

**The effects of co-culture on the expression of selected PKS genes in the lichenized fungus
*Xanthoparmelia taractica***

Hongli Si^{1#}, Yuemin Su^{1#}, Yichen Wang¹, Tanay Bose^{2,3*}, Runlei Chang^{1*}

1 College of Life Sciences, Shandong Normal University, Jinan 250014, China

2 Department of Biochemistry, Genetics and Microbiology, Faculty of Natural and Agricultural Sciences, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, South Africa

3 Department of Zoology and Entomology, Faculty of Natural and Agricultural Sciences, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, South Africa

Co-first authors

* Corresponding authors – Tanay.Bose@fabi.up.ac.za, changrunlei@163.com

ORCID IDs

H Si - 0000-0003-0585-3478

Y Su - 0000-0002-0849-4055

Y Wang - 0000-0001-6904-0594

T Bose - 0000-0002-2069-042X

R Chang - 0000-0001-6613-9062

Abstract

Lichens produce diverse secondary metabolites. A diversity of these compounds is synthesized by fungal polyketide synthases (PKSs). In this study, we catalogued the PKS genes from *Xanthoparmelia taractica*, a lichen with global distribution. To accomplish this, we isolated the symbionts to sequence the whole genome of the mycobiont and established an *in vitro* co-culture system for this lichen. We also added an endolichenic fungus, *Coniochaeta fibrosae*, to this co-culture to evaluate its effect on lichen symbiosis. The genome of the mycobiont *X. taractica* was around 43.1 Mb with 10,730 ORFs. Twenty-eight PKS genes were identified in the genome. These included 27 Type I and one Type III gene. Except for three PKS genes, XTPKS12, XTPKS18, and XTPKS22, the function of the majority of PKS genes remained unknown. We selected these genes for the expression analyses using a co-culture system. The co-culture system that included the mycobiont and the photobiont showed an early stage of lichenization because the fungi produced a hyphal network connecting and penetrating the algal cells. Also, XTPKS12 was down-regulated and XTPKS18 and XTPKS22 were modestly up-regulated. As predicted, *C. fibrosae* did not participate in the symbiosis. This study reconfirms that Type I is the most dominant PKS gene in lichenized fungi and the function of these genes might be influenced by symbiosis.

Keywords *Coniochaeta fibrosae*, expression analysis, functional annotation, genome sequencing, scanning electron microscopy

Introduction

Lichens are a symbiotic partnership between a mycobiont (fungal partner) and a photobiont (photoautotrophic partner). Lichens produce a plethora of secondary metabolites, which allows them to thrive in hostile environments (Goga et al. 2020). Various studies have predicted that these compounds assist lichens with habitat expansion and resilience to various biotic and abiotic stresses (Mitrović et al. 2011; Gaya et al. 2015; Solhaug and Gauslaa 2012). These secondary metabolites are also known to have antibacterial, antiproliferative, antioxidant, antiviral, anti-inflammatory, antiherbivore allelopathic, and photoprotective effects (Goga et al. 2020). Hence, lichens are widely used in ethnomedicine for the treatment of various diseases (Crawford 2015; Yang et al. 2021). A vast majority of these secondary metabolites are small aromatic polyketides synthesized by fungal polyketide synthases (PKSs) (Kealey et al. 2021; Kim et al. 2021b; Wang et al. 2018; Armaleo et al. 2011; Yousuf et al. 2014; Schmitt et al. 2008).

PKSs have been reported from a diverse range of bacteria, fungi, and plants (Austin and Noel 2003). PKSs are multidomain or multiprotein enzyme complexes that produce a wide range of natural compounds. These enzymes share structural and functional similarities with fatty acid synthase (Korman et al. 2010). PKSs are classified into four types based on their mechanisms and structure, which include enzyme catalysis, molecular recognition, protein-protein interaction, and domain organization. These are modular Type I, iterative Type I, Type II, and Type III (Shen 2003). Among these, Type I and II PKSs are large protein complexes, while Type III is a small homodimeric protein (Weng and Noel 2012). Concurrently, based on the presence of optional domains, fungal PKSs can be classified into three subgroups. These are including non-reducing (NR), partially reducing (PR), and reducing (R) PKSs (Gerasimova et al. 2022). Recent comparative genomic research of 29 lichenized fungi revealed that the number of PKS genes varies greatly between the groups of lichenized fungi. *Parmelia* sp., for example, has 70 Type I PKS genes, whereas only three genes were found in *Sclerophora sanguinea* (Kim et al. 2021b). Similarly, Gerasimova et al. (2022) examined 23 Ascomycota lichen-forming fungus genomes and observed a significant diversity of Type I PKS genes in *Bacidia rubella*. Despite the difficulty of linking PKS genes to their products in lichen-forming fungi, which is often due to a lack of established gene knockout or heterologous gene expression systems, some recent studies have been successful in achieving this, such as those by Okorafor et al. (2021), Kim et al. (2021a), Kim et al. (2021b), Kealey et al. (2021), Singh et al. (2022a).

The lack of well-established gene knockout or heterologous gene expression techniques in lichens has made it difficult to conclusively link a PKS gene to its corresponding metabolic product (Kealey et al. 2021). Even though over 1000 lichen metabolites have been reported,

only around ten have been effectively linked to the genes that encode them (Singh 2023). Currently, several PKS genes in lichen have been characterized using genome mining approach. For example, *CgrPKS16* from *Cladonia grayi* is engaged in the grayanic acid depsidone biosynthesis pathway (Armaleo et al. 2011), *NPPKS7* from *Nephromopsis pallescens* in usnic acid biosynthesis (Wang et al. 2018), and PKS *PFUR17_02294* and *atr1* from *Pseudevernia furfuracea* in lecanoric acid and atranorin production, respectively (Kim et al. 2021b; Kealey et al. 2021).

The presence of PKS genes in the genome of a lichenized fungus does not necessarily correspond to the production of secondary metabolites. A recent study identified 14 PKS genes in the genome of *Endocarpon pusillum*. However, no secondary metabolites were detected under conventional culture conditions. When the fungus was propagated in a vermiculite medium, however, three secondary metabolites were produced (Liu et al. 2018). Similarly, Jeong et al. (2021) demonstrated that *Cladonia metacorallifera* var. *reagens* requires fructose concentration in medium and light to produce cristazarin, whereas pH is important for *Cladonia uncialis* to produce usnic acid (Gunawardana et al. 2021). Certain secondary metabolites are only produced by the fungus when it is in a symbiotic relationship with an alga in a lichenized condition, such as hyperhomosekikaic acid (Yousuf et al. 2014). As a result, developing a co-culture system is critical for identifying some of the secondary metabolites produced by a lichenized fungus (Li et al. 2020).

The lichen *Xanthoparmelia taractica* has been reported to grow in various regions of China, Europe, and South and North America. However, no co-culture system for *X. taractica* and its algal partners has been established, as well as no genomic data for the lichenized fungal species is available. To address this, we established co-culture systems with the lichenized fungus (*X. taractica*), the endolichenic fungus (*Coniochaeta fibrosae*) and its algal partner (*Trebouxia* sp.), or either of the fungus and the alga. We also sequenced the whole genome of *X. taractica*, identified the PKS genes, and investigated the expression patterns of selected PKS in both co-culture and monoculture setups. We hypothesized: (1) the early phase of symbiosis will be observable in co-culture systems combining the lichenized fungus *X. taractica* and the alga, but not in the setup comprising endolichenic fungus *C. fibrosae* and the alga. This is because the former is known to form symbiotic partnerships with alga, while the latter is a non-lichenized fungal species that dwell within the lichen host, comparable to "endophytes"; and (2) when co-cultured with the alga, the expression pattern of selected PKS genes of the lichenized fungus *X. taractica* will vary compared to the pure culture of the fungus.

Material and method

Collection of lichen sample

In July 2019, samples of *Xanthoparmelia taractica* were collected from Balin Right Banner (14 98.8 m a.s.l., 44°13'45"N, 118°44'57"E), Chifeng City, Inner Mongolia Autonomous Region of China.

Isolation of fungi and alga from lichen thalli

To eliminate organic debris, lichen thalli were washed under running tap water. The thalli were then rinsed several times with sterile deionized water. Under a Leica Zoom 2000 stereomicroscope, the top cortical layer was scraped off with a sterile blade. After dissecting the medullary layer, tissues were rinsed with sterile deionized water.

For fungal isolation, pieces of medullary tissues were placed on the surface of 2 % potato dextrose agar (PDA; Qingdao Hope Bio-Technology Co., Ltd., China). All Petri plates were incubated at 25 °C for 14 days. Mycelia emerging from tissue pieces were sub-cultured onto new PDA plates. The single hyphal tip isolation technique was used to establish pure cultures of fungal isolates. All the plates were incubated in darkness at 25 °C.

