



## OPEN ACCESS

## EDITED BY

Francesco Vinale,  
University of Naples Federico II, Italy

## REVIEWED BY

John L. Sorensen,  
University of Manitoba, Canada  
Julia Gerasimova,  
Ludwig Maximilian University of  
Munich, Germany

## \*CORRESPONDENCE

Garima Singh  
gsingh458@gmail.com;  
garima.singh@unipd.it

## SPECIALTY SECTION

This article was submitted to  
Fungal Secondary Metabolites  
and Mycotoxins,  
a section of the journal  
Frontiers in Fungal Biology

RECEIVED 13 July 2022

ACCEPTED 30 August 2022

PUBLISHED 03 October 2022

## CITATION

Singh G, Dal Grande F and Schmitt I  
(2022) Genome mining as a  
biotechnological tool for the discovery  
of novel biosynthetic genes in lichens.  
*Front. Fungal Bio.* 3:993171.  
doi: 10.3389/ffunb.2022.993171

## COPYRIGHT

© 2022 Singh, Dal Grande and Schmitt.  
This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s)  
are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Genome mining as a biotechnological tool for the discovery of novel biosynthetic genes in lichens

Garima Singh<sup>1,2,3\*</sup>, Francesco Dal Grande<sup>1,2,3</sup>  
and Imke Schmitt<sup>1,2,4</sup>

<sup>1</sup>Senckenberg Biodiversity and Climate Research Centre (SBIK-F), Frankfurt am Main, Germany,

<sup>2</sup>LOEWE Center for Translational Biodiversity Genomics (TBG), Frankfurt am Main, Germany,

<sup>3</sup>Department of Biology, University of Padova, Padova, Italy, <sup>4</sup>Institute of Ecology, Diversity and  
Evolution, Goethe University, Frankfurt am Main, Germany

Natural products (NPs) and their derivatives are a major contributor to modern medicine. Historically, microorganisms such as bacteria and fungi have been instrumental in generating drugs and lead compounds because of the ease of culturing and genetically manipulating them. However, the ever-increasing demand for novel drugs highlights the need to bioprospect previously unexplored taxa for their biosynthetic potential. Next-generation sequencing technologies have expanded the range of organisms that can be explored for their biosynthetic content, as these technologies can provide a glimpse of an organism's entire biosynthetic landscape, without the need for cultivation. The entirety of biosynthetic genes can be compared to the genes of known function to identify the gene clusters potentially coding for novel products. In this study, we mine the genomes of nine lichen-forming fungal species of the genus *Umbilicaria* for biosynthetic genes, and categorize the biosynthetic gene clusters (BGCs) as "associated product structurally known" or "associated product putatively novel". Although lichen-forming fungi have been suggested to be a rich source of NPs, it is not known how their biosynthetic diversity compares to that of bacteria and non-lichenized fungi. We found that 25%–30% of biosynthetic genes are divergent as compared to the global database of BGCs, which comprises 1,200,000 characterized biosynthetic genes from plants, bacteria, and fungi. Out of 217 BGCs, 43 were highly divergent suggesting that they potentially encode structurally and functionally novel NPs. Clusters encoding the putatively novel metabolic diversity comprise polyketide synthases (30), non-ribosomal peptide synthetases (12), and terpenes (1). Our study emphasizes the utility of genomic data in bioprospecting microorganisms for their biosynthetic potential and in advancing the industrial application of unexplored taxa. We highlight the

untapped structural metabolic diversity encoded in the lichenized fungal genomes. To the best of our knowledge, this is the first investigation identifying genes coding for NPs with potentially novel properties in lichenized fungi.

#### KEYWORDS

secondary metabolites, Big-FAM, natural products, drug discovery, BiG-SLiCE, medicinal fungi, antiSMASH, MIBiG

## Introduction

Natural products (NPs) are structurally diverse molecules that are produced by nearly all organisms, including plants, fungi, and bacteria. Historically, NPs have played a key role in drug discovery owing to their broad range of pharmacological effects, encompassing antimicrobial, antitumor, and anti-inflammatory properties and protection against cardiovascular diseases (Newman and Cragg, 2012; Newman and Cragg, 2020). In recent decades, about 70% of new drugs have been developed from NPs or NP analogs (Newman and Cragg, 2012; Newman and Cragg, 2020). The demand for novel drugs, however, is ever increasing because of the emergence of antibiotic-resistant pathogens and new diseases, the existence of diseases for which no efficient treatments are available yet, and the need for current drugs to be replaced due to the toxicity or side-effects associated with their use (Demain, 2014; Chakraborty et al., 2021). One way to address global health threats and accelerate NP-based drug discovery efforts is to bioprospect unexplored taxa to assess their biosynthetic potential and to identify potentially novel drug leads.

The genes involved in the synthesis of NPs are often grouped together in biosynthetic gene clusters (BGCs) (Jensen, 2016; Calcott et al., 2018; Keller, 2019). BGCs typically have a core gene that codes for the backbone structure of the NP, and other genes that may be involved in the modification of the backbone or may have regulatory or transport-related functions (Aigle et al., 2014; Rigali et al., 2018; Keller, 2019; Kim et al., 2021). Depending on the core gene, BGCs are grouped into the following major classes: non-ribosomal peptide synthetases (NRPSs), polyketide synthases (PKSs), hybrid non-ribosomal peptide synthetase–polyketide synthases (NRPS–PKSs), terpenes, and ribosomally synthesized and post-translationally modified peptides (RiPPs). The conserved motifs of the core genes facilitate the bioinformatic detection of the clusters (Medema et al., 2011; Bertrand et al., 2018; Calchera et al., 2019; Kum and İnce, 2021).

Traditionally, a large proportion of NP-based drugs have been contributed by a few organisms, as drug discovery has mostly been restricted to culturable organisms (Newman et al.,

2003; Cragg and Newman, 2013; Yuan et al., 2016). In recent decades, the bioinformatic prediction of biosynthetic genes or BGCs (i.e., groups of two or more genes that are clustered together and are involved in the production of a secondary metabolite) has revolutionized NP-based drug discovery. This process is culture independent, and enables rapid identification of the entire biosynthetic landscape, including silent or unexpressed genes, from so far unexplored NP resources. Two tools have been vital to the bioinformatic approach to drug discovery: antiSMASH (Blin et al., 2019) and Minimum Information about a Biosynthetic Gene cluster (MIBiG) (Kautsar et al., 2020). antiSMASH includes one of the largest databases for BGC prediction (Blin et al., 2019), whereas MIBiG is a data repository that allows functional interpretation of target BGCs by comparison with BGCs with known functions (Kautsar et al., 2020). Recently, efforts have been made to cluster homologous BGCs into gene cluster families (GCFs) and to simultaneously identify novel BGCs (Kautsar et al., 2021a; Kautsar et al., 2021b). Two tools have been introduced to cluster BGCs into GCFs. BiG-FAM clusters structurally and functionally related BGCs into GCFs, and structurally identify the most diverse BGCs by comparing the query BGCs with about 1,200,000 BGCs in the BiG-FAM database (Kautsar et al., 2021a). BiG-SLiCE clusters homologous BGCs of a dataset into GCFs, without reference to an external database, to identify the unique BGCs in it (Kautsar et al., 2021b). Bioinformatic prediction and clustering of BGCs allow rapid identification of potentially novel drug leads, reducing the cost and time associated with drug discovery by early elimination of unpromising candidates.

