Title

The reference genome assembly of the bright cobblestone lichen, Acarospora socialis

Journal

Journal of Heredity, 114(6)

ISSN

0022-1503

Authors

Adams, Julia N

Escalona, Merly

Marimuthu, Mohan PA

et al.

Publication Date

2023-11-15

DOI

10.1093/jhered/esad052

Peer reviewed

Journal of Heredity, 2023, **114**, 707–714
https://doi.org/10.1093/jhered/esad052
Advance access publication 7 September 2023
Genome Resources

American Genetic Association

OXFORD

Genome Resources

The reference genome assembly of the bright cobblestone lichen, *Acarospora socialis*

Julia N. Adams[1,2], Merly Escalona[3], Mohan P.A. Marimuthu[4], Colin W. Fairbairn[5], Eric Beraut[5], William Seligmann[5], Oanh Nguyen[4], Noravit Chumchim[4] and Jason E. Stajich[2,6,*]

[1]Department of Botany and Plant Sciences, University of California Riverside, Riverside, CA 92521, United States,
[2]Department of Microbiology and Plant Pathology, University of California Riverside, Riverside, CA 92521, United States,
[3]Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, CA 95064, United States,
[4]DNA Technologies and Expression Analysis Core Laboratory, Genome Center, University of California-Davis, Davis, CA 95616, United States,
[5]Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA 95064, United States
[6]Institute for Integrative Genome Biology, University of California Riverside, Riverside, CA 92521, United States,

[*]Corresponding author:Department of Microbiology and Plant Pathology, University of California-Riverside, 900 University Ave, Riverside, CA 92521, United States. Email: jason.stajich@ucr.edu

Corresponding Editor: Kenneth Olsen

Abstract

Acarospora socialis, the bright cobblestone lichen, is commonly found in southwestern North America. This charismatic yellow lichen is a species of key ecological significance as it is often a pioneer species in new environments. Despite their ecological importance virtually no research has been conducted on the genomics of *A. socialis*. To address this, we used long-read sequencing to generate the first high-quality draft genome of *A. socialis*. Lichen thallus tissue was collected from Pinkham Canyon in Joshua Tree National Park, California and deposited in the UC Riverside herbarium under accession #295874. The de novo assembly of the mycobiont partner of the lichen was generated from Pacific Biosciences HiFi long reads and Dovetail Omni-C chromatin capture data. After removing algal and bacterial contigs, the fungal genome was approximately 31.2 Mb consisting of 38 scaffolds with contig and scaffold N50 of 2.4 Mb. The BUSCO completeness score of the assembled genome was 97.5% using the Ascomycota gene set. Information on the genome of *A. socialis* is important for California conservation purposes given that this lichen is threatened in some places locally by wildfires due to climate change. This reference genome will be used for understanding the genetic diversity, population genomics, and comparative genomics of *A. socialis* species. Genomic resources for this species will support population and landscape genomics investigations, exploring the use of *A. socialis* as a bioindicator species for climate change, and in studies of adaptation by comparing populations that occur across aridity gradients in California.

Key words: California Conservation Genomics Project, lichen genomics, metagenomics, mycobiont, symbiosis

Introduction

Lichens are a diverse group of symbiotic organisms found in nearly every terrestrial ecosystem from the poles (Kappen 2000; Sancho and Pintado 2004) to the tropics (Sipman and Harris 1989) and grow on diverse substrates, commonly found on rocks, soil, and tree bark (Nash 2008). In summary, lichens consist of the partnership between a mycobiont (fungus) and a photobiont [cyanobacteria (Büdel 1992), eukaryotic algae (Tschermak-Woess 1988; Gärtner 1992), or both; (Nash 2008; Honegger 2009)], along with numerous other microorganisms (Hawksworth and Grube 2020) such as bacteria (e.g. Cardinale et al. 2008; Bates et al. 2011), other fungi [e.g. endolichenic fungi (U'Ren et al. 2012), lichenicolous fungi (Diederich et al. 2018), and basidiomycete yeasts (Spribille et al. 2016)], and viruses (e.g. Merges et al. 2021). The primary mycobiont partner obtains carbohydrate-rich resources directly from their photosynthetic partner (Nash 2008; Lutzoni and Miadlikowska 2009), and as part of the symbiosis, the fungus protects the photobiont from desiccation and the damaging effects of solar radiation (Kranner and Lutzoni 1999). The lichen symbiosis is an evolutionarily and ecologically successful strategy for fungi (>20% of fungi are lichenized), with approximately 19,409 lichen-forming species currently known (Lücking et al. 2017). Lichens can be rich in secondary metabolites, which in some cases gives lichens their color (Holder et al. 2000; Nash 2008). Some species of lichens are extremely sensitive to their climatic environments and have been used for bioindicators of air pollution (Hawksworth 1970; Conti and Cecchetti 2001; Nimis et al. 2002; Jovan and Mccune 2006; Jovan 2008).