To isolate the alga, pieces of medullary tissues recovered as previously indicated were plated onto fresh Bold's Basal Medium (BBM) (Deason 1960). Thereafter, all Petri plates were incubated at 25 °C under a light/dark cycle (12/12 h) for 14 days. Algal colonies were sub-cultured on fresh BBM plates using the streak culture method to establish pure cultures. All the algal cultures were incubated at 25 °C under a light/dark cycle (12/12 h).

Identification of lichenized fungi and algae

Total genomic DNA from fungal and algal isolates was extracted using a modified CTAB method (Zhang et al. 2010). For molecular identification of the fungal and algal isolates, the complete internal transcribed spacer (ITS) was amplified using primer pair ITS1/ITS4 (White et al. 1990).

Each 25 µl PCR reaction included 12.5 µl of 2 × Taq Master Mix (buffer, dNTPs and Taq; Vazyme Biotech Co., Ltd, China), 9.5 µl of PCR grade H₂O, 1 µl each of forward and reverse primers, and 1 µl of DNA extract. PCR amplifications were conducted with an initial denaturation at 95 °C for 5 minutes, followed by 30 cycles of 95 °C for 30 s, annealing temperatures 54 °C for 60 s, 72 °C for 90 s, followed by a final elongation at 72 °C for 10 minutes. Positive amplifications were confirmed using agarose gel electrophoresis. All the PCR products were sequenced by Sangon Bioengineering (Shanghai) Co., Ltd. The resulting sequences were assembled using Geneious v. 10.2.2 (<https://www.geneious.com>).

Preliminary identification of fungal and algal isolates was done by using the BLAST algorithm (Altschul et al. 1990) available through the NCBI GenBank. Thereafter, sequences

of closely related fungal and algal species were retrieved from the GenBank using the distance tree feature (**Table S1-2**). All datasets were aligned using MAFFT v. 7 (Katoh and Standley 2013). If needed, manually adjusted using MEGA v. 7 (Kumar et al. 2016).

Phylogenetic analyses were done using software available through the CIPRES Science Gateway (Miller et al. 2010). Appropriate evolutionary models were selected using jModelTest2 (Darriba et al. 2012). Maximum likelihood (ML) phylogenetic analyses of the datasets were done using RAxML v. 8.2.2 (Stamatakis 2014). These analyses were carried out using the default settings and bootstrap values derived from 1,000 replicated searches. Phylogenetic trees were visualized using FigTree v. 1.4.3 (<http://tree.bio.ed.ac.uk/>).

Co-culturing of fungi and alga

Two fungal isolates and one algal isolate were co-cultured in this study. Isolates of lichenized fungi (423E) and alga (S420) were recovered in this study. Henceforth, these isolates will now be referred to as LF-423E and AGL-S420. The isolate of endolichenic fungus, *Coniochaeta fibrosae* (CGMCC 3.20304), was recovered in a previous study (Si et al. 2021) but from the same lichen thallus.

The fungal and algal cultures for this co-culturing experiment were prepared following the protocol suggested by Joneson et al. (2011). Fungal and algal isolates were sub-cultured separately in 100 ml each of malt yeast broth (MY, 20 g malt extract and 2 g yeast extract/L) and BBM broth, respectively. The flasks were placed in an incubator shaker at 120 rpm, 20 °C under a light/dark cycle (12/12 h) for four weeks. After the incubation period, 50 ml of each fungal and algal culture were pipetted into three polyethene tubes and centrifuged at 12 000 rpm for 10 min. The algal pellets were re-suspended into a 50 ml blend of BBM and MY broths in a 99:1 ratio. The fungal pellets were macerated using a tissue grinder before being resuspended in the 50 ml blend of media. In suspensions containing more than one organism, the mixes were made in ratios of 1:1 or 1:1:1.

The co-culturing trial included six set-ups. These are: 1) LF-423E only; 2) LF-423E + AGL-S420; 3) AGL-S420 only; 4) CGMCC3.20304 only; 5) CGMCC3.20304 + AGL-S420; and 6) LF-423E + AGL-S420 + CGMCC3.20304.

Sterile nitrocellulose membranes were placed on the surface of BBM and MY (99:1) agar plates. 500 µl of each of the above-mentioned suspensions were aspirated onto the surface of these nitrocellulose membranes and air-dried inside a laminar flow for 1 h. Each set-up was replicated thrice. All of the Petri plates were incubated at 20 °C for 21 days with a 12/12 h light/dark cycle.

Scanning electron microscopy

The interaction between fungal and algal isolates was observed using a scanning electron microscope. The nitrocellulose membrane was harvested and sliced into 1-2 cm strips from representative plates from set-up 2, 5 and 6 of the co-culture trial. Nitrocellulose strips with the organism were fixed using Karnovsky's fixative (2.5% glutaraldehyde and 1.0% paraformaldehyde in 0.1 M cacodylate buffer, pH 7.2) at 4 °C for 24 h. Fixed nitrocellulose strips were rinsed with 0.1 M cacodylate buffer (pH 7.2) followed by post-fixing with 1% OsO₄ dissolved in 0.1 M cacodylate for 2 hours at 4 °C. Post-fixed samples were gradually dehydrated with 70, 80, 90, and 100% ethanol. Thereafter, all the dehydrated samples were steeped in absolute ethanol for 30 min, and this step was repeated thrice. Samples were further dehydrated using absolute acetone for 20 min twice. All samples were then dried with the Leica EM CPD300 critical point dryer. The morphological characteristics of interactions between fungi and algae were observed using a Hitachi TM-3030 scanning electron microscope.

Genome sequencing and assembly of *Xanthoparmelia taractica*

DNA for the whole genome sequencing of *X. taractica* was extracted using E.Z.N.A.® HP Fungal DNA Kit (Omega Biotek). Library preparation and sequencing were done by Shanghai Personalbio Technology Co., Ltd. (China). Construction of the library was done using a TruSeq™ DNA Sample Prep Kit (Illumina, USA). Runs were conducted with 2 × 150 base paired-end reads with insert size 400 using the Illumina NovaSeq System. The quality of the raw sequence data was checked using FastQC v. 0.11.9 (Andrews 2010). Adapter sequences were trimmed using AdapterRemoval v. 2.0 (Schubert et al. 2016). SOAPec v. 2.0 was used for quality correction of all reads based on k-mer frequency, with the k-mer value set to 17 (Luo et al. 2012). High-quality reads from each sample were used for k-mer counting using Jellyfish v. 2.0.0 (Marçais and Kingsford 2011). The genome was assembled by a hybrid approach using the software Falcon (Chin et al. 2016) and CANU (Koren et al. 2017). The draft genome assembly was polished using Pilon V.1.18 (Walker et al. 2014). The resulting genome assembly was assessed using BUSCO v.3.0.2 (Simão et al. 2015).

The repeat sequence content of the genome was detected using RepeatMasker v. 4.0.5 (Tempel 2012) and the Repbase collection of repeats for fungus v. 20150807 (Kapitonov and Jurka 2008). RECON v. 1.0.8 (Bao and Eddy 2002) and RepeatScout v. 1.0.5 (Price et al. 2005) available through RepeatModeler v. 1.0.4 (Smit et al. 2015) were used to identify *de novo*-based repeat sequences. The transfer RNA (tRNA) and ribosomal RNA (rRNA) genes were identified using tRNAscan-SE v. 1.3.1 (Lowe and Eddy 1997) and RNAmmer v. 1.2 (Lagesen et al. 2007), respectively. The remaining non-coding RNA sequences (ncRNAs) were identified using the Rfam database for ncRNAs (Griffiths-Jones 2005).

Gene prediction of *Xanthoparmelia taractica*

The *ab initio* gene predictions were performed using the programs Augustus v. 3.03, glimmerHMM v. 3.0.1, and GeneMark-ES v. 4.35 (Stanke and Waack 2003; Majoros et al. 2004; Ter-Hovhannisyanyan et al. 2008). Homology-based gene prediction was done using Exonerate v2.2.0 (Slater and Birney 2005). The *ab initio* gene predictions and exonerate alignments were then combined using EvidenceModeller v1.1.1 (Haas et al. 2008) with a minimum intron length of 10 bp and weightings of CodingQuarry:1, GeneMark.hmm:1, protein exonerate:2.