Lichens are symbiotic organisms composed of fungal and photosynthetic partners (green algae or cyanobacteria, or both). It has been suggested that they are potentially rich sources of biosynthetic genes and NPs (Boustie and Grube, 2005; Shukla et al., 2010; Shrestha and St. Clair, 2013). Although the number of identified NPs per lichen-forming fungus (LFF) is typically fewer than five (Lumbsch, 1998), the number of BGCs in the genomes of LFF may range from 25 to 60 (Calchera et al., 2019). It is not known how BGCs from LFF relate in structure and function to BGCs from bacteria and non-lichenized fungi (i.e., if

a portion of the BGC landscape of LFF is distinct, and might serve as a source of NPs with novel therapeutic properties). Difficulties associated with the heterologous expression of LFF genes have so far restricted the application of LFF-derived NPs in the industry. Recently, two biosynthetic genes from LFF have been successfully heterologously expressed (Kealey et al., 2021; Kim et al., 2021). This, combined with advances in long-read sequencing technology, high quality genomes, and the low cost of sequencing, provides a promising way forward to discover LFF-derived NPs with novel pharmacological potential.

Here we mine and compare the long-read sequencing derived genomes of nine species of the lichenized fungal genus *Umbilicaria* to estimate the functional diversity of BGCs present in them. Specifically, we aim to answer the following questions: (1) what is the functional diversity of BGCs in *Umbilicaria*? and (2) what is the percentage of novel BGCs and species-specific BGCs in *Umbilicaria*?

## Materials and methods

### Dataset

The genomes of the following *Umbilicaria* species were used for this study: *U. deusta*, *U. freyi*, *U. grisea*, *U. subpolyphylla*, *U. hispanica*, *U. phaea*, *U. pustulata*, *U. muhlenbergii*, and *U. spodochroa*. Apart from *U. muhlenbergii*, which belongs to the BioProject PRJNA239196, all the other genomes are a part of BioProject PRJNA820300 (Table 1). The details of sample and genomic library preparation, as well as genome sequencing, for *U. muhlenbergii* are available in Park et al. (2014) and for the other eight *Umbilicaria* spp. in Singh et al. (2022). Briefly, all the genomes except *U. muhlenbergii* were generated via PacBio SMRT sequencing on the Sequel System II (Radboud University Medical Center (Radboudumc) in Nijmegen, the Netherlands) using the continuous long-read (CLR) mode or the circular consensus sequencing mode. The CLR reads were then processed into highly accurate consensus sequences (i.e.,

HiFi reads) and assembled into contigs using the assembler metaFlye v2.7 (Kolmogorov et al., 2019). The contigs were then scaffolded with Long Reads Scaffold (LRScf) v1.1.12 (github.com/shingocat/lrscaf; Qin et al., 2019). We used only binned Ascomycota reads for this study [extracted using blastx in DIAMOND (more-sensitive, frameshift 15, range-culling) on a custom database and following the MEGAN6 Community Edition pipeline (Huson et al., 2016).

Two metrics were used to evaluate the quality of the genomes: completeness and number of scaffolds. Completeness is the estimate of the fraction of genes present in the genome with respect to the expected gene content. Completeness is determined based on universally distributed orthologs. We used Benchmarking Universal Single-Copy Orthologs (BUSCO) (Simão et al., 2015) to estimate genome completeness. BUSCO estimates complete single-copy, duplicated, fragmented, and missing genes in the data. The number of scaffolds shows how fragmented the assembly is, with a larger number indicating a more fragmented assembly.

Using Funannotate v1.8.9 on the resulting assemblies to estimate the number of genes and proteins (Palmer and Stajich, 2019). Funannotate implements the algorithm evidence modeler for gene prediction, which uses several different gene prediction inputs (from Augustus, snap, GlimmerHMM, CodingQuarry, and GeneMark-ES/ET) (Borodovsky and Lomsadze, 2011). In the functional annotation step, Funannotate identifies Pfam domains, carbohydrate-active enzymes, secreted proteins, proteases (via MEROPS), BUSCO groups, gene ontology, InterPro terms, and fungal transcription factors.

### Biosynthetic gene cluster prediction and clustering: AntiSMASH

Biosynthetic gene clusters were predicted using antiSMASH (Antibiotics and SM Analysis Shell, v6.0), with scripts implemented in the Funannotate pipeline (Blin et al., 2019; Palmer and Stajich, 2019). We tested if a smaller genome size

TABLE 1 Voucher information of the genomes used in the study.

Organism	Sample ID	Sequencing technology	BioProject	BioSample	Genome accession no.
<i>U. deusta</i>	TBG_2334	PacBio Sequel II	PRJNA820300	SAMN26992774	JALILR000000000
<i>U. freyi</i>	TBG_2329	PacBio Sequel II	PRJNA820300	SAMN26992773	JALILQ000000000
<i>U. grisea</i>	TBG_2336	PacBio Sequel II	PRJNA820300	SAMN26992780	JALILX000000000
<i>U. hispanica</i>	TBG_2337	PacBio Sequel II	PRJNA820300	SAMN26992775	JALILS000000000
<i>U. muhlenbergii</i>	KoLRI No. LF000956	Illumina HiSeq	PRJNA239196	SAMN02650300	GCA_000611775.1
<i>U. phaea</i>	TBG_1112	PacBio Sequel II	PRJNA820300	SAMN26992776	JALILT000000000
<i>U. pustulata</i>	TBG_2345	PacBio Sequel II	PRJNA820300	SAMN26992777	JALILU000000000
<i>U. spodochroa</i>	TBG_2434	PacBio Sequel II	PRJNA820300	SAMN26992778	JALILV000000000
<i>U. subpolyphylla</i>	TBG_2324	PacBio Sequel II	PRJNA820300	SAMN26992779	JALILW000000000

was correlated with a smaller number of BGCs. A correlation coefficient near zero indicates no correlation and a coefficient close to 1 indicates a positive correlation.

## Biosynthetic gene cluster clustering into BiG-FAM gene cluster families

The homologous BGCs present in the *Umbilicaria* genomes were grouped into GCFs using BiG-FAM, which clusters structurally- and functionally-related BGCs, and identifies the structurally most divergent BGCs by comparing the query BGCs with the 1,225,071 BGCs in the BiG-FAM database. The 1,225,071 BGCs in BiG-FAM are clustered into 29,955 GCFs based on similar domain architectures. A GCF comprises closely related BGCs, potentially encoding the same or very similar compounds. By enabling such clustering, BiG-FAM establishes the degree of similarity of BGCs of a query taxon to currently known (functionally pre-characterized) fungal and bacterial BGCs. The antiSMASH job ID of each *Umbilicaria* species was used as input for BiG-FAM analysis.

