. Choi, and R. Verpoorte, "Metabolomics for bioactivity assessment of natural products," *Phyther. Res.*, vol. 25, no. 2, pp. 157–169, Feb. 2011, doi: 10.1002/PTR.3258.
- [199] W. A. Elkhateeb and G. M. Daba, "Lichens, an alternative drugs for modern diseases," *Int. J. Res. Pharm. Biosci.*, vol. 6, no. 10, pp. 5–9, 2019.
- [200] R. S. Hamida, M. A. Ali, N. E. Abdelmeguid, M. I. Al-Zaban, L. Baz, and M. M. Bin-Meferij, "Lichens—A Potential Source for Nanoparticles Fabrication: A Review on Nanoparticles Biosynthesis and Their Prospective Applications," *J. Fungi* 2021, Vol. 7, Page 291, vol. 7, no. 4, p. 291, Apr. 2021, doi:

## Bibliography

- 10.3390/JOF7040291.
- [201] N. Ahmad *et al.*, "Biosynthetic gene cluster synteny: Orthologous polyketide synthases in *Hypogymnia physodes*, *Hypogymnia tubulosa*, and *Parmelia sulcata*," *Microbiologyopen*, vol. 12, no. 5, Oct. 2023, doi: 10.1002/mbo3.1386.
- [202] B. Huang and Y. Zhang, "Teaching an old dog new tricks: Drug discovery by repositioning natural products and their derivatives," *Drug Discov. Today*, vol. 27, no. 7, pp. 1936–1944, Jul. 2022, doi: 10.1016/J.DRUDIS.2022.02.007.
- [203] P. Tiwari and H. Bae, "Endophytic Fungi: Key Insights, Emerging Prospects, and Challenges in Natural Product Drug Discovery," *Microorganisms*, vol. 10, no. 2, p. 360, Feb. 2022, doi: 10.3390/microorganisms10020360.
- [204] L. N. Aldrich *et al.*, "Discovery of Anticancer Agents of Diverse Natural Origin," *J. Nat. Prod.*, vol. 85, no. 3, pp. 702–719, Mar. 2022, doi: 10.1021/acs.jnatprod.2c00036.
- [205] D. J. Newman and G. M. Cragg, "Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019," *J. Nat. Prod.*, vol. 83, no. 3, pp. 770–803, Mar. 2020, doi: 10.1021/ACS.JNATPROD.9B01285.
- [206] G. Kumar and A. Kiran Tudu, "Tackling multidrug-resistant *Staphylococcus aureus* by natural products and their analogues acting as NorA efflux pump inhibitors," *Bioorg. Med. Chem.*, vol. 80, p. 117187, Feb. 2023, doi: 10.1016/J.BMC.2023.117187.
- [207] Z. Wang *et al.*, "A naturally inspired antibiotic to target multidrug-resistant pathogens," *Nat.* 2022 6017894, vol. 601, no. 7894, pp. 606–611, Jan. 2022, doi: 10.1038/s41586-021-04264-x.
- [208] F. E. Dayan, C. L. Cantrell, and S. O. Duke, "Natural products in crop protection," *Bioorg. Med. Chem.*, vol. 17, no. 12, pp. 4022–4034, Jun. 2009, doi: 10.1016/J.BMC.2009.01.046.
- [209] M. Thakur, I. K. Kasi, P. Islary, and S. K. Bhatti, "Nutritional and Health-Promoting Effects of Lichens Used in Food Applications," *Curr. Nutr. Rep.*, pp. 1–12, Aug. 2023, doi: 10.1007/S13668-023-00489-6/METRCS.
- [210] D. Joulain and R. Tabacchi, "Lichen extracts as raw materials in perfumery. Part 1: Oakmoss," *Flavour Fragr. J.*, vol. 24, no. 2, pp. 49–61, 2009, doi: 10.1002/FFJ.1916.
- [211] N. B. Sadeer and M. F. Mahomoodally, "Antibiotic Potentiation of Natural Products: A Promising Target to Fight Pathogenic Bacteria," *Curr. Drug Targets*, vol. 22, no. 5, pp. 555–572, Sep. 2020, doi: 10.2174/1389450121666200924113740.
- [212] M. I. Hutchings, A. W. Truman, and B. Wilkinson, "Antibiotics: past, present and future," *Curr. Opin. Microbiol.*, vol. 51, pp. 72–80, Oct. 2019, doi: 10.1016/j.mib.2019.10.008.
- [213] G. M. Cragg and D. J. Newman, "Biodiversity: A continuing source of novel drug leads," *Pure Appl. Chem.*, vol. 77, no. 1, pp. 7–24, Jan. 2005, doi: 10.1351/pac200577010007.
- [214] X. Meng *et al.*, "Developing fungal heterologous expression platforms to explore and improve the production of natural products from fungal biodiversity," *Biotechnol. Adv.*, vol. 54, p. 107866, Jan. 2022, doi: 10.1016/J.BIOTECHADV.2021.107866.
- [215] L. Katz and R. H. Baltz, "Natural product discovery: past, present, and future," *J. Ind. Microbiol. Biotechnol.*, vol. 43, no. 2–3, pp. 155–176, Mar. 2016, doi: 10.1007/S10295-015-1723-5.
- [216] H. N. Wang *et al.*, "Natural bioactive compounds from marine fungi (2017–2020)," <https://doi.org/10.1080/10286020.2021.1947254>, vol. 24, no. 3, pp. 203–230, 2021, doi: 10.1080/10286020.2021.1947254.
- [217] S. D. Crawford, "Lichens Used in Traditional Medicine," in *Lichen Secondary Metabolites*, Cham: Springer International Publishing, 2019, pp. 31–97. doi: 10.1007/978-3-030-16814-8\_2.
- [218] C. Y. Tan *et al.*, "α-Pyrone and Sterol Constituents of *Penicillium aurantiacobrunneum*, a Fungal Associate of the Lichen *Niebla homalea*," *J. Nat. Prod.*, vol. 82, no. 9, pp. 2529–2536, Sep. 2019, doi: 10.1021/acs.jnatprod.9b00340.
- [219] Y. Zhang, C. Y. Tan, R. W. Spjut, J. R. Fuchs, A. D. Kinghorn, and L. H. Rakatondraibe, "Specialized metabolites of the United States lichen *Niebla homalea* and their antiproliferative activities," *Phytochemistry*, vol. 180, p. 112521, Dec. 2020, doi: 10.1016/J.PHYTOCHEM.2020.112521.
- [220] E. K. Shwab and N. P. Keller, "Regulation of secondary metabolite production in filamentous ascomycetes," *Mycol. Res.*, vol. 112, no. 2, pp. 225–230, Feb. 2008, doi:

## Bibliography

- 10.1016/J.MYCRES.2007.08.021.
- [221] H. J. Bussink, A. Clark, and R. Oliver, "The *Cladosporium fulvum* Bap1 gene: Evidence for a novel class of Yap-related transcription factors with ankyrin repeats in phytopathogenic fungi," *Eur. J. Plant Pathol.*, vol. 107, no. 6, pp. 655–659, 2001, doi: <https://doi.org/10.1023/A:1017932721007>.
- [222] K. F. Pedley and J. D. Walton, "Regulation of cyclic peptide biosynthesis in a plant pathogenic fungus by a novel transcription factor," *Proc. Natl. Acad. Sci.*, vol. 98, no. 24, pp. 14174–14179, Nov. 2001, doi: 10.1073/PNAS.231491298.
- [223] E. M. Niehaus *et al.*, "Apicidin F: Characterization and Genetic Manipulation of a New Secondary Metabolite Gene Cluster in the Rice Pathogen *Fusarium fujikuroi*," *PLoS One*, vol. 9, no. 7, p. e103336, Jul. 2014, doi: 10.1371/JOURNAL.PONE.0103336.
- [224] J. H. Yu and N. Keller, "Regulation of Secondary Metabolism in Filamentous Fungi," <https://doi.org/10.1146/annurev.phyto.43.040204.140214>, vol. 43, pp. 437–458, Aug. 2005, doi: 10.1146/ANNUREV.PHYTO.43.040204.140214.
- [225] B. P. Knox and N. P. Keller, "Key Players in the Regulation of Fungal Secondary Metabolism," pp. 13–28, 2015, doi: 10.1007/978-1-4939-2531-5\_2.
- [226] Z. Luo *et al.*, "The PacC transcription factor regulates secondary metabolite production and stress response, but has only minor effects on virulence in the insect pathogenic fungus *Beauveria bassiana*," *Environ. Microbiol.*, vol. 19, no. 2, pp. 788–802, Feb. 2017, doi: 10.1111/1462-2920.13648.
- [227] W. Wang, X. Liang, Y. Li, P. Wang, and N. P. Keller, "Genetic Regulation of Mycotoxin Biosynthesis," *J. Fungi* 2023, Vol. 9, Page 21, vol. 9, no. 1, p. 21, Dec. 2023, doi: 10.3390/JOF9010021.
- [228] M. Brodhagen and N. P. Keller, "Signalling pathways connecting mycotoxin production and sporulation," *Mol. Plant Pathol.*, vol. 7, no. 4, pp. 285–301, Jul. 2006, doi: 10.1111/J.1364-3703.2006.00338.X.
- [229] W.-B. Shim and C. P. Woloshuk, "Regulation of Fumonisin B 1 Biosynthesis and Conidiation in *Fusarium verticillioides* by a Cyclin-Like (C-Type) Gene, FCC1," *Appl. Environ. Microbiol.*, vol. 67, no. 4, pp. 1607–1612, Apr. 2001, doi: 10.1128/AEM.67.4.1607-1612.2001.
- [230] P. Huang *et al.*, "Melanin Promotes Spore Production in the Rice Blast Fungus *Magnaporthe oryzae*," *Front. Microbiol.*, vol. 13, p. 843838, Feb. 2022, doi: 10.3389/fmicb.2022.843838.
- [231] J. W. Bok and N. P. Keller, "LaeA, a Regulator of Secondary Metabolism in *Aspergillus* spp.," *Eukaryot. Cell*, vol. 3, no. 2, pp. 527–535, Apr. 2004, doi: 10.1128/EC.3.2.527-535.2004.
- [232] Ö. Bayram and G. H. Braus, "Coordination of secondary metabolism and development in fungi: the velvet family of regulatory proteins," *FEMS Microbiol. Rev.*, vol. 36, no. 1, pp. 1–24, Jan. 2012, doi: 10.1111/j.1574-6976.2011.00285.x.
- [233] Ö. Bayram *et al.*, "VelB/VeA/LaeA complex coordinates light signal with fungal development and secondary metabolism," *Science (80-. )*, vol. 320, no. 5882, pp. 1504–1506, Jun. 2008, doi: 10.1126/SCIENCE.1155888.
- [234] R. Fischer, "Sex and Poison in the Dark," *Science (80-. )*, vol. 320, no. 5882, pp. 1430–1431, Jun. 2008, doi: 10.1126/science.1160123.
- [235] R. A. E. Butchko, D. W. Brown, M. Busman, B. Tudzynski, and P. Wiemann, "Lae1 regulates expression of multiple secondary metabolite gene clusters in *Fusarium verticillioides*," *Fungal Genet. Biol.*, vol. 49, no. 8, pp. 602–612, Aug. 2012, doi: 10.1016/J.FGB.2012.06.003.
- [236] H. B. Bode, B. Bethe, R. Höfs, and A. Zeeck, "Big Effects from Small Changes: Possible Ways to Explore Nature's Chemical Diversity," *ChemBioChem*, vol. 3, no. 7, p. 619, Jul. 2002, doi: 10.1002/1439-7633(20020703)3:7<619::AID-CBIC619>3.0.CO;2-9.
- [237] J. Fuchser and A. Zeeck, "Secondary Metabolites by Chemical Screening, 34. – Aspinolides and Aspinonene/Aspyrone Co-Metabolites, New Pentaketides Produced by *Aspergillus ochraceus*," *Liebigs Ann.*, vol. 1997, no. 1, pp. 87–95, Jan. 1997, doi: 10.1002/JLAC.199719970114.
- [238] P. D. Crittenden and N. Porter, "Lichen-forming fungi: potential sources of novel metabolites," *Trends Biotechnol.*, vol. 9, no. 1, pp. 409–414, Jan. 1991, doi: 10.1016/0167-7799(91)90141-4.
- [239] I. Yoshimura, T. Kurokawa, Y. Yamamoto, and Y. Kinoshita, "Development of Lichen Thalli in Vitro," *Bryologist*, vol. 96, no. 3, p. 412, 1993, doi: 10.2307/3243871.
- [240] B. A. Timsina, J. L. Sorensen, D. Weihrauch, and M. D. Piercey-Normore, "Effect of aposymbiotic conditions on colony growth and secondary metabolite production in the lichen-forming fungus *Ramalina*

## Bibliography

- dilacerata," *Fungal Biol.*, vol. 117, no. 11–12, pp. 731–743, Nov. 2013, doi: 10.1016/J.FUNBIO.2013.09.003.
- [241] A. M. Calvo, R. A. Wilson, J. W. Bok, and N. P. Keller, "Relationship between Secondary Metabolism and Fungal Development," *Microbiol. Mol. Biol. Rev.*, vol. 66, no. 3, pp. 447–459, Sep. 2002, doi: 10.1128/MMBR.66.3.447-459.2002.
- [242] B. C. Behera, N. Verma, A. Sonone, and U. Makhija, "Experimental studies on the growth and usnic acid production in 'lichen' *Usnea ghattensis* in vitro," *Microbiol. Res.*, vol. 161, no. 3, pp. 232–237, Jul. 2006, doi: 10.1016/J.MICRES.2005.08.006.
- [243] N. Verma, B. C. Behera, and A. Joshi, "Studies on nutritional requirement for the culture of lichen *Ramalina nervulosa* and *Ramalina pacifica* to enhance the production of antioxidant metabolites," *Folia Microbiol. (Praha)*, vol. 57, no. 2, pp. 107–114, Mar. 2012, doi: 10.1007/s12223-012-0100-2.
- [244] A. T. Fazio *et al.*, "Culture studies on the mycobiont isolated from *Parmotrema reticulatum* (Taylor) Choisy: metabolite production under different conditions," *Mycol. Prog.*, vol. 8, no. 4, pp. 359–365, Nov. 2009, doi: 10.1007/s11557-009-0609-1.
- [245] S. S. El-Hawary, M. H. A. Hassan, A. O. Hudhud, U. R. Abdelmohsen, and R. Mohammed, "Elicitation for activation of the actinomycete genome's cryptic secondary metabolite gene clusters," *RSC Adv.*, vol. 13, no. 9, pp. 5778–5795, Feb. 2023, doi: 10.1039/D2RA08222E.
- [246] E. Oppong-Danquah, M. Blümel, S. Scarpato, A. Mangoni, and D. Tasdemir, "Induction of Isochromanones by Co-Cultivation of the Marine Fungus *Cosmospora* sp. and the Phytopathogen *Magnaporthe oryzae*," *Int. J. Mol. Sci.*, vol. 23, no. 2, p. 782, Jan. 2022, doi: 10.3390/IJMS23020782/S1.
- [247] S. Xu, M. Li, Z. Hu, Y. Shao, J. Ying, and H. Zhang, "The Potential Use of Fungal Co-Culture Strategy for Discovery of New Secondary Metabolites," *Microorganisms*, vol. 11, no. 2, p. 464, Feb. 2023, doi: 10.3390/MICROORGANISMS11020464/S1.
- [248] S. Bertrand, N. Bohni, S. Schnee, O. Schumpp, K. Gindro, and J. L. Wolfender, "Metabolite induction via microorganism co-culture: A potential way to enhance chemical diversity for drug discovery," *Biotechnol. Adv.*, vol. 32, no. 6, pp. 1180–1204, Nov. 2014, doi: 10.1016/J.BIOTECHADV.2014.03.001.
- [249] A. Marmann, A. H. Aly, W. Lin, B. Wang, and P. Proksch, "Co-Cultivation—A Powerful Emerging Tool for Enhancing the Chemical Diversity of Microorganisms," *Mar. Drugs 2014, Vol. 12, Pages 1043-1065*, vol. 12, no. 2, pp. 1043–1065, Feb. 2014, doi: 10.3390/MD12021043.
- [250] T. Netzker *et al.*, "Microbial communication leading to the activation of silent fungal secondary metabolite gene clusters," *Front. Microbiol.*, vol. 6, no. MAR, p. 124427, Apr. 2015, doi: 10.3389/fmicb.2015.00299.

## 9 Additional Full-Length Publications

Comparative Genome-Wide Analysis of  
Two *Caryopteris x Clandonensis* Cultivars:  
Insights on the Biosynthesis of Volatile  
Terpenoids



*plants*

IMPACT  
FACTOR  
4.658

Indexed in:  
PubMed

Article

---

# Comparative Genome-Wide Analysis of Two *Caryopteris x Clandonensis* Cultivars: Insights on the Biosynthesis of Volatile Terpenoids

---

Manfred Ritz, Nadim Ahmad, Thomas Brueck and Norbert Mehlmer

Special Issue

Applications of Bioinformatics in Plant Resources and Omics

Edited by

Dr. Noe Fernandez-Pozo and Prof. Dr. M. Gonzalo Claros



<https://doi.org/10.3390/plants12030632>

## Article

# Comparative Genome-Wide Analysis of Two *Caryopteris x Clandonensis* Cultivars: Insights on the Biosynthesis of Volatile Terpenoids

Manfred Ritz <sup>†</sup> , Nadim Ahmad <sup>†</sup> , Thomas Brueck <sup>\*</sup>  and Norbert Mehlmer <sup>\*</sup> 

Werner Siemens Chair of Synthetic Biotechnology, Department of Chemistry,  
Technical University of Munich (TUM), 85748 Garching, Germany

\* Correspondence: brueck@tum.de (T.B.); norbert.mehlmer@tum.de (N.M.)

† These authors contributed equally to this work.

**Abstract:** *Caryopteris x Clandonensis*, also known as bluebeard, is an ornamental plant containing a large variety of terpenes and terpene-like compounds. Four different cultivars were subjected to a principal component analysis to elucidate variations in terpenoid-biosynthesis and consequently, two representative cultivars were sequenced on a genomic level. Functional annotation of genes as well as comparative genome analysis on long read datasets enabled the identification of cultivar-specific terpene synthase and cytochrome p450 enzyme sequences. This enables new insights, especially since terpenoids in research and industry are gaining increasing interest due to their importance in areas such as food preservation, fragrances, or as active ingredients in pharmaceutical formulations. According to BUSCO assessments, the presented genomes have an average size of 355 Mb and about 96.8% completeness. An average of 52,090 genes could be annotated as putative proteins, whereas about 42 were associated with terpene synthases and about 1340 with cytochrome p450 enzymes.

**Keywords:** reference genome; terpene synthases; *Caryopteris x clandonensis*; plant volatiles; long read sequencing; TPS subfamilies



**Citation:** Ritz, M.; Ahmad, N.; Brueck, T.; Mehlmer, N. Comparative Genome-Wide Analysis of Two *Caryopteris x Clandonensis* Cultivars: Insights on the Biosynthesis of Volatile Terpenoids. *Plants* **2023**, *12*, 632. <https://doi.org/10.3390/plants12030632>

Academic Editors: Noe Fernandez-Pozo and M. Gonzalo Claros

Received: 21 December 2022

Revised: 25 January 2023

Accepted: 27 January 2023

Published: 1 February 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Throughout the last decades, terpenes and terpenoids became more and more important in industrial applications. In the food industry terpenes are used, e.g., as flavoring compounds [1] or preservatives [2]. Due to its plant origin, the acceptance as a food additive is higher compared to chemical synthesis. In a pharmaceutical context the research and use of essential oils—with terpenes as their main components—range from anti-inflammatory [3], and immunomodulatory [4] to antiviral [5] and further indications [6–11]. The anti-cancer drug Taxol consists of a diterpenoid backbone [12] and is employed in different cancer treatments [13]. The success of this terpenoid surely is one of the reasons to further research terpenoids for pharmaceutical applications. Along with these applications, this class of molecules can be found throughout most organisms. Flowering plants show a vast diversity of terpenoids, which is a unique characteristic of the class *Angiospermae* [14]. In plants, they are used as a defense mechanism against biotic (e.g., herbivores or pests) and abiotic influences (e.g., radiation or climate stress) [15]. An example of a defense mechanism against biotic stress is the insect repellent activity of volatiles, such as p-menthane-3,8-diol from *Corymbia citriodora* [16,17]. This compound shows activity against the yellow fever mosquito *Aedes aegypti* [18]. *Caryopteris x clandonensis* essential oils also harbor a biological activity against these insects [19]. However, for these plants, the active agent is not yet identified. Additionally, terpenoids function as attractors for pollinators or as a possibility for energy storage [14].