Annotations of carbohydrate-active enzymes, antibiotic-related genes, and fungal virulence factors of *Xanthoparmelia taractica*

Carbohydrate-active enzymes (CAZymes) annotation was done using hmmscan v. 3.1b2 against CAZy database (Lombard et al. 2014). If the length of the alignment was greater than 80 then a stringent e-value cut-off was used $1e-5$, with coverage of 30%. However, if the length of the HMM alignment was less than 80 then the e-value was reduced to $1e-3$, with coverage of 30%. Antibiotic-related genes were identified using BLAST against the comprehensive antibiotic resistance database (McArthur et al. 2013) with an e-value cutoff of $1e-6$, identity > 45%, and coverage > 70%. Fungal virulence factors were also identified using BLAST against the database of fungal virulence factors (DFVF) (Lu et al. 2012) with an e-value cutoff of $1e-6$, identity > 45%, and coverage > 70%.

Gene functional annotation of *Xanthoparmelia taractica*

The protein sequences were functionally annotated using Diamond v. 0.9.10.111 (Buchfink et al. 2015) against the NCBI-nr database (NR) (release 2017.10.10) with an e-value cutoff of $1e-6$. The Swiss-Prot annotation was conducted by BLAST against the Swiss-Prot database (release 2017.11.22) with an e-value cutoff of $1e-6$. The functional annotation through orthology assignment was done by eggNOG-mapper v. 2 (Cantalapiedra et al. 2021) with the DIAMOND mapping mode (Buchfink et al. 2021) with an e-value cutoff of $1e-6$. Transport proteins were annotated by BLAST against the Transporter Classification Database (Saier et al. 2014). Protein domain families were annotated by using the Pfam database-A. Genes involved in host-pathogen interactions were identified by using the pathogen-host interactions (PHI) database (Urban et al. 2021). The KEGG Ortholog (KO) and pathway were annotated by KEGG orthology identifiers available through the web-based server KAAS (KEGG Automatic Annotation Server, v. 2.1) (Moriya et al. 2007) with a representative set: for eukaryotes, KO assignment methods were: bi-directional best hit. Gene Ontology (GO) was annotated by InterProScan (Quevillon et al. 2005) against the InterPro v. 66.0 protein signature database (Finn et al. 2017). Cytochrome P450 was identified by BLASTP (2.5.0+) against the fungal cytochrome P450 database (Park et al. 2008) with an e-value cutoff of $1e-5$.

Annotation of secondary metabolite biosynthesis gene clusters was performed using antiSMASH fungal version 6.0 (Blin et al. 2021) and the corresponding protein sequences of PKS genes from *X. taractica* were manually checked using BlastP. PKSs that were found in *X. taractica* and those from the NCBI were compared, and if there was less than a 70% similarity, they were excluded from the phylogenetic analysis. The final dataset included protein sequences of PKSs identified from *X. taractica* and those retrieved from the NCBI. The dataset was aligned using MAFFT and the ML trees were constructed and viewed using RaxML and FigTree, respectively.

PKS genes expression analyses

For the expression studies, the protein sequences of the *X. taractica* PKS genes that clustered with other PKSs from fungi with well-defined functions were considered. The β -tubulin (*BT*) gene was used as a reference.

Following the settings recommended by Thornton and Basu (2015), the primers for RT-qPCR were designed using PrimerQuest Tool (<https://sg.idtdna.com/PrimerQuest/Home/Index>). Primer T_m (°C): lowest melting temperature of 50 and optimal melting temperature of 60; Primer GC (%): minimum 50, optimum 55, and maximum 60; Primer size (nt): optimum 20 and maximum 24; GC Clamp (nt) 1; Amplicon Size (bp): smallest size 75, optimum size 100, and largest size 150. The presence of secondary structures (if any) and cross, self-dimers, and hairpin construction were verified using UNAFold (<http://www.idtdna.com/UNAFold>) and Beacon Designer (<http://www.premierbiosoft.com/qpcr/>), respectively. **Table S3** lists all of the primer pairings designed in this study.

Total RNA was extracted from set-ups 1 and 2 after 21 days using TRIzol reagent (Thermo Fisher Scientific) following the manufacturer's protocol. cDNA was synthesized using the SPARKscript II RT Plus Kit (Shandong Sparkjade Scientific Instruments Co., Ltd, China). Each 20 μ l of PCR reaction included 2 \times SYBR Green qPCR Mix 10 μ l, Primer F (10 μ M) 0.4 μ l, Primer R (10 μ M) 0.4 μ l, cDNA 1 μ l, and RNase Free water 8.2 μ l. For both PKS and *BT* gene regions, PCR amplifications were conducted with an initial denaturation at 95 °C for 3 min, followed by 45 cycles of 95 °C for 20 sec, 60 °C for 15 s, 95 °C for 5 s; melting curves at 65 °C for 1 min.

The dataset from the expression study (monoculture vs. co-culture) was statistically analysed using one-way analysis of variance (ANOVA) available through the R v. 4.2.2 (R Core Team 2022). A p-value ≤ 0.05 was considered significant.

Results

Identification of lichenized fungi and alga

Fungal isolate LF-423E recovered from the lichen *X. taractica* grew significantly slower than other fungal isolates, indicating that it was most likely a lichenized fungal species. Preliminary identification of the ITS sequences obtained from the fungal and algal isolates revealed that LF-423E was closely related to the fungus *X. taractica*, while algal strain AGL-S420 was closely related to an undescribed *Trebouxia* species. Maximum likelihood phylogenies using ITS sequences validated the identification of the isolates (**Fig. S1** and **S2**). The sequences generated in this study were submitted to GenBank under the accession numbers OP279582 (LF-423E) and OP279583 (AGL-S420).

SEM of fungal and algal co-cultivation setups

In either of the co-culture settings, no lichen-like structure was observed. Slow growth was observed in all instances (**Fig. 1 a, b, c**).

Algal cells (AGL-S420) formed several small aggregations on the nitrocellulose membrane (**Fig. 1 d-l**). Lichenized fungus (LF-423E) mycelia produced a reticulate structure that linked algal aggregations (**Fig. 1 d**). Fungal mycelia were also seen growing between algal cells and loosely enveloping them (**Fig. 1 e, f, g, h**). Appressoria were detected on the surface of algal cells in some situations. However, in the majority of cases, lichenized fungal mycelia entered algal cells without the formation of appressoria.

The set-up 5 involved co-culturing of the endolichenic fungus *C. fibrosae* (CGMCC 3.20304) and the alga (AGL-S420). Here, algal cells also form aggregations. However, the mycelia of *C. fibrosae* exclusively grew on the nitrocellulose membrane, without entering or enclosing the algal cells. Spores were observed in this set-up (**Fig. 1 i, j**).

We co-cultured lichenized fungus (LF-423E), endolichenic fungus (CGMCC 3.20304), and alga (AGL-S420) in setup 6. In this case, our observation was a combination of the preceding two observations. The mycelia of LF-423E penetrated and encased the algal cell clusters, but the mycelia of CGMCC 3.20304 exclusively proliferated and sporulated on the surface of the nitrocellulose membrane (**Fig. 1 k, l**).

Genome sequencing and assembly of *Xanthoparmelia taractica*

The *X. taractica* assembly consists of 1336 scaffolds with an N50 of 112, 782 bp. The computed genome size was roughly 43.1 Mb, with a 43.70% GC content. The genomic

completeness calculated by BUSCO was 98.6%. (**Table S4**). According to RepeatMasker annotation, the bulk of repeats is long terminal repeat (LTR) elements (8.19 %) and DNA elements (2.69 %) (**Table S5**). Annotated non-coding RNA included 61 copies of tRNA, 12 copies of rRNA, and 25 copies of various forms of ncRNAs (**Table S6**). The genome sequence was deposited at NCBI under the accession number JANQBS000000000.

Gene prediction of *Xanthoparmelia taractica*

This assembly was predicted to have 10,730 ORFs with an average gene length of 1593.9 bps. The total gene length was around 17 M bps, which accounts for 39.64% of the genome. The total exon number was 30,187 and the total exon length was around 16 M, which is 36.32% of the genome.

Annotations of carbohydrate-active enzymes, antibiotic-related genes, and fungal virulence factors of *Xanthoparmelia taractica*

CAZymes annotation revealed that there were no polysaccharide lyases in the genome, but there were seven carbohydrate-binding modules (**Fig. 2 a**). Annotation of antibiotic-related genes revealed two antibiotic resistance genes and one antibiotic biosynthesis gene. BLAST against DFVF found 255 fungal virulence factors.

Gene functional annotation of *Xanthoparmelia taractica*

The NCBI-nr database and the Swiss-Prot database were used to annotate 8883 and 6452 genes, respectively (**Table S7**). The eggNOG-mapper found 8278 genes and divided them into 25 categories. Among these, the functions of a majority of genes could not be ascertained (3247). Categories that include the highest number of genes with known functions were post-translational modification, protein turnover, chaperones (490); secondary metabolites biosynthesis (453), and carbohydrate transport and metabolism (441). The remaining categories had less than 400 genes. The lowest number of genes (4) were detected under the category nuclear structure and extracellular structures (**Fig. 2 b**).