## Quantification of biosynthetic gene cluster diversity and species-specific biosynthetic gene clusters in *Umbilicaria*: BiG-SLiCE

We used BiG-SLiCE (Kautsar et al., 2021b) to identify the most unique or species-specific BGCs within *Umbilicaria*. BiG-SLiCE 1.1.0. is a networking-based tool that assesses relationships of BGCs in the dataset (i.e., *Umbilicaria* BGCs in our study) and estimates their distance within the dataset to identify unique, species-specific BGCs. The resulting distance ( $d$ ) indicates how closely a given BGC is related to the other BGCs. BiG-SLiCE was run on the *Umbilicaria* BGC dataset (i.e., 217 BGCs from nine *Umbilicaria* spp.) using three different thresholds (400, 900, and 1800).

## Results

### Overview of biosynthetic gene clusters in the *Umbilicaria* genomes

*Umbilicaria* genomes contain 20–33 BGCs each, with the largest number of BGCs detected in *U. deusta* and the lowest in *U. phaea* (Figure 1A). We did not observe a correlation between genome size and number of BGCs (correlation coefficient = 0.10). *Umbilicaria* species contain an average of 13 PKS clusters and 4.2 NRPS clusters per species (Figure 1B), making a PKS to NRPS cluster proportion of 3.1. The most dominant classes of BGC in *Umbilicaria* are PKSs, which account for more than 50%

of the total number of BGCs, followed by terpene clusters (about 20%) and NRPS clusters (about 15%) (Figure 2A). In contrast, NRPSs are the most dominant class among fungal and bacterial BGCs (Figures 2B, C), amounting to about 42% and 30%, respectively.

## Biosynthetic gene cluster clustering: BiG-FAM

Of the total 217 BGCs found in nine *Umbilicaria* species, 18 (8%) BGCs obtained a BGC-to-GCFs pairing distance lower than 400, indicating that they potentially code for structurally very similar compounds known from the BGCs of their corresponding GCFs (Figures 3A, B). One hundred and fifty-six (72%) BGCs had a pairing distance of 400–900, suggesting that they share similar domain architectures with previously described BGCs in the BiG-FAM database. We describe the clusters belonging to above two groups as “associated product structurally known”. Forty-three (20%) BGCs had a pairing distance greater than 900, and are potentially BGCs encoding novel NPs (Figure 3A). We call these clusters “associated product putatively novel”. These BGCs belong to the classes terpenes (one BGC), NRPSs (12 BGCs), and PKSs (30 BGCs). The details of these BGCs and the sequence of the core gene are provided in Supplementary Information S1.

## Within-genus comparison of biosynthetic gene clusters: BiG-SLiCE

We identified species-specific BGCs within *Umbilicaria* using BiG-SLiCE. Out of 217 total BGCs, 159 (73%) grouped into 20 GCFs ( $d = 900$ ), suggesting that they are similar clusters shared by multiple species, whereas 58 BGCs (27%) had  $d > 900$ , indicating that they were only distantly related to other BGCs in *Umbilicaria*. Each *Umbilicaria* species contains 4–10 (6.45%–16.13%) unique species-specific BGCs (Supplementary Information 2A). In *U. deusta*, we detected two BGCs (both with PKSs as the core gene) that were extremely divergent ( $d > 1,800$ ) within the genus (Supplementary Information 2B). Of these BGCs, 15 were unique within *Umbilicaria*, as well divergent from the known BGCs present in the BiG-FAM database.

## Discussion

Lichens produce a large number of NPs, and they have even more BGCs (Meiser et al., 2017; Bertrand and Sorensen, 2018; Gerasimova et al., 2022). However, whether or not these BGCs encode hitherto unknown metabolically diverse chemical structures is not known. Here we quantify, for the first time,

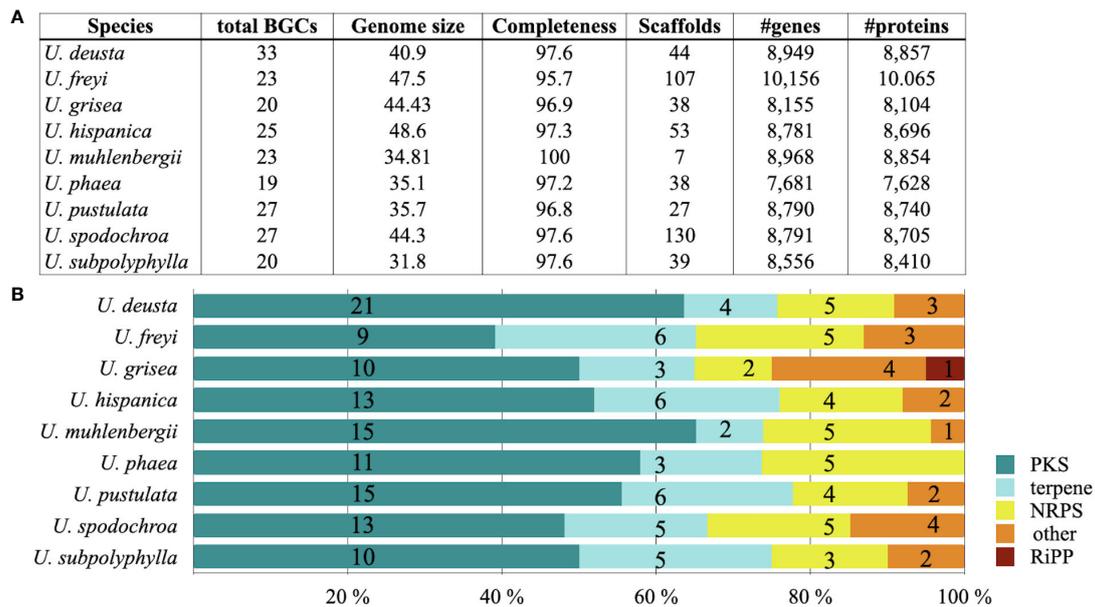


FIGURE 1

Genome quality metrics and diversity of biosynthetic genes in nine species of *Umbilicaria*. (A) Genome metrics, including the total number of biosynthetic gene clusters (BGCs) as predicted by antiSMASH, and the number of genes and proteins estimated by InterProScan and SignalP, as implemented in the Funannotate pipeline. (B) Diversity of BGCs associated with major natural product categories, indicated as percentages (colored bars) and absolute numbers (numbers on bars).