The extensive diversity of natural terpenes derives from the conserved evolution of terpene synthases (TPS) and terpene-modifying enzymes, such as cytochrome p450

enzymes [20,21]. Terpenes are divided into different classes defined by their backbone. The basis is two building blocks, isopentenylpyrophosphate (IPP) and dimethylallyl diphosphate (DMAPP) which are synthesized in plants via the mevalonate pathway. IPP is the activated form of an isoprene unit consisting of five C-atoms (C5), also called hemiterpene. These are connected to larger units forming monoterpenes (C10), sesquiterpenes (C15), diterpenes (C20) and higher terpene structures [22]. Further steps of increasing terpenoid diversity involve the promiscuity of TPS as well as the subsequent modification by cytochrome p450 enzymes, which may encompass hydroxylation, carboxylation, acetylation or peroxide linkage. Examples include the biosynthesis of p-menthane-3,8-diol [17], gibberellin [23], taxol [24], and artemisinin [25], respectively. This results in a vast pool of natural compounds which account for a multitude of possible applications [14,26].

In general, plant TPS are divided into eight subfamilies which are grouped into classes I, II and III. This separation is based on functional assessment, sequence likelihood and architecture of genes. Class I is comprised of copalyl diphosphate synthases (TPS-c), *ent*-kaurene synthases (TPS-e), other diterpene synthases (TPS-f) and lycopod specific (TPS-h). TPS-d is only included in class II, which is specific for *Gymnosperms*. Lastly, class III consists of TPS-a, cyclic monoterpene synthases and hemi-TPS (TPS-b) and acyclic mono-TPS (TPS-g), which are *Angiosperm* specific [27].

With the advent of state-of-the-art bioinformatic technologies, deciphering the molecular mechanisms involved in the formation of terpenoids has become significantly easier [28]. Furthermore, the possibility to produce terpenes recombinantly by means of biotechnological production systems, rather than chemical synthesis, makes it an ecological and cost-effective technology for the increasing demand for terpenes in industrial applications, despite open challenges [29].

The combination of cutting-edge bioinformatics and next-generation sequencing technologies provided by Pacific Biosciences, Oxford Nanopore and Illumina allows for the rapid generation of draft genomes as well as the annotation of valid gene models. In this context, long-read sequencing technologies will be highlighted, as they exhibit no amplification biases. Consequently, these technologies provide a reliable basis for de novo whole-genome assemblies. Openly accessible bioinformatic tools enable cost-efficient assemblies, annotations and secondary downstream analyses for a broad range of scientists, and are publicly available via [www.github.com](http://www.github.com) (accessed on 11 December 2022) [30]. Two of these are the Quality Assessment Tool for Genome Assemblies (QUAST) [31] and Benchmarking Universal Single-Copy Orthologs (BUSCO). The latter is employed to assess the completeness of the obtained genome assemblies. Here, conserved and species-specific gene sequences are curated in databases and detected via a match-making algorithm to check for the gene set completeness of the evaluated taxonomic group [32]. An investigated genome is classified as complete if respective single-copy orthologs are present in the assembly.

In this work, we present the genomes of two *Caryopteris x clandonensis* cultivars (Dark Knight and Pink Perfection) from the *Lamiaceae* family in high quality employing long-read sequencing. These plants display a wide range of different metabolic pathways in regard to terpenoid biosynthesis, as also seen in other plants of the order *Lamiales*, e.g., in *Jasminum sambac* [33]. To elucidate variations between these multivariate datasets a principal component analysis (PCA) was conducted. Based on evident differences in volatile compound composition the two cultivars, Dark Knight and Pink Perfection were compared on a genomic level. This submission will be the 12th whole genome sequence within *Lamiaceae*, consisting of about 4788 further species, making it a source for gene sequences and further experimental basis in plant and natural product focused biosynthesis research.

## 2. Results and Discussion

### 2.1. PCA Analysis of Volatile Compounds

Differences between the volatile compounds of four cultivars were investigated using a GC-MS Headspace analysis. Ten main volatile components visible between the cultivars



were selected, predominantly monoterpenoids and sesquiterpenoids, which are listed in Table 1. It has already been shown that there is a variety of monoterpene synthases that are able to catalyze ionization and isomerization starting from geranyl diphosphate [34]. Furthermore, the analysis of the cultivars revealed that a switch between pinene and limonene-derived compounds took place, which was sparsely synthesized in the other plants. In Table 1, these compounds are marked with an asterisk, one (\*) represents limonene-related terpenoids, and two (\*\*) represents pinene-related terpenoids. This especially is visible in the C4-C6 shift compared to the limonene backbone as seen in pinene (C4 to C6, see Figure S1). Similar substances could be identified as investigated previously for this plant species [19].

**Table 1.** Ten main volatile compounds of four *Caryopteris x clandonensis* cultivars, visible between the cultivars were selected and are hierarchically listed (top: higher concentration, bottom: lower concentration). GC-MS Headspace was performed and an identification with a NIST/EPA/NIH MS library version 2.0 was conducted. \* represents limonene-related terpenoids. \*\* represents pinene-related terpenoids.

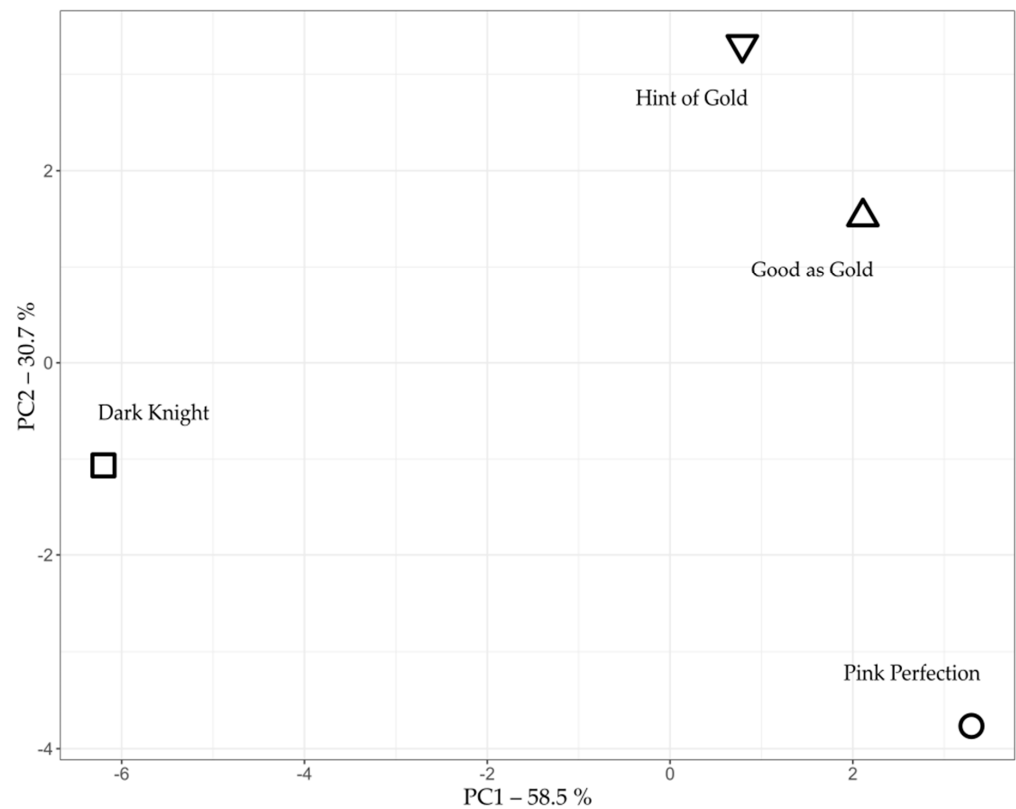
Dark Knight	Good as Gold	Hint of Gold	Pink Perfection
$\alpha$ -pinene **	D-limonene *	D-limonene *	D-limonene *
trans-pinocarveol **	Cubebol	Cubebol	cis-p-mentha-1(7),8-dien-2-ol *
Pinocarvone **	Carvone *	trans-carveol *	trans-p-mentha-2,8-dien-1-ol *
Caryophyllene oxide	trans-carveol *	Carvone *	Caryophyllene oxide
$\beta$ -pinene **	cis-p-mentha-1(7),8-dien-2-ol *	Caryophyllene oxide	trans-carveol *
(E,E)- $\alpha$ -farnesene	Caryophyllene oxide	trans-p-mentha-1(7),8-dien-2-ol *	cis-p-mentha-2,8-dien-1-ol *
$\alpha$ -campholenal	$\alpha$ -copaene	cis-p-mentha-1(7),8-dien-2-ol *	Carvone
$\alpha$ -copaene	$\beta$ -pinene **	cis-p-mentha-2,8-dien-1-ol *	$\alpha$ -pinene **
Caryophyllene	cis-p-mentha-2,8-dien-1-ol *	$\alpha$ -copaene	$\beta$ -pinene **
D-limonene *	trans-p-mentha-2,8-dien-1-ol *	trans-p-mentha-2,8-dien-1-ol *	Caryophyllene

As the plants are cultivars from *Caryopteris x clandonensis* a common base profile (e.g., caryophyllene, perillyl alcohol, sabinene, farnesene or campholenal) of volatiles was expected, see Table S1, as has been shown for other plants and their cultivars [35,36]. In this study, distinct differences between Dark Knight, Good as Gold, Hint of Gold, as well as Pink Perfection, can be shown.

To further investigate the variations in the compound profile found during the analysis, a principal component analysis (PCA) was performed (Figure 1). Good as Gold and Hint of Gold express high morphological and metabolic similarity (see Figure S2 and Table S1). This is also evident in Figure 1, as both cultivars are located close to each other. On the other hand, Dark Knight and Pink Perfection showed the highest deviation in volatile compound composition. Moreover, the switch between C1 and C6 as mentioned above results in an intriguing product spectrum. These data underline the variations between the cultivars and demonstrate a need for further investigations into the molecular makeup of underlying TPS and cytochrome p450 enzymes, which are key for generating the molecular diversity of plant-based terpenoid structures in plants [20]. Therefore, due to their distinct differences revealed in the PCA, the two cultivars, Dark Knight and Pink Perfection, were sequenced to elucidate genomic differences and identify unique and yet unknown genes.

## 2.2. Genome Sequencing and Quality Assessment

In Table 2, the sequencing metrics of the respective Sequel IIe runs are depicted. Details regarding sequencing quality reports can be found in Figure S3. Total bases were nearly twofold higher in Dark Knight than in Pink Perfection, the same as obtained HiFi reads and yield. However, the HiFi read length, read quality and number of passes are comparable in both sequencing runs. Deviations in sequencing parameters are closely related to utilized libraries and input DNA quality. As the read quality is well above Q20 both runs were subjected to further analyses.



**Figure 1.** A principal component analysis of four different *Caryopteris x clandonensis* cultivars, Dark Knight, Good as Gold, Hint of Gold and Pink Perfection regarding the area of their volatile compounds analyzed by GC-MS Headspace.

**Table 2.** Sequencing parameters of the PacBio Sequel IIe runs of *Caryopteris x clandonensis* cultivars Dark Knight and Pink Perfection.

Analysis Metric	Dark Knight	Pink Perfection
Total Bases (Gb)	444.13	229.43
HiFi Reads	1,823,939	843,632
HiFi Yield (Gb)	27.28	12.92
HiFi Read Length (mean, bp)	14,954	15,312
HiFi Read Quality (median)	Q35	Q34
HiFi Number of Passes (mean)	12	13

In this study, both genomes of Dark Knight and Pink Perfection were assembled using the IPA assembler with a consecutive duplicate purging and phasing step. A QAST analysis was conducted to assess assembly contiguity (see Table 3).

**Table 3.** Genome contiguity assessment based on statistics generated by using QAST.

Assembly	Dark Knight	Pink Perfection
# contigs	1183	782
Largest contig	29,672,976	31,977,049
Total length	366,625,098	344,117,456
Estimated reference length	300,000,000	300,000,000
GC (%)	31.50	31.77
N50	8,177,750	7,086,741
L50	13	14
# N's per 100 kbp	0.41	0.44

The number of assembled contigs diverged in both candidates (see Table 3). However, respective L50 values were small (13 for Dark Knight and 14 for Pink Perfection) compared to obtained N50 (8.2 Mb and 7.1 Mb respectively), which assures gene integrity with only low or no fragmentation. The total contig length of complete genomes corresponds to their size, which is comparable ( $3.44$  to  $3.66 \times 10^8$  bp), and the same as seen for GC content (31.5% and 31.77%). Furthermore, genome size was calculated using a k-mer-based analysis, with a k-mer size of 20. Results support the haploid genome size of ~355 Mb and estimated a diploid genome, see Figures S4 and S5. Based on the calculated genome size the coverage of Dark Knight and Pink Perfection resembles 74 and 38, respectively.

To assess the genome completeness and reliability of both genome sequences, a Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis was performed (see Figure 2). Both genomes were compared to the kingdom *Viridiplantae* and the clades *Embryophyta* and *Eudicotidae*, respectively. The selection of these lineages was based on the increasing grade of affiliation and the different accompanying BUSCO gene sets (in former order). For closer clades, more concise sequences are necessary in order to be identified as complete. In our case, even more affiliated clades show less deviation of completeness than expected in comparison to *Viridiplantae*. As the genomes were compared to different BUSCO datasets, the obtained results were depicted after normalization in Figure 2 to enable a concise comparison. Assessed genome completeness from the closest related clade (*Eudicotidae*) was 96.6% for Dark Knight and 96.8% for Pink Perfection, which were also compared to reference genomes of *Salvia splendens* (92.1%) [37] and *Sesamum indicum* (95.1%) [38]. The latter were only compared with the *Viridiplantae* database with BUSCO v2.0.1 and v3.0, whereas our data were analyzed by BUSCO v5.3.2. This may have caused the difference between 425 and 1440 BUSCO datasets, as frequent updates of the gene sets are necessary to improve BUSCO analysis [39]. The reference genomes were chosen due to the high prevalence in BLAST searches [40,41] using *Caryopteris x clandonensis* sequences. *S. splendens* appears to harbor mostly complete and duplicated BUSCOs, whereas *S. indicum* shows comparable results to the new genomes of Pink Perfection and Dark Knight with a majority of complete and single-copy BUSCOs. To interpret BUSCO results, it is necessary to understand duplicated BUSCOs and their nature, as these can be of biological or technical origin. In eukaryotic genomes, divergences in haplotypes often lead assemblers to form duplicates of high heterozygosity regions, resulting in contiguity issues and obstacles in further evaluation steps, such as gene annotation [42,43]. To circumvent these issues, tools such as “purge\_dups” are utilized to remove duplicate regions (haplotigs) from the assembly to assure genome contiguity [42]. A consecutive polishing of obtained contigs and haplotigs using phasing results in increased genome quality. Of the newly assembled genomes only 0.24–0.69%/0.71–2.67% are fragmented or missing, respectively. The absence of some BUSCO genes may be due to a loss of true genes or these may be existing as true gene duplications [43].

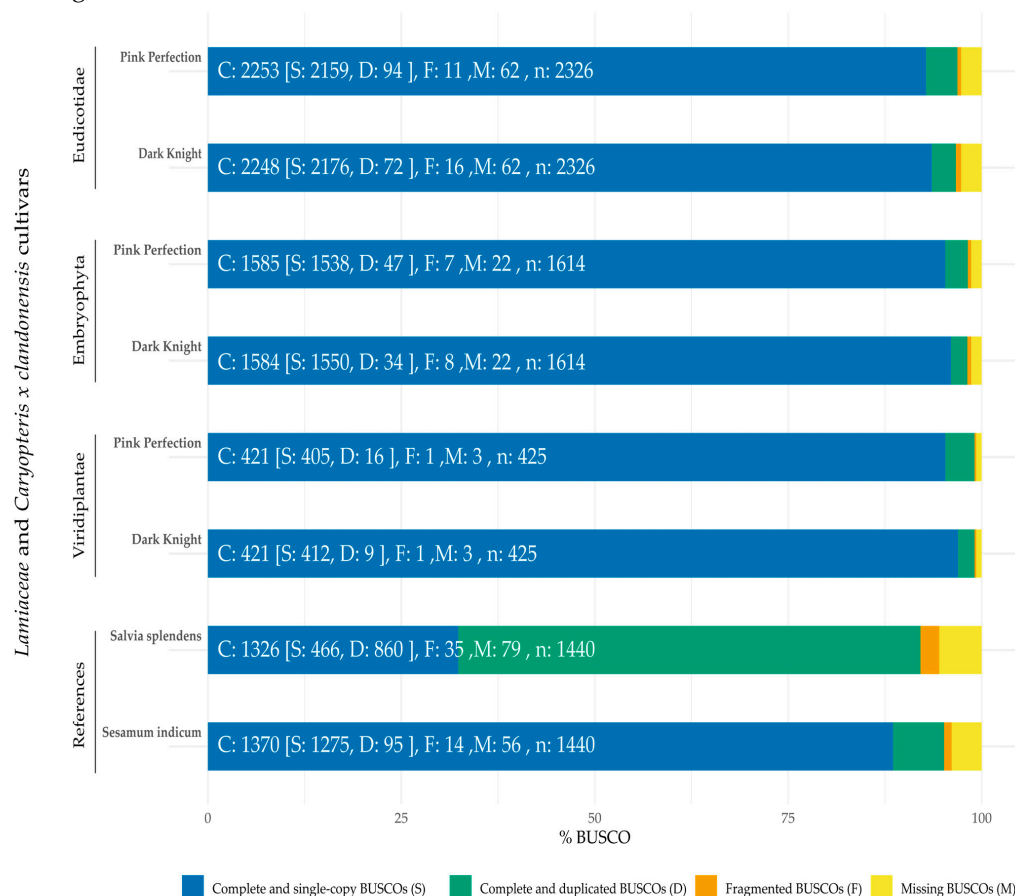
### 2.3. Evaluation of Structural Differences between Genome Assemblies

To concisely compare genomes, the collinear gene order also known as synteny or syntenic blocks needs to be assessed [44]. It plays an important role in visualizing matches between organisms [45].