Acarospora socialis H. Magn. (Acarosporaceae, Lecanoromycetes) is a yellow areolate to squamulose lichen commonly found in southwestern North America (Knudsen 2007), that is, across the Sonoran Desert regions of Arizona, southern California, Baja California, Baja California Sur, and Sonora, but can occur as far north as Oregon and additional

Received June 4, 2023; Accepted September 4, 2023

areas in the United States and Mexico (Knudsen 2007). Its bright yellow color is due to the production of a secondary metabolite known as rhizocarpic acid (Holder et al. 2000), also present in several other lichens, for example, *Rhizocarpon geographicum* (L.) DC. (Rhizocarpaceae), the map lichen. The primary photobiont partner is a chlorococcoid green alga in the genus *Trebouxia* (Duewer 1971; J.N. Adams et al. unpublished data). The bright cobblestone lichen is a wide-ranging foundational species of important ecological significance (Knudsen 2007), however its existence may be threatened in some places by wildfires (e.g. Miller et al. 2022) and climate change (e.g. Allen and Lendemer 2016; Nelson et al. 2022), which has been shown in previous work for other lichen species.

The generation of the reference genome of *A. socialis* will enhance ongoing state and federal conservation and management programs in California by focusing on a model lichen that is sensitive to air pollution, climate change, and wildfires. This knowledge will complement other projects in the California Conservation Genomics Project (CCGP) by establishing reference data for the mycobiont partner of this common yellow lichen necessary for documenting the diversity of *A. socialis* across the Southwest. These reference data will assist in development of conservation strategies informed by genomic diversity and support the utility of this lichen as a bioindicator of climate change and pollution.

Here, we present the first genome assembly of the lichen *A. socialis*. This is the first genome sequence and assembly of a lichen in the family Acarosporaceae sequenced directly from the lichen thallus using PacBio HiFi long-read sequencing and Dovetail Omni-C data. This reference genome assembly of *A. socialis* will provide a foundation for studying comparative, landscape, and population genomics of *Acarospora* throughout its range.

Methods

Biological materials

Acarospora socialis was collected at Pinkham Canyon (33.778278 N, 115.932556 W) in Joshua Tree National Park (JTNP), California on October 25, 2020. The lichen colony was fully exposed in the sun and was 11 inches above the ground from the center of the lichen colony. Photos of the lichen colony and the surrounding habitat and vegetation were taken with a Canon EOS D5 Mark 2 camera (Fig. 1). Fresh lichen tissue was sampled from volcanic rock (boulder) with a sterilized metal scraper into 12 separate tubes. The samples in tubes collected for DNA were immediately frozen in a cooler with dry ice. To reduce bacterial contamination, half of the samples (six total tubes) were surface-sterilized based on a modified protocol from Arnold et al. (2009). Briefly, each sample was washed for 2 min in sterile cell culture water, 10 s in 95% ethanol, and 2 min in 70% ethanol. Lichen samples were stored at −80 °C prior to DNA extractions. A voucher specimen from the collection was deposited in the University of California, Riverside (UCR) Herbarium under accession number UCR #295874.

High molecular weight DNA extraction

High molecular weight (HMW) genomic DNA (gDNA) was extracted from 350 mg of surface-sterilized whole thalli (DNAPC4-10) following a published protocol (Inglis et al. 2018), which utilizes a high-salt cetyltrimethylammonium bromide (CTAB) DNA extraction protocol. The extracted HMW DNA was further purified twice using the high-salt phenol-chloroform method (Pacific BioSciences—PacBio, Menlo Park, CA). The DNA purity (260/280 = 1.73 and 260/230 = 1.80) was assessed by absorbance ratios on a NanoDrop ND-1000 spectrophotometer. The DNA yield (3.4 µg) was quantified using QuantiFluor ONE dsDNA Dye