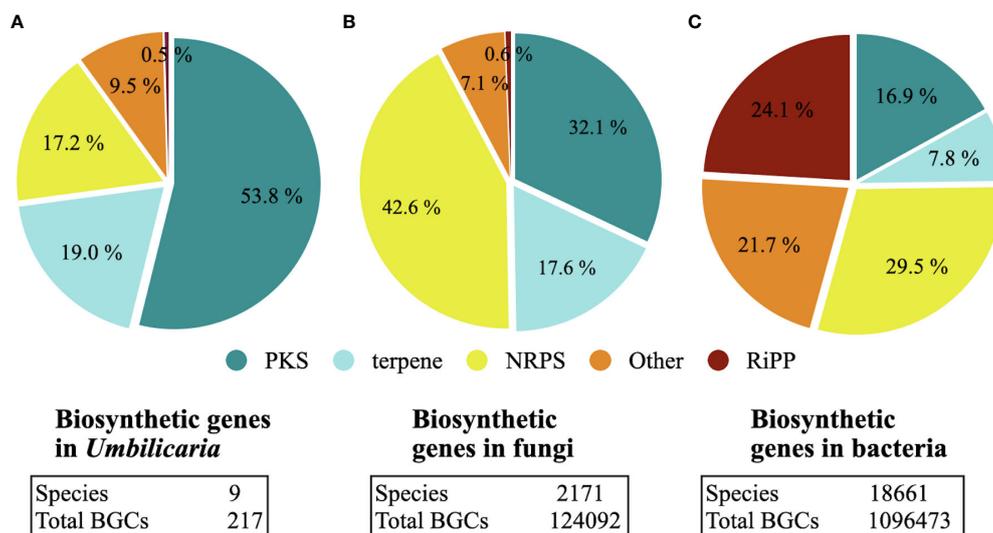


FIGURE 2

Biosynthetic gene clusters (BGCs) in (A) *Umbilicaria*, (B) the full fungal BGC dataset sensu Kautsar et al. 2021a, and (C) the full bacterial BGC dataset sensu Kautsar et al. 2021a. Polyketide synthases (PKSs) are the predominant class of BGCs in *Umbilicaria*, whereas in fungi and bacteria non-ribosomal peptide synthetases (NRPSs) are the most predominant BGC class. Although the number of publicly available lichen-forming fungal (LFF) genomes (> 50) is much smaller than the number of non-lichenized fungi (about 2,100), in all of the LFF genomes analyzed PKS clusters were the most common (see Discussion for details), suggesting that the predominance of PKSs, as observed here in the *Umbilicaria* dataset, is a common feature of LFF genomes.

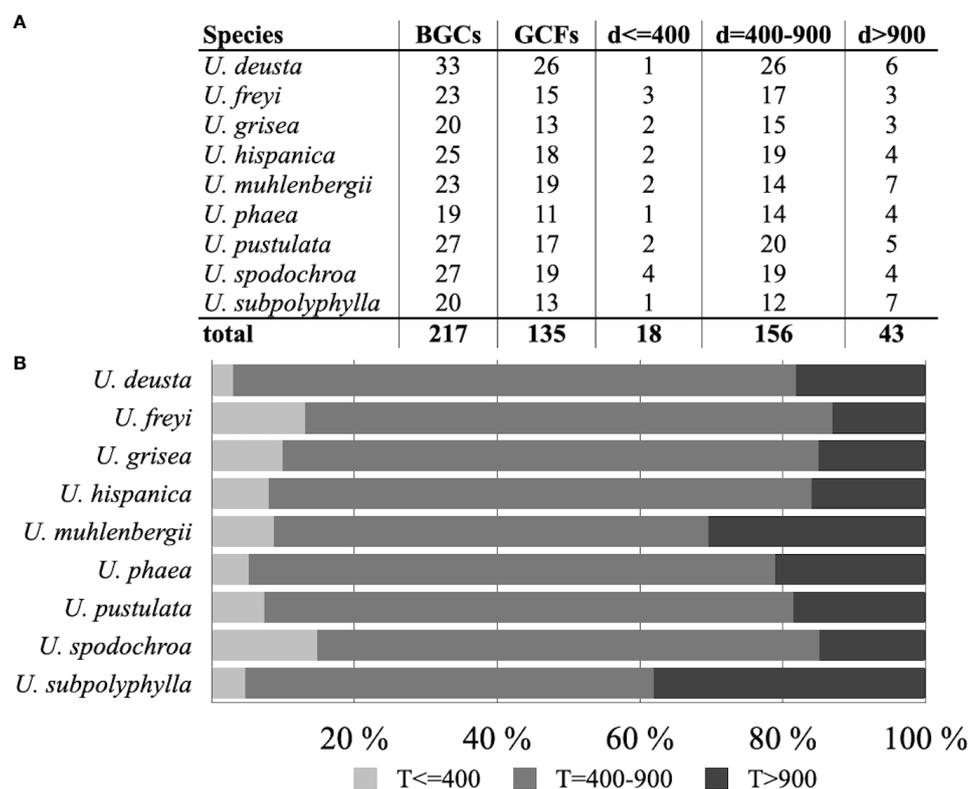


FIGURE 3

(A) Total biosynthetic gene clusters (BGCs) and gene cluster families (GCFs) as identified by BiG-FAM in *Umbilicaria*, along with the number of BGCs clustering into pre-characterized GCFs in BiG-FAM and their distance ( $d$ ) groups. Distance is a measure of how closely a given BGC is related to other BGCs ( $d \leq 400$  suggests that the cluster codes for a structurally and functionally similar NP;  $d = 400-900$  indicates that the BGC codes for a related but structurally and functionally divergent NP; and  $d > 900$  suggests that the BGC potentially codes for a novel NP). (B) Bar plots representing the percentage of BGCs in each *Umbilicaria* species with  $d \leq 400$ ,  $d = 400-900$ , and  $d > 900$ . Only a small proportion of BGCs in each species could be grouped into a pre-characterized GCF in the BiG-FAM database (21,678 species, 1,225,071 BGCs, and 29,955 GCFs), whereas a large proportion of BGCs are only distantly related to the pre-characterized BGCs. Approximately 15%–30% of BGCs could not be grouped into BiG-FAM GCFs and, therefore, potentially code for structurally and functionally divergent NPs.

the proportion of BGCs linked to putatively novel NPs in a group of closely related LFF. The identification of 23 clusters that can encode putatively novel compounds can provide useful insights for novel drug leads.

In this study, we mined the genomes of the *Umbilicaria* spp. to identify all BGCs (Figure 1), clustered the structurally novel BGCs and functionally similar BGCs into GCFs (Figures 3A, B), and identified gene clusters potentially coding for novel NPs (Figure 4; Supplementary Information 1). Using *Umbilicaria* spp. as a study system, we show that the LFF biosynthetic landscape is diverse from that of non-lichenized fungi and bacteria. The LFF biosynthetic landscape is particularly rich in PKSs (Figure 2), with a substantial portion of BGCs (about 28% in case of *Umbilicaria*) potentially coding for novel NPs (Figures 3A, B). To the best of our knowledge, this is the first investigation of this kind using state-of-the-art computational tools to determine the proportion of metabolic diversity in LFF potentially coding for novel compounds and to identify

candidate genes as a source of drug leads to enable drug discovery efforts to be prioritized.

## Biosynthetic potential and biosynthetic gene cluster diversity of *Umbilicaria* spp.

Although only PKS-derived NPs are reported from *Umbilicaria* species (gyrophoric acid, umbilicic acid, hiassic acid, etc.) (Posner et al., 1992; Davydov et al., 2017; Singh et al., 2021b), we found that the *Umbilicaria* BGC landscape is biosynthetically diverse and comprises three to five classes of NPs (Figures 1A, B). This is also the case for most other LFF; for instance, PKS-derived NPs are reported from *Bacidia* spp., *Cladonia* spp., *Endocarpon* spp., *Evernia prunastri*, *U. pustulata*, and *Pseudevernia furfuracea*, but all of them contain several PKS, NRPS, and terpene gene clusters (Calchera et al., 2019; Singh et al., 2021a; Singh et al., 2021b; Wang et al., 2021;

Gerasimova et al., 2022). All the above-mentioned studies show that the biosynthetic potential of LFF vastly exceeds their detectable chemical diversity. On average, LFF may contain up to 30–40 BGCs, but the number of identified compounds per species is usually fewer than 10 (Calchera et al., 2019; Pizarro et al., 2020; Singh et al., 2021a). This could be because most of the clusters are silent and do not synthesize the NP, or it could be simply because of the failure to detect the NP. Bioinformatic characterization of entire BGC landscape followed by identification of most distinct BGCs provides a way to estimate the novelty of all BGCs, including the unexpressed and silent ones.