In the fungal cytochrome P450 database, there were 10,480 putative genes with homologs, of which 5,413 had more than 30% similarity. Whereas, in the transport categorization database, there were 1,364 homologs with 1,085 genes sharing more than 30% similarity, with the most (399) categorized as electrochemical potential-driven transporters and the fewest (7) classified as transmembrane electron carriers (**Fig. 2 c**). Host-pathogen interactions database predicted that 2,314 genes were involved in this interaction. A majority of these genes were linked to loss of pathogenicity (222), reduced virulence (1123), and unaffected pathogenicity (989) (**Fig. 2 d**). Through KEGG annotation we identified 10,055 genes that were divided into eight categories. Among these, 3457 genes were associated with

BRITE functional-hierarchies, 820 with cellular processes, 568 with environmental information processing, 869 with genetic information processing, 1334 with human diseases, 1801 with metabolism, 1013 with organismal systems, and the functions for the remaining 193 remained unresolved (**Fig. 2 e**). InterProScan summarized the genes into three main categories, with 14683 genes that were involved in biological processes. Among these, 84 were involved with secondary metabolic activities. 10,922 and 11,787 were classified as cellular components, and molecular functions, respectively (**Fig. 2 f**).

PKS gene clustering and analyses

The antiSMASH annotation predicted 46 secondary metabolite gene clusters in the genome of *X. taractica*. These included nine NRPS, three NRPS-Type I PKS, nine terpene, and 25 PKS gene clusters. Among these, 28 genes represented PKS clusters. There was 27 Type I and the rest were Type III. In the BlastP (**Table 1**) and ML phylogenetic analyses (**Fig. S3**), XtPKS12 nested within a group that included a PKS with acyl transferase, acyl hydrolase, and lysophospholipase function identified from *Diplocarpon rosae* (PBP21839). Whereas, XtPKS18 and XtPKS22 grouped with PKS having 6-methylsalicylic acid synthase (MCJ1371537) and methylphloroacetophenone synthesis activities (QEO24563 and A0A0R8YWJ7), respectively. The remaining PKSs are grouped with uncharacterized proteins or hypothetical proteins (**Fig. S3**). Hence, genes XTPKS12, XTPKS18, and XTPKS22 were selected for the expression analyses.

PKS gene expression analyses

The PKS genes XTPKS12, XTPKS18, and XTPKS22 were expressed at low levels in *X. taractica* monoculture (**Fig. 3**). However, the expression pattern of these genes altered in the co-culture system. When *X. taractica* and alga were co-cultured, XTPKS12 was down-regulated (0.06 folds) while XTPKS18 and XTPKS22 were up-regulated (1.75 and 1.17 folds, respectively) compared to monoculture.

Statistical analyses of the expression data revealed that the expression of XTPKS18 and XTPKS22 was significantly different between monoculture and coculture ($p < 0.05$). There was, however, no significant difference in the expression of XTPKS12 between monoculture and coculture.

Discussion

In this study, we isolated a lichenized fungus, *Xanthoparmelia taractica*, and its algal partner, *Trebouxia* sp. from lichen samples collected from China in 2019. We sequenced the whole genome of the mycobiont, *X. taractica*. The genome was 43.1 Mb with 10730 ORFs. We detected 27 Type I and one Type III PKS gene. Except for three PKS genes, XTPKS12,

XTPKS18, and XTPKS22, the function of most of them remained unknown. We selected these three genes for downstream expression analyses using a co-culture system. The setups included the mycobiont, photobiont, and an endolichenic fungus, *Coniochaeta fibrosae*. The setups with the mycobiont and the photobiont showed an early stage of symbiosis. Hyphal networks of the mycobiont connected and penetrated the algal cells were observed in these setups. Moreover, XTPKS12 was down-regulated, and XTPKS18 and XTPKS22 were modestly up-regulated. *Coniochaeta fibrosae* did not participate in the symbiosis or had a significant effect on the functioning of the PKS genes tested.

The genome of *X. taractica* was around 43.1 MB in size, with a projected gene count of 10,730. There were no *Xanthoparmelia* genomes accessible in public repositories before the present study. So, we compared our genome size to that of other fungi from *Lecanorales*. We observed a significant difference in genome sizes in this order. *Cetradonia linearis*, for example, has the smallest genome (19.5 MB) whereas *Alectoria fallacina* has the largest (65.2 MB). However, there is no clear link between genome size and habitat choice. For example, *Niebla homalea* (50.6 MB), *Ramalina intermedia* (26.2 MB), and *X. taractica* (43.1 MB) are all rock-inhabiting lichens but with varying genome sizes. Similarly, a recent study that compared the genomes of 46 lichen-forming fungi from *Lecanoromycetes* also couldn't predict why the size of biosynthetic gene clusters varied between species (Pizarro et al. 2020).

Twenty-eight PKS genes were identified from the genome of *X. taractica*. This included one Type III PKS gene and 27 Type I PKS genes (including three NRPS-Type I PKS hybrid genes). Several recent studies indicate that lichenized fungi may have a higher number of PKS genes compared to other fungi (Song et al. 2022; Singh et al. 2022b; Gerasimova et al. 2022); however, this trend is not well established. The variety of unique secondary metabolites synthesized by lichenized fungi may be correlated with a larger number of PKS genes. These genes might have been gained by lichenized fungi as they evolved from non-lichenized ancestors (Song et al. 2022). Compared to other fungal lifestyles, lichenized fungi have a higher tolerance to harsh environments, such as high temperatures, intense UV light, and dry weather. The secondary metabolites produced by PKS genes in lichenized fungi may help the lichens adapt to unfavorable conditions (Devashree et al. 2021).

Early phases of lichenization were detected in a co-culture of *X. taractica* and *Trebouxia* sp. Similarly, to previous studies, we also found fungal hyphae entering algal cells, frequently through the development of appressoria, which is considered indicative of symbiont compatibility (Muggia et al. 2021; Ahmadjian 1964; Athukorala et al. 2014). The hyphae of lichenized fungus disintegrate the algal cells during nutrient exchange. This is analogous to the interactions between plants and pathogens. Hence, we identified 2,314 genes in *X. taractica*

that might be involved in host-pathogen interactions. One of which (scaffold25.t62) was orthologous to *MagB* and *GzGpa3* genes of *Magnaporthe oryzae* and *Fusarium graminearum*, respectively. A similar gene (*UmGPA3*) was also identified in the lichenized fungus *Umbilicaria muhlenbergii* that was homologous to *Ustilago maydis Gpa3* and *Candida albicans Gpa2* and is likely to be involved in the initial establishment of symbiosis (Wang et al. 2020). Concurrently, an *in vitro* resynthesis involving *Usnea hakonensis* and its algal partner revealed 305 fungal and 203 algal genes that promote lichenization through symbiotic interface establishment, nutrient flow, and many other mechanisms (Kono et al. 2020). Similarly, cellular ultrastructure and their biochemical components, such as wall fibrils, extracellular matrix, and lectins, can also promote lichenization (Roth and Goodenough 2021; Wang et al. 2014). This demonstrates that lichenization is a complex process triggered by a combination of factors, many of which are yet unknown.

The endolichenic fungus, *C. fibrosae*, was previously isolated from the same lichen thallus, from which we also recovered the mycobiont and the photobiont (Si et al. 2021). However, the outcomes from our co-culture study showed *C. fibrosae* doesn't influence the symbiosis between *X. taractica* and *Trebouxia* sp. In all setups that included *C. fibrosae*, the fungus profusely sporulated without penetrating the algal cells. However, we predict *C. fibrosae* is plausibly dependent on the photobiont for nutrition. This is because nutrient exchange can also happen through physical contact between organisms (without penetrating the algal cells), as seen in *Mortierella elongata* and *Nannochloropsis oceanica* (Du et al. 2019). Similarly, some lichenized fungi also adopt this mechanism for withdrawing nutrients from the photobiont, such as *Coenogonium subvirescens* (Honegger 2009). More in-depth studies are needed to demystify whether *C. fibrosae* is nutritionally dependent on *Trebouxia* sp. and what its mode of nutrient acquisition is.

