bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 # A candidate gene cluster for the bioactive natural product

2 # gyrophoric acid in lichen-forming fungi

3

4 Garima Singh[1,2,*], Anjuli Calchera[1], Dominik Merges[1], Jürgen Otte[1], Imke Schmitt[1,2,3],

5 Francesco Dal Grande[1,2]

6 [1] Senckenberg Biodiversity and Climate Research Centre (SBiK-F), 60325 Frankfurt am

7 Main, Germany * e-mail: garima.singh@senckenberg.de

8 [2] LOEWE Center for Translational Biodiversity Genomics (TBG), 60325 Frankfurt am Main,

9 Germany

10 [3] Institute of Ecology, Diversity and Evolution, Goethe University, 60325 Frankfurt am Main,

11 Germany

12

13

14 ## Key words

15 Biosynthetic genes, depsides, fungi, genome mining, long-read sequencing, microbial

16 biotechnology, PKS phylogeny, Secondary metabolites, *Umbilicaria*

17

18 ## Abstract

19 Natural products of lichen-forming fungi are structurally diverse and have a variety of

20 medicinal properties. Despite this, they a have limited implementation in industry, because the

21 corresponding genes remain unknown for most of the natural products. Here we implement a

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

22 long-read sequencing and bioinformatic approach to identify the biosynthetic gene cluster of
23 the bioactive natural product gyrophoric acid (GA). Using 15 high-quality genomes
24 representing nine GA-producing species of the lichen-forming fungal genus *Umbilicaria*, we
25 identify the most likely GA cluster and investigate cluster gene organization and composition
26 across the nine species. Our results show that GA clusters are promiscuous within
27 *Umbilicaria,* with only three genes that are conserved across species, including the PKS gene.
28 In addition, our results suggest that the same cluster codes for different but structurally similar
29 NPs, i.e., GA, umbilicaric acid and hiascic acid, bringing new evidence that lichen metabolite
30 diversity is also generated through regulatory mechanisms at the molecular level. Ours is the
31 first study to identify the most likely GA cluster, and thus provides essential information to
32 open new avenues for biotechnological approaches to producing and modifying GA and
33 similar lichen-derived compounds. We show that bioinformatics approaches are useful in
34 linking genes and potentially associated natural products. Genome analyses help unlocking
35 the pharmaceutical potential of organisms such as lichens, which are biosynthetically diverse
36 but slow growing, and difficult to cultivate due to their symbiotic nature.

37

38 Importance

39 The implementation of natural products in the pharmaceutical industry relies on the
40 possibility of modifying the natural product (NP) pathway to optimize yields and
41 pharmacological effects. Characterization of genes and pathways underlying natural product
42 biosynthesis is a major bottleneck for the use of natural products in the pharmaceutical
43 industry. Genome mining is a promising and relatively cost- and time-effective approach to
44 exploit unexplored NP resources for drug discovery. In this study, we identify the most likely

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

45 gene cluster for the lichen-forming fungal depside gyrophoric acid in nine *Umbilicaria*
46 species. This compound shows cytotoxic and antiproliferative properties against several
47 cancer cell lines, and is also a broad-spectrum antimicrobial agent. We identify the putative
48 GA cluster from nine *Umbilicaria* species. This information paves the way for generating GA
49 analogs with modified properties by selective activation/deactivation of genes.
50

51 Introduction

52 Natural products (NPs) and their derivatives/analogs constitute about 70% of modern
53 medicines (1, 2). NPs alone, however, i.e., unmodified molecules as produced by organisms
54 themselves in nature, constitute only a small portion of this. The vast majority, about 60-65%,
55 are derivatives and analogs of naturally-occurring substances, synthesized through
56 biotechnology or synthetic approaches (2, 3). The use of NPs in the pharmaceutical industry
57 relies on the ability to modify NP pathways in order to optimize yields and pharmacological
58 effects. Culture-dependent approaches to identifying/producing NPs are labor-intensive and
59 time-consuming, and not successful for every organism (4, 5). As a result, the biosynthetic
60 potential of many biosynthetically prolific organisms remains untapped. Information on the
61 genetic background and mechanisms of NP synthesis may thus contribute to fast-track NP-
62 based drug discovery (2, 6).

63 Lichens, symbiotic organisms composed of fungal and photosynthetic partners (green
64 algae or cyanobacteria, or both at the same time) (7–9), are a treasure chest of NPs (10–12).
65 So far, about 1,000 NPs with great structural and functional diversity have been reported from
66 lichen-forming fungi (LFF), and about 300-400 have been screened for bioactivity (11).
67 However, the genetic background of more than 97% of lichen NPs is unknown (13–16).

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

68 Lichen compounds have great pharmacological potential, encompassing antimicrobial,
69 antiproliferative, cytotoxic and antioxidant properties (11, 17–20). However, there are various
70 major bottlenecks for using lichen NPs in the pharmaceutical industry, including low yield in
71 nature, slow growth, tedious isolation/culturing methods, and a lack of understanding of their
72 genetic background. Targeted genome mining approaches integrate the latest DNA
73 sequencing technologies with computational advancements and large, publicly-available
74 databases of pre-identified BGCs to characterize genes coding for NPs (1, 21, 22). This
75 approach combines genome mining with the expected genetics of the NP to narrow down the
76 candidate biosynthetic genes.

77 In-silico approaches for linking natural products with their respective biosynthetic
78 gene clusters (BGCs) – genomic clusters of biosynthetic-related genes typically found in
79 fungi (23–25) – are becoming more common in LFF due to the increased availability of
80 genomic resources and databases (1, 21, 22), improvement of detection software and genome
81 mining tools, stabilizing PKS phylogenies, and information gained from recent successes in
82 the heterologous expression of *PKSs* from LFF (13, 14). For instance, the clade “Group I,
83 PKS 16” from Kim et al (13) is associated with the biosynthesis of orsellinic acid derivatives
84 (orcinol depsides and depsidones) such as lecanoric acid (14), grayanic acid (15), physodic
85 acid and olivetoric acid (16), whereas the clade “Group IX, PKS23” from Kim et al (13) is
86 associated with the biosynthesis of methylated orsellinic acid derivatives (β-orcinol depsides
87 and depsidones) such as atranorin. The cluster linked to usnic acid biosynthesis is also fairly
88 well understood (26, 27) and corresponds to “Group VI, PKS8” from Kim et al (11).

89 Here, we combine high-throughput long-read sequencing with a comparative
90 genomics approach to identify the putative cluster(s) linked to the synthesis of gyrophoric
91 acid (GA). GA is an NP produced by several LFF species, with a broad spectrum of

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

92 bioactivity (pharmacological properties such as anticancer and antimicrobial activity and
93 industrially-relevant properties such as usage as dyes (19, 28–30)), for which the molecular
94 mechanism and genetics of synthesis remain unknown. Identification of the GA gene cluster
95 would facilitate its production via biotechnology to optimize the yield as well as to generate
96 GA analogs with the desired pharmaceutical effect. For this study, we chose nine species of
97 GA producers belonging to the lichen-forming fungal genus *Umbilicaria* (Table 1). GA is the
98 most characteristic compound of this genus, and is found at high concentrations in all the
99 chosen species (28, 31–33). It is a depside containing three orsellinic acid-type rings joined
100 together by ester bonds (Fig. 1). Apart from GA, several other structurally-related depsides
101 such as umbilicaric acid, lecanoric acid and hiascic acid (Fig. 1) have also been reported from
102 *Umbilicaria,* but these usually constitute a minor fraction (<10%) of the total NPs detected
103 via HPLC (Fig. 1).
104 In the present study, we assembled highly contiguous long-read-based genomes of
105 LFF of the genus *Umbilicaria*, identified the biosynthetic gene clusters of all the species and
106 singled out candidate genes linked to GA biosynthesis.
107

108 ## Materials and methods

109 **Sampling and dataset**

110 We collected samples of the following eight *Umbilicaria* species: *U. deusta*, *U. freyi*, *U.*
111 *grisea*, *U. subpolyphylla*, *U. hispanica*, *U. phaea*, *U. pustulata*, and *U. spodochroa* for
112 genome sequencing (voucher information in Supplementary Table S1). When possible, we
113 sequenced two samples of the same species collected in different climatic zones. This was
114 done to take into account possible intra-specific variation in BGC content as recently shown

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

115 in Singh et al. (34). The genome of *U. muhlenbergii* was downloaded from the JGI database.
116 In addition, we sampled *Dermatocarpon miniatum* as a control, as it does not produce
117 depsides/depsidones.
118

119 **DNA extraction, library preparation and genome sequencing**

120 Lichen thalli were thoroughly washed with sterile water, and checked under the
121 stereomicroscope for the presence of possible contamination and other lichen thalli. DNA was
122 extracted from all the samples using a CTAB-based method (35) as presented in (36).
123 SMRTbell libraries were constructed according to the manufacturer's instructions of
124 the SMRTbell Express Prep Kit v. 2.0 following the Low DNA Input Protocol (Pacific
125 Biosciences, Menlo Park, CA). Total input for samples was approximately 170-800 ng.
126 Ligation with T-overhang SMRTbell adapters was performed at 20°C for 1 h or overnight.
127 Following ligation, the libraries were purified with a 0,45 x or 0,8 x AMPure PB bead clean
128 up step. The subsequent size selection step to remove SMRTbell templates <3 kb was
129 performed with 2,2 x of a 40% (v/ v) AMPure PB bead working solution.
130 SMRT sequencing was performed on the Sequel System II with the Sequel II
131 Sequencing Kit 2.0 using the continuous long read (CLR) mode or the circular consensus
132 sequencing (CCS) mode, 30 h movie time with no pre-extension and Software SMRTLINK
133 8.0. Each metagenomic library was sequenced on one SMART cell at the Medical Center
134 Nijmegen (the Netherlands), or at MPI Dresden.
135

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

136 **Genome assembly and annotation**

137 The continuous long reads (i.e. CLR reads) from the PacBio Sequel II CLR run were first
138 processed into highly accurate consensus sequences (i.e. HiFi reads) using PacBio tool CCS
139 v5.0.0 with default parameters (https://ccs.how). HiFi reads were then assembled into contigs
140 using the assembler metaFlye v2.7 (37). The resulting contigs were scaffolded with LRScaf
141 v1.1.12 (github.com/shingocat/lrscaf, (38)). The scaffolds were then taxonomically binned to
142 extract Ascomyocta reads with blastx using DIAMOND (--more-sensitive --frameshift 15 –
143 range-culling) on a custom database and following the MEGAN6 Community Edition
144 pipeline (39). All scaffolds assigned to Ascomycota were extracted as to represent the
145 *Umbilicaria* spp. Assembly statistics such as number of contigs, total length and N50 were
146 accessed with Assemblathon v2 (40) (Table 1). The completeness of the received mycobiont
147 bins (i.e. the fungal genomes) was estimated using Benchmarking Universal Single-Copy
148 Orthologs (BUSCO) analysis in BUSCO v4 (41).