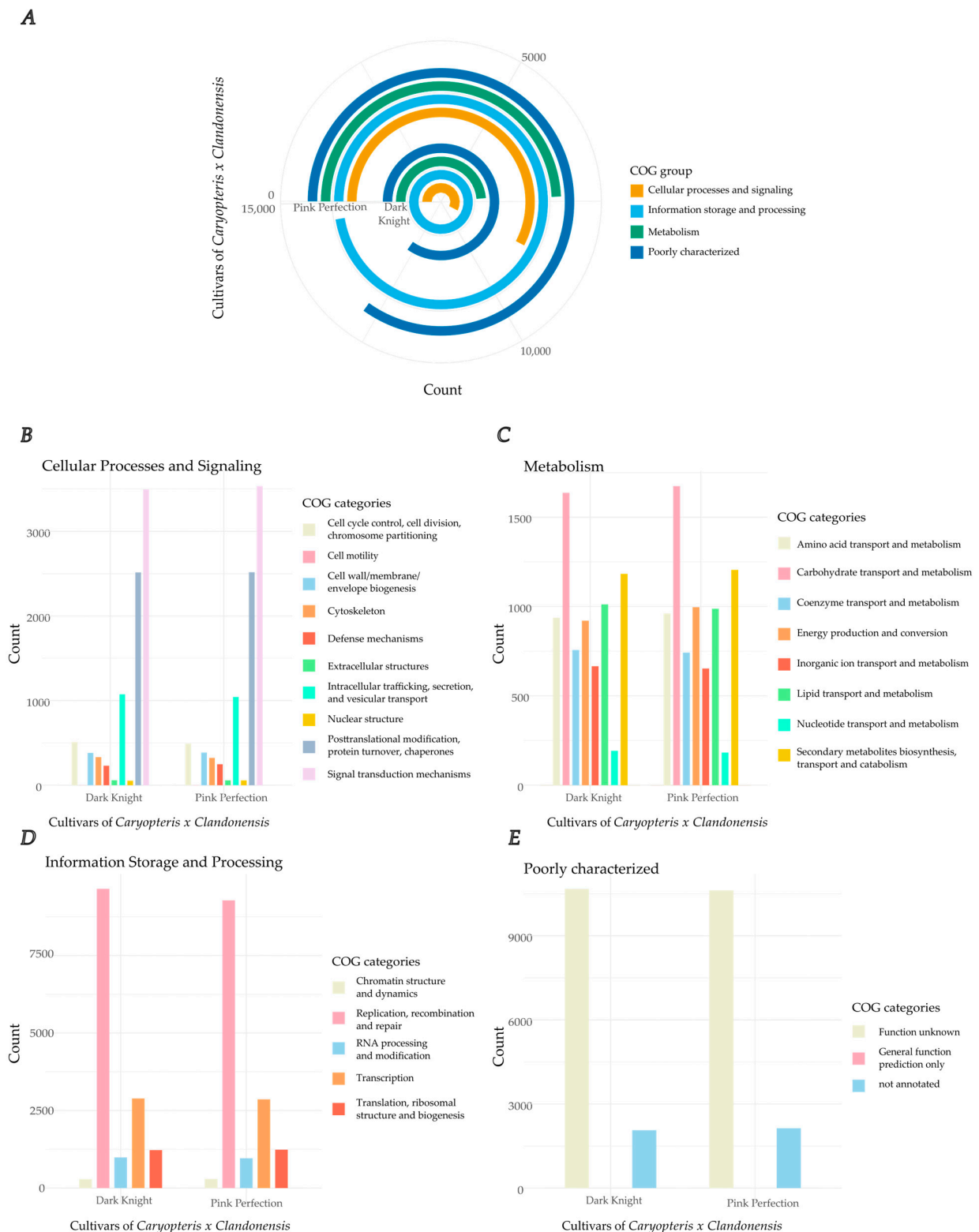
Investigating the synteny between cultivar genomes shows their close relation. Here, factors such as low contiguity and fragmentation have an effect on the analysis and lead to high error rates [46]. In our case, previously performed evaluations assured high contiguity and low fragmentation. Mauve was used to perform a multiple sequence alignment and applied to generate synteny blocks (Figure S6) [47]. Connections between these blocks reveal the high similarity within both genomes. This is typical for plant breeding, as specific traits are inherited from previous generations leading to inversions, duplications, or truncations in gene sets [48]. Furthermore, marker synteny can be used for phylogenetic analyses of cultivar evolution [49]. Thus, the plant samples seem to be closely related to the species *Caryopteris x clandonensis*.

#### 2.4. Gene Models and Functional Annotation

Gene models were computed using the presented genome assembly and a long-read IsoSeq database as hints via AUGUSTUS [50–53]. As a training set *Solanum lycopersicum* was chosen due to its ancestral relation to *Lamiaceae*. For the cultivars, a total of 52,865 (Dark Knight), and 51,315 (Pink Perfection) genes were predicted and resemble putative proteins. The Cluster of Orthologous Groups (COG) and Gene Ontology (GO) terms were evaluated for all cultivars. It is to mention that only ~81% of the predicted genes were annotated using COG and GO databases. Out of these ~30% are poorly characterized (Figure 3E) and only a fraction (30%) of those can be assigned with GO terms. In regard to the complete genomes, nearly 20% of the proposed gene models remain without an assigned function. Figure 3 shows the COG counts for the following categories: (3B) cellular processes and signaling (3C) information storage and processing (3D) metabolism and (3E) poorly characterized. Figure 3A combines all the aforementioned categories. The obtained results emphasize a strong similarity in the compared cultivars. Further data in regard to the exact amount of COG per category can be found in Tables S2 and S3. This finding is a further indicator of the completeness of the presented genomes, as different cultivars have a similar set of genes, only varying in small nucleotide polymorphisms or other structural variants, which distinguish them [54,55].



**Figure 2.** Comparison of BUSCO completeness of different cultivars of *Caryopteris x clandonensis* as well as *Salvia splendens* [37] and *Sesamum indicum* [38]. As the genomes were compared to other Benchmarking Universal Single-Copy Orthologs (BUSCO) datasets a normalization was performed to enable a comparison in genome completeness. Pink Perfection and Dark Knight were compared to the BUSCO datasets of *Viridiplantae*, *Embryophyta* and *Eudicotidae*, whereas *S. splendens* and *S. indicum* were compared to *Viridiplantae* only. Reference genomes were obtained from [37,38].



**Figure 3.** Annotation of gene sets for Cluster of Orthologous Groups (COG) for both cultivars, Dark Knight and Pink Perfection. (A) COG of two different cultivars of *Caryopteris x clandonensis*, Pink Perfection (outer ring) and Dark Knight (inner ring). Groups are divided in cellular processes and signaling, information storage and processing, metabolism, and a category for poorly characterized gene sets. (B) COG of cellular processes and signaling associated genes, total counts. (C) COG of metabolism-associated genes, total counts. (D) COG of information storage and processing associated genes, total counts. (E) COG of poorly characterized genes, total counts.

A closer look into the different groups reveals characteristic functions in the cultivars. Most genes identified and functionally annotated are associated with replication, recombination and repair, which make up about 20.5% of total annotated genes (Figure 3D) followed by signal transduction mechanism (~8%) (Figure 3A). Plants are exposed to endogenous and exogenous stresses such as chemicals or UV-radiation which can significantly alter DNA, thus there is high importance for repair mechanisms [56]. High redundancy of those ensures the safe replication of DNA with almost no errors [57].

In Figure 3C, proteins related to the COG category secondary metabolites biosynthesis, transport and catabolism, rank in second place within metabolism (2.8%). This category harbors TPS and cytochrome p450 enzymes. However, proteins associated with carbohydrate transport and metabolism are most abundant in this group as they are important for general metabolism and backbone synthesis.

Compared to about 29,458 with COG functionally annotated genes, 11,118 unique GO terms were assigned to 14,280 different genes (27% of total gene models). COG terms are ancestrally conserved regions, GO terminology in contrast proposes functional annotation of each hypothetical gene. A gene-set enrichment analysis was conducted with GO terms as a source for gene sets [58]. The following figures show the GO term clustering regarding the three main categories in plants: biological process (Figure 4), molecular function (Figure 5) and cellular components (Figure 6). For all three an analysis was conducted based on GO terms identified in Pink Perfection. Detailed data for Dark Knight and Pink Perfection can be found in Tables S4 and S5. The GO analysis was visualized using REVIGO [59]. Respective cluster position within the semantic space is irrelevant, as similar semantic terms are located in vicinity of each other in the plot [58].

In Figure 4, GO terms related to biological processes are depicted with their respective prevalence (dot size). In addition, some clusters with similar functions were grouped by circles into the main function of these GO terms, as can be seen, e.g., with “translation” in the bottom right corner. Incorporated into this cluster are the terms: protein modification process, DNA metabolic process, nucleobase-containing compound metabolic process, and protein metabolic process. The cluster organelle organization includes cytoskeleton organization, cytoplasm organization, and mitochondrion organization. Clustered with transport: ion transport, protein transport. The last cluster response to stress contains the GO terms response to a biotic stimulus, response to an abiotic stimulus, response to an external stimulus, and response to an endogenous stimulus. GO terms without clustering but still strongly prevalent in the PCA are biological, metabolic and biosynthetic processes.

For the GO analysis of the category molecular function, only one larger cluster was formed, which is nucleic acid binding. It incorporates the functions of DNA binding, RNA binding, and nucleotide binding. The two main components in this category are molecular function and catalytic activity.

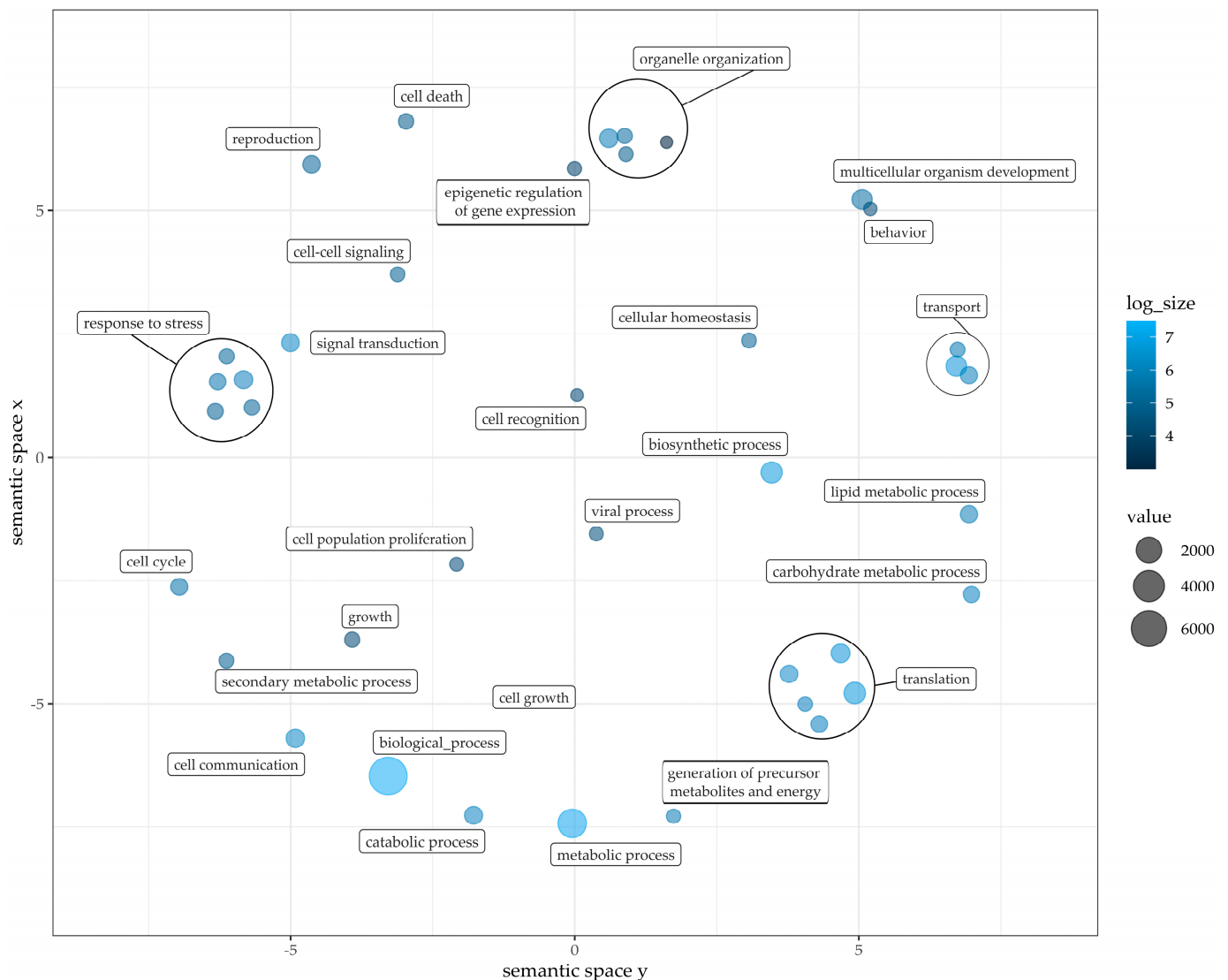
GO analysis in the category of cellular components yielded as the main results, intracellular anatomical structure and cellular components, as well as genes related to the cytoplasm. However, no semantic clustering was feasible based on the annotated GO terms.

### 2.5. Identification of Terpenoid Biosynthesis Enzymes

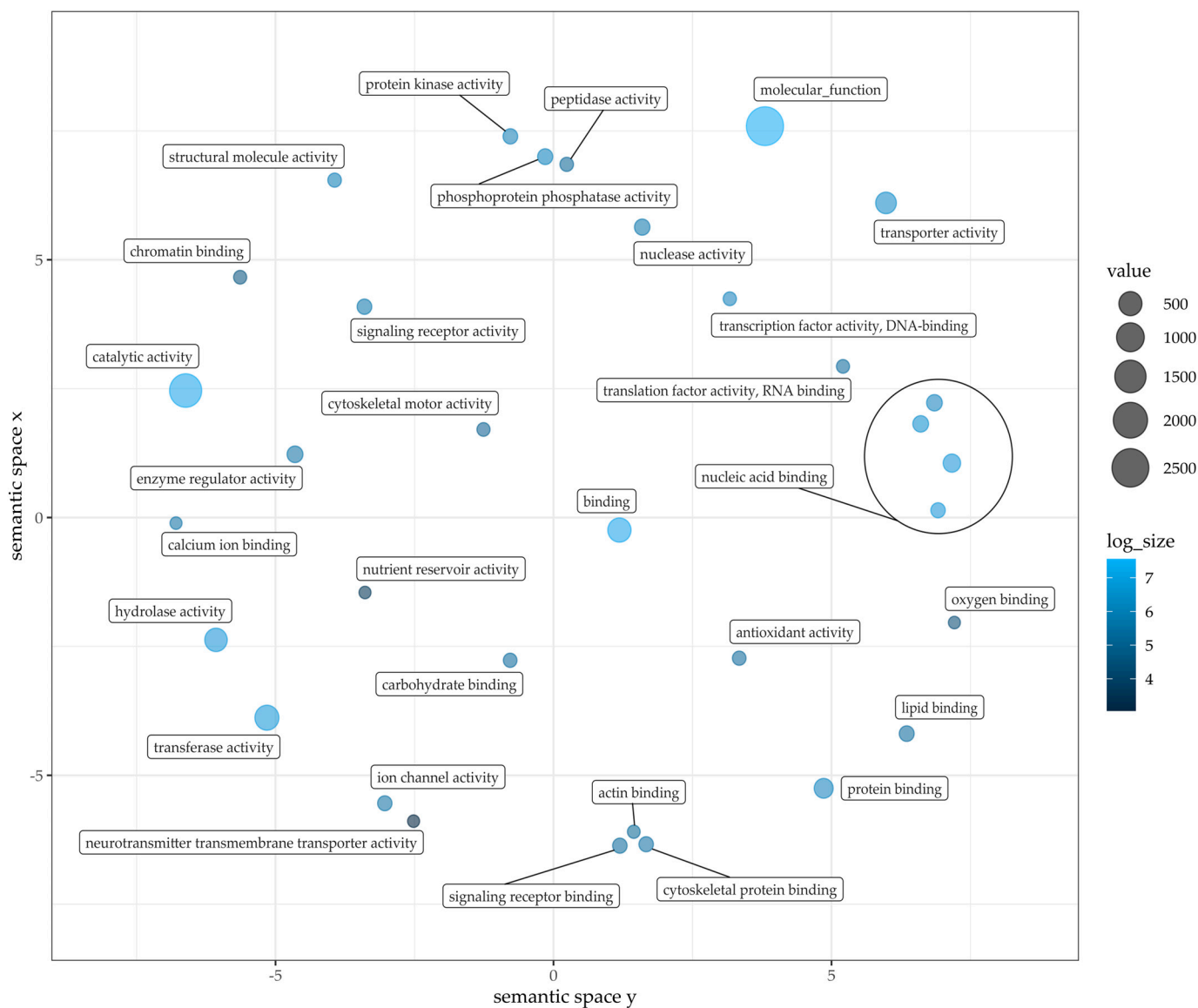
InterProScan predicts distinct protein domains and classifies them into families [60,61]. The seed files PF01397, PF03936 and IPR036965 are associated with TPS activity. In the annotated protein database, these seeds were used as homology motifs. For Dark Knight 43 and Pink Perfection 41 TPS were identified. The seed file IPR001128 is related to cytochrome p450 enzymes. Here, we were able to identify Dark Knight and Pink Perfection 1316 and 1363 sequences. Compared to other plants these findings are comparable, both for TPS and cytochrome p450 enzymes [62–65].

To investigate the similarity and the affiliation into TPS subfamilies regarding identified TPS, a phylogenetic tree was constructed. Analysis was based on multiple sequence alignment by Clustal Omega using default parameters (see Figure 7). To differentiate between TPS families, 55 selected sequences of representative plant species were utilized

as anchor sequences along with putative TPS from Dark Knight and Pink Perfection; the root was *Physcomitrella patens* (adapted from [66]). The multicolored clades belong to the different TPS subfamilies and are used as references, for a more detailed overview see the supplemental *lamiaceae* reference. Concise numbers of TPS subfamily distribution in both cultivars are shown in Table 4. The most prominent subfamilies are TPS-a (green), TPS-b (black) and TPS-c (purple), which is in line with the distribution in *Eudicots*, *Angiosperms* and land plants [67]. The subfamilies TPS-d and TPS-h are not present in the investigated cultivars. These findings are supported by the literature, as TPS-d clusters are derived from *Gymnosperm* species [63,68] and TPS-h are specific for *Selaginella moellendorffii* [67].



**Figure 4.** Gene Ontology term classification within biological processes of Pink Perfection. Clustered with response to stress: response to biotic stimulus, response to abiotic stimulus, response to external stimulus, response to endogenous stimulus. Clustered with translation: protein modification process, DNA metabolic process, nucleobase-containing compound metabolic process, protein metabolic process. Clustered with organelle organization: cytoskeleton organization, cytoplasm organization, mitochondrion organization. Clustered with transport: ion transport, protein transport. Figure was drafted employing REVIGO [59] and customized with R. Value and log size represents the counted GO terms across annotated gene models.

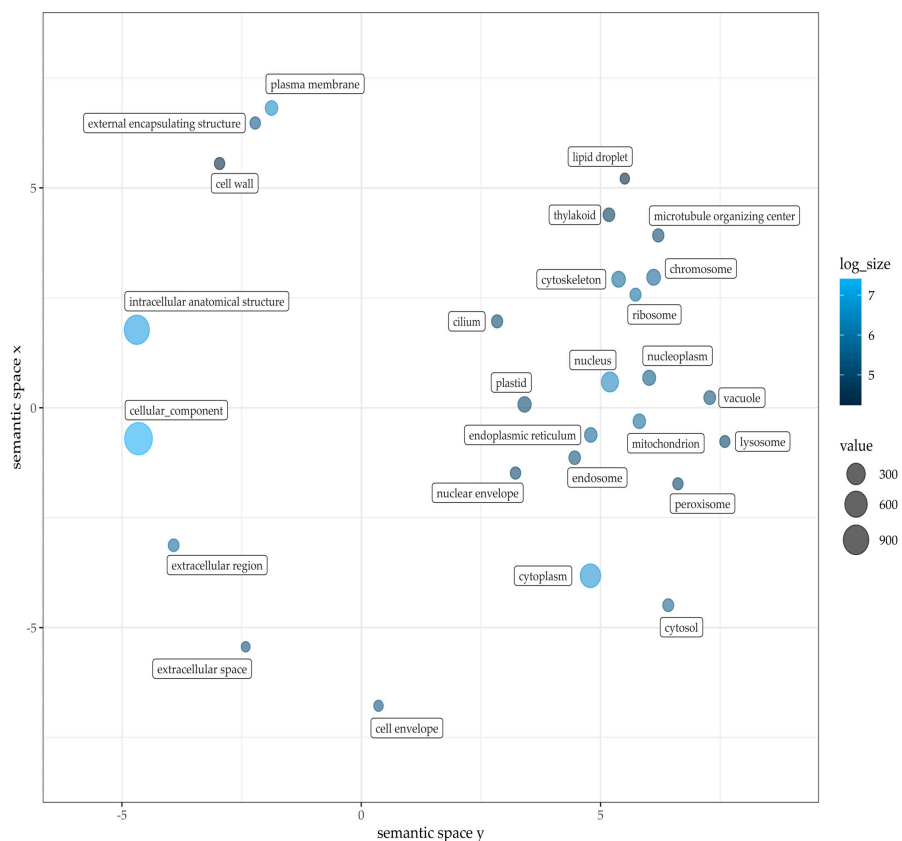


**Figure 5.** Gene Ontology term classification within molecular functions of Pink Perfection, clustered with nucleic acid binding: DNA binding, RNA binding, Nucleotide binding. Figure was drafted employing REVIGO [59] and customized with R. Value and log size represents the counted GO terms across annotated gene models.