## Biosynthetic gene cluster diversity of non-lichenized fungi compared with bacteria and non-lichenized fungi

We identified five classes of BGCs in the *Umbilicaria* genomes. PKSs were the most dominant class, accounting for about 50% of BGCs, followed by terpenes (19%) and NRPSs (14%) (Figures 1, 2A). BGCs, including PKS, typically make up the majority of BGCs in LFF, for instance about 60% in *E. prunastri*, 61% in *P. furfuracea*, 65% in *Cladonia* spp., 58% in *E. pusillum*, 46% in *Lobaria pulmonaria*, and 63% in *Ramalina peruviana* (Calchera et al., 2019; Kim et al., 2021; Singh et al., 2021a; Singh et al., 2021b).

Robey et al. (2021) identified 36,399 BGCs in 1,037 fungal genomes, which suggests that the average number of BGCs in a non-lichenized fungal genome is 35. This is lower than what has been reported from bacteria, with Liu et al. (2022) reporting 170,685 BGCs from 5,666 genomes (i.e., an average of 30 BGCs per genome). *Umbilicaria* species have, on average, 24 BGCs, which is lower than the average number of BGCs present in non-lichenized fungi and bacteria. However, *Umbilicaria* species, in general, are chemically not particularly diverse (Singh et al., 2022) and are, therefore, expected to have a smaller number of BGCs than other LFF.

Although the number of publicly accessible, good-quality genomes is somewhat lower for LFF (< 25), than for bacteria and non-lichenized fungi, the data available [nine *Umbilicaria* spp. genomes (Singh et al., 2022) plus nine other publicly available lichen genomes] suggest that the predominance of PKSs is a common feature of BGCs in LFF, accounting for more than 50% of the total number of BGCs. In contrast, NRPSs are the most prevalent BGC class in bacteria and non-lichenized fungi, accounting for about 30% and 42% of BGCs, respectively, followed by the PKSs (Figures 2B, C). This suggests that the biosynthetic potential of LFF is unique especially with respect to PKS diversity. In this regard, a recent study suggested that,

although bacteria and fungi may share a few NPs, they do not have an overlapping chemical space and, instead, have distinct biosynthetic potential (Robey et al., 2021). LFF, having a distinct BGC landscape, present a complementary source of NPs with promising medically relevant biosynthetic properties.

## *Umbilicaria* biosynthetic gene cluster: gene cluster families and novel natural products

Gene cluster families are the groups of BGCs that encode the same or very similar molecules. A total of 217 BGCs from nine *Umbilicaria* species were clustered into 135 unique GCFs (Figure 3A). This suggests that *Umbilicaria* spp. are potentially capable of synthesizing many structurally and functionally different NPs, although in nature only one compound class is typically detected (depsides, coded by the BGCs with PKS as the core gene).

Only a small fraction (8%) of *Umbilicaria* BGCs could be clustered with the pre-characterized BGCs (Figures 3A, B). About 71% of the BGCs were clustered to BiG-FAM GCFs with distance greater than 400–900, indicating that they were only distantly related in structure and function (Figures 3A, B). These BGCs are also interesting candidates to be investigated for their biosynthetic properties, as even a minor difference in the cluster and the chemistry of the final metabolites could cause a crucial difference in bioactivity related to function and the pharmacological potential of the product (Lautié et al., 2020).

Approximately 21% of BGCs were highly divergent ( $d > 900$ ) and are novel, potentially coding for structurally and functionally unique NPs, and could, therefore, be an interesting target for NP-based drug discovery (Figure 3B). The strikingly large number of novel BGCs in a single fungal genus adds to the mounting evidence that non-model and understudied taxa are an enormous, untapped source of novel NPs.

Genome mining for large genomic regions, such as fungal BGCs, works best when the genomes under study are complete and contiguous, as well as reliably annotated. Many publicly available LFF genomes do not fulfill these criteria, thus preventing a taxonomically broad study of biosynthetic novelty encoded in the genomes of LFF. We were surprised that even a “chemically boring” lichen taxon, such as the genus *Umbilicaria*, harbored 43 BGCs putatively encoding a diverse range of previously unknown NPs. This leads us to speculate that chemically more diverse taxa, for example, Lecanorales or Pertusariales, each of which includes hundreds of species, are even richer sources of BGCs with novel functions and of compounds with potential novel pharmaceutical applications.

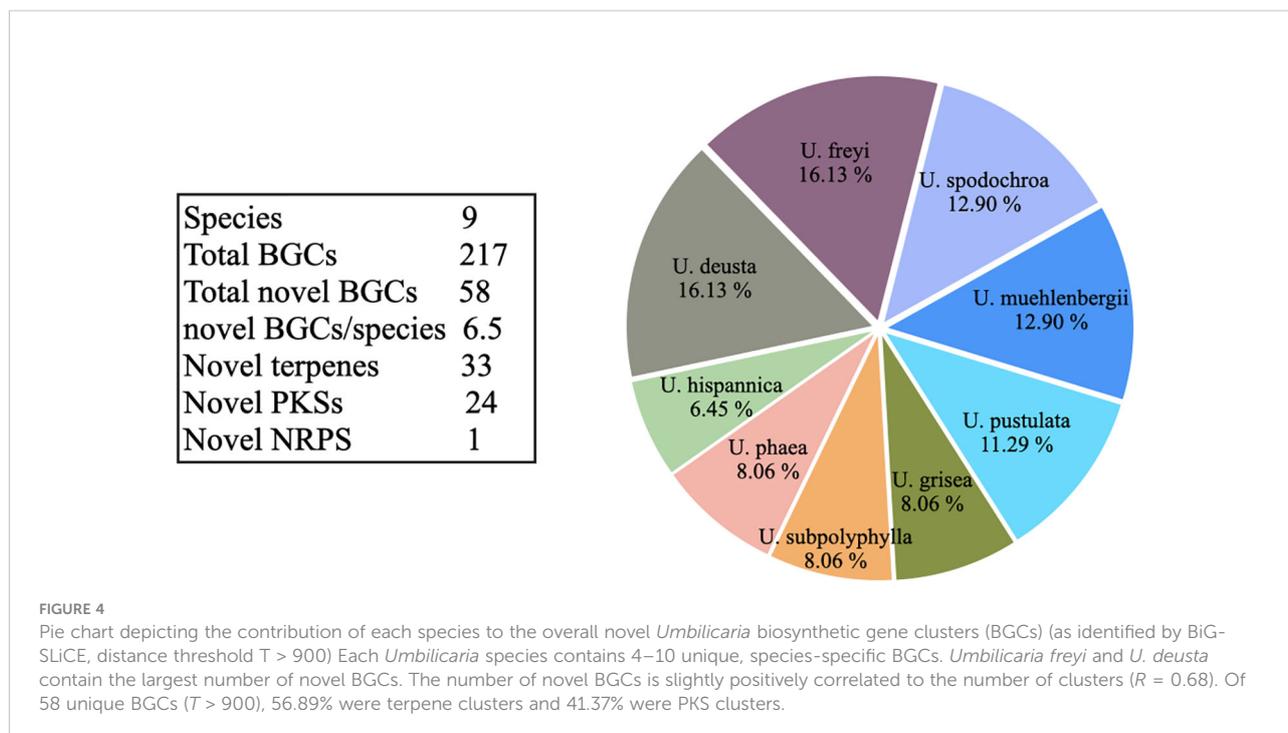
## Unique biosynthetic gene clusters within *Umbilicaria* spp.: BiG-SLiCE