149

150 **Identification and Annotations of Biosynthetic Gene Clusters**

151 Functional annotation of genomes, including genes, proteins and BGC prediction
152 (antiSMASH (antibiotics & SM Analysis Shell, v5.0)) was performed with scripts
153 implemented in the funannotate pipeline (42, 43). First, the genomes were masked for
154 repetitive elements, and then the gene prediction was performed using BUSCO2 to train
155 Augustus and self-training GeneMark-ES (41, 44). Functional annotation was done with
156 InterProScan (45), egg-NOG-mapper (46, 47) and BUSCO (41) with ascomycota_db models.
157 Secreted proteins were predicted using SignalP (48) as implemented in the funannotate
158 ‘annotate’ command.

159

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

160 **Selecting candidate gene clusters linked to GA biosynthesis**

161 We used the following criteria to select the candidate gene cluster associated with GA
162 synthesis in *Umbilicaria*: 1) it must contain a NR-PKS (as some of the structural features of a
163 NP can be directly inferred from the domain architecture of the *PKS*: *PKSs* without reducing
164 domains (*NR-PKSs*) are linked to non-reduced compounds such as gyrophoric acid, olivetoric
165 acid (16), physodic acid (16) and grayanic acid (15)), 2) it must be present in all the
166 *Umbilicaria* genomes, as all the species have GA as the major secondary metabolite (33), and
167 3) it must be closely related to the *PKSs* involved in the synthesis of orsellinic acid-based
168 compounds (15, 16), because orsellinic acid units constitute the building blocks of GA.
169

170 **Phylogenetic analyses**

171 We extracted the amino acid sequences of all the NR-PKS from the BGCs predicted by the
172 antiSMASH for all the *Umbilicaria* species and *Dermatocarpon miniatum* (Supplementary
173 Table S2). Additionally, the NR-PKS sequences of the following species were downloaded
174 from the previous publications and public databases: *Cladonia borealis, C. grayi*, *C.*
175 *macilenta*, *C. metacorallifera*, *C. rangiferina*, *C. uncialis*, *Pseudevernia furfuracea,*
176 *Stereocaulon alpinum* and *Umbilicaria muhlenbergii*. The final dataset contains amino acid
177 sequences of 229 NR-PKSs from 18 species belonging to four LFF genera. The sequences
178 were aligned using MAFFT as implemented in Geneious v5.4 (49, 50). Gaps were treated as
179 missing data. The maximum likelihood search was performed on the aligned sequences with
180 RAxML-HPC BlackBox v8.1.11 (51) on the Cipres Scientific gateway (52). Phylogenetic
181 trees were visualized using iTOL (53).
182

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

183 **BGC clustering and novel BGCs: BiG-SCAPE and CORASON**

184 We used BiG-SCAPE and CORASON (54) to identify the gene cluster networks and infer
185 evolutionary relationships among clusters of interest among different *Umbilicaria* spp. BiG-
186 SCAPE utilizes antiSMASH (42) and MIBiG databases (55) for inferring BGC sequence
187 similarity networks, whereas CORASON employs a phylogenomic approach to infer
188 evolutionary relationships between the clusters. BiG-SCAPE v1.0.1 was run in --auto mode,
189 to identify BGC families using antiSMASH output files (.gbk) as input. Networks were
190 generated using similarity thresholds of 0.25. The most likely GA cluster from all the
191 *Umbilicaria* spp. was examined for conservation and variation among different *Umbilicaria*
192 species using CORASON pipeline. The antiSMASH .gbk files of the corresponding clusters,
193 based on phylogenetic grouping, were used as input. The most-likely GA cluster from *U.*
194 *deusta* was used as reference to fish out the most closely-related clusters from the other
195 *Umbilicaria* spp.

196

197 # Results

198 **Genome sequencing, assembly and annotation**

199 The genome quality stats and assembly reports of all the genomes generated for this study are
200 presented in Table 1.

201

202 **Phylogenetic analysis**

203 To search for PKS genes involved in the synthesis of GA, we performed a phylogenetic
204 analysis by incorporating our sequences into the most comprehensive PKS dataset currently
205 available (Supplementary Table S2) (13, 16). NR-PKSs have been categorized into nine

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

206 groups based on protein sequence similarity and PKS domain architecture (13). We identified
207 a total of 110 NR-PKSs that were present in 15 *Umbilicaria* genomes (12 NR-PKSs on
208 average per species). Four NR-PKSs were common to all species: PKS15, PKS16, PKS20 and
209 a novel PKS clade (forming a monophyletic, supported clade to PKS33, Fig. 2). Only one
210 NR-PKS per species formed a supported monophyletic clade with PKS16 (Group I, i.e.,
211 orsellinic acid, depside and depsidone NR-PKSs) (Fig. 2). The most-likely NR-PKS for the
212 depsidone grayanic acid and the depsides olivetoric and physodic acid fall within this PKS
213 clade.

214

215 **Gyrophoric acid cluster**

216 The cluster most likely associated with GA is the cluster containing PKS16 (Fig. 2), as 1) it is
217 present in all *Umbilicaria* spp., 2) it contains an *NR-PKS* and 3) it forms a monophyletic
218 group with the clade "Group I, PKS 16" from Kim et al (11). This cluster contains about 11-
219 15 genes, including the core biosynthetic gene *NR-PKS* and a c*yt P450* (Fig. 3). The other
220 genes code for unidentified proteins. The *U. deusta* PKS16 cluster is presented as an example
221 of GA cluster (Fig. 3). The *PKS* is present in all *Umbilicaria* species investigated and displays
222 high homology across species.

223

224 **BGC clustering: BiG-SCAPE and CORASON**

225 BGC sequence similarity networks group gene clusters at multiple hierarchical levels. This
226 analysis implements a 'glocal' alignment mode that accurately groups both complete and
227 fragmented BGCs. The BGCs forming a supported monophyletic clade to PKS16 (Group I)
228 were then analyzed for conservation across species using CORASON. The CORASON

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

229 analysis showed that only three genes on the cluster were shared among the studied
230 *Umbilicaria* species: the core *PKS* and the two genes of unknown function/proteins adjacent
231 to the core gene (Fig. 4).
232

233 Discussion

234 Gyrophoric acid PKS

235 We found only one PKS per species forming a supported monophyletic clade to PKS16
236 (Group I, i.e., orsellinic acid and depside/depsidone PKSs) (Fig. 2). These are the most likely
237 GA PKSs.

238 The BGC associated with the biosynthesis of the following lichen depsides and
239 depsidones have been identified so far: atranorin (13), lecanoric- (14), grayanic- (15),
240 olivetoric- (16) and physodic acid (16). All these studies demonstrate that the PKS alone is
241 capable of synthesizing the backbone depside, whereas modifications such as methylation and
242 oxidation are made by enzymes coded by other genes in the cluster after the release of the
243 depside from the PKS. For instance, the synthesis of atranorin involves at least three genes
244 present within the atranorin cluster, but the core depside is coded only by the PKS (13). The
245 other two genes, a carboxyl methylase (O-methyltransferase) and a *cyt P450*, methylate the
246 carboxyl group and oxidize the methyl group (into -CHO), respectively, to produce the final
247 product atranorin (Fig. 1). As GA does not have side chain modifications (Fig. 1) we propose
248 that the *PKS* alone is involved in GA synthesis.

249 The depside PKSs identified so far code for didepsides, i.e. they contain two phenolic
250 rings joined with an ester bond, for example gyrayanic acid, atranorin, physodic acid and
251 olivetoric acid (13, 15, 16). Ours is the first study to identify the most-likely PKS associated

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

252 with a tridepside synthesis, i.e., three phenolic rings joined with two ester bonds. Our study
253 suggests that the *PKSs* coding for a didepside and a tridepside differ only in the length of the
254 sequence of the SAT domain. A tridepside *PKS* contains longer SAT coding sequence than
255 the didepside *PKS*. The number of ACP and PT domains is the same between the two.
256

257 **GA cluster in *Umbilicaria* spp.**

258 The most-likely GA cluster contains about 11-15 genes in different *Umbilicaria* spp.
259 (Fig. 3, 4). Interestingly, only three genes are common to all analyzed species, the *PKS* and
260 two genes of unknown function (with low sequence similarity to known genes) upstream and
261 downstream of the *PKS I* (Fig. 4). This suggests that these three genes form an integral part of
262 the GA cluster, whereas the other genes are facultative among GA producers. Differences
263 among the clusters synthesizing the same compound have been reported before, and have
264 been associated with species-specific BGC regulation or modifications to the depside released
265 by the PKS (26, 56).