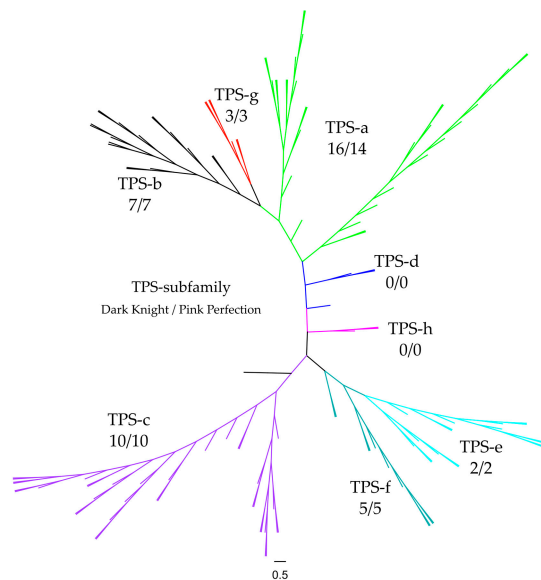
**Table 4.** Terpene synthase (TPS) subfamilies and their distribution in the *Caryopteris x clandonensis* cultivars Dark Knight and Pink Perfection. TPS-a, -b and -c show the highest prevalence in both cultivars.

TPS Subfamily	Dark Knight	Pink Perfection
a (green)	16	14
b (black)	7	7
c (purple)	10	10
d (blue)	-	-
e (turquoise)	2	2
f (petrol)	5	5
g (red)	3	3
h (pink)	-	-





**Figure 6.** Gene Ontology term classification within cellular components of Pink Perfection. Figure was drafted employing REVIGO [59] and customized with R. Value and log size represent the counted GO terms across annotated gene models.



**Figure 7.** Phylogenetic tree of putative terpene synthases (TPS) within *Caryopteris x clandonensis* cultivars Dark Knight (DK) and Pink Perfection (PP). TPS-a (green), TPS-b (black), TPS-c (purple), TPS-d (blue), TPS-e (turquoise), TPS-f (petrol), TPS-g (red), TPS-h (pink). For phylogenetic tree construction, TPS a-h of selected plant species were included to assure correct classification of identified TPS. Numbers below the respective TPS subfamily indicate the count of predicted TPS in the genomes of the cultivars.

### 3. Materials and Methods

#### 3.1. Plant Material

Four cultivars of *Caryopteris x clandonensis* were acquired from a local nursery (Foerster Pflanzen GmbH, Bietigheim-Bissingen, Germany) and grown to maturity in the open in a warm, moderate climate zone. After maturity, healthy leaves and blossoms were sampled and snap frozen in liquid nitrogen and stored at  $-80\text{ }^{\circ}\text{C}$  until preparation for transcriptome and genome sequencing. Fresh mature leaves were used for GC-MS headspace analysis of volatile compounds.

#### 3.2. GC-MS Analysis of Volatile Compounds

Fresh mature leaves were weighed in GC headspace vials and analyzed using a Trace GC-MS Ultra system with DSQII (Thermo Scientific, Waltham, MA, USA). Vials were incubated for 30 min at  $100\text{ }^{\circ}\text{C}$  and a TriPlus autosampler was used to inject  $500\text{ }\mu\text{L}$  of the sample in split mode onto a SGE BPX5 column (30 m, I.D.  $0.25\text{ mm}$ , film  $0.25\text{ }\mu\text{m}$ ); an injector temperature of  $280\text{ }^{\circ}\text{C}$  was used. The initial oven temperature was kept at  $50\text{ }^{\circ}\text{C}$  for 2.5 min. The temperature was increased with a ramp rate of  $10\text{ }^{\circ}\text{C}/\text{min}$  to  $320\text{ }^{\circ}\text{C}$  with a final hold for 5 min. Helium was used as a carrier gas with a flow rate of  $1.2\text{ mL}/\text{min}$  and a split ratio of 8. The MS chromatograms and spectra were recorded at  $70\text{ eV}$  (EI). Masses were detected between  $50\text{ }m/z$  and  $650\text{ }m/z$  in the positive mode [69]. Samples were measured in biological triplicates and the area average was used to compare peaks. Compounds were identified by spectral comparison with a NIST/EPA/NIH MS library version 2.0. To provide insight into the differentiation between plant samples a PCA was conducted.

#### 3.3. High Molecular Weight DNA Extraction and Library Preparation

High molecular weight genomic DNA (HMW gDNA) suitable for long-read sequencing was achieved using a plant-optimized CTAB—PCI extraction method based on different protocols [70–72];  $1\text{ g}$  of frozen, unthawed plant leaves were ground using a CryoMill (Retsch, Haan, Germany; three cycles, 6 min precool at  $5\text{ Hz}$ , disruption  $2:30\text{ min}$   $25\text{ Hz}$ , cooling between cycles  $0:30\text{ min}$  at  $5\text{ Hz}$ ). A CTAB extraction buffer (2% CTAB,  $100\text{ mM}$  Tris pH 8.0,  $20\text{ mM}$  EDTA,  $1.4\text{ M}$  NaCl) was supplemented with 2% PVP prior to usage and solved at  $60\text{ }^{\circ}\text{C}$ . The unthawed fine powder was mixed with  $5\text{ ml}$  buffer and incubated with  $200\text{ }\mu\text{L}$  Proteinase K (Qiagen, Venlo, The Netherlands) for 30 min at  $50\text{ }^{\circ}\text{C}$  and occasionally inverted. At room temperature,  $1\text{ mg}$  RNase A (Thermo Scientific, Waltham, MA, USA) was added and incubated for 10 min. The mixture was washed twice, saving and reusing the aqueous upper phase, with one volume PCI (25:24:1) and three times with chloroform ( $10,000\times g$ , 5 min,  $10\text{ }^{\circ}\text{C}$ ). To pellet the HMW gDNA, 30% PEG was added to the aqueous phase (1:4), inverted, incubated for 30 min on ice and spun for 30 min at  $12,000\times g$ ,  $10\text{ }^{\circ}\text{C}$ . The resulting shallow and colorless pellet was washed three times with 70% ethanol ( $5000\times g$ , 5 min,  $10\text{ }^{\circ}\text{C}$ ) and consequently, air dried at  $40\text{ }^{\circ}\text{C}$  and resuspended with  $100\text{ }\mu\text{L}$  elution buffer (Qiagen, Venlo, The Netherlands). Quality and size of the gDNA were assessed using a Qubit dsDNA HS Kit (Thermo Scientific, Waltham, MA, USA), a Nanodrop photometer (Implen, Munich, Germany) and a Femto Pulse system (Agilent, Santa Clara, CA, USA), respectively. If variations in DNA concentration between Qubit and Nanodrop were  $> 50\%$  an AMPure PB bead clean up or an electrophoretic clean up using a BluePippin system (Sage Science, Beverly, MA, USA) was performed;  $5\text{ }\mu\text{g}$  HMW gDNA were sheared in a gTube (Covaris, Woburn, MA, USA;  $1700\times g$ ) and used for whole genome library preparation using SMRTbell prep kit 3.0 (Pacific Biosciences, Menlo Park, CA, USA) according to the manufacturer's recommendations. Size selection of the resulting library was performed using AMPure PB beads. Libraries were stored at  $-20\text{ }^{\circ}\text{C}$ . Prior sequencing, primer and polymerase were bound using a Sequel II Binding Kit 3.2 (Pacific Biosciences, Menlo Park, CA, USA) according to the manufacturer's recommendations.

### 3.4. Genome Sequencing and Assembly

Sequencing was performed on a Sequel IIE (Pacific Biosciences, Menlo Park, CA, USA) with two hours pre-extension, two hours adaptive loading (target  $p1 + p2 = 0.95$ ) to an on-plate concentration of 85 pM, and 30 h movie time. The initial de novo genome assembly was performed using SMRT Link (v11.0.0+, Pacific Biosciences, Menlo Park, CA, USA) which uses Improved Phased Assembly (IPA) [73]. After polishing, the contigs were divided into primary and haplotype-associated contigs using `purge_dups` [74].

The assembled sequences can be found within the National Center for Biotechnology Information (NCBI). BioSample accession number: Dark Knight SAMN32308289, Pink Perfection SAMN32308290.

### 3.5. RNA Long Read IsoSeq

To increase the quality of the genome assembly, long-read transcripts were sequenced to add more depth and accuracy to the proposed gene models. For RNA extraction, frozen, unthawed leaves were ground using a CryoMill and an RNeasy Plant Mini Kit (Qiagen, Venlo, Niederlande). A Turbo DNA free Kit (Invitrogen) was used to further clean the RNA. The high-quality RNA was used to perform an IsoSeq library prep using SMRTbell prep kit 3.0 and Sequel II Binding Kit 3.2. (Pacific Biosciences, Menlo Park, CA, USA).

### 3.6. Bioinformatic and Statistical Analysis

Gene models were prepared through AUGUSTUS [50–53] using genomic data and long-read transcriptomic data as hints. Quality and completeness of the genome were estimated with QUAST (v5.2.0) [31] and BUSCO (v5.3.2) [39,43,75,76]. NCBI BLAST (v2.12.0+) [40,41] and InterProScan (v5.54-87) [60,61] were computed on a local computational unit. This analysis provided an annotation that was the basis for the determination of distinct protein families, in this case, terpene synthases and cytochrome p450 enzymes. EggNOG Mapper (v2.1.5) was used to determine COG and GO terms. Statistical analysis and figures were conducted using R (v4.2.1, revigo [59] and cateGORizer [77]). Synteny analysis was performed using Mauve [47] (v2.4.0) and Geneious Prime (Geneious). For k-mer analysis jellyfish (v2.3.0) [78] was used (k-mer size: 20). GenomeScope [79,80] was used for the visualization of k-mer frequencies. The following analyses were conducted using galaxy project [81]: BUSCO, QUAST, EggNOG, Jellyfish, and GenomeScope. If not further specified default parameters were used for analysis.

### 3.7. Identification of TPS and Cytochrome p450 Enzymes

Genes associated with these protein classes were found using InterProScan and the domain seed files IPR036965 (TPS activity) and IPR01128 (cytochrome p450 enzymes). The phylogenetic tree was constructed using a global alignment with Blosum62. As a genetic distance model, Jukes–Cantor was chosen along with Neighbor-Joining as the Tree building method. The outlier was *Physcomitrella patens*, XP\_024380398. Software used: Geneious Prime (Geneious).

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/plants12030632/s1>, Figure S1: Chemical structure of D-limonene backbone and difference to C6–C4 shift in  $\alpha$ -pinene, Figure S2: Cultivars of *Caryopteris x clandonensis* used in this manuscript, Figure S3: PacBio sequencing quality reports of different *Caryopteris x clandonensis* cultivars, Figure S4: GenomeScope profile of k-mer analysis of Dark Knight, Figure S4: GenomeScope profile of k-mer analysis of Pink Perfection, Figure S6: Synteny evaluation between the *Caryopteris x clandonensis* cultivars, Table S1: GC-MS Headspace data of TOP30 identified compounds via NIST database Table S2: Data Pink Perfection COG, Table S3: Data Dark Knight COG, Table S4: Data Pink Perfection GO cluster, Table S5: Data Dark Knight GO cluster, Supplemental Lamiaceae Reference: Phylogenetic tree references in FASTA format.

**Author Contributions:** Conceptualization, M.R., N.A. and N.M.; methodology, M.R. and N.A.; software, M.R., N.A. and N.M.; validation, M.R. and N.A.; formal analysis, M.R. and N.A.; investigation, M.R., N.A.; resources, T.B.; data curation, M.R. and N.A.; writing—original draft preparation, M.R. and N.A.; writing—review and editing, M.R., N.A., N.M. and T.B.; visualization, M.R.; supervision, N.M. and T.B.; project administration, N.M. and T.B.; funding acquisition, T.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the German Federal Ministry of Education and Research, grant number 031B0824A.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data available in a publicly accessible repository. The data presented in this study are openly available in National Center for Biotechnology Information (NCBI). BioSample accession number: Dark Knight SAMN32308289, Pink Perfection SAMN32308290.

**Acknowledgments:** M.R., N.A., N.M. and T.B. gratefully acknowledge the support of colleagues at the Werner Siemens Chair for Synthetic Biotechnology during conducting experiments and writing this manuscript. Furthermore, all authors want to thank Foerstner Pflanzen GmbH for providing plant materials.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

- Caputi, L. Use of terpenoids as natural flavouring compounds in food industry. *Recent Pat. Food Nutr. Agric.* **2011**, *3*, 9–16. [[CrossRef](#)] [[PubMed](#)]
- Masyita, A.; Sari, R.M.; Astuti, A.D.; Yasir, B.; Rumata, N.R.; Emran, T.B.; Nainu, F.; Simal-Gandara, J. Terpenes and terpenoids as main bioactive compounds of essential oils, their roles in human health and potential application as natural food preservatives. *Food Chem. X* **2022**, *13*, 100217. [[CrossRef](#)] [[PubMed](#)]
- da Silva, G.L.; Luft, C.; Lunardelli, A.; Amaral, R.H.; Melo, D.A.D.S.; Donadio, M.V.; Nunes, F.B.; DE Azambuja, M.S.; Santana, J.C.; Moraes, C.M.; et al. Antioxidant, analgesic and anti-inflammatory effects of lavender essential oil. *An. Acad. Bras. Cienc.* **2015**, *87*, 1397–1408. [[CrossRef](#)] [[PubMed](#)]
- Mediratta, P.; Sharma, K.; Singh, S. Evaluation of immunomodulatory potential of *Ocimum sanctum* seed oil and its possible mechanism of action. *J. Ethnopharmacol.* **2002**, *80*, 15–20. [[CrossRef](#)] [[PubMed](#)]
- da Silva, J.K.R.; Figueiredo, P.L.B.; Byler, K.G.; Setzer, W.N. Essential oils as antiviral agents, potential of essential oils to treat SARS-CoV-2 infection: An in-silico investigation. *Int. J. Mol. Sci.* **2020**, *21*, 3426. [[CrossRef](#)] [[PubMed](#)]
- Abdollahi, M.; Karimpour, H.; Monsef-Esfehani, H.R. Antinociceptive effects of *Teucrium polium* L. total extract and essential oil in mouse writhing test. *Pharmacol. Res.* **2003**, *48*, 31–35. [[CrossRef](#)] [[PubMed](#)]
- Đorđević, S.; Petrović, S.; Dobrić, S.; Milenković, M.; Vučićević, D.; Žižić, S.; Kukić, J. Antimicrobial, anti-inflammatory, anti-ulcer and antioxidant activities of *Carlina acanthifolia* root essential oil. *J. Ethnopharmacol.* **2007**, *109*, 458–463. [[CrossRef](#)]
- Cowen, D.; Wolf, A.; Paige, B.H. Toxoplasmic encephalomyelitis. *Arch. Neurol. Psychiatry* **1942**, *48*, 689–739. [[CrossRef](#)]
- Jantan, I.; Ping, W.O.; Visuvalingam, S.D.; Ahmad, N.W. Larvicidal activity of the essential oils and methanol extracts of Malaysian plants on *Aedes aegypti*. *Pharm. Biol.* **2008**, *41*, 234–236. [[CrossRef](#)]
- Cox-Georgian, D.; Ramadoss, N.; Dona, C.; Basu, C. Therapeutic and medicinal uses of terpenes. *Med. Plants Farm Pharm.* **2019**, *67*, 333–359. [[CrossRef](#)]
- Sicora, O. The ethanolic stem extract of *Caryopteris x Clandonensis* Posseses antiproliferative potential by blocking breast cancer cells in mitosis. *Farmacologia* **2019**, *67*, 1077–1082. [[CrossRef](#)]
- Wani, M.C.; Taylor, H.L.; Wall, M.E.; Coggon, P.; Mcphail, A.T. Plant antitumor agents. VI. The isolation and structure of Taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *J. Am. Chem. Soc.* **1971**, *93*, 2325–2327. [[CrossRef](#)]
- Weaver, B.A. How Taxol/paclitaxel kills cancer cells. *Mol. Biol. Cell* **2014**, *25*, 2677–2681. [[CrossRef](#)]
- Pichersky, E.; Raguso, R.A. Why do plants produce so many terpenoid compounds? *New Phytol.* **2018**, *220*, 692–702. [[CrossRef](#)]
- Holopainen, J.K.; Himanen, S.J.; Yuan, J.S.; Chen, F.; Stewart, C.N. *Ecological Functions of Terpenoids in Changing Climates*; Ramawat, K., Mérillon, J.M., Eds.; Natural Products; Springer: Berlin/Heidelberg, Germany, 2013. [[CrossRef](#)]
- Drapeau, J.; Rossano, M.; Touraud, D.; Obermayr, U.; Geier, M.; Rose, A.; Kunz, W. Green synthesis of para-Menthane-3,8-diol from *Eucalyptus citriodora*: Application for repellent products. *Comptes Rendus Chim.* **2011**, *14*, 629–635. [[CrossRef](#)]
- Lee, S.Y.; Kim, S.H.; Hong, C.Y.; Park, S.Y.; Choi, I.G. Biotransformation of (–)- $\alpha$ -pinene and geraniol to  $\alpha$ -terpineol and p-menthane-3,8-diol by the white rot fungus, *Polyporus brumalis*. *J. Microbiol.* **2017**, *53*, 462–467. [[CrossRef](#)]