Biosynthetic gene clusters that are unique to one species are candidates for interesting NPs (Navarro-Muñoz et al., 2020; Kautsar et al., 2021b; Robey et al., 2021). On average, each *Umbilicaria* species contains seven species-specific BGCs. *U. deusta* and *U. freyi* have the greatest number of novel BGCs, whereas *U. hispanica* contains the fewest (Figure 4). This suggests that even closely-related species (i.e., species within a single genus) contain diverse biosynthetic potential. Species- or strain-specific biosynthetic potential has already been demonstrated for LFF, for example in *U. pustulata* (Singh et al., 2021b) and *P. furfuracea* (Singh et al., 2021a), and it is rather common among fungi (Alam et al., 2021; Robey et al., 2021; Singh et al., 2021b). For instance, the majority (57%) of the BGCs in *Streptomyces* are strain specific (Choudoir et al., 2018). The unique BGCs within *Umbilicaria* belong to the BGC classes PKSs, terpenes, and NRPSs, as well as to the indoles (Supplementary Information S2). Notably, of these classes, only PKS-derived NPs have been well studied in LFF. Several studies have shown PKS-derived NPs to have diverse pharmacological properties (Manojlović et al., 2012; Cardile et al., 2017; Ingelfinger et al., 2020).

Two PKSs obtained a pairing distance greater than 1800. These PKSs were the most divergent (Supplementary Information S2)

within *Umbilicaria* and are “orphan (i.e., clusters for which corresponding metabolite cannot be predicted). Recently, several orphan clusters have been activated to synthesize a compound with useful pharmacological properties; for example, the antibiotic holomycin gene cluster from the marine bacterium *Photobacterium galathea* (Mattern et al., 2015; Shi et al., 2019; Ziko et al., 2019; Buijs et al., 2020). The novel and orphan clusters reported in this study are potentially interesting source of molecules with unique pharmacological properties and may novel serve as drug leads.

About 17% of fungal BGCs, 8% of bacterial BGCs, and 19% of LFF BGCs are terpenes (Figure 2). Terpenes are pharmaceutically extremely versatile, having antimicrobial, anti-inflammatory, neurodegenerative, and cytotoxic properties (Jaeger and Cuny, 2016; Cox-Georgian et al., 2019; Guimarães et al., 2019; Jiang et al., 2020; Yang et al., 2020; Del Prado-Audelo et al., 2021). Among the most common plant-derived terpenes and terpenoids are curcumin and eucalyptus oil. Although several studies have reported the pharmacological properties of fungal terpenes, such studies on LFF-derived terpenes are lacking, even though LFF genomes contain higher number of terpenes. In this study, we also report structurally and functionally unique terpenes as promising candidates to be investigated for their pharmaceutical potential.



## Data availability statement

Publicly available datasets were analyzed in this study. These data can be found here: National Center for Biotechnology Information (NCBI) BioProject, <https://www.ncbi.nlm.nih.gov/bioproject/>, with accession: PRJNA820300 and Figshare, <https://figshare.com/>, with accession: 10.6084/m9.figshare.19625997. The datasets supporting the conclusions of this article are available in the GenBank repository, accession PRJNA820300, under the accession numbers JALILQ000000000-JALILY000000000. The lichen samples of the corresponding *Umbilicaria* spp. are available as BioSamples SAMN27294873–SAMN27294881 and the mycobiont samples are available as BioSamples SAMN26992773–SAMN26992781. The antiSMASH files of *Umbilicaria* spp. are available at Figshare (DOI: [10.6084/m9.figshare.19625997](https://doi.org/10.6084/m9.figshare.19625997)).

## Author contributions

GS analyzed and interpreted the data, generated the figures and tables, and wrote the manuscript. FG analyzed the data and assisted with the bioinformatic parts of the study. IS interpreted the data, co-prepared the figures, and co-wrote the manuscript. All authors read and approved the final manuscript.

## Funding

This research was funded by LOEWE-Centre TBG, funded by the Hessen State Ministry of Higher Education, Research and the Arts (HMWK).

## References

- Aigle, B., Lautru, S., Spittler, D., Dickschat, J. S., Challis, G. L., Leblond, P., et al. (2014). Genome mining of streptomyces ambofaciens. *J. Ind. Microbiol. Biotechnol.* 41, 251–263. doi: [10.1007/s10295-013-1379-y](https://doi.org/10.1007/s10295-013-1379-y)
- Alam, K., Islam, M. M., Li, C., Sultana, S., Zhong, L., Shen, Q., et al. (2021). Genome mining of pseudomonas species: Diversity and evolution of metabolic and biosynthetic potential. *Molecules* 26, 7524. doi: [10.3390/molecules26247524](https://doi.org/10.3390/molecules26247524)
- Bertrand, R. L., Abdel-Hameed, M., and Sorensen, J. L. (2018). Lichen biosynthetic gene clusters part II: Homology mapping suggests a functional diversity. *J. Nat. Prod.* 81, 732–748. doi: [10.1021/acs.jnatprod.7b00770](https://doi.org/10.1021/acs.jnatprod.7b00770)
- Bertrand, R. L., and Sorensen, J. L. (2018). A comprehensive catalogue of polyketide synthase gene clusters in lichenizing fungi. *J. Ind. Microbiol. Biotechnol.* 45, 1067–1081. doi: [10.1007/s10295-018-2080-y](https://doi.org/10.1007/s10295-018-2080-y)
- Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., et al. (2019). antiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* 47, W81–W87. doi: [10.1093/nar/gkz310](https://doi.org/10.1093/nar/gkz310)
- Borodovsky, M., and Lomsadze, A. (2011). Eukaryotic gene prediction using GeneMark.hmm-e and GeneMark-ES. *Curr. Protoc. Bioinfo.* 4, 1–10. doi: [10.1002/0471250953.bi0406s35](https://doi.org/10.1002/0471250953.bi0406s35)
- Boustie, J., and Grube, M. (2005). Lichens—a promising source of bioactive secondary metabolites. *Plant Genet. Resour.* 3, 273–287. doi: [10.1079/PGR200572](https://doi.org/10.1079/PGR200572)

## Acknowledgments

We thank Professor Marnix Medema and Dr Satria Kautsar for their support with the BiG-SLiCE program.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/ffunb.2022.993171/full#supplementary-material>

### SUPPLEMENTARY INFORMATION S1

Most divergent BGCs in *Umbilicaria*, as identified by BiG-FAM, along with the cluster information and sequence.