266 The most-likely GA cluster also contains a *cyt P450* (Fig. 3, 4), which has been
267 associated with depsidone production or oxidation of an acyl chain (13, 15, 16). However, the
268 location and orientation of the *cyt P450* in the putative *Umbilicaria* GA cluster is different
269 from a typical depsidone cluster (Fig. 3) (15). In the GA cluster, the *cyt P450* is not located
270 next to and has the same orientation as the *PKS,* whereas in a depsidone cluster, *cyt P450* lies
271 next to the *PKS*, in opposite orientation. Such organization is suggestive of genes being
272 regulated and co-expressed by the same promoter (15). This is the case for the depsidone
273 grayanic acid synthesis (in *Cladonia grayi*), which involves the synthesis of the depside
274 intermediate by *PKS* followed by oxidation of the released depside into depsidone (15). The
275 *PKS* and *cyt P450* form the integral part of depsidone synthesis (57) whereas the depside is

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

276 coded by the *PKS* alone, with the exception of the side chain modifications (13, 14).
277 Therefore, despite being part of the GA cluster, the *cyt P450* does not seem to be involved in
278 GA synthesis or in the synthesis of umbilicaric- and/or lecanoric acid reported from
279 *Umbilicaria* spp. analyzed in this study. The synthesis of hiascic acid however would require
280 the hydroxylation of a methyl group by cyt P450 enzyme after the depside is released from
281 the PKS (Fig. 1, the OH group in bold in hiascic acid). The lower proportion of hiascic acid as
282 compared to GA could be because the *cyt P450* may not be co-expressed with the *PKS*.
283 Even though the functions of most of the genes identified in the present study are
284 unknown, our study provides novel insights into GA cluster composition and organization
285 across different species (Fig. 4). This information is crucial in order to open the way for future
286 genetic manipulation of the GA biosynthetic pathway that may be aimed at increasing
287 structural diversity and/or yield of the products, as well as in order to generate analogs with
288 novel properties.
289
290 **One cluster, different compounds**
291 Variation in cluster composition reflects the potential to produce diverse NPs. Apart from GA,
292 other depsides related in structure to GA, i.e., lecanoric-, umbilicaric- and hiascic acid (Fig. 1)
293 are often reported in *Umbilicaria* spp. as minor metabolites (31). Interestingly, we found only
294 one orcinol-depside *PKS* in *Umbilicaria* spp (Fig. 2). This strongly indicates that all the
295 *Umbilicaria* depsides are coded by the same PKS cluster. One cluster coding for different,
296 structurally-related compounds has also been reported previously (16, 56, 58). For instance, in
297 the case of the antifungal drug caspofungin acetate, a semisynthetic derivative of the NP
298 pneumocandins from the fungus *Glarea lozoyensis*, selective inactivation of different genes in
299 this biosynthetic gene cluster generates 13 different analogues, some of them with elevated

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

300 antifungal activity relative to the original compound and its semisynthetic derivative (59).
301 Similarly, the aspyridone biosynthetic cluster from *Aspergillus nidulans* produces eight
302 different compounds in a heterologous host (58). These studies show that a single PKS cluster
303 is capable of producing different compounds depending upon which genes are co-expressed
304 and on the available starters. In lichens, a single *PKS* has been associated with the synthesis of
305 olivetoric and physodic acid (16) and the same *PKS* has been shown to be involved in the
306 synthesis of lecanoric acid in a heterologous host (14). We propose that the same PKS cluster
307 is most likely involved in the synthesis of GA, umbilicaric- (an additional methyl group, Fig.
308 1), hiascic- (additional hydroxyl group, Fig. 1), and lecanoric acid (didepside with no side
309 chains, Fig. 1) in *Umbilicaria*. It is possible, however, that in nature only GA is synthesized in
310 members of the genus *Umbilicaria*, and the co-occurring minor compound lecanoric acid is a
311 hydrolysis product of GA (57).

312 Interestingly, although umbilicaric acid is reported from some *Umbilicaria* species (*U.*
313 *grisea*, *U. freyi*, *U. mühlenbergii* and *U. subpolyphylla* (31, 61)), O-methyltransferase (OMT)
314 was not identified in the depside-related BGC of any *Umbilicaria* species (Fig. 3). OMT
315 would be required for the methylation of oxygen to produce umbilicaric acid (Fig. 1). Its
316 absence from depside-related BGCs suggests that an external OMT, e.g. from other BGCs,
317 might be involved in the production of umbilicaric acid in *Umbilicaria*. This could explain the
318 lower amounts of umbilicaric acid as compared to GA found in these species (31). In contrast,
319 when the O-methylated compound is the major secondary metabolite, as in the case of
320 grayanic acid and atranorin, OMT is an integral part of the BGC and is co-expressed along
321 with the other crucial genes for grayanic acid production, i.e., the *PKS* and *cyt P450* (13, 15).

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

322

323 **Future perspectives**

324 Advances in long-read sequencing and in computational approaches to genome mining not
325 only enable linking biosynthetic genes to NPs but also provide an overview of the entire gene
326 cluster composition and organization. Ours is the first study to identify the most-likely GA
327 cluster, which is essential for opening up avenues for biotechnological approaches to
328 producing and modifying this compound and possibly other lichen compounds. In particular,
329 this information can be applied to generate novel NP analogs with improved pharmacological
330 properties via synthetic biology, biotechnology and combinatorial biosynthesis approaches.
331 This paves the way to an entirely new horizon for utilizing these understudied taxa for
332 pharmacological industry and drug discovery.

333

334 # Acknowledgements

335 We thank Prof. Daniele Armaleo (Duke University) for his inputs on the domain composition
336 of didepsides and tridepsides.

337

338 # Figure legends

339 **Figure 1** Chemical structures and nomenclature. Structure of a lichen depside, atranorin, GA
340 and other depsides produced by *Umbilicaria* spp.

341 **Figure 2** NR-PKS phylogeny of lichen-forming fungi. This is a maximum-likelihood tree
342 based on amino acid sequences of NR-PKSs from nine *Umbilicaria* spp., six *Cladonia* spp.,
343 *Dermatocarpon miniatum*, *Stereocaulon alpinum* and *Pseudevernia furfuracea*. Branches in

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

344 bold indicate bootstrap support >70%. Green color clades represent the PKSs common to all
345 nine *Umbilicaria* spp. used in this study. PKS groups are based on Kim et al. (13).
346 **Figure 3** Gyrophoric acid cluster from *Umbilicaria deusta* as predicted by antiSMASH.
347 Colored boxes indicate genes. Genes in grey represent genes coding for unknown proteins
348 **Figure 4** CORASON-based PKS phylogeny to elucidate evolutionary relationships and
349 cluster organization of GA cluster in *Umbilicaria* spp. The bar plot below depicts the
350 percentage of *Umbilicaria* species in which a particular gene is present.
351
352 **Table 1:** Genome quality and annotation statistics
353 **Supplementary Tables**
354 1. Voucher information table of the *Umbilicaria* samples collected for this study.
355 2. Dataset used for the phylogenetic analysis, along with the amino acid sequences of the
356 PKS
357

358 # References

359 1. Newman DJ, Cragg GM. 2016. Natural products as sources of new drugs from 1981 to
360 2014. J Nat Prod. American Chemical Society.
361 2. Newman DJ, Cragg GM. 2020. Natural products as sources of new drugs over the
362 nearly four decades from 01/1981 to 09/2019. J Nat Prod. American Chemical Society.
363 3. Newman DJ, Cragg GM. 2007. Natural products as sources of new drugs over the last
364 25 years. J Nat Prod.
365 4. Xu W, Klumbys E, Ang EL, Zhao H. 2020. Emerging molecular biology tools and
366 strategies for engineering natural product biosynthesis. Metab Eng Commun. Elsevier

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

367 B.V.

368 5. Palazzotto E, Tong Y, Lee SY, Weber T. 2019. Synthetic biology and metabolic
369 engineering of actinomycetes for natural product discovery. Biotechnol Adv. Elsevier
370 Inc.

371 6. Atanasov AG, Zotchev SB, Dirsch VM, Orhan IE, Banach M, Rollinger JM, Barreca
372 D, Weckwerth W, Bauer R, Bayer EA, Majeed M, Bishayee A, Bochkov V, Bonn GK,
373 Braidy N, Bucar F, Cifuentes A, D'Onofrio G, Bodkin M, Diederich M, Dinkova-
374 Kostova AT, Efferth T, El Bairi K, Arkells N, Fan TP, Fiebich BL, Freissmuth M,
375 Georgiev MI, Gibbons S, Godfrey KM, Gruber CW, Heer J, Huber LA, Ibanez E,
376 Kijjoa A, Kiss AK, Lu A, Macias FA, Miller MJS, Mocan A, Müller R, Nicoletti F,
377 Perry G, Pittalà V, Rastrelli L, Ristow M, Russo GL, Silva AS, Schuster D, Sheridan
378 H, Skalicka-Woźniak K, Skaltsounis L, Sobarzo-Sánchez E, Bredt DS, Stuppner H,
379 Sureda A, Tzvetkov NT, Vacca RA, Aggarwal BB, Battino M, Giampieri F, Wink M,
380 Wolfender JL, Xiao J, Yeung AWK, Lizard G, Popp MA, Heinrich M, Berindan-
381 Neagoe I, Stadler M, Daglia M, Verpoorte R, Supuran CT. 2021. Natural products in
382 drug discovery: advances and opportunities. Nat Rev Drug Discov. Nature Research.

383 7. Hawksworth DL, Honegger R. 1994. The lichen thallus: a symbiotic phenotype of
384 nutritionally specialized fungi and its response to gall producers., p. 77–98. *In*
385 Williams, MAJ (ed.), Plant Galls: Organisms, Interactions, Populations. Clarendon
386 Press, Oxford.

387 8. Ahmadjian V. 1993. The Lichen Symbiosis. John Wiley and Sons., New York.

388 9. Ahmadjian V. 1982. Algal/fungal symbioses., p. 179–233. *In* Round, FE Chapman, DJ
389 (ed.), Progress in Phycological Research. Elsevier Biomedical Press, Amsterdam.

390 10. Calchera A, Dal Grande F, Bode HB, Schmitt I. 2019. Biosynthetic gene content of the

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

391 "perfume lichens" *Evernia prunastri* and *Pseudevernia furfuracea*. Molecules 24:203.
392 11. Goga M, Elečko J, Marcinčinová M, Ručová D, Bačkorová M, Bačkor M. 2020.
393 Lichen metabolites: an overview of some secondary metabolites and their biological
394 potential, p. 175–209. *In* .
395 12. Boustie J, Grube M. 2005. Lichens—a promising source of bioactive secondary
396 metabolites. Plant Genet Resour 3:273–287.
397 13. Kim W, Liu R, Woo S, Kang K Bin, Park H, Yu YH, Ha H-H, Oh S-Y, Yang JH, Kim
398 H, Yun S-H, Hur J-S. 2021. Linking a gene cluster to atranorin, a major cortical
399 substance of lichens, through genetic dereplication and heterologous expression. MBio
400 e0111121.
401 14. Kealey JT, Craig JP, Barr PJ. 2021. Identification of a lichen depside polyketide
402 synthase gene by heterologous expression in *Saccharomyces cerevisiae*. Metab Eng
403 Commun e00172.
404 15. Armaleo D, Sun X, Culberson C. 2011. Insights from the first putative biosynthetic
405 gene cluster for a lichen depside and depsidone. Mycologia 103:741–754.
406 16. Singh G, Armaleo D, Dal Grande F, Schmitt I. 2021. Depside and depsidone synthesis
407 in lichenized fungi comes into focus through a genome-wide comparison of the
408 olivetoric acid and physodic acid chemotypes of *Pseudevernia furfuracea*.
409 Biomolecules 11:1445.
410 17. Barbero M, Artuso E, Prandi C. 2017. Fungal anticancer metabolites: synthesis towards
411 drug discovery. Curr Med Chem 25:141–185.
412 18. Stanojković T. 2015. Investigations of lichen secondary metabolites with potential
413 anticancer activity, p. 127–146. *In* Lichen Secondary Metabolites: Bioactive Properties
414 and Pharmaceutical Potential. Springer International Publishing.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