18. Drapeau, J.; Verdier, M.; Touraud, D.; Kröckel, U.; Geier, M.; Rose, A.; Kunz, W. Effective insect repellent formulation in both surfactantless and classical microemulsions with a long-lasting protection for human beings. *Chem. Biodivers.* **2009**, *6*, 934–947. [[CrossRef](#)]
19. Blythe, E.; Tabanca, N.; Demirci, B.; Bernier, U.; Agramonte, N.; Ali, A.; Khan, I. Composition of the essential oil of Pink Chablis™ bluebeard (*Caryopteris × clandonensis* 'Durio') and its biological activity against the yellow fever mosquito *Aedes aegypti*. *Nat. Volatiles Essent. Oils* **2015**, *2*, 11–21.
20. Bathe, U.; Tissier, A. Cytochrome P450 enzymes: A driving force of plant diterpene diversity. *Phytochemistry* **2019**, *161*, 149–162. [[CrossRef](#)]
21. Hernandez-Ortega, A.; Vinaixa, M.; Zebec, Z.; Takano, E.; Scrutton, N.S. A toolbox for diverse Oxyfunctionalisation of monoterpenes. *Sci. Rep.* **2018**, *8*, 1–8. [[CrossRef](#)]
22. Mabou, F.D.; Belinda, I.; Yossa, N. TERPENES: Structural classification and biological activities. *IOSR J. Pharm. Biol. Sci. e-ISSN* **2021**, *16*, 2319–7676.
23. Nett, R.S.; Montanares, M.; Marcassa, A.; Lu, X.; Nagel, R.; Charles, T.C.; Hedden, P.; Rojas, M.C.; Peters, R.J. Elucidation of gibberellin biosynthesis in bacteria reveals convergent evolution. *Nat. Chem. Biol.* **2016**, *13*, 69–74. [[CrossRef](#)] [[PubMed](#)]
24. Wang, T.; Li, L.; Zhuang, W.; Zhang, F.; Shu, X.; Wang, N.; Wang, Z. Recent research progress in Taxol biosynthetic pathway and acylation reactions mediated by *Taxus Acyltransferases*. *Molecules* **2021**, *26*, 2855. [[CrossRef](#)] [[PubMed](#)]
25. Wen, W.; Yu, R. Artemisinin biosynthesis and its regulatory enzymes: Progress and perspective. *Pharmacogn. Rev.* **2011**, *5*, 189–194. [[CrossRef](#)]
26. Gershenzon, J.; Dudareva, N. The function of terpene natural products in the natural world. *Nat. Chem. Biol.* **2007**, *3*, 408–414. [[CrossRef](#)]
27. Zhang, X.; Niu, M.; da Silva, J.A.T.; Zhang, Y.; Yuan, Y.; Jia, Y.; Xiao, Y.; Li, Y.; Fang, L.; Zeng, S.; et al. Identification and functional characterization of three new terpene synthase genes involved in chemical defense and abiotic stresses in *Santalum album*. *BMC Plant Biol.* **2019**, *19*, 115. [[CrossRef](#)]
28. Sharma, V.; Sarkar, I.N. Bioinformatics opportunities for identification and study of medicinal plants. *Brief. Bioinform.* **2013**, *14*, 238–250. [[CrossRef](#)]
29. Helfrich, E.J.N.; Lin, G.-M.; Voigt, C.A.; Clardy, J. Bacterial terpene biosynthesis: Challenges and opportunities for pathway engineering. *Beilstein J. Org. Chem.* **2019**, *15*, 2889–2906. [[CrossRef](#)]
30. Amarsinghe, S.L.; Su, S.; Dong, X.; Zappia, L.; Ritchie, M.E.; Gouil, Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* **2020**, *21*, 1–16. [[CrossRef](#)]
31. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [[CrossRef](#)]
32. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)]
33. Chen, G.; Mostafa, S.; Lu, Z.; Du, R.; Cui, J.; Wang, Y.; Liao, Q.; Lu, J.; Mao, X.; Chang, B.; et al. The Jasmine (*Jasminum sambac*) genome provides insight into the biosynthesis of flower fragrances and Jasmonates. *Genom. Proteom. Bioinform.* **2022**, in press. [[CrossRef](#)]
34. Degenhardt, J.; Köllner, T.G.; Gershenzon, J. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* **2009**, *70*, 1621–1637. [[CrossRef](#)]
35. Zhu, X.; Li, Q.; Li, J.; Luo, J.; Chen, W.; Li, X. Comparative study of volatile compounds in the fruit of two banana cultivars at different ripening stages. *Molecules* **2018**, *23*, 2456. [[CrossRef](#)]
36. Cramer, A.-C.J.; Mattinson, D.S.; Fellman, J.K.; Baik, B.-K. Analysis of volatile compounds from various types of barley cultivars. *J. Agric. Food Chem.* **2005**, *53*, 7526–7531. [[CrossRef](#)]
37. Dong, A.-X.; Xin, H.-B.; Li, Z.-J.; Liu, H.; Sun, Y.-Q.; Nie, S.; Zhao, Z.-N.; Cui, R.-F.; Zhang, R.-G.; Yun, Q.-Z.; et al. High-quality assembly of the reference genome for scarlet sage, *Salvia splendens*, an economically important ornamental plant. *Gigascience* **2018**, *7*, giy068. [[CrossRef](#)]
38. Li, C.; Li, X.; Liu, H.; Wang, X.; Li, W.; Chen, M.-S.; Niu, L.-J. Chromatin architectures are associated with response to dark treatment in the oil crop *Sesamum indicum*, based on a high-quality genome assembly. *Plant Cell Physiol.* **2020**, *61*, 978–987. [[CrossRef](#)]
39. Manni, M.; Berkeley, M.R.; Seppely, M.; A Simão, F.; Zdobnov, E.M. BUSCO Update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **2021**, *38*, 4647–4654. [[CrossRef](#)]
40. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
41. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [[CrossRef](#)]
42. Guan, D.; A McCarthy, S.; Wood, J.; Howe, K.; Wang, Y.; Durbin, R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **2020**, *36*, 2896–2898. [[CrossRef](#)] [[PubMed](#)]
43. Manni, M.; Berkeley, M.R.; Seppely, M.; Zdobnov, E.M. BUSCO: Assessing genomic data quality and beyond. *Curr. Protoc.* **2021**, *1*, e323. [[CrossRef](#)] [[PubMed](#)]

44. Tang, H.; Lyons, E.; Pedersen, B.; Schnable, J.C.; Paterson, A.H.; Freeling, M. Screening synteny blocks in pairwise genome comparisons through integer programming. *BMC Bioinform.* **2011**, *12*, 102. [[CrossRef](#)] [[PubMed](#)]
45. Lee, J.; Hong, W.-Y.; Cho, M.; Sim, M.; Lee, D.; Ko, Y.; Kim, J. Synteny Portal: A web-based application portal for synteny block analysis. *Nucleic Acids Res.* **2016**, *44*, W35–W40. [[CrossRef](#)] [[PubMed](#)]
46. Liu, D.; Hunt, M.; Tsai, I.J. Inferring synteny between genome assemblies: A systematic evaluation. *BMC Bioinform.* **2018**, *19*, 1–13. [[CrossRef](#)]
47. Darling, A.E.; Mau, B.; Perna, N.T. progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **2010**, *5*, e11147. [[CrossRef](#)]
48. Arús, P.; Toshiya, Y.; Elisabeth, D.; Abbott, A.G. Synteny in the Rosaceae. In *Plant Breeding Reviews*; John and Wiley and Sons: Hoboken, NJ, USA, 2010; pp. 175–211.
49. Devos, K.M.; Moore, G.; Gale, M.D. Conservation of marker synteny during evolution. *Euphytica* **1995**, *85*, 67–372. [[CrossRef](#)]
50. Hoff, K.J.; Lomsadze, A.; Borodovsky, M.; Stanke, M. Whole-genome annotation with BRAKER. *Methods Mol. Biol.* **2019**, *1962*, 65–95. [[CrossRef](#)]
51. Hoff, K.J.; Lange, S.; Lomsadze, A.; Borodovsky, M.; Stanke, M. BRAKER1: Unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS: Table 1. *Bioinformatics* **2016**, *32*, 767–769. [[CrossRef](#)]
52. Brúna, T.; Hoff, K.J.; Lomsadze, A.; Stanke, M.; Borodovsky, M. BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.* **2021**, *3*, lqaa108. [[CrossRef](#)]
53. Stanke, M.; Schöffmann, O.; Morgenstern, B.; Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinform.* **2006**, *7*, 62. [[CrossRef](#)]
54. Yang, Z.; Ge, X.; Yang, Z.; Qin, W.; Sun, G.; Wang, Z.; Li, Z.; Liu, J.; Wu, J.; Wang, Y.; et al. Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. *Nat. Commun.* **2019**, *10*, 2989. [[CrossRef](#)]
55. Liu, S.; An, Y.; Tong, W.; Qin, X.; Samarina, L.; Guo, R.; Xia, X.; Wei, C. Characterization of genome-wide genetic variations between two varieties of tea plant (*Camellia sinensis*) and development of InDel markers for genetic research. *BMC Genom.* **2019**, *20*, 1–16. [[CrossRef](#)]
56. Chatterjee, N.; Walker, G.C. Mechanisms of DNA damage, repair, and mutagenesis. *Environ. Mol. Mutagen.* **2017**, *58*, 235–263. [[CrossRef](#)] [[PubMed](#)]
57. Raina, A.; Sahu, P.K.; Laskar, R.A.; Rajora, N.; Sao, R.; Khan, S.; Ganai, R.A. Mechanisms of genome maintenance in plants: Playing it safe with breaks and bumps. *Front. Genet.* **2021**, *12*, 675686. [[CrossRef](#)]
58. Lim, C.; Pratama, M.Y.; Rivera, C.; Silvestro, M.; Tsao, P.S.; Maegdefessel, L.; Gallagher, K.A.; Maldonado, T.; Ramkhalawon, B. Linking single nucleotide polymorphisms to signaling blueprints in abdominal aortic aneurysms. *Sci. Rep.* **2022**, *12*, 20990. [[CrossRef](#)]
59. Supek, F.; Bošnjak, M.; Škunca, N.; Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **2011**, *6*, e21800. [[CrossRef](#)]
60. Blum, M.; Chang, H.-Y.; Chuguransky, S.; Grego, T.; Kandasamy, S.; Mitchell, A.; Nuka, G.; Paysan-Lafosse, T.; Qureshi, M.; Raj, S.; et al. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* **2021**, *49*, D344–D354. [[CrossRef](#)]
61. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240. [[CrossRef](#)]
62. Butler, J.B.; Freeman, J.; Potts, B.M.; Vaillancourt, R.; Grattapaglia, D.; Silva-Junior, O.B.; Simmons, B.; Healey, A.L.; Schmutz, J.; Barry, K.; et al. Annotation of the *Corymbia terpena* synthase gene family shows broad conservation but dynamic evolution of physical clusters relative to *Eucalyptus*. *Heredity* **2018**, *121*, 87–104. [[CrossRef](#)]
63. Warren, R.L.; Keeling, C.I.; Yuen, M.M.S.; Raymond, A.; Taylor, G.A.; Vandervalk, B.P.; Mohamadi, H.; Paulino, D.; Chiu, R.; Jackman, S.D.; et al. Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. *Plant J.* **2015**, *83*, 189–212. [[CrossRef](#)] [[PubMed](#)]
64. Jia, K.-H.; Liu, H.; Zhang, R.-G.; Xu, J.; Zhou, S.-S.; Jiao, S.-Q.; Yan, X.-M.; Tian, X.-C.; Shi, T.-L.; Luo, H.; et al. Chromosome-scale assembly and evolution of the tetraploid *Salvia splendens* (*Lamiaceae*) genome. *Hortic. Res.* **2021**, *8*, 1–15. [[CrossRef](#)] [[PubMed](#)]
65. Chen, Z.; Vining, K.J.; Qi, X.; Yu, X.; Zheng, Y.; Liu, Z.; Fang, H.; Li, L.; Bai, Y.; Liang, C.; et al. Genome-wide analysis of terpene synthase gene family in *Mentha longifolia* and catalytic activity analysis of a single terpene synthase. *Genes* **2021**, *12*, 518. [[CrossRef](#)] [[PubMed](#)]
66. Hamilton, J.P.; Godden, G.T.; Lanier, E.; Bhat, W.W.; Kinser, T.J.; Vaillancourt, B.; Wang, H.; Wood, J.C.; Jiang, J.; Soltis, P.S.; et al. Generation of a chromosome-scale genome assembly of the insect-repellent terpenoid-producing *Lamiaceae* species, *Callicarpa americana*. *Gigascience* **2020**, *9*, gaa093. [[CrossRef](#)] [[PubMed](#)]
67. Chen, F.; Tholl, D.; Bohlmann, J.; Pichersky, E. The family of terpene synthases in plants: A mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.* **2011**, *66*, 212–229. [[CrossRef](#)]
68. Shalev, T.J.; Yuen, M.M.S.; Gesell, A.; Yuen, A.; Russell, J.H.; Bohlmann, J. An annotated transcriptome of highly inbred *Thuja plicata* (*Cupressaceae*) and its utility for gene discovery of terpenoid biosynthesis and conifer defense. *Tree Genet. Genomes* **2018**, *14*, 35. [[CrossRef](#)]
69. Ringel, M.; Reinbold, M.; Hirte, M.; Haack, M.; Huber, C.; Eisenreich, W.; Masri, M.A.; Schenk, G.; Guddat, L.W.; Loll, B.; et al. Towards a sustainable generation of pseudopterisin-type bioactives. *Green Chem.* **2020**, *22*, 6033–6046. [[CrossRef](#)]

70. Inglis, P.W.; Pappas, M.; Resende, L.V.; Grattapaglia, D. Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLoS ONE* **2018**, *13*, e0206085. [[CrossRef](#)]
71. Healey, A.; Furtado, A.; Cooper, T.; Henry, R.J. Protocol: A simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods* **2014**, *10*, 21. [[CrossRef](#)]
72. Rogers, S.O.; Bendich, A.J. Extraction of total cellular DNA from plants, algae and fungi. *Plant Mol. Biol. Man.* **1994**, *2*, 183–190. [[CrossRef](#)]
73. GitHub—PacificBiosciences/pbipa: Improved Phased Assembler. Available online: <https://github.com/PacificBiosciences/pbipa> (accessed on 11 December 2022).
74. GitHub—dfguan/purge\_dups: Haplotypic Duplication Identification Tool. Available online: [https://github.com/dfguan/purge\\_dups](https://github.com/dfguan/purge_dups) (accessed on 11 December 2022).
75. Kriventseva, E.V.; Kuznetsov, D.; Tegenfeldt, F.; Manni, M.; Dias, R.; A Simão, F.; Zdobnov, E.M. OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* **2019**, *47*, D807–D811. [[CrossRef](#)] [[PubMed](#)]
76. Seppey, M.; Manni, M.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol. Biol.* **2019**, *1962*, 227–245. [[CrossRef](#)] [[PubMed](#)]
77. Hu, Z.-L.; Bao, J.; Reecy, J. CateGORizer: A web-based program to batch analyze gene ontology classification categories. *Online J. Bioinform.* **2008**, *9*, 108–112. Available online: <http://www.animalgenome.org/bioinfo/tools/catego/> (accessed on 11 December 2022).
78. Marçais, G.; Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **2011**, *27*, 764–770. [[CrossRef](#)]
79. Vurture, G.W.; Sedlazeck, F.J.; Nattestad, M.; Underwood, C.J.; Fang, H.; Gurtowski, J.; Schatz, M.C. GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics* **2017**, *33*, 2202–2204. [[CrossRef](#)]
80. Ranallo-Benavidez, T.R.; Jaron, K.S.; Schatz, M.C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **2020**, *11*, 1432. [[CrossRef](#)]
81. Afgan, E.; Baker, D.; Batut, B.; van den Beek, M.; Bouvier, D.; Čech, M.; Chilton, J.; Clements, D.; Coraor, N.; Grüning, B.A.; et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **2018**, *46*, W537–W544. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Differential RNA-Seq Analysis Predicts  
Genes Related to Terpene Tailoring in  
*Caryopteris x Clandonensis*





*plants*

IMPACT  
FACTOR  
**4.658**

Indexed in:  
**PubMed**

Article

---

# Differential RNA-Seq Analysis Predicts Genes Related to Terpene Tailoring in *Caryopteris* × *clandonensis*

---

Manfred Ritz, Nadim Ahmad, Thomas Brueck and Norbert Mehlmer

Special Issue

Recent Advances in Plant Genomics and Transcriptome Analysis

Edited by

Dr. Nam-Soo Kim



<https://doi.org/10.3390/plants12122305>

## Article

# Differential RNA-Seq Analysis Predicts Genes Related to Terpene Tailoring in *Caryopteris* × *clandonensis*

Manfred Ritz <sup>†</sup>, Nadim Ahmad <sup>†</sup>, Thomas Brueck <sup>\*†</sup> and Norbert Mehlmer <sup>\*†</sup>

Werner Siemens Chair of Synthetic Biotechnology, Department of Chemistry, Technical University of Munich (TUM), 85748 Garching, Germany; manfred.ritz@tum.de (M.R.); nadim.ahmad@tum.de (N.A.)

\* Correspondence: brueck@tum.de (T.B.); norbert.mehlmer@tum.de (N.M.)

† These authors contributed equally to this work.

**Abstract:** Enzymatic terpene functionalization is an essential part of plant secondary metabolite diversity. Within this, multiple terpene-modifying enzymes are required to enable the chemical diversity of volatile compounds essential in plant communication and defense. This work sheds light on the differentially transcribed genes within *Caryopteris* × *clandonensis* that are capable of functionalizing cyclic terpene scaffolds, which are the product of terpene cyclase action. The available genomic reference was subjected to further improvements to provide a comprehensive basis, where the number of contigs was minimized. RNA-Seq data of six cultivars, Dark Knight, Grand Bleu, Good as Gold, Hint of Gold, Pink Perfection, and Sunny Blue, were mapped on the reference, and their distinct transcription profile investigated. Within this data resource, we detected interesting variations and additionally genes with high and low transcript abundancies in leaves of *Caryopteris* × *clandonensis* related to terpene functionalization. As previously described, different cultivars vary in their modification of monoterpenes, especially limonene, resulting in different limonene-derived molecules. This study focuses on predicting the cytochrome p450 enzymes underlying this varied transcription pattern between investigated samples. Thus, making them a reasonable explanation for terpenoid differences between these plants. Furthermore, these data provide the basis for functional assays and the verification of putative enzyme activities.

**Keywords:** terpene biosynthesis; cytochrome p450; *Caryopteris* × *clandonensis*; long read sequencing; transcriptomics; chemical diversity; volatile compound



**Citation:** Ritz, M.; Ahmad, N.; Brueck, T.; Mehlmer, N. Differential RNA-Seq Analysis Predicts Genes Related to Terpene Tailoring in *Caryopteris* × *clandonensis*. *Plants* **2023**, *12*, 2305. <https://doi.org/10.3390/plants12122305>

Academic Editors: Andreas W. Ebert and Nam-Soo Kim

Received: 25 April 2023

Revised: 17 May 2023

Accepted: 7 June 2023

Published: 13 June 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

*Caryopteris* × *clandonensis* is an ornamental plant, also known as “Bluebeard”, which is phylogenetically classified in the *Lamiaceae* family. It is easily cultivated and rich in volatile compounds. These, and other molecules detected and described, are terpenes, e.g.,  $\alpha$ -copaene, limonene, or  $\delta$ -cadinene [1], terpene derivatives, e.g., keto-glycosides, clandonosides, and harpagides [2], as well as the pyrano-juglon derivative  $\alpha$ -caryopteron [3]. The species’ essential oil was found to display mosquito-repellent activity; however, the active agent for this mode of action was not yet detected [4]. The *Lamiaceae* family is known to harbor an interesting and valuable profile in secondary metabolites, including terpenoids, flavonoids, and phenylpropanoids [5–7]. These compounds play important roles in the plant’s interaction with its environment [8,9] as for the defense against abiotic and biotic stresses [10]. They also harbor potential in pharmaceutical or industrial applications, as seen for taxol [11], menthol [12], malvidin [13], isoliquiritigenin [14] or umbelliferone [15]. In general, terpenes and terpenoids are a molecule class, which is produced in vast varieties by flowering plants [16] and is involved in a wide range of biological activities. Essential oils and their monoterpenes, such as  $\alpha$ -pinene and limonene, were investigated in terms of their anti-inflammatory and virucidal activity in recent studies [17–19]. Moreover, other terpenoids employ antibacterial properties [20] while others act as insecticides [4], are used

as allelochemicals [21], or as attractants for pollinators [22]. The backbone of plant-derived terpenes is produced via the mevalonate pathway. For this, the precursors dimethylallyl diphosphate (DMAPP) and the functional isomer isopentenyl pyrophosphate (IPP) can be connected via isoprenyl diphosphate synthases (IDS) to form larger units of terpenes. IPP consists of five C-atoms (hemiterpene) whereas, through condensation of IPP and DMAPP via IDS monoterpenes (C10), sesquiterpenes (C15), diterpenes (C20), and higher terpene structures are built [23]. Further tailoring of these basic terpenes is conducted by terpene synthases (TPS) and cytochrome p450 enzymes (CYPs). Plant TPS mediate complex carbocation reactions, resulting in various cyclic structures of higher terpenes [24,25]. These can be divided into eight subfamilies (TPSa-h) which can be clade- or even species-specific [26]. The first step in tailoring monoterpenes is hydroxylation. Subsequently, CYPs are mediating a plethora of further reactions to enhance the functionalization (carboxylation, acetylation or forming peroxides) [27,28]. Due to their promiscuity towards substrates, only a few enzymes are necessary to yield various terpenoid structures and, therefore, differences in their functions and modality [29]. Multiple sequences of different source organisms are available in curated databases [30,31]. These allow easy access to the genetic information on these enzymes. With CYPs occurring in all living organisms [30], the enzymes, similarly to TPS, are divided for better identification, whereas specific CYP families are reserved for each type of organism. Plant CYP families can be found in CYP71-99 and CYP701-999, and in a four-digit scheme from CYP7001-9999 [32]. The categorization into these classes is dependent on sequence similarity. The same family (Arabic number) needs matching amino acids  $\geq 40\%$  and the subfamily (Arabic letter)  $\geq 55\%$  [33]. Therefore, the CYP76S40 [34] is the 40th individual enzyme from the CYP76S subfamily and the CYP76 family. This way, after annotation, contaminating sequences can be discarded solely due to their classification in a non-plant CYP family.