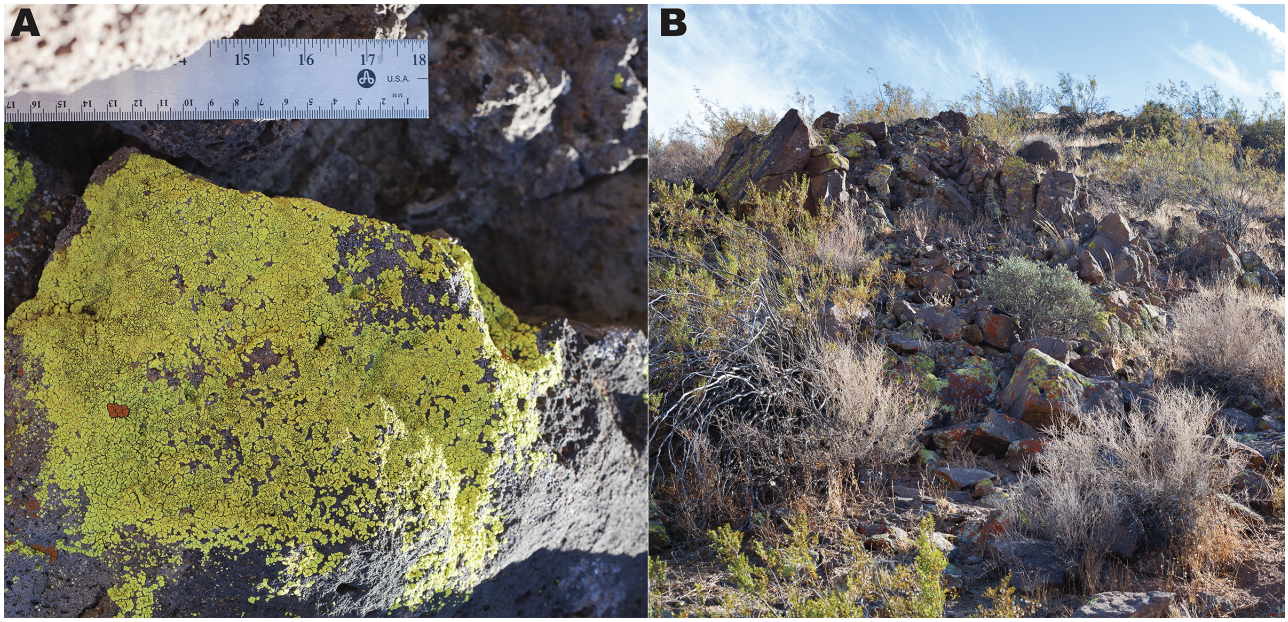

Fig. 1. *Acarospora socialis*, the bright cobblestone lichen. A) An individual in Joshua Tree National Park, Pinkham Canyon, California. The ruler (mm) shows the scale of the lichen. B) The surrounding vegetation where the lichen *A. socialis* was collected at Pinkham Canyon. The vegetation is dominated by creosote bush (*Larrea tridentata*) in a Creosote bush scrub community. Photos were taken by R. W. Adams.

assay (Promega, Madison, WI). The size distribution of the HMW gDNA was estimated using the Femto Pulse system (Agilent, Santa Clara, CA) and found that 35% of the total fragments were 30 kb or more.

PacBio HiFi library preparation and sequencing

The HiFi SMRTbell library was constructed using the SMRTbell Express Template Prep Kit v2.0 (PacBio, Cat. #100-938-900) according to the manufacturer's instructions. HMW gDNA was sheared to a target DNA size distribution between 12 and 20 kb. The sheared gDNA was concentrated using 1.8× of AMPure PB beads (PacBio Cat. #100-265-900) for the removal of single-strand overhangs at 37 °C for 15 min, followed by further enzymatic steps of DNA damage repair at 37 °C for 30 min, end repair and A-tailing at 20 °C for 10 min and 65 °C for 30 min, and ligation of overhang adapter v3 at 20 °C for 60 min. The SMRTbell library was purified and concentrated with 0.45× Ampure PB beads for size selection with 40% diluted AMPure PB beads (PacBio, Cat. #100-265-900) to remove short SMRTbell templates, <3 kb. The 12 to 20 kb average HiFi SMRTbell library was sequenced at UC Davis DNA Technologies Core (Davis, CA) using 0.5 8 M SMRT cell, Sequel II sequencing chemistry 2.0, and 30-h movies each on a PacBio Sequel II sequencer.

Omni-C library preparation and sequencing

The Omni-C library was prepared using the Dovetail Omni-C Kit (Dovetail Genomics, Scotts Valley, CA) according to the manufacturer's protocol with slight modifications. First, specimen tissue was thoroughly ground with a mortar and pestle while cooled with liquid nitrogen. Nuclear isolation was then performed using published methods (Workman et al. 2018). Subsequently, chromatin was fixed in place in the nucleus and digested under various conditions of DNase I until a suitable fragment length distribution of DNA molecules was obtained. Chromatin ends were repaired and ligated to a biotinylated bridge adapter followed by proximity ligation of adapter containing ends. After proximity ligation, crosslinks were reversed, and the DNA purified from proteins. Purified DNA was treated to remove biotin that was not internal to ligated fragments. An NGS library was generated using an NEB Ultra II DNA Library Prep kit (New England Biolabs, Ipswich, MA) with an Illumina compatible y-adaptor. Biotin-containing fragments were then captured using streptavidin beads. The post capture product was split into two replicates prior to PCR enrichment to preserve library complexity with each replicate receiving unique dual indices. The library was sequenced at the UC Santa Cruz Paleogenomics laboratory (Santa Cruz, CA) on an Illumina NextSeq 550 platform (Illumina, San Diego, CA) to generate approximately 30 million, 2 × 150 bp read pairs.

Genome assembly

The draft genome of *A. socialis* was assembled following the protocol Version 5.0 from the CCGP, specific for haploid genomes, as outlined on Table 1. As with other CCGP assemblies, our goal is to produce a high-quality and highly contiguous assembly using PacBio HiFi reads and Omni-C data while minimizing manual curation.