415 19. Ingelfinger R, Henke M, Roser L, Ulshöfer T, Calchera A, Singh G, Parnham MJ,
416 Geisslinger G, Fürst R, Schmitt I, Schiffmann S. 2020. Unraveling the pharmacological
417 potential of lichen extracts in the context of cancer and inflammation with a broad
418 screening approach. Front Pharmacol 11:1322.
419 20. Molnár K, Farkas E. 2010. Current results on biological activities of lichen secondary
420 metabolites: A review. Zeitschrift fur Naturforsch - Sect C J Biosci. Verlag der
421 Zeitschrift fur Naturforschung.
422 21. Van Santen JA, Kautsar SA, Medema MH, Linington RG. 2021. Microbial natural
423 product databases: Moving forward in the multi-omics era. Nat Prod Rep. Royal
424 Society of Chemistry.
425 22. Bachmann BO, Van Lanen SG, Baltz RH. 2014. Microbial genome mining for
426 accelerated natural products discovery: Is a renaissance in the making? J Ind Microbiol
427 Biotechnol. J Ind Microbiol Biotechnol.
428 23. Hoffmeister D, Keller NP. 2007. Natural products of filamentous fungi: Enzymes,
429 genes, and their regulation. Nat Prod Rep.
430 24. Keller NP. 2019. Fungal secondary metabolism: regulation, function and drug
431 discovery. Nat Rev Microbiol. Nature Publishing Group.
432 25. Brakhage AA. 2013. Regulation of fungal secondary metabolism. Nat Rev Microbiol.
433 26. Pizarro D, Divakar PK, Grewe F, Crespo A, Dal Grande F, Lumbsch HT. 2020.
434 Genome-wide analysis of biosynthetic gene cluster reveals correlated gene loss with
435 absence of usnic acid in lichen-forming fungi. Genome Biol Evol 12:1858–1868.
436 27. Abdel-Hameed M, Bertrand RL, Piercey-Normore MD, Sorensen JL. 2016. Putative
437 identification of the usnic acid biosynthetic gene cluster by de novo whole-genome
438 sequencing of a lichen-forming fungus. Fungal Biol 120:306–316.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

439 28. Nguyen K-H, Chollet-Krugler M, Gouault N, Tomasi S. 2013. UV-protectant
440 metabolites from lichens and their symbiotic partners. Nat Prod Rep 30:1490.
441 29. Lohezic-Le Devehat F, Legouin B, Couteau C, Boustie J, Coiffard L. 2013. Lichenic
442 extracts and metabolites as UV filters. J Photochem Photobiol B Biol 120:17–28.
443 30. Buçukoglu TZ, Albayrak S, Halici MG, Tay T. 2013. Antimicrobial and antioxidant
444 activities of extracts and lichen acids obtained from some *Umbilicaria* species from
445 central Anatolia, Turkey. J Food Process Preserv 37:1103–1110.
446 31. Posner B, Feige GB, Huneck S. 1992. Studies on the chemistry of the lichen genus
447 *Umbilicaria* hoffm. Zeitschrift fur Naturforsch - Sect C J Biosci 47:1–9.
448 32. Narui T, Culberson CF, Culberson WL, Johnson A, Shibata S. 1996. A contribution to
449 the chemistry of the lichen family umbilicariaceae (Ascomycotina). Bryologist 99:199–
450 211.
451 33. Davydov EA, Peršoh D, Rambold G. 2017. Umbilicariaceae (lichenized Ascomycota)
452 – Trait evolution and a new generic concept. Taxon 66:1282–1303.
453 34. Singh G, Calchera A, Schulz M, Drechsler M, Bode HB, Schmitt I, Dal Grande F.
454 2021. Climate-specific biosynthetic gene clusters in populations of a lichen-forming
455 fungus. Environ Microbiol 00:1462-2920.15605.
456 35. Rubio-Piña JA, Zapata-Pérez O. 2011. Isolation of total RNA from tissues rich in
457 polyphenols and polysaccharides of mangrove plants. Electron J Biotechnol 14.
458 36. Merges D, Dal Grande F, Greve C, Otte J, Schmitt I. 2021. Virus diversity in
459 metagenomes of a lichen symbiosis (*Umbilicaria phaea*): complete viral genomes,
460 putative hosts and elevational distributions. Environ Microbiol 23:6637–6650.
461 37. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone
462 reads using repeat graphs. Nat Biotechnol 37:540–546.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

463 38. Qin M, Wu S, Li A, Zhao F, Feng H, Ding L, Ruan J. 2019. LRScaf: Improving draft
464 genomes using long noisy reads. BMC Genomics 20:955.
465 39. Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, Ruscheweyh H-J, Tappu
466 R. 2016. MEGAN Community Edition - Interactive exploration and analysis of large-
467 scale microbiome sequencing data. PLOS Comput Biol 12:e1004957.
468 40. Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I, Boisvert S,
469 Chapman JA, Chapuis G, Chikhi R, Chitsaz H, Chou W-C, Corbeil J, Del Fabbro C,
470 Docking TR, Durbin R, Earl D, Emrich S, Fedotov P, Fonseca NA, Ganapathy G,
471 Gibbs RA, Gnerre S, Godzaridis É, Goldstein S, Haimel M, Hall G, Haussler D, Hiatt
472 JB, Ho IY, Howard J, Hunt M, Jackman SD, Jaffe DB, Jarvis ED, Jiang H, Kazakov S,
473 Kersey PJ, Kitzman JO, Knight JR, Koren S, Lam T-W, Lavenier D, Laviolette F, Li
474 Y, Li Z, Liu B, Liu Y, Luo R, MacCallum I, MacManes MD, Maillet N, Melnikov S,
475 Naquin D, Ning Z, Otto TD, Paten B, Paulo OS, Phillippy AM, Pina-Martins F, Place
476 M, Przybylski D, Qin X, Qu C, Ribeiro FJ, Richards S, Rokhsar DS, Ruby JG,
477 Scalabrin S, Schatz MC, Schwartz DC, Sergushichev A, Sharpe T, Shaw TI, Shendure
478 J, Shi Y, Simpson JT, Song H, Tsarev F, Vezzi F, Vicedomini R, Vieira BM, Wang J,
479 Worley KC, Yin S, Yiu S-M, Yuan J, Zhang G, Zhang H, Zhou S, Korf IF. 2013.
480 Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate
481 species. Gigascience 2:10.
482 41. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. 2015.
483 BUSCO: assessing genome assembly and annotation completeness with single-copy
484 orthologs. Bioinformatics 31:3210–3212.
485 42. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T.
486 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

487 Nucleic Acids Res 47:W81–W87.
488 43. Palmer J, Stajich J. 2019. Funannotate v1.7.4. Zenodo
489 https://doi.org/https://doi.org/10.5281/ZENODO.2604804.
490 44. Borodovsky M, Lomsadze A. 2011. Eukaryotic gene prediction using GeneMark.hmm-
491 E and GeneMark-ES. Curr Protoc Bioinforma CHAPTER:Unit-4.610.
492 45. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005.
493 InterProScan: Protein domains identifier. Nucleic Acids Res 33:W116.
494 46. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H,
495 Mende DR, Letunic I, Rattei T, Jensen LJ, Von Mering C, Bork P. 2019. EggNOG 5.0:
496 A hierarchical, functionally and phylogenetically annotated orthology resource based
497 on 5090 organisms and 2502 viruses. Nucleic Acids Res 47:D309–D314.
498 47. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Von Mering C,
499 Bork P. 2017. Fast genome-wide functional annotation through orthology assignment
500 by eggNOG-mapper. Mol Biol Evol 34:2115–2122.
501 48. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak
502 S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions
503 using deep neural networks. Nat Biotechnol 37:420–423.
504 49. Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in
505 accuracy of multiple sequence alignment. Nucleic Acids Res 33:511–8.
506 50. Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J,
507 Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A.
508 2011. Geneious. 5.4.
509 51. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic
510 analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–90.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

511 52. Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for
512 inference of large phylogenetic trees, p. 1–8. *In* Proceedings of the Gateway
513 Computing Environments Workshop (GCE)New Orlean.
514 53. Letunic I, Bork P. 2021. Interactive tree of life (iTOL) v5: An online tool for
515 phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296.
516 54. Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH,
517 Parkinson EI, De Los Santos ELC, Yeong M, Cruz-Morales P, Abubucker S, Roeters
518 A, Lokhorst W, Fernandez-Guerra A, Cappelini LTD, Goering AW, Thomson RJ,
519 Metcalf WW, Kelleher NL, Barona-Gomez F, Medema MH. 2020. A computational
520 framework to explore large-scale biosynthetic diversity. Nat Chem Biol 16:60–68.
521 55. Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, Van Der Hooft JJJ,
522 Van Santen JA, Tracanna V, Suarez Duran HG, Pascal Andreu V, Selem-Mojica N,
523 Alanjary M, Robinson SL, Lund G, Epstein SC, Sisto AC, Charkoudian LK, Collemare
524 J, Linington RG, Weber T, Medema MH. 2020. MIBiG 2.0: A repository for
525 biosynthetic gene clusters of known function. Nucleic Acids Res 48:D454–D458.
526 56. Martinet L, Naômé A, Deflandre B, Maciejewska M, Tellatin D, Tenconi E,
527 Smargiasso N, De Pauw E, Van Wezel GP, Rigali S. 2019. A single biosynthetic gene
528 cluster is responsible for the production of bagremycin antibiotics and ferroverdin iron
529 chelators. MBio 10:e01230-19.
530 57. Armaleo D. 1995. Factors affecting depside and depsidone biosynthesis in a cultured
531 lichen fungus. Crypt Bot 5:14–21.
532 58. Wasil Z, Pahirulzaman KAK, Butts C, Simpson TJ, Lazarus CM, Cox RJ. 2013. One
533 pathway, many compounds: Heterologous expression of a fungal biosynthetic pathway
534 reveals its intrinsic potential for diversity. Chem Sci 4:3845–3856.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