One approach to elucidate variations in the enzymatic makeup and investigate the sequences underlying terpene diversity is to compare differentially expressed gene (DEG) products at a quantitative level using modern bioinformatics tools. Differences in the metabolite profile exist during different stages of plant growth [35]. Different genes are regulated from seedlings to mature plants to translate their genomic information into proteins and interact in plant differentiation, protection or communication, depending on their developmental state [36]. During plant breeding, deletion, duplications, mutations or fragmentations can occur. Therefore, a distinct set of genes varies in its nucleotide code and their transcription or translation rate, resulting in different phenotypes in the mature plant [37]. The data can be levied and evaluated regarding efficacy to investigate these differences. The number of transcripts does not solely result in higher protein outcome, but also in, respectively, higher concentrations of secondary metabolites. Therefore, differential expression analysis can identify genes or gene products responsible for either the stress response mechanisms observed for abiotic stressors, such as drought or radiation, or as has been shown for biotic stressors, such as pests and plant reactions to herbivores [38]. Typical DEG experiments harness the up- and down-regulation of genes after induction or shock, e.g., during exposure to chemicals [39] or different environments [40]. Another possibility is the investigation of specific traits of plant cultivars due to their variations between hybrid plants [41]. Previously, the variations in *Caryopteris*  $\times$  *clandonensis* volatile compound setup was investigated, and a difference in the synthesis of limonene-derived molecules (LDM) was observed [1]. The cultivar Dark Knight was detected to harbor a low amount, whereas Pink Perfection shows high amounts of LDM. These variations were discovered without a distinct change in their TPS or CYP makeup.

To that end, we show that the identification of terpene variety between different plant cultivars can be pursued on a molecular level using a quantitative bioinformatics method such as RNA-Seq analysis. Furthermore, we focus on terpene functionalization enzymes, especially cytochrome p450 enzymes, to elucidate the mechanisms behind the variations in monoterpene modifications as seen for limonene [1].

## 2. Results and Discussion

### 2.1. RNA Sequencing and Mapping Quality

Samples subjected to short-read sequencing were taken from leaves of six *Caryopteris × clandonensis* cultivars known to show differences in their LDM profile, Dark Knight (DK), Grand Bleu (GB), Good as Gold (GG), Hint of Gold (HG), Sunny Blue (SB), and Pink Perfection (PP). Sequencing was performed using an Illumina NovaSeq platform, which generated about 20 million raw reads in bases for each sample. The reads were processed to remove low-quality reads, bases, and adapter sequences, resulting in the clean reads used for downstream analysis. After this purification step, a loss of 5.0 to 14.9 million bases was seen between the samples. In Table 1, the run as well as cleaning and mapping statistics are summarized. The Q20 and Q30 scores indicate the sequencing quality, with Q30 indicating a lower error rate than Q20. This experiment's high Q20 and Q30 scores suggest that the sequencing quality was highly sufficient, with only a few sequencing errors. Moreover, the clean reads exhibit a slight increase in quality scores, persistent throughout all samples.

**Table 1.** Statistics of short Illumina reads used for mapping on the reference genome (NCBI SAMN32308290 (Pink Perfection, PP)). A paired-end run was employed on a NovaSeq6000 SP (2 × 150 bp) for sequencing.

<i>Caryopteris × clandonensis</i> Cultivar		Raw Reads in Bases		Q20 in %	Q30 in %	Clean Reads in Bases		Q20 in %	Q30 in %	Totally Mapped in %	Uniquely Mapped in %
		Unique	Duplicate			Unique	Duplicate				
Dark Knight	R1	24,501,785	19,238,555	99.95	94.76	13,072,273	29,945,355	99.99	95.08	87.8	79.3
	R2	26,380,719	17,359,621	99.25	87.90	16,204,470	26,813,158	99.46	88.25		
Grand Bleu	R1	17,917,215	51,971,129	99.85	93.75	11,552,426	57,659,626	99.98	94.21	85.8	76.8
	R2	18,808,258	51,080,086	99.51	92.12	13,260,446	55,951,606	99.68	92.42		
Good as Gold	R1	22,797,327	27,074,692	99.60	94.52	13,160,322	31,359,084	99.95	95.08	86.7	75.4
	R2	25,142,438	24,729,581	99.35	89.41	16,112,061	28,407,345	99.54	89.76		
Hint of Gold	R1	20,547,645	20,953,229	99.89	94.67	15,165,071	33,935,373	99.98	95.06	86.4	77.2
	R2	23,044,700	18,456,174	99.38	88.40	18,084,814	31,015,630	99.56	88.81		
Sunny Blue	R1	20,535,582	25,022,034	99.96	94.96	13,908,152	27,181,140	99.99	95.35	87.0	80.5
	R2	22,771,085	22,786,531	99.39	88.30	16,573,745	24,515,547	99.56	88.60		
Pink Perfection	R1	25,751,312	28,610,858	99.96	94.23	12,295,625	26,046,846	99.99	94.60	87.7	82.0
	R2	29,512,685	24,849,485	99.40	90.42	14,649,539	23,692,932	99.58	90.70		

The available genome sequences from *Caryopteris × clandonensis* PP [1] were subjected to further cleaning and improvement steps to curb the influence of contamination. A binning algorithm, MetaBAT2 [42], usually used for metagenomic data, was used on the long-read assembly of the genome and differentiated into 40 bins. The completeness and contiguity were checked and, in summary, the 782 scaffolds/848 contigs, which add up to 344 Mb with a genome completeness score of 96.8%, were reduced to 53 scaffolds/88 contigs, which add up to 298 Mb and a BUSCO score of 96.5%. The utilized BUSCO gene sets belonged to the closest affiliate *Eudicotidae*. Detailed information can be found in Table S1. This refined genome was used as a reference for mapping the short-read sequences. A preliminary mapping of DK transcripts on the respective long-read genomic data, compared to mapping the transcripts on the PP genomic data, revealed an increased assignment of unique reads. Thus, the genome of *Caryopteris × clandonensis* PP was chosen as a mapping reference for both cultivars, DK and PP, resulting in a more comprehensive downstream analysis. The exact mapping counts for the different methods can be found in Table S2.

The percentages of reads mapped to the reference genome, as seen in Table 1, indicate the data accuracy and low presence of contaminating DNA. The amount of uniquely mapped reads is also an important metric, as it indicates the proportion of reads that map to a unique location in the reference genome. A high percentage of uniquely mapped reads (greater than 70%) is desirable, reducing the possibility of mapping errors or ambiguous mapping locations [43]. In our setting, we were able to accurately map between 85.8% and 87.8% of the sequences, indicating that a large proportion of the reads were successfully located on the provided genome. Furthermore, the percentage of uniquely mapped reads ranged from 75.4% to 82.0%, which is reasonably high and suggests that the quality of the sequencing reads was sufficient to allow for exact mapping and is suited for downstream

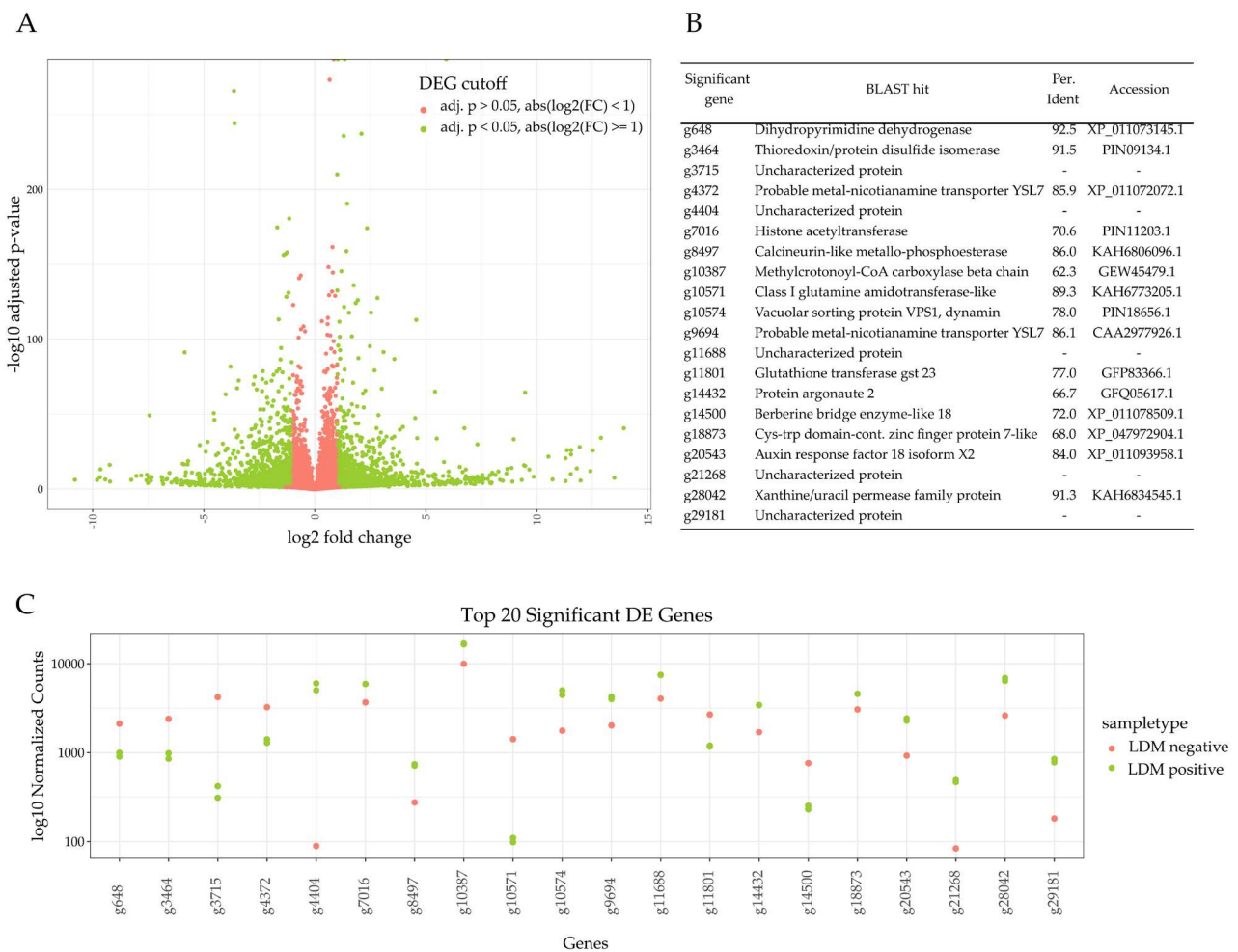
analysis. The observed duplication rates varied between 5.7% and 11.3%, and are well-known in plant transcript mapping due to transcript isoforms [44].

## 2.2. Identification of DEG

To identify the mechanism behind the modification of LDM, we wanted to focus on the DEGs between the cultivars of *Caryopteris × clandonensis*. Therefore, the mapping data were subset and pooled into highly LDM-positive (SB, PP) and highly LDM-negative (DK, GB) cultivars. The cultivars GG and HG were neither highly LDM-positive nor highly LDM-negative, therefore both were disregarded during the initial DEG analysis. From the 29,210 predicted genes in the mapping reference, 23,477 were observed to map in all investigated sets. The DEGs were filtered using a log<sub>2</sub> fold-change cutoff of absolute values greater than 1, and an adjusted *p*-value of a minimum of 0.05, thereby the values for each cultivar were transcribed at least two-fold. The values fitting these parameters are highlighted in green; those which were disregarded during further analysis, because of not fitting the parameters, are shown in red. Compared to the genes close to the middle, there are a few genes with high fold-changes in LDM-positive plants, compared to LDM-negative and those with significantly higher or lower transcript abundance. After filtering the DEGs between LDM-positive and LDM-negative cultivars, 3305 genes were identified, as seen in Figure 1A. For 100 genes, no Pfam class [45] and, for a further 168, no EggNOG [46] description, could be assigned. Regarding the DEGs, a closer look reveals the 20 most diverged genes, which can be seen in Figure 1B,C. Half of the annotated genes are still uncharacterized, or their distinct function is unknown, according to the cluster of orthologous groups. Interestingly, the genes associated with metal transport and metal binding are differentially transcribed, as seen for g4372, g9694, and g8497. These functions are known to be responsible for catalyzing redox reactions in plants [47,48]. Examining DEGs further, g14432 is associated with the protein argonaute family and g1887 is a zinc finger-like protein, whereas g3464 is a thioredoxin/disulfide isomerase. These proteins regulate biological processes [49], as well as responses to abiotic stresses such as drought stress [50,51]. In general, these DEGs describe the effects on the primary metabolism and stress response of plants; however, they do not show any direct participation in tailoring secondary metabolites within the plants. CYPs, in particular, are iron-binding; however, a connection between the upregulation of metal-transporting proteins and CYPs cannot be drawn from this data. The biosynthesis of LDM is not artificially induced in one cultivar or silenced in the other. Thus, a specific and significant transcription of related terpenetailoring genes cannot be observed. To elucidate these mechanisms, it is necessary to take a closer look into the DEGs of CYPs [28,52].

## 2.3. Terpene Tailoring through CYPs between Plant Cultivars

The identified 3305 DEGs can be further filtered into genes related to CYPs due to conserved domains and the corresponding CYP Pfam class. Here, the domain PF00067 was integrated into IPR001128. Both domains are indicators for sequences associated with the cytochrome p450 superfamily (IPR036396) [53]. This homology-based search allowed the identification of 70 putative sequences with different total lengths. Assuming a minimum size of 29 kDa for a CYP, 61 genes remain. From a statistical point of view, the average size of this pool amounts to a median of 1485 nucleotides, corresponding to the average size of a translated protein of 54.5 kDa. This is also reported in the literature, with an average plant CYP molecular mass between 45 and 62 kDa [54,55]. In regards to the identification of LDM-modifying enzymes, this subset is necessary to obtain a detailed overview into CYPs. These enzymes are known to play a huge part in terpene diversity in plants [56]. They are able to catalyze the hydroxylation of different backbones due to their substrate promiscuity [29,57]. Therefore, the transcript abundance of specific CYPs may reveal the mechanism behind LDM variances in this plant.

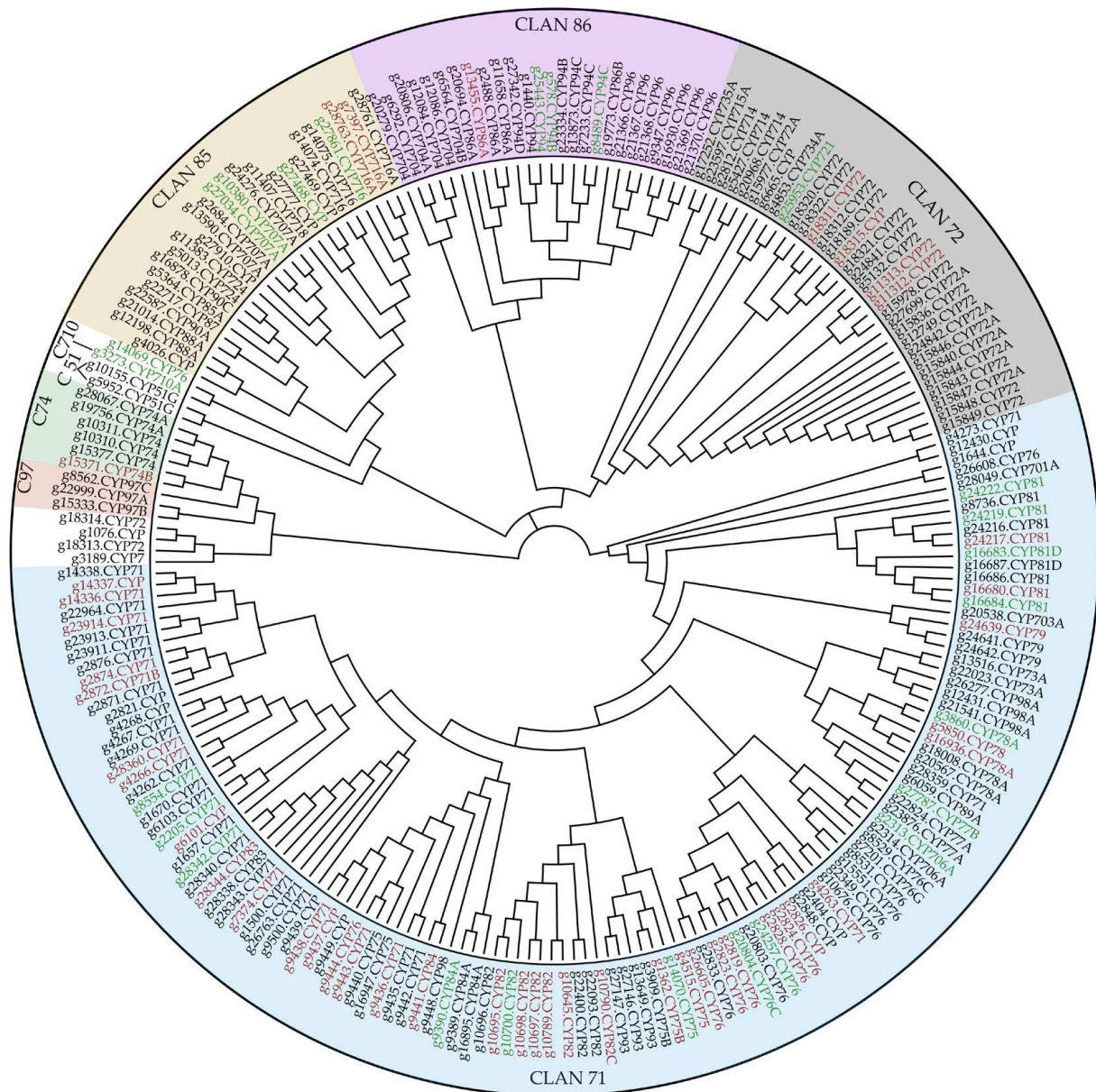


**Figure 1.** Differential expressed genes (DEGs) of *Caryopteris* × *clandonensis* cultivars highly producing limonene-derived molecules (LDM-positive) and cultivars which produce lower amounts of LDM (LDM-negative). Cultivars used for LDM-positive subset: Sunny Blue and Pink Perfection, and for LDM-negative subset Dark Knight and Grand Bleu. (A) The volcano plot of DEG was identified between the LDM-positive vs. LDM-negative plant cultivar subsets. Absolute log<sub>2</sub> fold-change cutoff was set to 1 and an adjusted *p*-value of 0.05 was used to assign the DEGs; values fitting these parameters are highlighted in green and those which were disregarded during further analysis are shown in red. (B) Top 20 most significantly transcribed genes and their respective description, including BLAST search percentage identity and determined accession for the putative assignment. (C) log<sub>10</sub> normalized counts of the top 20 significant DEG in this setup. Genes from LDM-positive samples are displayed in green, those corresponding to LDM-negative samples are highlighted in red.

Out of all the 23,477 mapped genes, 221 CYPs were detected, whereas 61 showed differences in transcript abundance. In Figure 2, all identified CYPs are visualized in an unrooted phylogenetic tree. CYPs with high transcript abundance in LDM-positive cultivars are highlighted in green, whereas CYPs with low transcript abundance are represented in red.