Briefly, we removed remnant adapter sequences from the PacBio HiFi dataset using HiFiAdapterFilt (Sim et al. 2022) and generated the initial haploid assembly with the filtered PacBio HiFi reads and the Omni-C data using HiFiasm (Cheng et al. 2022). We checked for contamination using the BlobToolKit framework (Challis et al. 2020) and removed contigs that did not belong to the phylum Ascomycota. We then scaffolded the assembly using the Omni-C data with SALSA (Ghurye et al. 2017, 2019). The assembly was manually curated by generating and analyzing a contact map and cutting at the gaps where misassemblies were identified and no further joins were made after this step. To generate the contact map, we aligned the Omni-C data against the assembly with BWA-MEM (Li 2013), identified ligation junctions and generated Omni-C pairs using pairtools (Goloborodko et al. 2018). We generated a multi-resolution Omni-C matrix with cooler (Abdennur and Mirny 2020) and balanced it with hicExplorer (Ramírez et al. 2018). We used HiGlass (Kerpedjiev et al. 2018) and the PretextSuite (https://github.com/wtsi-hpag/PretextView; https://github.com/wtsi-hpag/PretextMap; https://github.com/wtsi-hpag/PretextSnapshot) to visualize the contact map.

Genome size estimation and quality assessment

To estimate the genome size, we generated k-mer counts database from the PacBio HiFi reads using meryl (https://github.com/marbl/meryl) and provided this to GenomeScope 2.0 (Ranallo-Benavidez et al. 2020) to estimate genome features including sequencing error, genome size, heterozygosity, and repeat content. To obtain general contiguity metrics, we ran QUAST (Gurevich et al. 2013). To evaluate the genome quality and completeness we used BUSCO (Manni et al. 2021) with the ascomycota database (ascomycota_odb10) which contains 1,706 genes. Assessment of base level accuracy (QV) and k-mer completeness was performed using the previously generated meryl database and mercury (Rhie et al. 2020). We further estimated genome assembly accuracy via BUSCO gene set frameshift analysis using the pipeline described in (Korlach et al. 2017).

Following data availability and quality metrics established by Rhie et al. (2021), we used the derived genome quality notation $x.y.P.Q.C$, where x = log10[contig NG50]; y = log10[scaffold NG50]; P = log10 [phased block NG50]; Q = Phred base accuracy QV (quality value); C = % genome represented by the first "n" scaffolds, following a karyotype of n = 24, inferred from ancestral taxa (Genome On A Tree, Release 2022.07.21).

Mitochondrial contig identification and assembly

A draft *A. socialis* mitochondrial (MT) partial genome of ~16 kb was constructed from Illumina sequence data generated as part of the CCGP population genomics project using AAFTF (Stajich and Palmer 2019) to run NOVOplasty (Dierckxsens et al. 2017). This draft MT genome was searched against our lichen genome assembly of 79 scaffolds using BLASTN (Altschul et al. 1990) and Minimap2 (Li 2018) to identify scaffolds as mitochondrial. The ~16 kb draft genome aligned with 99.8% sequence identity to several scaffolds and the longest of those, scaffold 23 at ~44 kb, was chosen as a basis to start building the complete MT genome. Scaffold 23 was then searched against all 79 scaffolds of the assembly via BLASTN and two long scaffolds (scaffolds 36 and 37) were chosen that exhibited characteristics of partially aligning to the ends of scaffold 23, but also extending the sequence region of interest. A consensus

Table 1. Assembly pipeline and software used.

Assembly	Software and options[§]	Version
Filtering PacBio HiFi adapters	HiFiAdapterFilt	Commit 64d1c7b
K-mer counting	Meryl (k = 21)	1
Estimation of genome size and heterozygosity	GenomeScope	2
De novo assembly (contiging)	HiFiasm (Hi-C Mode, -l0 --n-hap 1)	0.16.1-r375
Scaffolding		
Omni-C Scaffolding	SALSA (-DNASE, -i 20, -p yes)	2
Omni-C Contact map generation		
Short-read alignment	BWA-MEM (-5SP)	0.7.17-r1188
SAM/BAM processing	samtools	1.11
SAM/BAM filtering	pairtools	0.3.0
Pairs indexing	pairix	0.3.7
Matrix generation	cooler	0.8.10
Matrix balancing	hicExplorer (hicCorrectmatrix correct --filterThreshold -2 4)	3.6
Contact map visualization	HiGlass	2.1.11
	PretextMap	0.1.4
	PretextView	0.1.5
	PretextSnapshot	0.0.3
Genome quality assessment		
Basic assembly metrics	QUAST (--est-ref-size)	5.0.2
Assembly completeness	BUSCO (-m geno, -l ascomycota)	5.0.0
	Merqury	2020-01-29
Contamination screening		
Local alignment tool	BLAST + (blastn, -db nt, -outfmt "6 qseqid staxids bitscore std," -max_target_seqs 1, -max_hsps 1, -evalue 1e-25	2.10
General contamination screening	BlobToolKit (PacBio HiFi Data for coverage and BUSCODB = ascomycota; TAXID = 1307805)	2.3.3
Mitochondrial identification and assembly		
Automatic assembly of the fungi	AAFTF	0.4.0
Organelle assembler	NOVOplasty	4.3.1
Basic local alignment search tool	BLASTN	2.13.0
Sequence alignment program	Minimap2	2.24
Sequence assembly program	CAP3	2015-02-10
Molecular genetics analysis	MEGA	11