535 59. Chen L, Yue Q, Zhang X, Xiang M, Wang C, Li S, Che Y, Ortiz-López FJ, Bills GF,
536 Liu X, An Z. 2013. Genomics-driven discovery of the pneumocandin biosynthetic gene
537 cluster in the fungus *Glarea lozoyensis*. BMC Genomics 14.
538 60. Leuckert C. 1985. Probleme der Flechten-Chemotaxonomie -- Stoffkombinationen und
539 ihre taxonomische Wertung. Problems of lichen chemotaxonomy -- patterns of
540 substances and their taxonomical valuation. Ber Dtsch Bot Ges 98:401–408.
541 61. Davydov EA, Blum OB, Kashevarov GP, Grakhov VP. 2019. Umbilicaria
542 subpolyphylla Oxner: The correct name for U. iberica Sancho & Krzewicka and its
543 bipolar distribution pattern. Lichenologist. Cambridge University Press.
544
545 2. Newman DJ, Cragg GM. 2020. Natural products as sources of new drugs over the
546 nearly four decades from 01/1981 to 09/2019. J Nat Prod. American Chemical Society.
547 3. Newman DJ, Cragg GM. 2007. Natural products as sources of new drugs over the last
548 25 years. J Nat Prod.
549 4. Atanasov AG, Zotchev SB, Dirsch VM, Orhan IE, Banach M, Rollinger JM, Barreca
550 D, Weckwerth W, Bauer R, Bayer EA, Majeed M, Bishayee A, Bochkov V, Bonn GK,
551 Braidy N, Bucar F, Cifuentes A, D'Onofrio G, Bodkin M, Diederich M, Dinkova-
552 Kostova AT, Efferth T, El Bairi K, Arkells N, Fan TP, Fiebich BL, Freissmuth M,
553 Georgiev MI, Gibbons S, Godfrey KM, Gruber CW, Heer J, Huber LA, Ibanez E,
554 Kijjoa A, Kiss AK, Lu A, Macias FA, Miller MJS, Mocan A, Müller R, Nicoletti F,
555 Perry G, Pittalà V, Rastrelli L, Ristow M, Russo GL, Silva AS, Schuster D, Sheridan
556 H, Skalicka-Woźniak K, Skaltsounis L, Sobarzo-Sánchez E, Bredt DS, Stuppner H,
557 Sureda A, Tzvetkov NT, Vacca RA, Aggarwal BB, Battino M, Giampieri F, Wink M,
558 Wolfender JL, Xiao J, Yeung AWK, Lizard G, Popp MA, Heinrich M, Berindan-

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

559 Neagoe I, Stadler M, Daglia M, Verpoorte R, Supuran CT. 2021. Natural products in
560 drug discovery: advances and opportunities. Nat Rev Drug Discov. Nature Research.
561 5. Hawksworth DL, Honegger R. 1994. The lichen thallus: a symbiotic phenotype of
562 nutritionally specialized fungi and its response to gall producers., p. 77–98. *In*
563 Williams, MAJ (ed.), Plant Galls: Organisms, Interactions, Populations. Clarendon
564 Press, Oxford.
565 6. Ahmadjian V. 1993. The Lichen Symbiosis. John Wiley and Sons., New York.
566 7. Ahmadjian V. 1982. Algal/fungal symbioses., p. 179–233. *In* Round, FE Chapman, DJ
567 (ed.), Progress in Phycological Research. Elsevier Biomedical Press, Amsterdam.
568 8. Calchera A, Dal Grande F, Bode HB, Schmitt I. 2019. Biosynthetic gene content of the
569 "perfume lichens" *Evernia prunastri* and *Pseudevernia furfuracea*. Molecules 24:203.
570 9. Goga M, Elečko J, Marcinčinová M, Ručová D, Bačkorová M, Bačkor M. 2020.
571 Lichen metabolites: an overview of some secondary metabolites and their biological
572 potential, p. 175–209. *In* .
573 10. Boustie J, Grube M. 2005. Lichens—a promising source of bioactive secondary
574 metabolites. Plant Genet Resour 3:273–287.
575 11. Kim W, Liu R, Woo S, Kang K Bin, Park H, Yu YH, Ha H-H, Oh S-Y, Yang JH, Kim
576 H, Yun S-H, Hur J-S. 2021. Linking a gene cluster to atranorin, a major cortical
577 substance of lichens, through genetic dereplication and heterologous expression. MBio
578 e0111121.
579 12. Kealey JT, Craig JP, Barr PJ. 2021. Identification of a lichen depside polyketide
580 synthase gene by heterologous expression in *Saccharomyces cerevisiae*. Metab Eng
581 Commun e00172.
582 13. Armaleo D, Sun X, Culberson C. 2011. Insights from the first putative biosynthetic

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

583 gene cluster for a lichen depside and depsidone. Mycologia 103:741–754.

584 14. Singh G, Armaleo D, Dal Grande F, Schmitt I. 2021. Depside and depsidone synthesis
585 in lichenized fungi comes into focus through a genome-wide comparison of the
586 olivetoric acid and physodic acid chemotypes of *Pseudevernia furfuracea*.
587 Biomolecules 11:1445.

588 15. Barbero M, Artuso E, Prandi C. 2017. Fungal anticancer metabolites: synthesis towards
589 drug discovery. Curr Med Chem 25:141–185.

590 16. Stanojković T. 2015. Investigations of lichen secondary metabolites with potential
591 anticancer activity, p. 127–146. *In* Lichen Secondary Metabolites: Bioactive Properties
592 and Pharmaceutical Potential. Springer International Publishing.

593 17. Ingelfinger R, Henke M, Roser L, Ulshöfer T, Calchera A, Singh G, Parnham MJ,
594 Geisslinger G, Fürst R, Schmitt I, Schiffmann S. 2020. Unraveling the pharmacological
595 potential of lichen extracts in the context of cancer and inflammation with a broad
596 screening approach. Front Pharmacol 11:1322.

597 18. Molnár K, Farkas E. 2010. Current results on biological activities of lichen secondary
598 metabolites: A review. Zeitschrift fur Naturforsch - Sect C J Biosci. Verlag der
599 Zeitschrift fur Naturforschung.

600 19. Van Santen JA, Kautsar SA, Medema MH, Linington RG. 2021. Microbial natural
601 product databases: Moving forward in the multi-omics era. Nat Prod Rep. Royal
602 Society of Chemistry.

603 20. Bachmann BO, Van Lanen SG, Baltz RH. 2014. Microbial genome mining for
604 accelerated natural products discovery: Is a renaissance in the making? J Ind Microbiol
605 Biotechnol. J Ind Microbiol Biotechnol.

606 21. Pizarro D, Divakar PK, Grewe F, Crespo A, Dal Grande F, Lumbsch HT. 2020.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

607 Genome-wide analysis of biosynthetic gene cluster reveals correlated gene loss with
608 absence of usnic acid in lichen-forming fungi. Genome Biol Evol 12:1858–1868.
609 22. Abdel-Hameed M, Bertrand RL, Piercey-Normore MD, Sorensen JL. 2016. Putative
610 identification of the usnic acid biosynthetic gene cluster by de novo whole-genome
611 sequencing of a lichen-forming fungus. Fungal Biol 120:306–316.
612 23. Nguyen K-H, Chollet-Krugler M, Gouault N, Tomasi S. 2013. UV-protectant
613 metabolites from lichens and their symbiotic partners. Nat Prod Rep 30:1490.
614 24. Lohezic-Le Devehat F, Legouin B, Couteau C, Boustie J, Coiffard L. 2013. Lichenic
615 extracts and metabolites as UV filters. J Photochem Photobiol B Biol 120:17–28.
616 25. Buçukoglu TZ, Albayrak S, Halici MG, Tay T. 2013. Antimicrobial and antioxidant
617 activities of extracts and lichen acids obtained from some *Umbilicaria* species from
618 central Anatolia, Turkey. J Food Process Preserv 37:1103–1110.
619 26. Posner B, Feige GB, Huneck S. 1992. Studies on the chemistry of the lichen genus
620 *Umbilicaria* hoffm. Zeitschrift fur Naturforsch - Sect C J Biosci 47:1–9.
621 27. Narui T, Culberson CF, Culberson WL, Johnson A, Shibata S. 1996. A contribution to
622 the chemistry of the lichen family umbilicariaceae (Ascomycotina). Bryologist 99:199–
623 211.
624 28. Davydov EA, Peršoh D, Rambold G. 2017. Umbilicariaceae (lichenized Ascomycota)
625 – Trait evolution and a new generic concept. Taxon 66:1282–1303.
626 29. Singh G, Calchera A, Schulz M, Drechsler M, Bode HB, Schmitt I, Dal Grande F.
627 2021. Climate-specific biosynthetic gene clusters in populations of a lichen-forming
628 fungus. Environ Microbiol 00:1462-2920.15605.
629 30. Rubio-Piña JA, Zapata-Pérez O. 2011. Isolation of total RNA from tissues rich in
630 polyphenols and polysaccharides of mangrove plants. Electron J Biotechnol 14.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

631 31. Merges D, Dal Grande F, Greve C, Otte J, Schmitt I. 2021. Virus diversity in
632 metagenomes of a lichen symbiosis (*Umbilicaria phaea*): complete viral genomes,
633 putative hosts and elevational distributions. Environ Microbiol 23:6637–6650.
634 32. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone
635 reads using repeat graphs. Nat Biotechnol 37:540–546.
636 33. Qin M, Wu S, Li A, Zhao F, Feng H, Ding L, Ruan J. 2019. LRScaf: Improving draft
637 genomes using long noisy reads. BMC Genomics 20:955.
638 34. Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, Ruscheweyh H-J, Tappu
639 R. 2016. MEGAN Community Edition - Interactive exploration and analysis of large-
640 scale microbiome sequencing data. PLOS Comput Biol 12:e1004957.
641 35. Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I, Boisvert S,
642 Chapman JA, Chapuis G, Chikhi R, Chitsaz H, Chou W-C, Corbeil J, Del Fabbro C,
643 Docking TR, Durbin R, Earl D, Emrich S, Fedotov P, Fonseca NA, Ganapathy G,
644 Gibbs RA, Gnerre S, Godzaridis É, Goldstein S, Haimel M, Hall G, Haussler D, Hiatt
645 JB, Ho IY, Howard J, Hunt M, Jackman SD, Jaffe DB, Jarvis ED, Jiang H, Kazakov S,
646 Kersey PJ, Kitzman JO, Knight JR, Koren S, Lam T-W, Lavenier D, Laviolette F, Li
647 Y, Li Z, Liu B, Liu Y, Luo R, MacCallum I, MacManes MD, Maillet N, Melnikov S,
648 Naquin D, Ning Z, Otto TD, Paten B, Paulo OS, Phillippy AM, Pina-Martins F, Place
649 M, Przybylski D, Qin X, Qu C, Ribeiro FJ, Richards S, Rokhsar DS, Ruby JG,
650 Scalabrin S, Schatz MC, Schwartz DC, Sergushichev A, Sharpe T, Shaw TI, Shendure
651 J, Shi Y, Simpson JT, Song H, Tsarev F, Vezzi F, Vicedomini R, Vieira BM, Wang J,
652 Worley KC, Yin S, Yiu S-M, Yuan J, Zhang G, Zhang H, Zhou S, Korf IF. 2013.
653 Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate
654 species. Gigascience 2:10.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