In monoculture, the expression levels of all tested PKS genes from *X. taractica* were substantially low even after 21 days. When *X. taractica* and *Trebouxia* sp. were co-cultured, XTPKS18 and XTPKS22 were significantly up-regulated, while XTPKS12 was slightly down-regulated. There are a few plausible explanations for this. The life cycle stage of the mycobiont could have affected the expression of genes XTPKS12, XTPKS18, and XTPKS22. In a co-culture study of *Gyalolechia flavorubescens* and *Trebouxia gelatinosa*, several PKS genes were highly expressed at 12, 24, 48, and 72 h whereas other PKS genes were expressed after four to six weeks (Song et al. 2022). Besides this, the expression of PKS genes can also be affected by cultural conditions. This was confirmed by a transcriptomic study of secondary metabolite genes in *Cladonia rangiferina* where expression of PKS genes was substantially influenced by culture conditions, microbial media, and abiotic factors such as light (Sveshnikova and Piercey-Normore 2021). The expression of some PKS genes is also known to vary between isolates of

the same fungus. For example, in *Cladonia macilenta*, PKS21 is responsible for the biosynthesis of biruloquinone, which gets exclusively upregulated in the purple isolate but not in the white variety (Kim et al. 2021a). So, in *X. taractica* the regulation of PKS genes could influence by one of the above scenarios or a mixture of them.

In this work, we successfully co-cultured *X. taractica* with the alga *Trebouxia* sp. and sequenced the whole genome of the lichenized fungus. We were unable to determine the involvement of the endolichenic fungal, *C. fibrosae*, in the lichen thallus or how this fungus obtains nutrients from the algae. Similarly, we identified 28 PKS genes in the genome of the mycobiont. For a majority of these PKS genes, their putative functions couldn't be identified during the BlastP searches and phylogenetic analyses. This demonstrates that our present understanding of the biology and genetics of lichenized fungi is sparse. Nonetheless, several promising discoveries from recent investigations have progressed our knowledge substantially (Pizarro et al. 2020; Armaleo et al. 2019; Keuler et al. 2020; Tagirdzhanova et al. 2021; Allen and Lendemer 2022; Resl et al. 2022; Merges et al. 2021). The information gleaned from this and subsequent investigations will help us better understand lichen symbiosis.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not Applicable

Availability of data and materials

All sequence data generated in this study (the genome sequence of *X. taractica*, **Table S1** and **S2**) are available at NCBI GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>).

Conflict of interest

The authors declare that there is no conflict of interest.

Funding

500 This study was funded by The National Natural Science Foundation of China (Project
501 No.: 31600100) and Open Fund for Instruments and Equipment of Shandong Normal
University (2023.1, 2023.9).

Author contributions

All authors contributed equally to the conception, writing and preparation of this manuscript. Primary isolations and preliminary identification of lichenized and algal species were completed by H Si, Y Su, and Y Wang. Molecular and phylogenetic analyses were completed by R Chang, T Bose, Y Su, and Y Wang. The co-culture systems were carried out by H Si, R Chang and Y Su. The manuscript was written by R Chang and T Bose, and all authors commented on previous versions of the manuscript. This study was supervised by H Si and R Chang. All authors read and approved the final manuscript.

Acknowledgement

The authors would like to thank Prof. Lisong Wang of the Kunming Institute of Botany, The Chinese Academy of Sciences and Prof. Zuntian Zhao of Shandong Normal University for their assistance in identifying lichen specimens.

References

- Ahmadjian V (1964) Further studies on lichenized fungi. *The Bryologist* 67:87-98. <https://doi.org/10.2307/3241190>
- Allen JL, Lendemer JC (2022) A call to reconceptualize lichen symbioses. *Trends in Ecology & Evolution* 37:582-589. <https://doi.org/10.1016/j.tree.2022.03.004>
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215:403-410. [https://doi.org/10.1016/s0022-2836\(05\)80360-2](https://doi.org/10.1016/s0022-2836(05)80360-2)
- Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom,
- Armaleo D, Müller O, Lutzoni F, Andrésson ÓS, Blanc G, Bode HB, Collart FR, Dal Grande F, Dietrich F, Grigoriev IV, Joneson S, Kuo A, Larsen PE, Logsdon JM, Lopez D, Martin F, May SP, McDonald TR, Merchant SS, Miao V, Morin E, Oono R, Pellegrini M, Rubinstein N, Sanchez-Puerta MV, Savelkoul E, Schmitt I, Slot JC, Soanes D, Szövényi P, Talbot NJ, Veneault-Fourrey C, Xavier BB (2019) The lichen symbiosis re-viewed through the genomes of *Cladonia grayi* and its algal partner *Asterochloris glomerata*. *BMC Genomics* 20:605. <https://doi.org/10.1186/s12864-019-5629-x>
- Armaleo D, Sun X, Culbertson C (2011) Insights from the first putative biosynthetic gene cluster for a lichen depside and depsidone. *Mycologia* 103:741-754. <https://doi.org/10.3852/10-335>
- Athukorala SNP, Huebner E, Piercey-Normore MD (2014) Identification and comparison of

- the 3 early stages of resynthesis for the lichen *Cladonia rangiferina*. *Canadian Journal of Microbiology* 60:41-52. <https://doi.org/10.1139/cjm-2013-0313>
- Austin MB, Noel JP (2003) The chalcone synthase superfamily of type III polyketide synthases. *Natural Product Reports* 20:79-110. <https://doi.org/10.1039/B100917F>
- Bao Z, Eddy SR (2002) Automated de novo identification of repeat sequence families in sequenced genomes. *Genome research* 12:1269-1276. <https://doi.org/10.1101/gr.88502>
- Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema Marnix H, Weber T (2021) antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research* 49:W29-W35. <https://doi.org/10.1093/nar/gkab335>
- Buchfink B, Reuter K, Drost H-G (2021) Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods* 18:366 - 368. <https://doi.org/10.1038/s41592-021-01101-x>
- Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 12:59-60. <https://doi.org/10.1038/nmeth.3176>
- Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J (2021) eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution* 38:5825-5829. <https://doi.org/10.1093/molbev/msab293>
- Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A (2016) Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods* 13:1050-1054. <https://doi.org/10.1038/nmeth.4035>
- Crawford SD (2015) Lichens Used in Traditional Medicine. In: Ranković B (ed) *Lichen Secondary Metabolites: Bioactive Properties and Pharmaceutical Potential*. Springer International Publishing, Cham, pp 27-80. https://doi.org/10.1007/978-3-319-13374-4_2
- Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9:772-772. <https://doi.org/10.1038/nmeth.2109>
- Deason TR (1960) *Exploratory studies of Texas soil algae*. University of Texas Publications 6022:70
- Devashree, Pandey A, Dikshit A (2021) Lichens: Fungal symbionts and their secondary metabolites. In: Singh J, Gehlot P (eds) *New and Future Developments in Microbial Biotechnology and Bioengineering*. Elsevier, pp 107-115. <https://doi.org/10.1016/B978-0-12-821005-5.00007-7>
- Du ZY, Zienkiewicz K, Vande Pol N, Ostrom NE, Benning C, Bonito GM (2019) Algal-fungal symbiosis leads to photosynthetic mycelium. *Elife* 8:e47815. <https://doi.org/10.7554/eLife.47815>
- Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang HY, Dosztányi Z, El-Gebali S, Fraser M, Gough J, Haft D, Holliday GL, Huang H, Huang X, Letunic I, Lopez R, Lu S, Marchler-Bauer A, Mi H, Mistry J, Natale DA, Necci M, Nuka G, Orengo CA, Park Y, Pesseat S, Piovesan D, Potter SC, Rawlings ND, Redaschi N, Richardson L, Rivoire C, Sangrador-Vegas A, Sigrist C, Sillitoe I, Smithers B, Squizzato S, Sutton G, Thanki N, Thomas PD, Tosatto SC, Wu CH, Xenarios I, Yeh LS, Young SY, Mitchell AL (2017) InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Research* 45:D190-d199. <https://doi.org/10.1093/nar/gkw1107>
- Gaya E, Fernández-Brime S, Vargas R, Lachlan Robert F, Gueidan C, Ramírez-Mejía M, Lutzoni F (2015) The adaptive radiation of lichen-forming Teloschistaceae is associated with sunscreens pigments and a bark-to-rock substrate shift. *Proceedings of the National Academy of Sciences* 112:11600-11605. <https://doi.org/10.1073/pnas.1507072112>
- Gerasimova JV, Beck A, Werth S, Resl P (2022) High diversity of Type I polyketide genes in