To allocate the putative CYPs to their distinct family or subfamily, the Pfam-classified CYP sequences were subjected to a BLAST search using a custom CYP database [54]. The sequences were assigned to the same subfamily if the percent identity was above 55%, and to the same family if greater than 40%. Eight CYP clans were highlighted within the found enzymes, CLAN51, CLAN71, CLAN72, CLAN74, CLAN710, CLAN85, CLAN86, and CLAN 97. This highlights that the major classes 71 and 72 are found to be involved, primarily, in the terpene tailoring of different terpene classes [28]. For CYP71, a variety of monoterpene modifications are described [34,58–61]. In our setting, most DEGs were observed in this clan. The enzymes related to CYP72 are described as tailoring triterpenoids

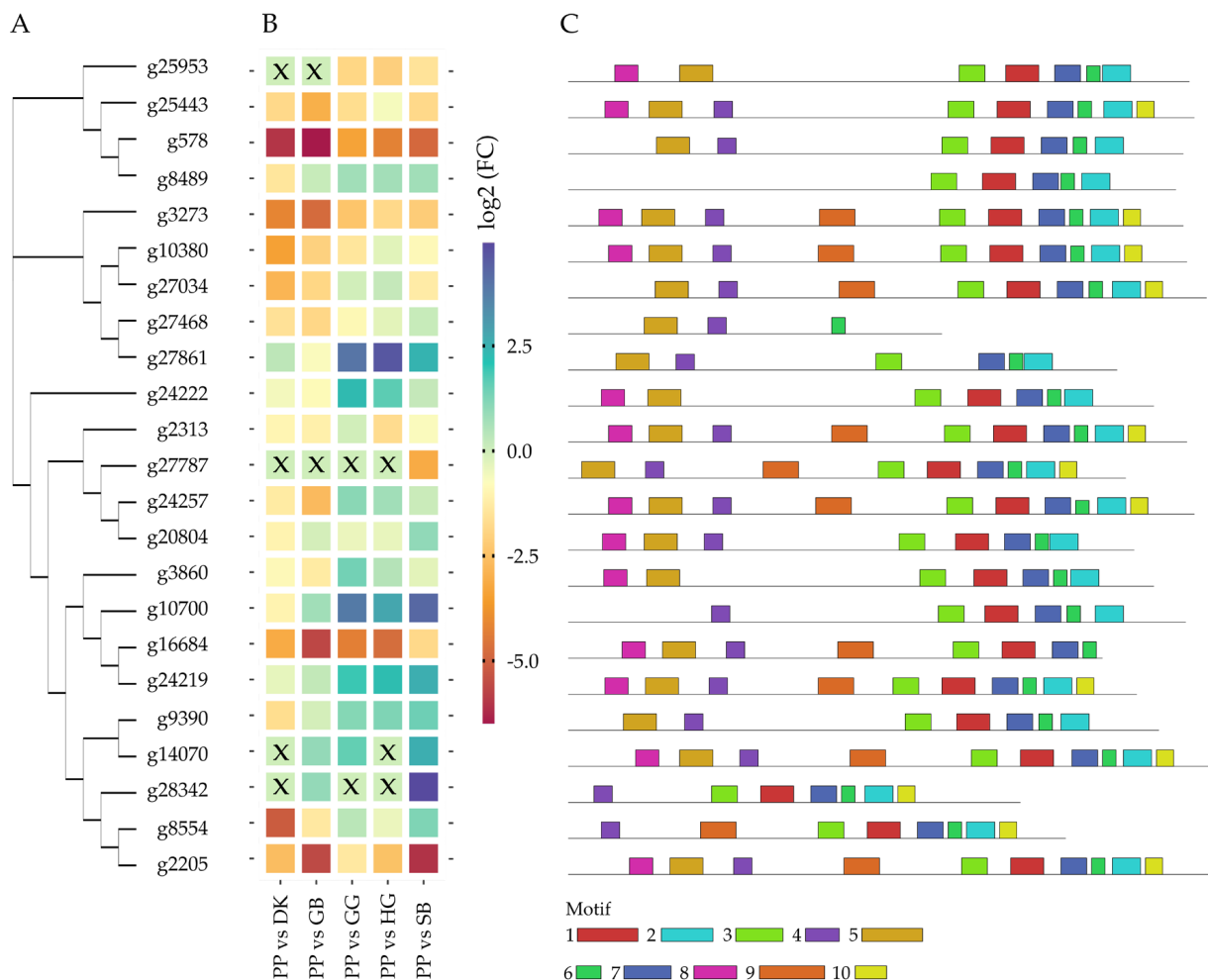
as saponins, characterized within plant defensive mechanisms against biotic stressors such as herbivores or microbes [38,62].



**Figure 2.** Phylogenetic tree of all transcribed cytochrome p450 (CYP) enzymes within the six investigated cultivars. Clan localization is highlighted on the outer ring. Differentially expressed genes (DEGs) were marked in green for a high transcript abundance in limonene-derived-molecules-positive cultivars and red for low transcript abundance, as seen in their fold-change differences. The tree was constructed using the following parameters: Global alignment with a Blosum62 cost matrix, Genetic distance model Jukes-Cantor, Neighbor-Joining and no outgroup was used, gap open penalty was set to 12, and gap extension penalty to 3 during pairwise alignments.

DEGs with high transcript abundance in LDM-positive samples were used to compare the genes between all sequenced cultivars. PP and SB were considered highly LDM-positive, whereas DK and GB were LDM-negative. GG and HG were in between and, therefore, were excluded in the initial DEG analysis. For the comparison of CYPs between the four previously mentioned samples and the two latter samples, the CYPs found in LDM-positive and LDM-negative samples were searched in GG and HG, and the normalized counts of all samples were compared. PP was chosen, due to its LDM profile, as a setpoint to compare

the transcript abundance between all samples. In, the results of a comparative approach are visualized. The phylogenetic distance between the identified CYPs is shown in 3A. Three clusters can be differentiated, with the first seen in the upper part consisting of 4 genes (g25953, g25443, g578, g8489), the second in the middle (g3273, g10380, g27034, g27468, g27861), and the third cluster with 14 genes (g24222, g2313, g27787, g24257, g20804, g3860, g10700, g16684, g24219, g9390, g14070, g28342, g8554, g2205) at the bottom. In Section 3B, the fold-change between the cultivars is visualized; boxes marked with X were transcripts with no mapping results in the respective cultivar. The clusters do not share a similar transcript abundance pattern, nor do the genes that are closely related. However, investigating the recurring, fixed-length patterns inside the sequences led to the discovery of five motifs shared among all sequences. Figure 3C visualizes the motifs and their distribution in the sequence. The exact motif sequences are presented in Table S3. A closer look also reveals distinct recurring, CYP-specific domains [63]. The conserved regions were reviewed extensively [38] and can be confirmed in this dataset. Starting with the proline-rich membrane hinge (motif 8), which is part of the membrane anchor, another conserved motif, which is important for the correct function of CYPs, is the site for oxygen binding and activation, A/G-G-X-E/D-T-T/S (motif 3). This is followed by the E-R-R triad and P(E)R(F) domain. Furthermore, the heme-binding site, with cysteine as the main ligand to the heme, C-X-G (motif 2), which is necessary for the typical redox reaction of CYPs [64], as well as the ERR triad (motif 6) and the (P(E)R(F)) sequence (motif 6), can be differentiated among the discovered 10 motifs.



**Figure 3.** Analysis of differentially expressed cytochrome p450 enzymes (CYP) in different plant cultivars of *Caryopteris × clandonensis*, Dark Knight (DK), Grand Bleu (GB), Good as Gold (GG), Hint



of Gold (HG), Sunny Blue (SB), and Pink Perfection (PP). (A) Phylogenetic analysis of CYP sequences with highly abundant transcripts regarding limonene-derived molecules (LDM) within the cultivars, using Neighbor-joining method. (B) Heatmap of normalized transcript counts between distinct cultivars. X represents enzymes with no transcripts in respective cultivars. The color palette displays genes with high transcripts abundance in red to light-yellow colors, high transcript abundance is depicted in light-green to blue (C) Identification of recurring, fixed-length patterns (motifs) identified in LDM-positive transcripts. Motifs 1 to 10 are illustrated as colored boxes, to distinguish the motifs between the different genes. Sequences can be found in Table S3.

Regarding the production of LDM, the genes g8554, g27861, g10700, and g24222 show an interesting pattern compared to the highly LDM-positive cultivar PP, which makes them candidates for further functional characterization to prove their LDM-producing potential.

The candidate genes were further investigated in terms of their putative function. The initial estimates, using sequence and structural homology, consider g2422 and g8554 to be involved in the hydroxylation of cinnamic acid, whereas g27861 and g10700 display unknown activity towards flavonoids, sterols, and ferruginol. This substrate promiscuity is known for CYPs, as they are able to catalyze different ligands [57,65], thus making functional characterization using prokaryotic, yeast, or plant expression systems indispensable to support claims on putative functions.

### 3. Materials and Methods

#### 3.1. Plant Material

Cultivars of *Caryopteris × clandonensis*, DK, GB, GG, HG, SB, and PP, were acquired from a local nursery (Foerstner Pflanzen GmbH, Bietigheim-Bissingen). DK and GB were investigated to show a highly LDM-negative profile, whereas SB and PP show a highly LDM-positive profile. GG and HG showed a non-conclusive profile in between. After growing to maturity in the open in a warm, moderate climate zone, healthy leaves were sampled and snap-frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until RNA preparation for RNA-Seq.

#### 3.2. Genomic Resource

The reference genome of *Caryopteris × clandonensis* used in this study was obtained from NCBI SAMN32308290 (PP). The raw data were assembled as previously described [1] and subjected to further refinements. For further processing, the reference was cleared from possible contaminations, and scaled down from 783 contigs to 53 contigs using Metabat2 (v2.15) [42], keeping the genome completeness with 96.5% at a high level according to BUSCO (v5.3.2) [66] analysis (2326 BUSCO groups, lineage dataset: *Eudicotidae*). Gene model prediction was conducted using AUGUSTUS [67–70]. To detect repetitive sequences, such as tandem repeats or transposable elements, soft masking was employed using Red (v2018.09.10) [71].

#### 3.3. RNA Preparation and Short Read Sequencing

High-quality RNA was extracted using the RNeasy Plant Mini Kit (Qiagen, Venlo, The Netherlands) according to the manufacturer's protocol. To ensure RNA integrity, the Bioanalyzer RNA 6000 assay kit (Agilent, Santa Clara, CA, USA) was employed to yield an average RNA Integrity Number of 7.7. The library preparation was performed using the Illumina stranded mRNA prep kit with IDT for Illumina UD Indexes, Plate A. Corresponding adapter was the Illumina Nextera Adapter (CTGTCTCTTATACATCT). Library preparation was performed according to the manufacturer's protocol with a shortened fragmentation time from 8 min (protocol) to 2 min (this study). Sequencing was performed at the Helmholtz Munich (HMGU) by the Genomics Core Facility on a NovaSeq6000 SP ( $2 \times 150$  bp). For each sample, two lanes were loaded and an average of 22 Mio fragments were yielded. The corresponding lanes of each sample were concatenated tail-to-head (v8.25) [72]. The combined short reads were subjected to comprehensive quality control steps. Every step was analyzed with FastQC (v0.11.9) [73] and the necessity of another trimming step was evaluated. Sequences shorter than 20 bp minimum length and with a

quality phred score beneath 20 were extracted from the paired-end read data. The Illumina Nextera Adapter was used to trim each read pair using Cutadapt (v.4.0) [74]. The first 10 bases were cut from the sequences, due to their sequence GC content, using Trimmomatic (v0.38) and headcrop parameter [75].

### 3.4. Mapping and Annotation of Aligned Reads

Refined short reads were mapped on the clean reference genome using STAR (v2.7.10b) [76], 140 bases were chosen as the length of the genomic sequence around annotated junctions. EggNOG (v2.1.5) [46,77] was employed to evaluate the function of the differentially expressed genes using Pfam, GO, and COG databases. MEME suite (v5.5.1) [78] was used for identification of motifs within sequences of interest. Visualizations were built in R. Except for STAR; all sequencing analyses were conducted using galaxy project [79]. Analysis was based on reference-based RNA-Seq data analysis [80,81]. The detection of CYPs was performed using a homology-based search, using the conserved domain PF00067, which was integrated to IPR001128. Both domains are indicators for a sequence association with the cytochrome p450 superfamily (IPR036396) [53]. CYP-family classification was performed using a BLAST search [82] and a custom database [83].

### 3.5. Evaluation of Differential Gene Expression between Aerial Plant Parts

Aligned transcripts were counted using FeatureCounts (v3.16) [84], normalized, and differentially investigated with DESeq2 (v1.34.0) [85–87]. An adjusted *p*-value below 0.05, and a fold-change greater than 2 and below 0.5, was used to determine the most differentially expressed genes in this dataset.

## 4. Conclusions

This study provides a basis for further CYP research in *Caryopteris × clandonensis*, especially regarding LDM. Furthermore, the reference genome was subjected to a cleaning step, resulting in a decrease from 782 scaffolds to 53 scaffolds. Six cultivars were subjected to an RNA analysis, which gradually neared the prediction of 4 possible LDM tailoring CYPs out of 24, which were differentially expressed, and showed high transcript abundance, compared to the other cultivars. Furthermore, the classification and phylogenetic analysis of all mapped CYPs were conducted and they showed a distinct clustering in CYP CLAN71 and 72. All essential and conserved motifs could be recognized within these sequences. However, experimentally focused research for functional characterization needs to be conducted in order to identify the exact predicted function of these enzymes. A further *in silico* step can include the prediction of docking and catalysis sites within a three-dimensional structural model, as well as through molecular dynamic techniques and free energy calculations [88,89].

In general, this approach can be used to detect further mechanisms and pathways in plants, which show valuable medicinal effects. The biotechnological production of artemisinin [90] and taxol [11] is a popular example of the possibilities in medicinal plant research. There are already several approaches used, which combine omics approaches to identify substances of interest [91–93].

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/plants12122305/s1>, Table S1: BUSCO assessment and assembly statistics, Table S2: Mapping statistics on the genomic reference of Pink Perfection, Table S3: Motif sequences of identified reoccurring patterns.

**Author Contributions:** Conceptualization, M.R., N.A. and N.M.; methodology, M.R. and N.A.; software, M.R., N.A. and N.M.; validation, M.R. and N.A.; formal analysis, M.R. and N.A.; investigation, M.R. and N.A.; resources, T.B.; data curation, M.R. and N.A.; writing—original draft preparation, M.R.; writing—review and editing, M.R., N.A., N.M. and T.B.; visualization, M.R.; supervision, N.M. and T.B.; project administration, N.M. and T.B.; funding acquisition, T.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the German Federal Ministry of Education and Research, grant number 031B0824A.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are available in a publicly accessible repository. The refined genome data presented in this study are openly available at the National Center for Biotechnology Information (NCBI). BioSample accession number: Pink Perfection SAMN32308290.

**Acknowledgments:** The authors want to gratefully acknowledge the support of Christine Wurmser, (Chair of Animal Physiology and Immunology, TUM School of Life Sciences, Technical University of Munich) for her support in the library preparation and the handling of Illumina sequencing, and Foerstner Pflanzen GmbH, for providing plant materials. Furthermore, the authors want to acknowledge the support of the following colleagues at the Werner Siemens-Chair for Synthetic Biotechnology: Nathanael Arnold, Kevin Heieck, Zora Rerop, Selina Engelhart-Straub, and further colleagues for their support during conducting experiments and writing this manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Ritz, M.; Ahmad, N.; Brueck, T.; Mehlmer, N. Comparative Genome-Wide Analysis of Two Caryopteris × Clandonensis Cultivars: Insights on the Biosynthesis of Volatile Terpenoids. *Plants* **2023**, *12*, 632. [[CrossRef](#)] [[PubMed](#)]
2. Hannedouche, S.; Jacquemond-Collet, I.; Fabre, N.; Stanislas, E.; Moulis, C. Iridoid keto-glycosides from Caryopteris × Clandonensis. *Phytochemistry* **1999**, *51*, 767–769. [[CrossRef](#)]
3. Matsumoto, T.; Mayer, C.; Eugster, C.H.  $\alpha$ -Caryopteron, ein neues Pyrano-juglon aus Caryopteris clandonensis. *Helv. Chim. Acta* **1969**, *52*, 808–812. [[CrossRef](#)]
4. Blythe, E.K.; Tabanca, N.; Demirci, B.; Bernier, U.R.; Agramonte, N.M.; Ali, A.; Baser, H.C.; Khan, I.A. Composition of the essential oil of Pink Chablis™ bluebeard (Caryopteris × clandonensis ‘Durio’) and its biological activity against the yellow fever mosquito *Aedes aegypti*. *Nat. Volatiles Essent. Oils* **2015**, *2*, 11–21.
5. Abdelaty, N.A.; Attia, E.Z.; Hamed, A.N.E.; Desoukey, S.Y. A review on various classes of secondary metabolites and biological activities of Lamiaceae (*Labiatae*) (2002–2018). *J. Adv. Biomed. Pharm. Sci.* **2021**, *4*, 16–31. [[CrossRef](#)]
6. Siciliano, T.; Bader, A.; Vassallo, A.; Braca, A.; Morelli, I.; Pizza, C.; De Tommasi, N. Secondary metabolites from *Ballota undulata* (*Lamiaceae*). *Biochem. Syst. Ecol.* **2005**, *33*, 341–351. [[CrossRef](#)]
7. Mimica-Dukic, N.; Bozin, B.; Mentha, L. Species (*Lamiaceae*) as Promising Sources of Bioactive Secondary Metabolites. *Curr. Pharm. Des.* **2008**, *14*, 3141–3150. [[CrossRef](#)]
8. Kliebenstein, D.J. Secondary metabolites and plant/environment interactions: A view through *Arabidopsis thaliana* tinged glasses. *Plant. Cell Environ.* **2004**, *27*, 675–684. [[CrossRef](#)]
9. Boncan, D.A.T.; Tsang, S.S.K.; Li, C.; Lee, I.H.T.; Lam, H.M.; Chan, T.F.; Hui, J.H.L. Terpenes and Terpenoids in Plants: Interactions with Environment and Insects. *Int. J. Mol. Sci.* **2020**, *21*, 7382. [[CrossRef](#)]
10. Holopainen, J.K.; Himanen, S.J.; Yuan, J.S.; Chen, F.; Stewart, C.N. Ecological functions of terpenoids in changing climates. *Nat. Prod.* **2013**, *1*, 2913–2940.
11. Wang, T.; Li, L.; Zhuang, W.; Zhang, F.; Shu, X.; Wang, N.; Wang, Z. Recent Research Progress in Taxol Biosynthetic Pathway and Acylation Reactions Mediated by Taxus Acyltransferases. *Molecules* **2021**, *26*, 2855. [[CrossRef](#)] [[PubMed](#)]
12. Kamatou, G.P.P.; Vermaak, I.; Viljoen, A.M.; Lawrence, B.M. Menthol: A simple monoterpene with remarkable biological properties. *Phytochemistry* **2013**, *96*, 15–25. [[CrossRef](#)]
13. Khoo, H.E.; Azlan, A.; Tang, S.T.; Lim, S.M. Anthocyanidins and anthocyanins: Colored pigments as food, pharmaceutical ingredients, and the potential health benefits. *Food Nutr. Res.* **2017**, *61*, 1361779. [[CrossRef](#)]
14. Selvaraj, B.; Kim, D.W.; Huh, G.; Lee, H.; Kang, K.; Lee, J.W. Synthesis and biological evaluation of isoliquiritigenin derivatives as a neuroprotective agent against glutamate mediated neurotoxicity in HT22 cells. *Bioorg. Med. Chem. Lett.* **2020**, *30*, 127058. [[CrossRef](#)]
15. Mazimba, O. Umbelliferone: Sources, chemistry and bioactivities review. *Bull. Fac. Pharm. Cairo Univ.* **2017**, *55*, 223–232. [[CrossRef](#)]
16. Pichersky, E.; Raguso, R.A. Why do plants produce so many terpenoid compounds? *New Phytol.* **2018**, *220*, 692–702. [[CrossRef](#)] [[PubMed](#)]
17. Lešnik, S.; Furlan, V.; Bren, U. Rosemary (*Rosmarinus officinalis* L.): Extraction techniques, analytical methods and health-promoting biological effects. *Phytochem. Rev.* **2021**, *20*, 1273–1328. [[CrossRef](#)]