655 36. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. 2015.
656 BUSCO: assessing genome assembly and annotation completeness with single-copy
657 orthologs. Bioinformatics 31:3210–3212.
658 37. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T.
659 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline.
660 Nucleic Acids Res 47:W81–W87.
661 38. Palmer J, Stajich J. 2019. Funannotate v1.7.4. Zenodo
662 https://doi.org/https://doi.org/10.5281/ZENODO.2604804.
663 39. Borodovsky M, Lomsadze A. 2011. Eukaryotic gene prediction using GeneMark.hmm-
664 E and GeneMark-ES. Curr Protoc Bioinforma CHAPTER:Unit-4.610.
665 40. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005.
666 InterProScan: Protein domains identifier. Nucleic Acids Res 33:W116.
667 41. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H,
668 Mende DR, Letunic I, Rattei T, Jensen LJ, Von Mering C, Bork P. 2019. EggNOG 5.0:
669 A hierarchical, functionally and phylogenetically annotated orthology resource based
670 on 5090 organisms and 2502 viruses. Nucleic Acids Res 47:D309–D314.
671 42. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Von Mering C,
672 Bork P. 2017. Fast genome-wide functional annotation through orthology assignment
673 by eggNOG-mapper. Mol Biol Evol 34:2115–2122.
674 43. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak
675 S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions
676 using deep neural networks. Nat Biotechnol 37:420–423.
677 44. Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in
678 accuracy of multiple sequence alignment. Nucleic Acids Res 33:511–8.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

679 45. Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J,
680 Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A.
681 2011. Geneious. 5.4.
682 46. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic
683 analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–90.
684 47. Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for
685 inference of large phylogenetic trees, p. 1–8. *In* Proceedings of the Gateway
686 Computing Environments Workshop (GCE)New Orlean.
687 48. Letunic I, Bork P. 2021. Interactive tree of life (iTOL) v5: An online tool for
688 phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296.
689 49. Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH,
690 Parkinson EI, De Los Santos ELC, Yeong M, Cruz-Morales P, Abubucker S, Roeters
691 A, Lokhorst W, Fernandez-Guerra A, Cappelini LTD, Goering AW, Thomson RJ,
692 Metcalf WW, Kelleher NL, Barona-Gomez F, Medema MH. 2020. A computational
693 framework to explore large-scale biosynthetic diversity. Nat Chem Biol 16:60–68.
694 50. Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, Van Der Hooft JJJ,
695 Van Santen JA, Tracanna V, Suarez Duran HG, Pascal Andreu V, Selem-Mojica N,
696 Alanjary M, Robinson SL, Lund G, Epstein SC, Sisto AC, Charkoudian LK, Collemare
697 J, Linington RG, Weber T, Medema MH. 2020. MIBiG 2.0: A repository for
698 biosynthetic gene clusters of known function. Nucleic Acids Res 48:D454–D458.
699 51. Martinet L, Naômé A, Deflandre B, Maciejewska M, Tellatin D, Tenconi E,
700 Smargiasso N, De Pauw E, Van Wezel GP, Rigali S. 2019. A single biosynthetic gene
701 cluster is responsible for the production of bagremycin antibiotics and ferroverdin iron
702 chelators. MBio 10:e01230-19.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

703 52. Armaleo D. 1995. Factors affecting depside and depsidone biosynthesis in a cultured
704 lichen fungus. Crypt Bot 5:14–21.
705 53. Wasil Z, Pahirulzaman KAK, Butts C, Simpson TJ, Lazarus CM, Cox RJ. 2013. One
706 pathway, many compounds: Heterologous expression of a fungal biosynthetic pathway
707 reveals its intrinsic potential for diversity. Chem Sci 4:3845–3856.
708 54. Chen L, Yue Q, Zhang X, Xiang M, Wang C, Li S, Che Y, Ortiz-López FJ, Bills GF,
709 Liu X, An Z. 2013. Genomics-driven discovery of the pneumocandin biosynthetic gene
710 cluster in the fungus *Glarea lozoyensis*. BMC Genomics 14.
711 55. Chen L, Li Y, Yue Q, Loksztejn A, Yokoyama K, Felix EA, Liu X, Zhang N, An Z,
712 Bills GF. 2016. Engineering of new pneumocandin side-chain analogues from *Glarea*
713 *lozoyensis* by mutasynthesis and evaluation of their antifungal activity. ACS Chem
714 Biol 11:2724–2733.
715 56. Li Y, Chen L, Yue Q, Liu X, An Z, Bills GF. 2015. Genetic manipulation of the
716 pneumocandin biosynthetic pathway for generation of analogues and evaluation of
717 their antifungal activity. ACS Chem Biol 10:1702–1710.
718 57. Leuckert C. 1985. Probleme der Flechten-Chemotaxonomie -- Stoffkombinationen und
719 ihre taxonomische Wertung. Problems of lichen chemotaxonomy -- patterns of
720 substances and their taxonomical valuation. Ber Dtsch Bot Ges 98:401–408.
721 58. Davydov EA, Blum OB, Kashevarov GP, Grakhov VP. 2019. Umbilicaria
722 subpolyphylla Oxner: The correct name for U. iberica Sancho & Krzewicka and its
723 bipolar distribution pattern. Lichenologist. Cambridge University Press.
724
725 2. Newman DJ, Cragg GM. 2020. Natural products as sources of new drugs over the
726 nearly four decades from 01/1981 to 09/2019. J Nat Prod. American Chemical Society.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

727 3. Newman DJ, Cragg GM. 2007. Natural products as sources of new drugs over the last
728 25 years. J Nat Prod.
729 4. Atanasov AG, Zotchev SB, Dirsch VM, Orhan IE, Banach M, Rollinger JM, Barreca
730 D, Weckwerth W, Bauer R, Bayer EA, Majeed M, Bishayee A, Bochkov V, Bonn GK,
731 Braidy N, Bucar F, Cifuentes A, D'Onofrio G, Bodkin M, Diederich M, Dinkova-
732 Kostova AT, Efferth T, El Bairi K, Arkells N, Fan TP, Fiebich BL, Freissmuth M,
733 Georgiev MI, Gibbons S, Godfrey KM, Gruber CW, Heer J, Huber LA, Ibanez E,
734 Kijjoa A, Kiss AK, Lu A, Macias FA, Miller MJS, Mocan A, Müller R, Nicoletti F,
735 Perry G, Pittalà V, Rastrelli L, Ristow M, Russo GL, Silva AS, Schuster D, Sheridan
736 H, Skalicka-Woźniak K, Skaltsounis L, Sobarzo-Sánchez E, Bredt DS, Stuppner H,
737 Sureda A, Tzvetkov NT, Vacca RA, Aggarwal BB, Battino M, Giampieri F, Wink M,
738 Wolfender JL, Xiao J, Yeung AWK, Lizard G, Popp MA, Heinrich M, Berindan-
739 Neagoe I, Stadler M, Daglia M, Verpoorte R, Supuran CT. 2021. Natural products in
740 drug discovery: advances and opportunities. Nat Rev Drug Discov. Nature Research.
741 5. Hawksworth DL, Honegger R. 1994. The lichen thallus: a symbiotic phenotype of
742 nutritionally specialized fungi and its response to gall producers., p. 77–98. *In*
743 Williams, MAJ (ed.), Plant Galls: Organisms, Interactions, Populations. Clarendon
744 Press, Oxford.
745 6. Ahmadjian V. 1993. The Lichen Symbiosis. John Wiley and Sons., New York.
746 7. Ahmadjian V. 1982. Algal/fungal symbioses., p. 179–233. *In* Round, FE Chapman, DJ
747 (ed.), Progress in Phycological Research. Elsevier Biomedical Press, Amsterdam.
748 8. Calchera A, Dal Grande F, Bode HB, Schmitt I. 2019. Biosynthetic gene content of the
749 "perfume lichens" *Evernia prunastri* and *Pseudevernia furfuracea*. Molecules 24:203.
750 9. Goga M, Elečko J, Marcinčinová M, Ručová D, Bačkorová M, Bačkor M. 2020.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

751 Lichen metabolites: an overview of some secondary metabolites and their biological
752 potential, p. 175–209. *In* .
753 10. Boustie J, Grube M. 2005. Lichens—a promising source of bioactive secondary
754 metabolites. Plant Genet Resour 3:273–287.
755 11. Kim W, Liu R, Woo S, Kang K Bin, Park H, Yu YH, Ha H-H, Oh S-Y, Yang JH, Kim
756 H, Yun S-H, Hur J-S. 2021. Linking a gene cluster to atranorin, a major cortical
757 substance of lichens, through genetic dereplication and heterologous expression. MBio
758 e0111121.
759 12. Kealey JT, Craig JP, Barr PJ. 2021. Identification of a lichen depside polyketide
760 synthase gene by heterologous expression in *Saccharomyces cerevisiae*. Metab Eng
761 Commun e00172.
762 13. Armaleo D, Sun X, Culberson C. 2011. Insights from the first putative biosynthetic
763 gene cluster for a lichen depside and depsidone. Mycologia 103:741–754.
764 14. Singh G, Armaleo D, Dal Grande F, Schmitt I. 2021. Depside and depsidone synthesis
765 in lichenized fungi comes into focus through a genome-wide comparison of the
766 olivetoric acid and physodic acid chemotypes of *Pseudevernia furfuracea*.
767 Biomolecules 11:1445.
768 15. Barbero M, Artuso E, Prandi C. 2017. Fungal anticancer metabolites: synthesis towards
769 drug discovery. Curr Med Chem 25:141–185.
770 16. Stanojković T. 2015. Investigations of lichen secondary metabolites with potential
771 anticancer activity, p. 127–146. *In* Lichen Secondary Metabolites: Bioactive Properties
772 and Pharmaceutical Potential. Springer International Publishing.
773 17. Ingelfinger R, Henke M, Roser L, Ulshöfer T, Calchera A, Singh G, Parnham MJ,
774 Geisslinger G, Fürst R, Schmitt I, Schiffmann S. 2020. Unraveling the pharmacological