- Bacidia rubella* as revealed by the comparative analysis of 23 lichen genomes. *Journal of Fungi* 8:449
- Goga M, Elečko J, Marcinčinová M, Ručová D, Bačkorová M, Bačkor M (2020) Lichen metabolites: an overview of some secondary metabolites and their biological potential. In: Mérillon J-M, Ramawat KG (eds) *Co-Evolution of Secondary Metabolites*. Springer International Publishing, Cham, pp 175-209. https://doi.org/10.1007/978-3-319-96397-6_57
- Griffiths - Jones S (2005) Annotating non - coding RNAs with Rfam. *Current protocols in bioinformatics* 9:12.15. 11-12.15. 12. <https://doi.org/10.1093/nar/gki081>
- Gunawardana W, GDI, Sveshnikova N, Arsenaault A, Piercey-Normore MD (2021) Effect of environmental factors on polyketide synthase gene expression, usnic acid biosynthesis, and ground cover of a lichen-fungus, *Cladonia uncialis*. *Botany* 99:735-745. 10.1139/cjb-2021-0009
- Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR (2008) Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome biology* 9:R7-R7. <https://doi.org/10.1186/gb-2008-9-1-r7>
- Honegger R (2009) Lichen-forming fungi and their photobionts. vol 5. pp 307-333. https://doi.org/10.1007/978-3-540-87407-2_16
- Jeong M-H, Park C-H, Kim JA, Choi ED, Kim S, Hur J-S, Park S-Y (2021) Production and activity of cristazarin in the lichen-forming fungus *Cladonia metacorallifera*. *Journal of Fungi* 7:601. doi:10.3390/jof7080601
- Joneson S, Armaleo D, Lutzoni F (2011) Fungal and algal gene expression in early developmental stages of lichen-symbiosis. *Mycologia* 103:291-306. <https://doi.org/10.3852/10-064>
- Kapitonov VV, Jurka J (2008) A universal classification of eukaryotic transposable elements implemented in Repbase. *Nature Reviews Genetics* 9:411-412. <https://doi.org/10.1038/nrg2165-c1>
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30:772-780. <https://doi.org/10.1093/molbev/mst010>
- Kealey JT, Craig JP, Barr PJ (2021) Identification of a lichen depside polyketide synthase gene by heterologous expression in *Saccharomyces cerevisiae*. *Metabolic engineering communications* 13:e00172-e00172. <https://doi.org/10.1016/j.mec.2021.e00172>
- Keuler R, Garretson A, Saunders T, Erickson RJ, St. Andre N, Grewe F, Smith H, Lumbsch HT, Huang J-P, St. Clair LL, Leavitt SD (2020) Genome-scale data reveal the role of hybridization in lichen-forming fungi. *Scientific Reports* 10:1497. <https://doi.org/10.1038/s41598-020-58279-x>
- Kim W, Jeong M-H, Yun S-H, Hur J-S (2021a) Transcriptome analysis identifies a gene cluster for the biosynthesis of biruloquinone, a rare phenanthraquinone, in a lichen-forming fungus *Cladonia macilenta*. *Journal of Fungi* 7:398. <https://doi.org/10.3390/jof7050398>
- Kim W, Liu R, Woo S, Kang KB, Park H, Yu YH, Ha H-H, Oh S-Y, Yang JH, Kim H (2021b) Linking a gene cluster to atranorin, a major cortical substance of lichens, through genetic dereplication and heterologous expression. *MBio* 12:e01111-01121. <https://doi.org/10.1128/mbio.01111-21>
- Kono M, Kon Y, Ohmura Y, Satta Y, Terai Y (2020) In vitro resynthesis of lichenization reveals the genetic background of symbiosis-specific fungal-algal interaction in *Usnea hakonensis*. *BMC Genomics* 21:671. <https://doi.org/10.1186/s12864-020-07086-9>
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome research* 27:722-736. <https://doi.org/10.1101/gr.215087.116>
- Korman T, Ames B, Tsai S (2010) Structural enzymology of polyketide synthase: The structure–sequence–function correlation. vol 1. pp 305-345.

- <https://doi.org/10.1016/B978-008045382-8.00020-4>
- Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33:1870-1874. <https://doi.org/10.1093/molbev/msw054>
- Lagesen K, Hallin P, Rødland EA, Stårfeldt H-H, Rognes T, Ussery DW (2007) RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research* 35:3100-3108. <https://doi.org/10.1093/nar/gkm160>
- Li T, Jiang L, Hu Y, Paul JT, Zuniga C, Zengler K, Betenbaugh MJ (2020) Creating a synthetic lichen: Mutualistic co-culture of fungi and extracellular polysaccharide-secreting cyanobacterium *Nostoc* PCC 7413. *Algal Research* 45:101755. <https://doi.org/10.1016/j.algal.2019.101755>
- Liu R, Jiang S, Li E, Wei J (2018) A preliminary study on the secondary metabolites from the mycobiont and phycobiont of lichen *Endocarpon pusillum*. *Mycosystema* 37:246-255. <https://doi.org/10.13346/j.mycosystema.170216>
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* 42:D490-495. <https://doi.org/10.1093/nar/gkt1178>
- Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* 25:955-964. <https://doi.org/10.1093/nar/25.5.955>
- Lu T, Yao B, Zhang C (2012) DFVF: database of fungal virulence factors. *Database* 2012:bas032. <https://doi.org/10.1093/database/bas032>
- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1:2047-2217X-2041-2018. <https://doi.org/10.1186/2047-217X-1-18>
- Majoros WH, Pertea M, Salzberg SL (2004) TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20:2878-2879. <https://doi.org/10.1093/bioinformatics/bth315>
- Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27:764-770. <https://doi.org/10.1093/bioinformatics/btr011>
- McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L, Kalan L, King AM, Koteva K, Morar M, Mulvey MR, O'Brien JS, Pawlowski AC, Piddock LJ, Spanogiannopoulos P, Sutherland AD, Tang I, Taylor PL, Thaker M, Wang W, Yan M, Yu T, Wright GD (2013) The comprehensive antibiotic resistance database. *Antimicrob Agents Chemother* 57:3348-3357. <https://doi.org/10.1128/aac.00419-13>
- Merges D, Dal Grande F, Greve C, Otte J, Schmitt I (2021) Virus diversity in metagenomes of a lichen symbiosis (*Umbilicaria phaea*): complete viral genomes, putative hosts and elevational distributions. *Environ Microbiol* 23:6637-6650. <https://doi.org/10.1111/1462-2920.15802>
- Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Paper presented at the Gateway Computing Environments Workshop (GCE), New Orleans, LA, 14 Nov. 2010
- Mitrović T, Stamenković S, Cvetković V, Tošić S, Stanković M, Radojević I, Stefanović O, Čomić L, Đačić D, Ćurčić M (2011) Antioxidant, antimicrobial and antiproliferative activities of five lichen species. *International journal of molecular sciences* 12:5428-5448. <https://doi.org/10.3390/ijms12085428>
- Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* 35:W182-185. <https://doi.org/10.1093/nar/gkm321>
- Muggia L, Coleine C, De Carolis R, Cometto A, Selbmann L (2021) *Antarctolichenia onofrii* gen. nov. sp. nov. from Antarctic endolithic communities untangles the evolution of rock-inhabiting and lichenized fungi in Arthoniomycetes. *Journal of Fungi* 7.

- <https://doi.org/10.3390/jof7110935>
- Okorafor IC, Chen M, Tang Y (2021) High-titer production of olivetolic acid and analogs in engineered fungal host using a nonplant biosynthetic pathway. *ACS synthetic biology* 10:2159-2166
- Park J, Lee S, Choi J, Ahn K, Park B, Park J, Kang S, Lee YH (2008) Fungal cytochrome P450 database. *BMC Genomics* 9:402. <https://doi.org/10.1186/1471-2164-9-402>
- Pizarro D, Divakar PK, Grewe F, Crespo A, Dal Grande F, Lumbsch HT (2020) Genome-wide analysis of biosynthetic gene cluster reveals correlated gene loss with absence of usnic acid in lichen-forming fungi. *Genome Biology and Evolution* 12:1858-1868. <https://doi.org/10.1093/gbe/evaa189>
- Price AL, Jones NC, Pevzner PA (2005) De novo identification of repeat families in large genomes. *Bioinformatics* 21:i351-i358. <https://doi.org/10.1093/bioinformatics/bti1018>
- Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R (2005) InterProScan: protein domains identifier. *Nucleic Acids Research* 33:W116-120. <https://doi.org/10.1093/nar/gki442>
- R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Vienna, Austria.
- Resl P, Bujold AR, Tagirdzhanova G, Meidl P, Freire Rallo S, Kono M, Fernández-Brime S, Guðmundsson H, Andrésón ÓS, Muggia L, Mayrhofer H, McCutcheon JP, Wedin M, Werth S, Willis LM, Spribille T (2022) Large differences in carbohydrate degradation and transport potential among lichen fungal symbionts. *Nature Communications* 13:2634. <https://doi.org/10.1038/s41467-022-30218-6>
- Roth R, Goodenough U (2021) Lichen I. Solo fungal and algal partners. *Algal Research* 58:102334. <https://doi.org/10.1016/j.algal.2021.102334>
- Saier MH, Jr., Reddy VS, Tamang DG, Västermark A (2014) The transporter classification database. *Nucleic Acids Research* 42:D251-258. <https://doi.org/10.1093/nar/gkt1097>
- Schmitt I, Kautz S, Lumbsch HT (2008) 6-MSAS-like polyketide synthase genes occur in lichenized ascomycetes. *Mycological Research* 112:289-296. <https://doi.org/10.1016/j.mycres.2007.08.023>
- Schubert M, Lindgreen S, Orlando L (2016) AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Research Notes* 9:88. <https://doi.org/10.1186/s13104-016-1900-2>
- Shen B (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Current Opinion in Chemical Biology* 7:285-295. [https://doi.org/10.1016/S1367-5931\(03\)00020-6](https://doi.org/10.1016/S1367-5931(03)00020-6)
- Si H, Su Y, Zheng X, Ding M, Bose T, Chang R (2021) Phylogenetic and morphological analyses of *Coniochaeta* isolates recovered from Inner Mongolia and Yunnan revealed three new endolichenic fungal species. *Mycology* 83:105-121. <https://doi.org/10.3897/mycokeys.83.71140>
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210-3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Singh G (2023) Linking lichen metabolites to genes: emerging concepts and lessons from molecular biology and metagenomics. *Journal of Fungi* 9:160. doi:10.3390/jof9020160
- Singh G, Calchera A, Merges D, Valim H, Otte J, Schmitt I, Dal Grande F (2022a) A candidate gene cluster for the bioactive natural product gyrophoric acid in lichen-forming fungi. *Microbiology Spectrum* 10:e00109-00122
- Singh G, Dal Grande F, Schmitt I (2022b) Genome mining as a biotechnological tool for the discovery of novel biosynthetic genes in lichens. *bioRxiv:2022.2005.2004.490581*
- Slater GSC, Birney E (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:31. <https://doi.org/10.1186/1471-2105-6-31>
- Smit A, Hubley R, Green P (2015) RepeatModeler Open-1.0. 2008–2015. Seattle, USA: Institute for Systems Biology Available from: <http://www.repeatmasker.org>, Last Accessed May 1:2018
- Solhaug KA, Gauslaa Y (2012) Secondary lichen compounds as protection against excess solar