Software citations are listed in the text.
[§]Options detailed for non-default parameters.

assembly was then generated from these three scaffolds (23, 36, and 37) with CAP3 (Huang and Madan 1999) which produced a 99,036 bp MT genome with repeating regions at the ends of the contig where the circular structure joins and overlaps. These overlapping regions were identified and edited using BLASTN, and MEGA (Tamura et al. 2021). The complete MT genome assembly is 83,434 bp.

Results

Sequencing data

The Omni-C and PacBio HiFi sequencing libraries generated 16.8 million read pairs and 694.9 thousand reads respectively. The latter yielded 123 fold coverage (N50 read length 7,389 bp; minimum read length 51 bp; mean read length 6,196 bp; maximum read length of 41,285 bp) based on the size estimated by Genomescope 2.0 of 34.7 Mb. Based on PacBio HiFi reads, we estimated 0.452% sequencing error rate. The k-mer spectrum based on PacBio HiFi reads show (Fig. 2A) a unimodal distribution with a single peak at ~123.

Nuclear genome assembly

The final (glAcaSoci1) genome assembly size is close to the estimated value from Genomescope (Fig. 2A, Pflug et al. 2020). The assembly consists of 37 scaffolds spanning 31.1 Mb with contig and scaffold N50 of 2.4 Mb, longest contig and scaffold of 4.07 Mb. Detailed assembly statistics are reported in tabular form in Table 2, and graphical representation for the assembly in Fig. 2B. The assembly has a BUSCO completeness score of 97.5% using the Ascomycota gene set, a per-base quality (QV) of 59.83, a k-mer completeness of 98.01 and a frameshift indel QV of 47.72. Telomere statistics were calculated with the "assess" tool in AAFTF (Stajich and Palmer 2019) and the number of telomeres found were 9 forward and 8 reverse.

The initial contig assembly consisted of 932 contigs, from which we kept 79 contigs belonging to Ascomycota. From