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

775 potential of lichen extracts in the context of cancer and inflammation with a broad
776 screening approach. Front Pharmacol 11:1322.
777 18. Molnár K, Farkas E. 2010. Current results on biological activities of lichen secondary
778 metabolites: A review. Zeitschrift fur Naturforsch - Sect C J Biosci. Verlag der
779 Zeitschrift fur Naturforschung.
780 19. Van Santen JA, Kautsar SA, Medema MH, Linington RG. 2021. Microbial natural
781 product databases: Moving forward in the multi-omics era. Nat Prod Rep. Royal
782 Society of Chemistry.
783 20. Bachmann BO, Van Lanen SG, Baltz RH. 2014. Microbial genome mining for
784 accelerated natural products discovery: Is a renaissance in the making? J Ind Microbiol
785 Biotechnol. J Ind Microbiol Biotechnol.
786 21. Pizarro D, Divakar PK, Grewe F, Crespo A, Dal Grande F, Lumbsch HT. 2020.
787 Genome-wide analysis of biosynthetic gene cluster reveals correlated gene loss with
788 absence of usnic acid in lichen-forming fungi. Genome Biol Evol 12:1858–1868.
789 22. Abdel-Hameed M, Bertrand RL, Piercey-Normore MD, Sorensen JL. 2016. Putative
790 identification of the usnic acid biosynthetic gene cluster by de novo whole-genome
791 sequencing of a lichen-forming fungus. Fungal Biol 120:306–316.
792 23. Nguyen K-H, Chollet-Krugler M, Gouault N, Tomasi S. 2013. UV-protectant
793 metabolites from lichens and their symbiotic partners. Nat Prod Rep 30:1490.
794 24. Lohezic-Le Devehat F, Legouin B, Couteau C, Boustie J, Coiffard L. 2013. Lichenic
795 extracts and metabolites as UV filters. J Photochem Photobiol B Biol 120:17–28.
796 25. Buçukoglu TZ, Albayrak S, Halici MG, Tay T. 2013. Antimicrobial and antioxidant
797 activities of extracts and lichen acids obtained from some *Umbilicaria* species from
798 central Anatolia, Turkey. J Food Process Preserv 37:1103–1110.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

799 26. Posner B, Feige GB, Huneck S. 1992. Studies on the chemistry of the lichen genus
800 *Umbilicaria* hoffm. Zeitschrift fur Naturforsch - Sect C J Biosci 47:1–9.
801 27. Narui T, Culberson CF, Culberson WL, Johnson A, Shibata S. 1996. A contribution to
802 the chemistry of the lichen family umbilicariaceae (Ascomycotina). Bryologist 99:199–
803 211.
804 28. Davydov EA, Peršoh D, Rambold G. 2017. Umbilicariaceae (lichenized Ascomycota)
805 – Trait evolution and a new generic concept. Taxon 66:1282–1303.
806 29. Singh G, Calchera A, Schulz M, Drechsler M, Bode HB, Schmitt I, Dal Grande F.
807 2021. Climate-specific biosynthetic gene clusters in populations of a lichen-forming
808 fungus. Environ Microbiol 00:1462-2920.15605.
809 30. Rubio-Piña JA, Zapata-Pérez O. 2011. Isolation of total RNA from tissues rich in
810 polyphenols and polysaccharides of mangrove plants. Electron J Biotechnol 14.
811 31. Merges D, Dal Grande F, Greve C, Otte J, Schmitt I. 2021. Virus diversity in
812 metagenomes of a lichen symbiosis (*Umbilicaria phaea*): complete viral genomes,
813 putative hosts and elevational distributions. Environ Microbiol 23:6637–6650.
814 32. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone
815 reads using repeat graphs. Nat Biotechnol 37:540–546.
816 33. Qin M, Wu S, Li A, Zhao F, Feng H, Ding L, Ruan J. 2019. LRScaf: Improving draft
817 genomes using long noisy reads. BMC Genomics 20:955.
818 34. Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, Ruscheweyh H-J, Tappu
819 R. 2016. MEGAN Community Edition - Interactive exploration and analysis of large-
820 scale microbiome sequencing data. PLOS Comput Biol 12:e1004957.
821 35. Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I, Boisvert S,
822 Chapman JA, Chapuis G, Chikhi R, Chitsaz H, Chou W-C, Corbeil J, Del Fabbro C,

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

823 Docking TR, Durbin R, Earl D, Emrich S, Fedotov P, Fonseca NA, Ganapathy G,
824 Gibbs RA, Gnerre S, Godzaridis É, Goldstein S, Haimel M, Hall G, Haussler D, Hiatt
825 JB, Ho IY, Howard J, Hunt M, Jackman SD, Jaffe DB, Jarvis ED, Jiang H, Kazakov S,
826 Kersey PJ, Kitzman JO, Knight JR, Koren S, Lam T-W, Lavenier D, Laviolette F, Li
827 Y, Li Z, Liu B, Liu Y, Luo R, MacCallum I, MacManes MD, Maillet N, Melnikov S,
828 Naquin D, Ning Z, Otto TD, Paten B, Paulo OS, Phillippy AM, Pina-Martins F, Place
829 M, Przybylski D, Qin X, Qu C, Ribeiro FJ, Richards S, Rokhsar DS, Ruby JG,
830 Scalabrin S, Schatz MC, Schwartz DC, Sergushichev A, Sharpe T, Shaw TI, Shendure
831 J, Shi Y, Simpson JT, Song H, Tsarev F, Vezzi F, Vicedomini R, Vieira BM, Wang J,
832 Worley KC, Yin S, Yiu S-M, Yuan J, Zhang G, Zhang H, Zhou S, Korf IF. 2013.
833 Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate
834 species. Gigascience 2:10.

835 36. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. 2015.
836 BUSCO: assessing genome assembly and annotation completeness with single-copy
837 orthologs. Bioinformatics 31:3210–3212.

838 37. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T.
839 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline.
840 Nucleic Acids Res 47:W81–W87.

841 38. Palmer J, Stajich J. 2019. Funannotate v1.7.4. Zenodo
842 https://doi.org/https://doi.org/10.5281/ZENODO.2604804.

843 39. Borodovsky M, Lomsadze A. 2011. Eukaryotic gene prediction using GeneMark.hmm-
844 E and GeneMark-ES. Curr Protoc Bioinforma CHAPTER:Unit-4.610.

845 40. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005.
846 InterProScan: Protein domains identifier. Nucleic Acids Res 33:W116.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

847 41. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H,
848 Mende DR, Letunic I, Rattei T, Jensen LJ, Von Mering C, Bork P. 2019. EggNOG 5.0:
849 A hierarchical, functionally and phylogenetically annotated orthology resource based
850 on 5090 organisms and 2502 viruses. Nucleic Acids Res 47:D309–D314.
851 42. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Von Mering C,
852 Bork P. 2017. Fast genome-wide functional annotation through orthology assignment
853 by eggNOG-mapper. Mol Biol Evol 34:2115–2122.
854 43. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak
855 S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions
856 using deep neural networks. Nat Biotechnol 37:420–423.
857 44. Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in
858 accuracy of multiple sequence alignment. Nucleic Acids Res 33:511–8.
859 45. Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J,
860 Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A.
861 2011. Geneious. 5.4.
862 46. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic
863 analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–90.
864 47. Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for
865 inference of large phylogenetic trees, p. 1–8. *In* Proceedings of the Gateway
866 Computing Environments Workshop (GCE)New Orlean.
867 48. Letunic I, Bork P. 2021. Interactive tree of life (iTOL) v5: An online tool for
868 phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296.
869 49. Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH,
870 Parkinson EI, De Los Santos ELC, Yeong M, Cruz-Morales P, Abubucker S, Roeters

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

871 A, Lokhorst W, Fernandez-Guerra A, Cappelini LTD, Goering AW, Thomson RJ,
872 Metcalf WW, Kelleher NL, Barona-Gomez F, Medema MH. 2020. A computational
873 framework to explore large-scale biosynthetic diversity. Nat Chem Biol 16:60–68.

874 50. Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, Van Der Hooft JJJ,
875 Van Santen JA, Tracanna V, Suarez Duran HG, Pascal Andreu V, Selem-Mojica N,
876 Alanjary M, Robinson SL, Lund G, Epstein SC, Sisto AC, Charkoudian LK, Collemare
877 J, Linington RG, Weber T, Medema MH. 2020. MIBiG 2.0: A repository for
878 biosynthetic gene clusters of known function. Nucleic Acids Res 48:D454–D458.

879 51. Martinet L, Naômé A, Deflandre B, Maciejewska M, Tellatin D, Tenconi E,
880 Smargiasso N, De Pauw E, Van Wezel GP, Rigali S. 2019. A single biosynthetic gene
881 cluster is responsible for the production of bagremycin antibiotics and ferroverdin iron
882 chelators. MBio 10:e01230-19.

883 52. Armaleo D. 1995. Factors affecting depside and depsidone biosynthesis in a cultured
884 lichen fungus. Crypt Bot 5:14–21.

885 53. Wasil Z, Pahirulzaman KAK, Butts C, Simpson TJ, Lazarus CM, Cox RJ. 2013. One
886 pathway, many compounds: Heterologous expression of a fungal biosynthetic pathway
887 reveals its intrinsic potential for diversity. Chem Sci 4:3845–3856.

888 54. Chen L, Yue Q, Zhang X, Xiang M, Wang C, Li S, Che Y, Ortiz-López FJ, Bills GF,
889 Liu X, An Z. 2013. Genomics-driven discovery of the pneumocandin biosynthetic gene
890 cluster in the fungus *Glarea lozoyensis*. BMC Genomics 14.