- radiation and herbivores. In: Lüttge U, Beyschlag W, Büdel B, Francis D (eds) *Progress in Botany* 73. Springer, pp 283-304. https://doi.org/10.1007/978-3-642-22746-2_11
- Song H, Kim K-T, Park S-Y, Lee G-W, Choi J, Jeon J, Cheong K, Choi G, Hur J-S, Lee Y-H (2022) A comparative genomic analysis of lichen-forming fungi reveals new insights into fungal lifestyles. *Scientific Reports* 12:10724. <https://doi.org/10.1038/s41598-022-14340-5>
- Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312-1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Stanke M, Waack S (2003) Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19:ii215-ii225. <https://doi.org/10.1093/bioinformatics/btg1080>
- Sveshnikova N, Piercey-Normore MD (2021) Transcriptome comparison of secondary metabolite biosynthesis genes expressed in cultured and lichenized conditions of *Cladonia rangiferina*. *Diversity* 13:529. <https://doi.org/10.3390/d13110529>
- Tagirdzhanova G, Saary P, Tingley JP, Díaz-Escandón D, Abbott DW, Finn RD, Spribille T (2021) Predicted input of uncultured fungal symbionts to a lichen symbiosis from metagenome-assembled genomes. *Genome Biology and Evolution* 13:evab047. <https://doi.org/10.1093/gbe/evab047>
- Tempel S (2012) Using and understanding RepeatMasker. *Mobile genetic elements*. Springer, pp 29-51. https://doi.org/10.1007/978-1-61779-603-6_2
- Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M (2008) Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome research* 18:1979-1990. <https://doi.org/10.1101/gr.081612.108>
- Thornton B, Basu C (2015) Rapid and simple method of qPCR primer design. *PCR Primer Design*. Springer, pp 173-179. https://doi.org/10.1007/978-1-4939-2365-6_13
- Urban M, Cuzick A, Seager J, Wood V, Rutherford K, Venkatesh Shilpa Y, Sahu J, Iyer SV, Khamari L, De Silva N, Martinez MC, Pedro H, Yates Andrew D, Hammond-Kosack Kim E (2021) PHI-base in 2022: a multi-species phenotype database for Pathogen–Host Interactions. *Nucleic Acids Research* 50:D837-D847. <https://doi.org/10.1093/nar/gkab1037>
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>
- Wang Y, Geng C, Yuan X, Hua M, Tian F, Li C (2018) Identification of a putative polyketide synthase gene involved in usnic acid biosynthesis in the lichen *Nephromopsis pallescens*. *PLOS ONE* 13:e0199110-e0199110. <https://doi.org/10.1371/journal.pone.0199110>
- Wang Y, Wei X, Bian Z, Wei J, Xu J-R (2020) Coregulation of dimorphism and symbiosis by cyclic AMP signaling in the lichenized fungus *Umbilicaria muhlenbergii*. *Proceedings of the National Academy of Sciences* 117:23847-23858. <https://doi.org/10.1073/pnas.2005109117>
- Wang YY, Liu B, Zhang XY, Zhou QM, Zhang T, Li H, Yu YF, Zhang XL, Hao XY, Wang M, Wang L, Wei JC (2014) Genome characteristics reveal the impact of lichenization on lichen-forming fungus *Endocarpon pusillum* Hedwig (Verrucariales, Ascomycota). *BMC Genomics* 15:34. <https://doi.org/10.1186/1471-2164-15-34>
- Weng J-K, Noel JP (2012) Structure–function analyses of plant type III polyketide synthases. In: Hopwood DA (ed) *Methods in Enzymology*, vol 515. Academic Press, pp 317-335. <https://doi.org/10.1016/B978-0-12-394290-6.00014-8>
- White TJ, Bruns T, Lee S, Taylor JW (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) *PCR protocols: a guide to methods and applications*. Academic Press, San Diego (California), pp 315-322. <https://doi.org/10.1016/B978-0-12-372180-8.50042-1>
- Yang M-X, Devkota S, Wang L-S, Scheidegger C (2021) Ethnolichenology—the use of lichens

- in the himalayas and southwestern parts of China. *Diversity* 13:330. <https://doi.org/10.3390/d13070330>
- Yousuf S, Choudhary MI, Atta ur R (2014) Lichens: chemistry and biological activities. In: Atta ur R (ed) *Studies in Natural Products Chemistry*, vol 43. Elsevier, pp 223-259. <https://doi.org/10.1016/B978-0-444-63430-6.00007-2>
- Zhang Y-J, Zhang S, Liu X, Wen HA, Wang M (2010) A simple method of genomic DNA extraction suitable for analysis of bulk fungal strains. *Letters in Applied Microbiology* 51:114-118. <https://doi.org/10.1111/j.1472-765X.2010.02867.x>

Table 1 The list of PKS gene annotated in *Xanthoparmelia taractica*.