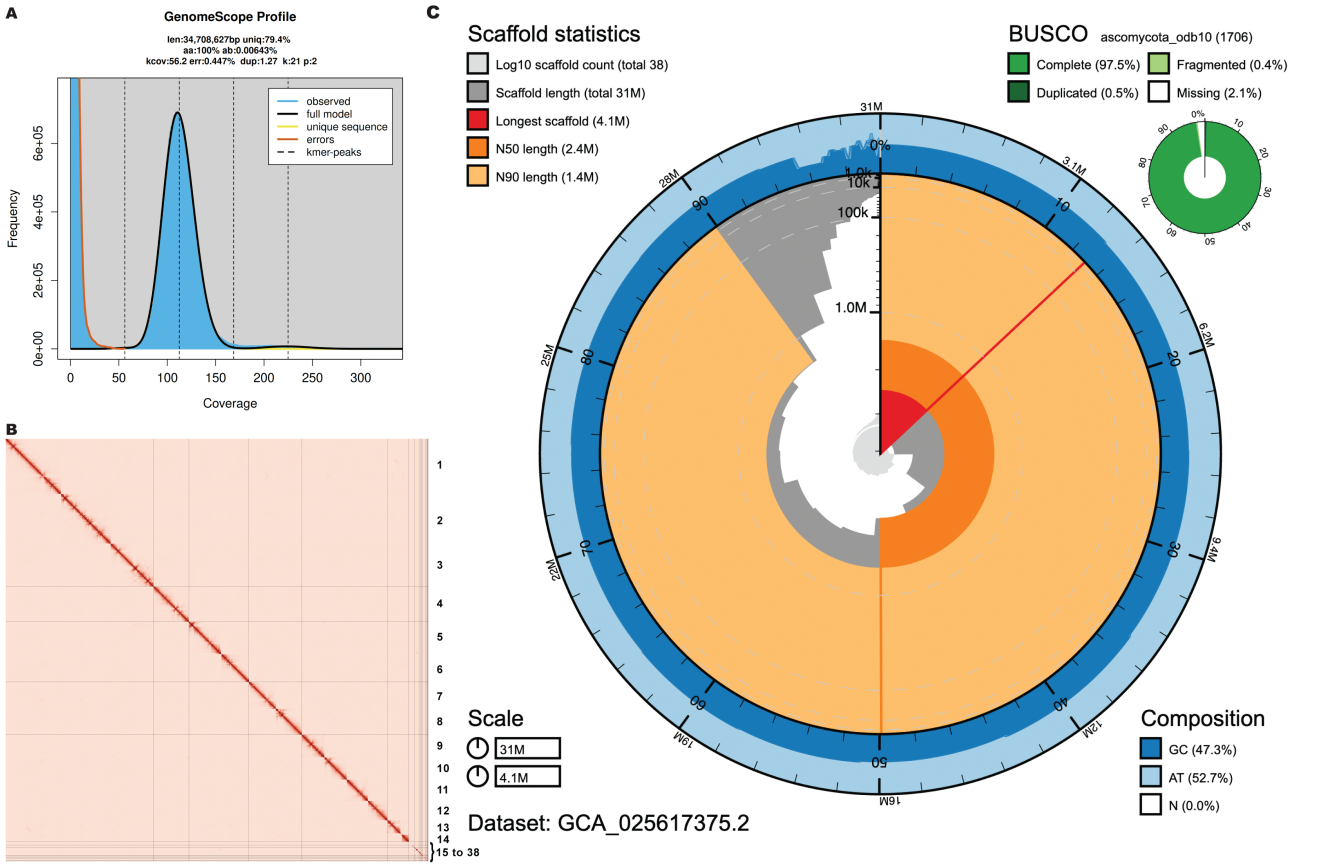

Fig. 2. A) K-mer spectrum. B) Omni-C contact map to give an idea of the level of chromosome assembly. C) BlobToolKit Snailplot showing N50 metrics for *A. socialis* assembly glAcaSoci1 and BUSCO scores for the Ascomycota set of orthologues.

the contigs removed, 722 were identified as bacteria, 11 to chlorophyta, 4 to Planctomycetes and 4 as Firmicutes, 10 as other, and 102 as no-hit. Detailed information on the specific species can be found on Supplementary Tables 1, 2 and Figures. During manual curation we generated a single break in a contig and after mitochondrial genome assembly, we removed 42 contigs that were considered as mitochondrion and replaced with the single MT genome assembly.

The Omni-C contact maps show that the assembly is highly contiguous (Fig. 2C). We have deposited the resulting assembly into NCBI GenBank (see Table 2 and Data Availability for details).

Mitochondrial genome assembly

We assembled a complete mitochondrial (MT) genome for *A. socialis,* which is the first complete MT genome deposited in the family Acarosporaceae. The assembled mitochondrial sequence has a length of 83,434 bp. The base composition of the final assembly version is *A* = 36.5%, *C* = 15.1%, *G* = 12.9%, *T* = 35.5%.

Discussion

We have sequenced the first genome of the species *A. socialis*, representing the initial step toward understanding and conducting further studies on population, landscape, and comparative genomics of this lichen and more broadly on the diversification of the genus *Acarospora*. In addition, we successfully sequenced the lichen mycobiont genome directly from the thallus of *A. socialis* using long-read PacBio HiFi sequencing and Omni-C chromatin capture data, which is one of the first times this methodology has been used to achieve a near chromosome resolution assembly of a lichenized fungi. This approach enabled us to avoid culturing of the mycobiont, which is often slow and/or unsuccessful. Our results show a promising roadmap for generating high-quality mycobiont genomes directly from the thallus. This genome will serve as an important reference in the context of conservation genomics research where lichens, as foundational species in California and the US Southwest, are increasingly under threat from wildfires and drought driven by climate change.

The genome assembly consists of 38 scaffolds spanning 31.2 Mb with contig and scaffold N50 of 2.4 Mb. The most closely related species previously sequenced was the genome of *Acarospora strigata*, which was sequenced via Illumina HiSeq technology from a culture and has an assembled genome of 27 Mb, with 938 scaffolds, and an N50 of 81 kb (McDonald et al. 2013).

Previous published genomes or metagenomes of lichenized fungi were sequenced with short Illumina reads and range in size from 24.2 Mb (*Caloplaca ligustica*) to 59 Mb (*Physcia* cf. *stellaris*) (McDonald et al. 2013; Llewellyn et al. 2023). Greshake Tzovaras et al. (2020) used a metagenome approach to sample the thallus of the lichen *Umbiliaria pustulata* with whole-genome shotgun sequencing combining both Illumina short reads and PacBio long reads. The study found that the genome size of the mycobiont to

Table 2. Sequencing and assembly statistics, and accession numbers.