### SUPPLEMENTARY INFORMATION S2

Most distantly related BGCs within *Umbilicaria*, as identified by BiG-SLiCE, along with the cluster information.

- Buijs, Y., Isbrandt, T., Zhang, S.-D., Larsen, T. O., and Gram, L. (2020). The antibiotic andrimid produced by vibrio corallilyticus increases expression of biosynthetic gene clusters and antibiotic production in photobacterium galathea. *Front. Microbiol.* 113276. doi: [10.3389/fmicb.2020.622055](https://doi.org/10.3389/fmicb.2020.622055)

Calchera, A., Dal Grande, F., Bode, H. B., and Schmitt, I. (2019). Biosynthetic gene content of the “perfume lichens” evernia prunastri and *Pseudevernia furfuracea*. *Molecules* 24, 203. doi: [10.3390/molecules24010203](https://doi.org/10.3390/molecules24010203)

Calcott, M. J., Ackerley, D. F., Knight, A., Keyzers, R. A., and Owen, J. G. (2018). Secondary metabolism in the lichen symbiosis. *Chem. Soc. Rev.* 47, 1730–1760. doi: [10.1039/C7CS00431A](https://doi.org/10.1039/C7CS00431A)

Cardile, V., Graziano, A. C. E., Avola, R., Piovano, M., and Russo, A. (2017). Potential anticancer activity of lichen secondary metabolite physodic acid. *Chem. Biol. Interact.* 263, 36–45. doi: [10.1016/j.cbi.2016.12.007](https://doi.org/10.1016/j.cbi.2016.12.007)

Chakraborty, K., Kizhakkekalam, V. K., Joy, M., and Chakraborty, R. D. (2021). A leap forward towards unraveling newer anti-infective agents from an unconventional source: A draft genome sequence illuminating the future promise of marine heterotrophic *Bacillus* sp. against drug-resistant pathogens. *Mar. Biotechnol.* 23, 790–808. doi: [10.1007/s10126-021-10064-1](https://doi.org/10.1007/s10126-021-10064-1)

Choudoir, M. J., Pepe-Ranney, C., and Buckley, D. H. (2018). Diversification of secondary metabolite biosynthetic gene clusters coincides with lineage divergence in streptomyces. *Antibiotics* 7, 1–15. doi: [10.3390/antibiotics7010012](https://doi.org/10.3390/antibiotics7010012)