PKS	Domains*	Classification	Closely related sequences		Similarity %
			GenBank Acc. No.	Taxa	
XtPKS1	AT	Type I PKS	OBT45611	<i>Pseudogymnoascus</i> sp.	34.58
XtPKS2	AT-DH-ER-KR-PP-C-A-PP-TD	NRPS-Type I PKS	XP_007802712	<i>Endocarpon pusillum</i>	50.07
XtPKS3	AT-PP-TD	Type I PKS	TVY93233	<i>Lachnellula willkommii</i>	44.99
XtPKS4	KS-AT-DH-CMT-KR-PP-C-A-PP-TD	NRPS-Type I PKS	XP_037164498	<i>Letharia columbiana</i>	88.91
XtPKS5	KS-AT-DH-PP-PP-CMT	Type I PKS	CAF9911258	<i>Imshaugia aleurites</i>	76.71
XtPKS6	KS-AT-DH-ER-KR-PP	Type I PKS	CAF9913835	<i>Alectoria fallacina</i>	77.71
XtPKS7	KS-AT-PT-ACP-PP-TD	Type I PKS	GAQ03707	<i>Aspergillus lentulus</i>	60.64
XtPKS8	KS-AT-PT-PP	Type I PKS	QIX11496	<i>Cladonia metacorallifera</i>	70.52
XtPKS9	KS-AT-DH-ER-KR-PP	Type I PKS	CAF9932618	<i>Imshaugia aleurites</i>	88.89
XtPKS10	SAT-KS-AT-PT-ACP-TE	Type I PKS	AGI60158	<i>Dolichousnea longissima</i>	82.88
XtPKS11	KS-AT-PT-PP-PP-TE	Type I PKS	TVY85230	<i>Lachnellula suecica</i>	44.40
XtPKS12	AT	Type I PKS	PBP21839	<i>Diplocarpon rosae</i>	81.32
XtPKS13	KS-AT-DH-CMT-ER-KR-PP	Type I PKS	XP_033541655	<i>Lindgomyces ingoldianus</i>	43.20
XtPKS14	AT-DH-ER-KR-PP	Type I PKS	CAF9937189	<i>Alectoria fallacina</i>	64.08
XtPKS15	KS-AT-DH-CMT-KR-PP-C-A-PP-TD	NRPS-Type I PKS	CAF9926813	<i>Alectoria fallacina</i>	85.31
XtPKS16	KS-AT-DH-CMT-KR-PP-TD	Type I PKS	XP_038741986	<i>Colletotrichum karsti</i>	58.88
XtPKS17	SAT-KS-AT-PT-PP-ACP-TE	Type I PKS	CAF9911758	<i>Alectoria fallacina</i>	85.67
XtPKS18	AT	Type I PKS	CAF9929810	<i>Heterodermia speciosa</i>	81.19
XtPKS19	KS-AT-DH-ER-KR-PP	Type I PKS	XP_033573117	<i>Mytilinidion resinicola</i>	46.41
XtPKS20	DH-CMT-ER-KR-PP	Type I PKS	CAF9941811	<i>Imshaugia aleurites</i>	77.55
XtPKS21	KS-AT-PT-ACP-PP-TE	Type I PKS	AII32482	<i>Xanthoparmelia substrigosa</i>	95.56
XtPKS22	SAT-KS-AT-DH-PP-CMT	Type I PKS	AGI60156	<i>Dolichousnea longissima</i>	82.93
XtPKS23	Not available	Type III PKS	KAG6999366	<i>Physcia stellaris</i>	62.35
XtPKS24	KS-AT-DH-ER-KR-PP	Type I PKS	PQE17449	<i>Rutstroemia</i> sp.	69.30
XtPKS25	KS-AT-DH	Type I PKS	CAF9930981	<i>Alectoria fallacina</i>	64.01
XtPKS26	KS-AT-PT-ACP-TE	Type I PKS	KAF4632819	<i>Cudoniella acicularis</i>	41.65
XtPKS27	KS-AT-DH-KR-PP	Type I PKS	CAF9932618	<i>Imshaugia aleurites</i>	55.02
XtPKS28	KS-AT-PP-CMT	Type I PKS	ABS58604	<i>Xanthoparmelia semiviridis</i>	95.52

*A: Adenylation; AT: Acyltransferase; ACP: Acyl carrier protein with an SH group on the

cofactor, a serine-attached 4'-phosphopantetheine; C: Condensation; CMT: Carbon methyltransferase; DH: Dehydratase; ER: Enoylreductase; KS: Keto-synthase with an SH group on a cysteine side-chain; KR: Ketoreductase; MT: Methyltransferase O- or C- (α or β); PP: Phosphopantetheine; PT: Product template; SAT: Starter unit : ACP transacylase in aflatoxin biosynthesis; SH: PLP-dependent cysteine lyase; TD: Terminal reductase; TE: Thioesterase.

Figure legends

Fig. 1 Co-culture of mycobiont (*Xanthoparmelia taractica*), photobiont (*Trebouxia* sp.), and endolichenic fungi (*Coniochaeta fibrosae*). Cultures of **a**: mycobiont *Xanthoparmelia taractica*; **b**: mycobiont *Xanthoparmelia taractica* + photobiont *Trebouxia* sp.; **c**: photobiont *Trebouxia* sp. on nitrocellulose membrane placed on the surface of Bold's basal medium and malt extract yeast (99:1) agar plates. **d**: Co-culture of mycobiont and photobiont. Algal cells are enclosed in hyphae and form many small aggregations on the nitrocellulose membrane. White arrows = algal cell clusters enclosed by hyphae, black arrow = fungal hyphae on the nitrocellulose membrane; **e, f, g, h**: Co-culture of mycobiont and photobiont. The hyphae of *X. taractica* were observed to penetrate algal cells in this co-culture setup. White arrow = appressoria; **i, j**: Co-culture of endolichenic fungi and photobiont. In this set-up the fungal hyphae didn't penetrate the algal cells. White arrow = conidia produced by endolichenic fungus; **k, l**: Co-culture of mycobiont, photobiont, and endolichenic fungus. The hyphae of mycobiont penetrated photobiont, but endolichenic fungus did not. Scale bars: d = 50 μ m; e-l = 5 μ m.

Fig. 2 Carbohydrate-active enzymes (CAZymes) identification and gene functional annotation of *Xanthoparmelia taractica*. **a**: CAZymes identified in the genome which were divided into six functional classes. These were Glycosyl Transferases (GT), Polysaccharide Lyases (PL), Carbohydrate Esterases (CE), Auxiliary Activities (AA), Carbohydrate-Binding Modules (CMB), and Glycoside Hydrolases (GH); **b**: eggNOG classifications of assembled unigenes. The functional annotations were divided into 25 categories, corresponding to clusters of orthologous groups (COGs). A = RNA processing and modification, B = Chromatin structure and dynamics, C = Energy production and conversion, D = Cell cycle control, cell division, chromosome partitioning, E = Amino acid transport and metabolism, F = Nucleotide transport and metabolism, G = Carbohydrate transport and metabolism, H = Coenzyme transport and metabolism, I = Lipid transport and metabolism, J = Translation, ribosomal structure and biogenesis, K = Transcription, L = Replication, recombination and repair, M = Cell wall/membrane/envelope biogenesis, N = Cell motility, O = Posttranslational modification, protein turnover, chaperones, P = Inorganic ion transport and metabolism, Q = Secondary metabolites biosynthesis, transport and catabolism, R = General function prediction only, S = Function unknown, T = Signal transduction mechanisms, U = Intracellular trafficking, secretion, and vesicular transport, V = Defense mechanisms, W = Extracellular structures, Y = Nuclear structure, Z = Cytoskeleton; **c**: Transport proteins annotation of assembled unigenes. Unigenes were summarized into seven main categories. A = Channels/Pores, B = Electrochemical

911 Potential-driven Transporters, C = Primary Active Transporters, D = Group Translocators, E = Transmembrane Electron Carriers, F = Accessory Factors Involved in Transport, G = Incompletely Characterized Transport Systems; **d**: Pathogen-host interactions (PHI) annotation of assembled unigenes. Unigenes were summarized into nine main categories. A = Chemistry target: resistance to chemical, B = Chemistry target: sensitivity to chemical, C = Effector (plant avirulence determinant), D = Enhanced antagonism, E = Increased virulence (hypervirulence), F = Lethal, G = Loss of pathogenicity, H = Reduced virulence, I = Unaffected pathogenicity; **e**: Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology annotation of assembled unigenes. The unigenes were summarized into eight main categories; **f**: Gene ontology (GO) classification of assembled unigenes. Unigenes were summarized into three main categories.

Fig. 3 RT-qPCR expression analysis of PKS12, PKS18, and PKS22 in *Xanthoparmelia taractica* monoculture and with *Trebouxia sp.* in co-culture systems. Bars indicate standard errors, while an asterisk is equivalent to a p-value ≤ 0.05 .

Supplementary figure legends

Fig. S1 Maximum likelihood phylogenetic analyses of *Xanthoparmelia* species using complete ITS region. Isolates recovered in this study are in bold font and highlighted in light green. ML bootstrap support values ≥ 75 are indicated at the nodes.

Fig. S2 Maximum likelihood phylogenetic analyses of *Trebouxia* using complete ITS region. Isolates recovered in this study are in bold font and highlighted in light green. ML bootstrap support values ≥ 75 are indicated at the nodes.

Fig. S3 PKSs identified in *Xanthoparmelia taractica* and closely related protein sequences were phylogenetically analyzed using maximum likelihood. PKSs identified in this study are in bold font. Clades highlighted in green indicate PKSs with well-defined functions, whereas those without distinct functions are highlighted in blue. ML bootstrap support values ≥ 75 are indicated at the nodes.

Supplementary Table legends

Table S1 GenBank accession numbers of complete ITS sequences of *Xanthoparmelia* species used for the maximum likelihood phylogenetic analysis. Two species of *Parmelia* served as the outgroup taxa.

Table S2 GenBank accession numbers of complete ITS sequences of *Trebouxia* species used for the maximum likelihood phylogenetic analysis. Two species of *Symbiochloris* served as the outgroup taxa.

Table S3 RT-qPCR primers designed in this study for the expression analysis of selected PKS genes. Forward and reverse primers are denoted by the letters F and R in the primer name.

Table S4 Summary of BUSCO statistics for the *Xanthoparmelia taractica* genome.

Table S5 List of repeats identified in the genome of *Xanthoparmelia taractica*.

Table S6 Non-coding RNA identified from the genome of *Xanthoparmelia taractica*.

Table S7 A summary of the functional annotations of the *Xanthoparmelia taractica* genome.