Bio Projects and Vouchers	CCGP NCBI BioProject		PRJNA720569			
	Genera NCBI BioProject		PRJNA766244			
	Species NCBI BioProject		PRJNA766244			
	NCBI BioSample		SAMN29043977			
	Specimen identification		UCR:JNA_AS_CCGP158			
	NCBI Genome accessions					
	Assembly accession		JAMYGW000000000			
	Genome sequences		GCA_025617375.2			
Genome sequence	PacBio HiFi reads	Run	1 PACBIO_SMRT (Sequel II) run: 694,908 spots, 4.3 G bases, 1.8 Gb			
		Accession	SRR23194840			
	Omni-C Illumina reads	Run	2 ILLUMINA (Illumina NovaSeq 6000) runs: 32.4M spots, 9.8G bases, 3.2 Gb			
		Accession	SRR23194838, SRR23194839			
Genome Assembly Quality Metrics	Assembly identifier (Quality code*)		glAcaSoci1.1(6.6.6.Q59.C98)			
	HiFi Read coverage§		123×			
	Number of contigs		38			
	Contig N50 (bp)		2,402,664			
	Contig NG50§		2,032,088			
	Longest Contigs		4,079,479			
	Number of scaffolds		38			
	Scaffold N50		2,402,664			
	Scaffold NG50§		2,032,088			
	Largest scaffold		4,079,479			
	Size of final assembly (bp)		31,188,520			
	Phased block NG50§		2,032,088			
	Gaps per Gbp (#Gaps)		160 (5)			
	Indel QV (Frame shift)		47.72282579			
	Base pair QV		59.83			
	k-mer completeness		98.01			
	BUSCO completeness (ascomycota) *n* = 1,706	C	S	D	F	M
		97.50%	97.00%	0.50%	0.40%	2.10%
	1 Organelle sequence		CM055765			

*Assembly quality code *x.y.P.Q.C*, where, *x* = log10[contig NG50]; *y* = log10[scaffold NG50]; *P* = log10 [phased block NG50]; *Q* = Phred base accuracy QV (quality value); C = % genome represented by the first "*n*" scaffolds, following a karyotype of *n* = 24, inferred from ancestral taxa (Genome On A Tree, Release 2022.07.21). BUSCO scores. (C)omplete and (S)ingle; (C)omplete and (D)uplicated; (F)ragmented and (M)issing BUSCO genes. *n*, number of BUSCO genes in the set/database; Bp, base pairs.
§Read coverage and NGx statistics have been calculated based on a genome size of 34.7 Mb.

be 33 Mb contained on 43 scaffolds with a scaffold N50 of 1.8 Mb, which is similar to the size, N50, and number of scaffolds of the 31.2 Mb *A. socialis* genome. Long read sequencing applied to both these projects likely is necessary to achieve this degree of continuity. Lichen metagenomes of *Hypogymnia physodes* (58.4 Mb) and *Hypogymnia tubulosa* (39.8 Mb) sequenced with Pacific Biosciences (Ahmad et al. 2023) range in size. The size variation may reflect differences in genomic content but could contain additional symbiont contigs or ploidy. The use of Omni-C scaffolding with PacBio sequencing provides a useful improvement to produce high-quality genomes with long contiguous scaffolds and a high degree of completeness.

The distribution of *A. socialis* is widespread throughout California however the increasing frequency and intensity of wildfires due to climate change has eradicated some populations in local areas. Lichens are extremely sensitive to wildfires (Romagni and Gries 1997), which are periodical, but increasingly more common in California. Increased fire frequency negatively impacts long-lived and slow growing

lichens, as it can take decades for lichens to recolonize an area after intense fires (Miller et al. 2022). These factors make it important to document the genomes of *A. socialis* for conservation and biodiversity studies before increased fire frequencies lead to the disappearance of this common and foundational lichen species in local areas of California. This reference genome will help to identify traits associated with drought tolerance and fire resistance in this species. This genomic resource fulfills the goals of the CCGP and will enable conservation scientists and policymakers to make decisions and targeted conservation plans that can mitigate the impact of climate change on *A. socialis*.

Supplementary material

Supplementary material is available at *Journal of Heredity* online.

Acknowledgments

PacBio Sequel II library prep and sequencing was carried out at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01. Deep sequencing of Omni-C libraries used the Novaseq S4 sequencing platforms at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant. We thank the staff at the UC Davis DNA Technologies and Expression Analysis Cores and the UC Santa Cruz Paleogenomics Laboratory for their diligence and dedication to generating high-quality sequence data. We thank R.W. Adams for help in the field to collect *A. socialis* from Pinkham Canyon with J.N. Adams and for taking the photos in Figure 1. We also thank K. Knudsen for assistance in identifying field sites at Pinkham Canyon in Joshua Tree National Park for J.N. Adams. We also thank T. Salvato for help in depositing the UCR voucher (#295874) of *A. socialis* from Pinkham Canyon into the UCR Herbarium. T. Salvato and A. Sanders also helped to identify the plants in Fig. 1B at Pinkham Canyon. We also thank Dr. C. Ettinger, Dr. M. Dal Forno, M. Yacoub, S. Sharma, and two anonymous reviewers for comments and suggestions on this manuscript. *Acarospora socialis* was collected with the following permit number (JOTR-2017-SCI-0040) at Joshua Tree National Park. Computations were performed using the computer clusters and data storage resources of the UC Riverside High Performance Comptational Cluster, which were funded by grants from NSF (MRI-2215705, MRI-1429826) and NIH (1S10OD016290-01A1).