891 55. Chen L, Li Y, Yue Q, Loksztejn A, Yokoyama K, Felix EA, Liu X, Zhang N, An Z,
892 Bills GF. 2016. Engineering of new pneumocandin side-chain analogues from *Glarea*
893 *lozoyensis* by mutasynthesis and evaluation of their antifungal activity. ACS Chem
894 Biol 11:2724–2733.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

895 56. Li Y, Chen L, Yue Q, Liu X, An Z, Bills GF. 2015. Genetic manipulation of the
896 pneumocandin biosynthetic pathway for generation of analogues and evaluation of
897 their antifungal activity. ACS Chem Biol 10:1702–1710.
898 57. Leuckert C. 1985. Probleme der Flechten-Chemotaxonomie -- Stoffkombinationen und
899 ihre taxonomische Wertung. Problems of lichen chemotaxonomy -- patterns of
900 substances and their taxonomical valuation. Ber Dtsch Bot Ges 98:401–408.
901

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

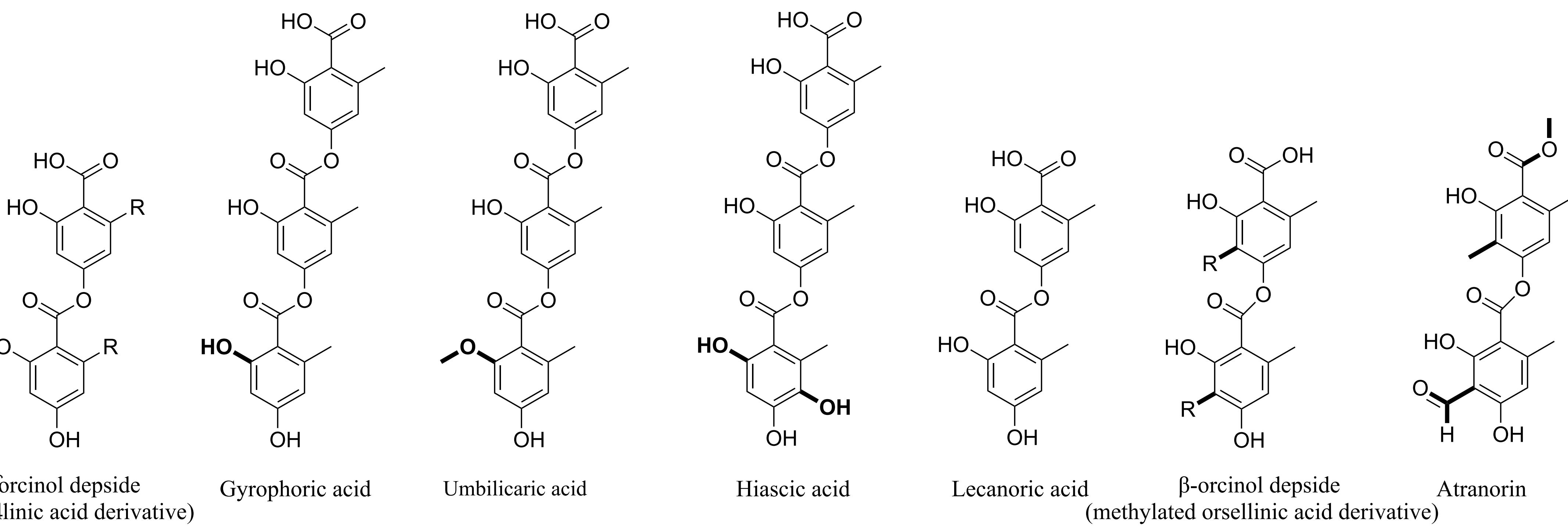
Table 1 Genome quality and annotation statistics

Species	TBG number	ccs HiFi yield	# scaffolds	N50	Completeness	assembly size (Mb)	# Genes	# Proteins
Umbilicaria freyi	TBG_2329	47.39%	113	2575640	95.7	53.4	10,156	10,065
Umbilicaria freyi	TBG_2330	46.41%	54	2043163	85.9	50	8,848	8,773
Umbilicaria deusta	TBG_2334	47.86%	47	1669916	97.6	41.6	8,949	8,857
Umbilicaria deusta	TBG_2335	43.54%	42	1865302	90.2	37.4	8,194	8,049
Umbilicaria hispanica	TBG_2322	38.71%	130	3125324	96.8	43.4	9,111	9,021
Umbilicaria hispanica	TBG_2337	54.22%	60	4226768	97.3	41.9	8,781	8,696
Umbilicaria pustulata	TBG_2333	33%	26	2620629	97.3	37.7	9,569	9,503
Umbilicaria pustulata	TBG_2345	32.26%	31	2364512	96.8	35.7	8,790	8,740
Umbilicaria spodochroa	TBG_2434	34.20%	139	993216	97.0	40.9	8,791	8,705
Umbilicaria spodochroa	TBG_2435	40.93%	97	1249424	97.1	40.1	8,612	8,507
Umbilicaria subpolyphylla	TBG_2323	41.14%	190	1544375	99.6	58.2	16,993	16,852
Umbilicaria subpolyphylla	TBG_2324	33.68%	42	1514392	97.6	33.7	8,556	8,410
Umbilicaria grisea	TBG_2336	42.54%	40	1822796	96.9	44.43	NA	NA
Dermatocarpon miniatum	TBG_2326	29.36%	28	5077191	98.1	63.5	9,273	9,189
Dermatocarpon miniatum	TBG_2331	26.28%	22	4245366	98.4	49.8	7,938	7,871

902

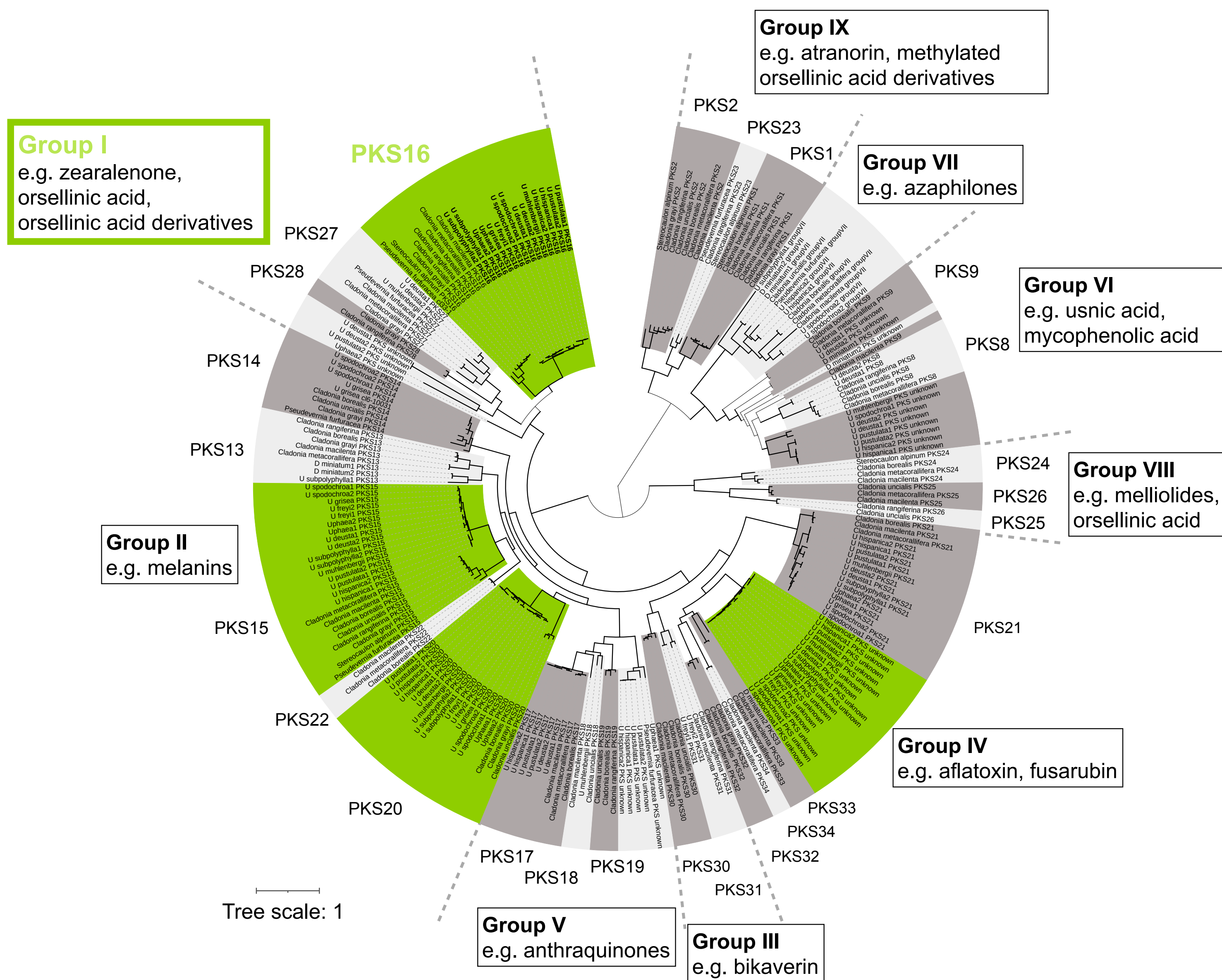
bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 1



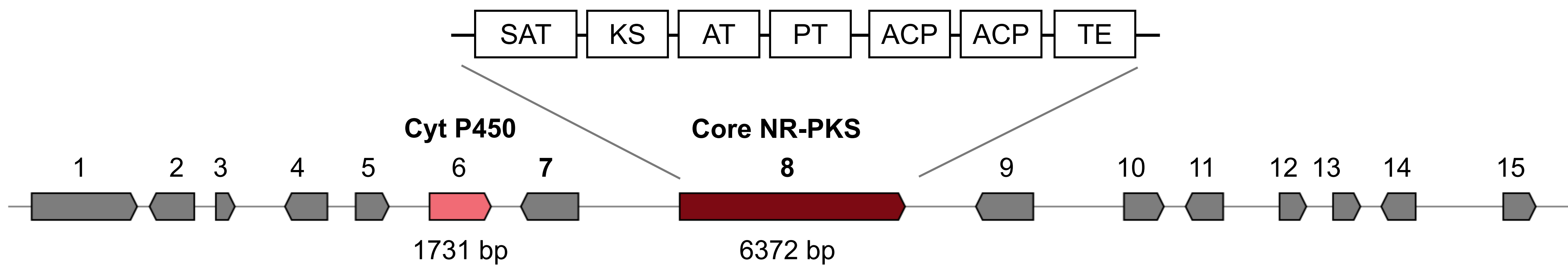
bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 2



bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 3



bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.475839; this version posted January 16, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 4