- Cox-Georgian, et al., (2019). Therapeutic and medicinal uses of terpenes *Med. Plants: From Farm to Pharm.* 333–359. doi: 10.1007/978-3-030-31269-5\_15
- Cragg, G. M., and Newman, D. J. (2013). Natural products: A continuing source of novel drug leads. *Biochim. Biophys. Acta - Gen. Subj.* 1830, 3670–3695. doi: 10.1016/j.bbagen.2013.02.008
- Davydov, E. A., Peršoh, D., and Rambold, G. (2017). Umbilicariaceae (lichenized ascomycota) – trait evolution and a new generic concept. *Taxon* 66, 1282–1303. doi: 10.12705/666.2
- Del Prado-Audelo, M. L., Cortés, H., Caballero-Floran, I. H., González-Torres, M., Escutia-Guadarrama, L., Bernal-Chávez, S. A., et al. (2021). Therapeutic applications of terpenes on inflammatory diseases. *Front. Pharmacol.* 12, 122114. doi: 10.3389/fphar.2021.704197
- Demain, A. L. (2014). Importance of microbial natural products and the need to revitalize their discovery. *J. Ind. Microbiol. Biotechnol.* 41, 185–201. doi: 10.1007/s10295-013-1325-z
- Gerasimova, J. V., Beck, A., Werth, S., and Resl, P. (2022). High diversity of type I polyketide genes in *Bacidia rubella* as revealed by the comparative analysis of 23 lichen genomes. *J. Fungi* 8, 449. doi: 10.3390/jof8050449
- Guimarães, A. C., Meireles, L. M., Lemos, M. F., Guimarães, M. C. C., Endringer, D. C., Fronza, M., et al. (2019). Antibacterial activity of terpenes and terpenoids present in essential oils. *Molecules* 24, 2471. doi: 10.3390/molecules24132471
- Huson, D. H., Beier, S., Flade, I., Górski, A., El-Hadidi, M., Mitra, S., et al. (2016). MEGAN community edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* 12, e1004957. doi: 10.1371/journal.pcbi.1004957
- Ingelfinger, R., Henke, M., Roser, L., Ulshöfer, T., Calchera, A., Singh, G., et al. (2020). Unraveling the pharmacological potential of lichen extracts in the context of cancer and inflammation with a broad screening approach. *Front. Pharmacol.* 11, 11322. doi: 10.3389/fphar.2020.01322
- Jaeger, R., and Cuny, E. (2016). Terpenoids with special pharmacological significance: A review. *Nat. Prod. Commun.* 11, 1373–1390. doi: 10.1177/1934578x1601100946
- Jensen, P. R. (2016). Natural products and the gene cluster revolution. *Trends Microbiol.* 24, 968–977. doi: 10.1016/j.tim.2016.07.006
- Jiang, M., Wu, Z., Guo, H., Liu, L., and Chen, S. (2020). A review of terpenes from marine-derived fungi: 2015–2019. *Mar. Drugs* 18, 321. doi: 10.3390/md18060321
- Kautsar, S. A., Blin, K., Shaw, S., Navarro-Muñoz, J. C., Terlouw, B. R., van der Hoof, J. J. J., et al. (2020). MIBiG 2.0: A repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.* 48, D454–D458. doi: 10.1093/nar/gkz882
- Kautsar, S. A., Blin, K., Shaw, S., Weber, T., and Medema, M. H. (2021a). BiG-FAM: The biosynthetic gene cluster families database. *Nucleic Acids Res.* 49, D490–D497. doi: 10.1093/nar/gkaa812
- Kautsar, S. A., van der Hoof, J. J. J., De Ridder, D., and Medema, M. H. (2021b). BiG-SLiCE: A highly scalable tool maps the diversity of 1.2 million biosynthetic gene clusters. *Gigascience* 10, 1–17. doi: 10.1093/gigascience/giaa154
- Kealey, J. T., Craig, J. P., and Barr, P. J. (2021). Identification of a lichen depside polyketide synthase gene by heterologous expression in *Saccharomyces cerevisiae*. *Metab. Eng. Commun.* 13, e00172. doi: 10.1016/j.mec.2021.e00172
- Keller, N. P. (2019). Fungal secondary metabolism: Regulation, function and drug discovery. *Nat. Rev. Microbiol.* 17, 167–180. doi: 10.1038/s41579-018-0121-1
- Kim, W., Liu, R., Woo, S., Bin Kang, K., Park, H., Yu, Y. H., et al. (2021). Linking a gene cluster to atranorin, a major cortical substance of lichens, through genetic dereplication and heterologous expression. *MBio* 12:e0111121. doi: 10.1128/mbio.01111-21
- Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37, 540–546. doi: 10.1038/s41587-019-0072-8
- Kum, E., and İnce, E. (2021). Genome-guided investigation of secondary metabolites produced by a potential new strain *Streptomyces* BA2 isolated from an endemic plant rhizosphere in Turkey. *Arch. Microbiol.* 203, 2431–2438. doi: 10.1007/s00203-021-02210-z
- Lautié, E., Russo, O., Ducrot, P., and Boutin, J. A. (2020). Unraveling plant natural chemical diversity for drug discovery purposes. *Front. Pharmacol.* 11, 11322. doi: 10.3389/fphar.2020.00397
- Liu, M., Li, Y., and Li, H. (2022). Deep learning to predict the biosynthetic gene clusters in bacterial genomes. *J. Mol. Biol.* 434, 167597 (Boca Raton). doi: 10.1016/j.jmb.2022.167597
- Lumbsch, H. T. (1998). “Chemical fungal taxonomy: An overview,” in *Chemical fungal taxonomy*. Eds. J. C. Frisvad, P. D. Bridge and D. K. Arora (Boca Raton: CRC Press), 1–18. doi: 10.1201/9781003064626-1
- Manojlović, N., Ranković, B., Kosanić, M., Vasiljević, P., and Stanojković, T. (2012). Chemical composition of three *Parmelia* lichens and antioxidant, antimicrobial and cytotoxic activities of some their major metabolites. *Phytomedicine* 19, 1166–1172. doi: 10.1016/j.phymed.2012.07.012
- Mattern, D. J., Schoeler, H., Weber, J., Novohradská, S., Kraibooj, K., Dahse, H. M., et al. (2015). Identification of the antiphagocytic tryptacin gene cluster in the human-pathogenic fungus *Aspergillus fumigatus*. *Appl. Microbiol. Biotechnol.* 99, 10151–10161. doi: 10.1007/s00253-015-6898-1
- Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., et al. (2011). antiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* 39, W339–W346. doi: 10.1093/nar/gkr466
- Meiser, A., Otte, J., Schmitt, I., and Grande, F. D. (2017). Sequencing genomes from mixed DNA samples - evaluating the metagenome skimming approach in lichenized fungi. *Sci. Rep.* 7, 1–13. doi: 10.1038/s41598-017-14576-6
- Navarro-Muñoz, J. C., Selem-Mojica, N., Mullowney, M. W., Kautsar, S. A., Tryon, J. H., Parkinson, E. I., et al. (2020). A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* 16, 60–68. doi: 10.1038/s41589-019-0400-9
- Newman, D. J., and Cragg, G. M. (2012). Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.* 75, 311–335. doi: 10.1021/np200906s
- Newman, D. J., and Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* 83, 770–803. doi: 10.1021/acs.jnatprod.9b01285
- Newman, D. J., Cragg, G. M., and Snader, K. M. (2003). Natural products as sources of new drugs over the period 1981–2002. *J. Nat. Prod.* 66, 1022–1037. doi: 10.1021/np0300096l
- Palmer, J., and Stajich, J. (2019). Nextgenus/funannotate: Funannotate v1.5.3 (Version 1.5.3). *Zenodo*. doi: 10.5281/zenodo.2604804
- Park, S. Y., Choi, J., Lee, G. W., Jeong, M. H., Kim, J. A., Oh, S. O., et al. (2014). Draft genome sequence of *Umbilicaria muelhenbergii* KoLRLIF000956, a lichen-forming fungus amenable to genetic manipulation. *Genome Announc.* 2, e00357. doi: 10.1128/genomeA.00357-14
- Pizarro, D., Divakar, P. K., Grewe, F., Crespo, A., Dal Grande, F., and Lumbsch, H. T. (2020). Genome-wide analysis of biosynthetic gene cluster reveals correlated gene loss with absence of usnic acid in lichen-forming fungi. *Genome Biol. Evol.* 12, 1858–1868. doi: 10.1093/gbe/evaa189
- Posner, B., Feige, G. B., and Huneck, S. (1992). Studies on the chemistry of the lichen genus *Umbilicaria* hoffm. *Z. fur Naturforsch. - Sect. C J. Biosci.* 47, 1–9. doi: 10.1515/znc-1992-1-202
- Qin, M., Wu, S., Li, A., Zhao, F., Feng, H., Ding, L., et al. (2019). LRScarf: Improving draft genomes using long noisy reads. *BMC Genomics* 20, 955. doi: 10.1186/s12864-019-6337-2
- Rigali, S., Anderssen, S., Naômé, A., and van Wezel, G. P. (2018). Cracking the regulatory code of biosynthetic gene clusters as a strategy for natural product discovery. *Biochem. Pharmacol.* 153, 24–34. doi: 10.1016/j.bcp.2018.01.007
- Robey, M. T., Caesar, L. K., Drott, M. T., Keller, N. P., and Kelleher, N. L. (2021). An interpreted atlas of biosynthetic gene clusters from 1,000 fungal genomes. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2020230118. doi: 10.1073/pnas.2020230118
- Shi, J., Zeng, Y. J., Zhang, B., Shao, F. L., Chen, Y. C., Xu, X., et al. (2019). Comparative genome mining and heterologous expression of an orphan NRPS gene cluster direct the production of ashimides. *Chem. Sci.* 10, 3042–3048. doi: 10.1039/c8sc05670f
- Shrestha, G., and St. Clair, L. L. (2013). Lichens: A promising source of antibiotic and anticancer drugs. *Phytochem. Rev.* 12, 229–244. doi: 10.1007/s11101-013-9283-7
- Shukla, V., Joshi, G. P., and Rawat, M. S. M. (2010). Lichens as a potential natural source of bioactive compounds: A review. *Phytochem. Rev.* 9, 303–314. doi: 10.1007/s11101-010-9189-6
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Singh, G., Armaleo, D., Dal Grande, F., and 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 *Pseudevernia furfuracea*. *Biomolecules* 11, 1445. doi: 10.3390/biom11101445
- Singh, G., Calchera, A., Merges, D., Otte, J., Schmitt, I., and Grande, F. D. (2022). A candidate gene cluster for the bioactive natural product gyrophoric acid in lichen-forming fungi *Microbiology Spectrum* 10, e00109–22. doi: 10.1128/spectrum.00109-22
- Singh, G., Calchera, A., Schulz, M., Drechsler, M., Bode, H. B., and Schmitt, I. (2021b). Climate-specific biosynthetic gene clusters in populations of a lichen-forming fungus. *Environ. Microbiol.* 23, 4260–4275. doi: 10.1111/1462-2920.15605

Wang, J., Nielsen, J., and Liu, Z. (2021). Synthetic biology advanced natural product discovery. *Metabolites* 11, 785. doi: 10.3390/metabo11110785

Yang, W., Chen, X., Li, Y., Guo, S., Wang, Z., and Yu, X. (2020). Advances in pharmacological activities of terpenoids. *Nat. Prod. Commun.* 15, 1–13. doi: 10.1177/1934578X20903555

Yuan, H., Ma, Q., Ye, L., and Piao, G. (2016). The traditional medicine and modern medicine from natural products. *Molecules* 21, 559. doi: 10.3390/molecules21050559

Ziko, L., Saqr, A. H. A., Ouf, A., Gimpel, M., Aziz, R. K., Neubauer, P., et al. (2019). Antibacterial and anticancer activities of orphan biosynthetic gene clusters from Atlantis II red Sea brine pool. *Microb. Cell Fact.* 18, 56. doi: 10.1186/s12934-019-1103-3