18. Furlan, V.; Bren, U. Helichrysum italicum: From Extraction, Distillation, and Encapsulation Techniques to Beneficial Health Effects. *Foods* **2023**, *12*, 802. [\[CrossRef\]](#)
19. Fadilah, N.Q.; Jittmittraphap, A.; Leaugwutiwong, P.; Pripdeevech, P.; Dhanushka, D.; Mahidol, C.; Ruchirawat, S.; Kittakoop, P. Virucidal Activity of Essential Oils From Citrus x aurantium L. Against Influenza A Virus H1N1: Limonene as a Potential Household Disinfectant Against Virus. *Nat. Prod. Commun.* **2022**, *17*, 1934578X211072713. [\[CrossRef\]](#)
20. Chassagne, F.; Samarakoon, T.; Porras, G.; Lyles, J.T.; Dettweiler, M.; Marquez, L.; Salam, A.M.; Shabih, S.; Farrokhi, D.R.; Quave, C.L. A Systematic Review of Plants With Antibacterial Activities: A Taxonomic and Phylogenetic Perspective. *Front. Pharmacol.* **2021**, *11*, 2069. [\[CrossRef\]](#)
21. Islam, A.K.M.M.; Suttiyut, T.; Anwar, M.P.; Juraimi, A.S.; Kato-Noguchi, H. Allelopathic Properties of Lamiaceae Species: Prospects and Challenges to Use in Agriculture. *Plants* **2022**, *11*, 1478. [\[CrossRef\]](#)
22. Byers, K.J.R.P.; Bradshaw, H.D.; Riffell, J.A. Three floral volatiles contribute to differential pollinator attraction in monkeyflowers (*Mimulus*). *J. Exp. Biol.* **2014**, *217*, 614–623. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Nagel, R.; Schmidt, A.; Peters, R.J. Isoprenyl diphosphate synthases: The chain length determining step in terpene biosynthesis. *Planta* **2018**, *249*, 9–20. [\[CrossRef\]](#)
24. Dickschat, J.S. Bacterial Diterpene Biosynthesis. *Angew. Chem. Int. Ed.* **2019**, *58*, 15964–15976. [\[CrossRef\]](#)
25. Bohlmann, J.; Meyer-Gauen, G.; Croteau, R. Plant terpenoid synthases: Molecular biology and phylogenetic analysis. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 4126–4133. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Chen, F.; Tholl, D.; Bohlmann, J.; Pichersky, E. The family of terpene synthases in plants: A mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.* **2011**, *66*, 212–229. [\[CrossRef\]](#)
27. Karunanithi, P.S.; Zerbe, P. Terpene Synthases as Metabolic Gatekeepers in the Evolution of Plant Terpenoid Chemical Diversity. *Front. Plant Sci.* **2019**, *10*, 1166. [\[CrossRef\]](#)
28. Liu, X.; Zhu, X.; Wang, H.; Liu, T.; Cheng, J.; Jiang, H. Discovery and modification of cytochrome P450 for plant natural products biosynthesis. *Synth. Syst. Biotechnol.* **2020**, *5*, 187. [\[CrossRef\]](#)
29. Foti, R.S.; Honaker, M.; Nath, A.; Pearson, J.T.; Buttrick, B.; Isoherranen, N.; Atkins, W.M. Catalytic vs. Inhibitory Promiscuity in Cytochrome P450s: Implications for Evolution of New Function. *Biochemistry* **2011**, *50*, 2387. [\[CrossRef\]](#)
30. Fischer, M.; Knoll, M.; Sirim, D.; Wagner, F.; Funke, S.; Pleiss, J.; Bateman, A. The Cytochrome P450 Engineering Database: A navigation and prediction tool for the cytochrome P450 protein family. *Bioinformatics* **2007**, *23*, 2015–2017. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Nelson, D.R. The Cytochrome P450 Homepage. *Hum. Genom.* **2009**, *4*, 59. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Nelson, D.R. Cytochrome P450 nomenclature, 2004. *Methods Mol. Biol.* **2006**, *320*, 1–10. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Rasool, S.; Mohamed, R. Plant cytochrome P450s: Nomenclature and involvement in natural product biosynthesis. *Protoplasma* **2015**, *253*, 1197–1209. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Krause, S.T.; Liao, P.; Crocoll, C.; Boachon, B.; Förster, C.; Leidecker, F.; Wiese, N.; Zhao, D.; Wood, J.C.; Buell, C.R.; et al. The biosynthesis of thymol, carvacrol, and thymohydroquinone in Lamiaceae proceeds via cytochrome P450s and a short-chain dehydrogenase. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2110092118. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Gupta, P.; Geniza, M.; Naithani, S.; Phillips, J.L.; Haq, E.; Jaiswal, P. Chia (*Salvia hispanica*) Gene Expression Atlas Elucidates Dynamic Spatio-Temporal Changes Associated With Plant Growth and Development. *Front. Plant Sci.* **2021**, *12*, 667678. [\[CrossRef\]](#)
36. Li, H.; Li, J.; Dong, Y.; Hao, H.; Ling, Z.; Bai, H.; Wang, H.; Cui, H.; Shi, L. Time-series transcriptome provides insights into the gene regulation network involved in the volatile terpenoid metabolism during the flower development of lavender. *BMC Plant Biol.* **2019**, *19*, 313. [\[CrossRef\]](#)
37. Lichman, B.R.; Godden, G.T.; Buell, C.R. Gene and genome duplications in the evolution of chemodiversity: Perspectives from studies of Lamiaceae. *Curr. Opin. Plant Biol.* **2020**, *55*, 74–83. [\[CrossRef\]](#)
38. Bak, S.; Beisson, F.; Bishop, G.; Hamberger, B.; Höfer, R.; Paquette, S.; Werck-Reichhart, D. Cytochromes P450. *Arab. Book* **2011**, *9*, e0144. [\[CrossRef\]](#)
39. Xie, Y.; Ye, S.; Wang, Y.; Xu, L.; Zhu, X.; Yang, J.; Feng, H.; Yu, R.; Karanja, B.; Gong, Y.; et al. Transcriptome-based gene profiling provides novel insights into the characteristics of radish root response to Cr stress with next-generation sequencing. *Front. Plant Sci.* **2015**, *6*, 202. [\[CrossRef\]](#)
40. Manzano, A.; Carnero-Diaz, E.; Herranz, R.; Medina, F.J. Recent transcriptomic studies to elucidate the plant adaptive response to spaceflight and to simulated space environments. *iScience* **2022**, *25*, 104687. [\[CrossRef\]](#)
41. Howlader, J.; Robin, A.H.K.; Natarajan, S.; Biswas, M.K.; Sumi, K.R.; Song, C.Y.; Park, J.-I.; Nou, I.-S. Transcriptome Analysis by RNA-Seq Reveals Genes Related to Plant Height in Two Sets of Parent-hybrid Combinations in Easter lily (*Lilium longiflorum*). *Sci. Rep.* **2020**, *10*, 9082. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Kang, D.D.; Li, F.; Kirton, E.; Thomas, A.; Egan, R.; An, H.; Wang, Z. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **2019**, *7*, e7359. [\[CrossRef\]](#)
43. Conesa, A.; Madrigal, P.; Tarazona, S.; Gomez-Cabrero, D.; Cervera, A.; McPherson, A.; Szczesniak, M.W.; Gaffney, D.J.; Elo, L.L.; Zhang, X.; et al. A survey of best practices for RNA-seq data analysis. *Genome Biol.* **2016**, *17*, 13. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Chaudhary, S.; Khokhar, W.; Jabre, I.; Reddy, A.S.N.; Byrne, L.J.; Wilson, C.M.; Syed, N.H. Alternative splicing and protein diversity: Plants versus animals. *Front. Plant Sci.* **2019**, *10*, 708. [\[CrossRef\]](#)
45. Finn, R.D.; Coghill, P.; Eberhardt, R.Y.; Eddy, S.R.; Mistry, J.; Mitchell, A.L.; Potter, S.C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; et al. The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **2015**, *44*, D279–D285. [\[CrossRef\]](#) [\[PubMed\]](#)

46. Huerta-Cepas, J.; Szklarczyk, D.; Heller, D.; Hernández-Plaza, A.; Forslund, S.K.; Cook, H.; Mende, D.R.; Letunic, I.; Rattei, T.; Jensen, L.J.; et al. eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **2019**, *47*, D309. [[CrossRef](#)]
47. Curie, C.; Cassin, G.; Couch, D.; Divol, F.; Higuchi, K.; Le Jean, M.; Misson, J.; Schikora, A.; Czernic, P.; Mari, S. Metal movement within the plant: Contribution of nicotianamine and yellow stripe 1-like transporters. *Ann. Bot.* **2009**, *103*, 1–11. [[CrossRef](#)]
48. Ishimaru, Y.; Masuda, H.; Bashir, K.; Inoue, H.; Tsukamoto, T.; Takahashi, M.; Nakanishi, H.; Aoki, N.; Hirose, T.; Ohsugi, R.; et al. Rice metal-nicotianamine transporter, OsYSL2, is required for the long-distance transport of iron and manganese. *Plant J.* **2010**, *62*, 379–390. [[CrossRef](#)]
49. Li, Z.; Li, W.; Guo, M.; Liu, S.; Liu, L.; Yu, Y.; Mo, B.; Chen, X.; Gao, L. Origin, evolution and diversification of plant ARGONAUTE proteins. *Plant J.* **2022**, *109*, 1086–1097. [[CrossRef](#)]
50. Zhang, Z.; Liu, X.; Li, R.; Yuan, L.; Dai, Y.; Wang, X. Identification and functional analysis of a protein disulfide isomerase (AtPDI1) in *Arabidopsis thaliana*. *Front. Plant Sci.* **2018**, *9*, 913. [[CrossRef](#)]
51. Finkelstein, R. Abscisic Acid Synthesis and Response. *Arab. Book* **2013**, *11*, e0166. [[CrossRef](#)] [[PubMed](#)]
52. Liao, W.; Zhao, S.; Zhang, M.; Dong, K.; Chen, Y.; Fu, C.; Yu, L. Transcriptome assembly and systematic identification of novel cytochrome P450s in *taxus chinensis*. *Front. Plant Sci.* **2017**, *8*, 1468. [[CrossRef](#)]
53. Degtyarenko, K.N. Structural domains of P450-containing monooxygenase systems. *Protein Eng. Des. Sel.* **1995**, *8*, 737–747. [[CrossRef](#)] [[PubMed](#)]
54. Vasav, A.P.; Barvkar, V.T. Phylogenomic analysis of cytochrome P450 multigene family and their differential expression analysis in *Solanum lycopersicum* L. suggested tissue specific promoters. *BMC Genom.* **2019**, *20*, 116. [[CrossRef](#)] [[PubMed](#)]
55. Wegrzyn, G.; Schachner, M.; Gabbiani, G.; Minerdi, D.; Savoi, S.; Sabbatini, P. Role of Cytochrome P450 Enzyme in Plant Microorganisms & Communication: A Focus on Grapevine. *Int. J. Mol. Sci.* **2023**, *24*, 4695. [[CrossRef](#)]
56. Bathe, U.; Tissier, A. Cytochrome P450 enzymes: A driving force of plant diterpene diversity. *Phytochemistry* **2019**, *161*, 149–162. [[CrossRef](#)]
57. Hansen, C.C.; Nelson, D.R.; Møller, B.L.; Werck-Reichhart, D. Plant cytochrome P450 plasticity and evolution. *Mol. Plant* **2021**, *14*, 1244–1265. [[CrossRef](#)]
58. Haudenschild, C.; Schalk, M.; Karp, F.; Croteau, R. Functional Expression of Regiospecific Cytochrome P450 Limonene Hydroxylases from Mint (*Mentha* spp.) in *Escherichia coli* and *Saccharomyces cerevisiae*. *Arch. Biochem. Biophys.* **2000**, *379*, 127–136. [[CrossRef](#)]
59. Lupien, S.; Karp, F.; Wildung, M.; Croteau, R. Regiospecific cytochrome P450 limonene hydroxylases from mint (*Mentha*) species: cDNA isolation, characterization, and functional expression of (-)-4S-limonene-3-hydroxylase and (-)-4S-limonene-6-hydroxylase. *Arch. Biochem. Biophys.* **1999**, *368*, 181–192. [[CrossRef](#)]
60. Chen, X.; Zhang, C.; Too, H.P. Multienzyme Biosynthesis of Dihydroartemisinic Acid. *Molecules* **2017**, *22*, 1422. [[CrossRef](#)]
61. Wu, Y.; Hillwig, M.L.; Wang, Q.; Peters, R.J. Parsing a multifunctional biosynthetic gene cluster from rice: Biochemical characterization of CYP71Z6 & 7. *FEBS Lett.* **2011**, *585*, 3446. [[CrossRef](#)]
62. Sawai, S.; Saito, K. Triterpenoid Biosynthesis and Engineering in Plants. *Front. Plant Sci.* **2011**, *585*, 3446–3451. [[CrossRef](#)] [[PubMed](#)]
63. Chen, Z.; Qi, X.; Yu, X.; Zheng, Y.; Liu, Z.; Fang, H.; Li, L.; Bai, Y.; Liang, C.; Li, W. Genome-Wide Analysis of Terpene Synthase Gene Family in *Mentha longifolia* and Catalytic Activity Analysis of a Single Terpene Synthase. *Genes* **2021**, *12*, 518. [[CrossRef](#)] [[PubMed](#)]
64. Zhang, W.; Liu, Y.; Yan, J.; Cao, S.; Bai, F.; Yang, Y.; Huang, S.; Yao, L.; Anzai, Y.; Kato, F.; et al. New reactions and products resulting from alternative interactions between the P450 enzyme and redox partners. *J. Am. Chem. Soc.* **2014**, *136*, 3640–3646. [[CrossRef](#)] [[PubMed](#)]
65. Hernandez-Ortega, A.; Vinaixa, M.; Zebec, Z.; Takano, E.; Scrutton, N.S. A Toolbox for Diverse Oxyfunctionalisation of Monoterpenes OPEN. *Sci. Rep.* **2018**, *8*, 14396. [[CrossRef](#)]
66. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)]
67. Hoff, K.J.; Lomsadze, A.; Borodovsky, M.; Stanke, M. Whole-Genome Annotation with BRAKER. *Methods Mol. Biol.* **2019**, *1962*, 65. [[CrossRef](#)]
68. Hoff, K.J.; Lange, S.; Lomsadze, A.; Borodovsky, M.; Stanke, M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **2016**, *32*, 767–769. [[CrossRef](#)]
69. Brůna, T.; Hoff, K.J.; Lomsadze, A.; Stanke, M.; Borodovsky, M. BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.* **2021**, *3*, lqaa108. [[CrossRef](#)]
70. Stanke, M.; Schöffmann, O.; Morgenstern, B.; Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinform.* **2006**, *7*, 62. [[CrossRef](#)] [[PubMed](#)]
71. Girgis, H.Z. Red: An intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. *BMC Bioinform.* **2015**, *16*, 227. [[CrossRef](#)]
72. Grüning, B.; Yusuf, D.; Houwaart, T.; Anika; Miladi, M.; Gu, Q.; Batut, B.; Soranzo, N.; Gamaleldin, H.; Von Kuster, G.; et al. *Bgruening/Galaxytools: September Release 2019*; Zenodo: Geneva, Switzerland, 2018. [[CrossRef](#)]

73. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. Available online: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 24 March 2023).
74. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **2011**, *17*, 10. [[CrossRef](#)]
75. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]
76. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [[CrossRef](#)] [[PubMed](#)]
77. Huerta-Cepas, J.; Forslund, K.; Coelho, L.P.; Szklarczyk, D.; Jensen, L.J.; Von Mering, C.; Bork, P. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **2017**, *34*, 2115–2122. [[CrossRef](#)] [[PubMed](#)]
78. Bailey, T.L.; Johnson, J.; Grant, C.E.; Noble, W.S. The MEME Suite. *Nucleic Acids Res.* **2015**, *43*, W39–W49. [[CrossRef](#)] [[PubMed](#)]
79. Afgan, E.; Baker, D.; Batut, B.; Van Den Beek, M.; Bouvier, D.; Ech, M.; Chilton, J.; Clements, D.; Coraor, N.; Grüning, B.A.; et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **2018**, *46*, W537–W544. [[CrossRef](#)]
80. Batut, B.; Freeberg, M.; Heydarian, M.; Erxleben, A.; Videm, P.; Blank, C.; Doyle, M.; Soranzo, N.; van Heusden, P.; Delisle, L. Reference-Based RNA-Seq Data Analysis (Galaxy Training Materials). Available online: <https://training.galaxyproject.org/training-material/topics/transcriptomics/tutorials/ref-based/tutorial.html#citing-this-tutorial> (accessed on 25 March 2023).
81. Batut, B.; Hiltmann, S.; Bagnacani, A.; Baker, D.; Bhardwaj, V.; Blank, C.; Bretaudeau, A.; Brillet-Guéguen, L.; Čech, M.; Chilton, J.; et al. Community-Driven Data Analysis Training for Biology. *Cell Syst.* **2018**, *6*, 752–758.e1. [[CrossRef](#)] [[PubMed](#)]
82. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
83. Kweon, O.; Kim, S.J.; Kim, J.H.; Nho, S.W.; Bae, D.; Chon, J.; Hart, M.; Baek, D.H.; Kim, Y.C.; Wang, W.; et al. CYPminer: An automated cytochrome P450 identification, classification, and data analysis tool for genome data sets across kingdoms. *BMC Bioinform.* **2020**, *21*, 160. [[CrossRef](#)]
84. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**, *30*, 923–930. [[CrossRef](#)] [[PubMed](#)]
85. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)]
86. Zhu, A.; Ibrahim, J.G.; Love, M.I. Heavy-tailed prior distributions for sequence count data: Removing the noise and preserving large differences. *Bioinformatics* **2019**, *35*, 2084–2092. [[CrossRef](#)] [[PubMed](#)]
87. Bioinformatics Training at the Harvard Chan Bioinformatics Core. Available online: <https://hbctraining.github.io/main/> (accessed on 19 April 2023).
88. Pantiora, P.; Furlan, V.; Matiadis, D.; Mavroidi, B.; Perperopoulou, F.; Papageorgiou, A.C.; Sagnou, M.; Bren, U.; Pelecanou, M.; Labrou, N.E. Monocarbonyl Curcumin Analogues as Potent Inhibitors against Human Glutathione Transferase P1-1. *Antioxidants* **2023**, *12*, 63. [[CrossRef](#)] [[PubMed](#)]
89. Kores, K.; Kolenc, Z.; Furlan, V.; Bren, U. Inverse Molecular Docking Elucidating the Anticarcinogenic Potential of the Hop Natural Product Xanthohumol and Its Metabolites. *Foods* **2022**, *11*, 1253. [[CrossRef](#)]
90. Wen, W.; Yu, R. Artemisinin biosynthesis and its regulatory enzymes: Progress and perspective. *Pharmacogn. Rev.* **2011**, *5*, 189. [[CrossRef](#)]
91. Sun, W.; Xu, Z.; Song, C.; Chen, S. Herbgenomics: Decipher molecular genetics of medicinal plants. *Innovation* **2022**, *3*, 100322. [[CrossRef](#)]
92. Alami, M.M.; Ouyang, Z.; Zhang, Y.; Shu, S.; Yang, G.; Mei, Z.; Wang, X. The Current Developments in Medicinal Plant Genomics Enabled the Diversification of Secondary Metabolites' Biosynthesis. *Int. J. Mol. Sci.* **2022**, *23*, 15932. [[CrossRef](#)]
93. Cheng, Q.Q.; Ouyang, Y.; Tang, Z.Y.; Lao, C.C.; Zhang, Y.Y.; Cheng, C.S.; Zhou, H. Review on the Development and Applications of Medicinal Plant Genomes. *Front. Plant Sci.* **2021**, *12*, 2981. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.