Funding

This work was supported by the California Conservation Genomics Project, with funding provided to the University of California by the State of California, State Budget Act of 2019 [UC Award ID RSI-19-690224]. J.N. Adams was supported by a National Science Foundation (NSF) Graduate Research Fellowship and a Lewis and Clark grant from the American Philosophical Society. J.E.S. is a CIFAR Fellow in the program Fungal Kingdom: Threats and Opportunities and supported by U.S. Department of Agriculture, National Institute of Food and Agriculture Hatch projects CA-R-PPA-211-5062-H.

Data Availability

Data generated for this study are available under NCBI BioProject PRJNA720569. Raw sequencing data for sample UCR:JNA_AS_CCGP158 (NCBI BioSample SAMN29043977) are deposited in the NCBI Sequence Read Archive (SRA) under SRR23194838-SRR23194840. Assembly scripts and other data for the analyses presented can be found at the following GitHub repository: github.com/ccgproject/ccgp_assembly.

References

Abdennur N, Mirny LA. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics*. 2020:36:311–316.

Ahmad N, Calchera A, Otte J, Schmitt I, Brueck T, Mehlmer N. Biosynthetic potential of *Hypogymnia* holobionts: insights into secondary metabolite pathways. *J Fungi*. 2023:9:546.

Allen JL, Lendemer JC. Climate change impacts on endemic, high elevation lichens in a biodiversity hotspot. *Biodivers Conserv*. 2016:25:555–568.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990:215:403–410.

Arnold AE, Miadlikowska J, Higgins KL, Sarvate SD, Gugger P, Way A, Hofstetter V, Kauff F, Lutzoni F. A phylogenetic estimation of trophic transition networks for Ascomycetous fungi: are lichens cradles of symbiotrophic fungal diversification? *Syst Biol*. 2009:58:283–297.

Bates ST, Cropsey GWG, Caporaso JG, Knight R, Fierer N. Bacterial communities associated with the lichen symbiosis. *Appl Environ Microbiol*. 2011:77:1309–1314.

Büdel B. Taxonomy of lichenized procaryotic blue–green algae. In: Reisser W, editor. *Algae and symbioses*. Bristol, UK: Biopress Limited; 1992. p. 301–324.

Cardinale M, Vieira de Castro J Jr, Müller H, Berg G, Grube M. In situ analysis of the bacterial community associated with the reindeer lichen *Cladonia arbuscula* reveals predominance of Alphaproteobacteria. *FEMS Microbiol Ecol*. 2008:66:63–71.

Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. BlobToolKit—interactive quality assessment of genome assemblies. *G3 Genes Genomes Genet*. 2020:10:1361–1374.

Cheng H, Jarvis ED, Fedrigo O, Koepfli KP, Urban L, Gemmell NJ, Li H. Haplotype-resolved assembly of diploid genomes without parental data. *Nat Biotechnol*. 2022:40:1332–1335.

Conti ME, Cecchetti G. Biological monitoring: lichens as bioindicators of air pollution assessment—a review. *Environ Pollut*. 2001:114:471–492.

Diederich P, Lawrey JD, Ertz D. The 2018 classification and checklist of lichenicolous fungi, with 2000 non-lichenized, obligately lichenicolous taxa. *Bryologist*. 2018:121:340–425.

Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res*. 2017:45:e18.

Duewer EAH. *The taxonomy and morphology of the lichen genus Acarospora and its phycosymbiont Trebouxia [Doctoral dissertation Thesis]*. The University of Arizona; 1971.

Gärtner G. Taxonomy of symbiotic eukaryotic algae. In: Reisser W, editor. *Algae and symbioses*. Bristol, UK: Biopress Limited; 1992. p. 325–338.

Ghurye J, Pop M, Koren S, Bickhart D, Chin CS. Scaffolding of long read assemblies using long range contact information. *BMC Genom*. 2017:18:527.

Ghurye J, Rhie A, Walenz BP, Schmitt A, Selvaraj S, Pop M, Phillippy AM, Koren S. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Comput Biol*. 2019:15:e1007273.

Goloborodko A, Abdennur N, Venev S, Brandão HB, Fudenberg G. 2018. *mirnylab/pairtools: v0.2.0*. https://zenodo.org/record/1490831 (2018 November 18)

Greshake Tzovaras B, Segers FHID, Bicker A, Dal Grande F, Otte J, Anvar SY, Hankeln T, Schmitt I, Ebersberger I. What is in *Umbilicaria pustulata*? A metagenomic approach to reconstruct the holo-genome of a lichen. *Genome Biol Evol*. 2020:12:309–324.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013:29:1072–1075.

Hawksworth DL. Lichens as litmus for air pollution: a historical review. *Int J Environ Stud*. 1970:1:281–296.

Hawksworth DL, Grube M. Lichens redefined as complex ecosystems. *New Phytol*. 2020:227:1281–1283.

Holder JM, Wynn-Williams DD, Rull Perez F, Edwards HGM. Raman spectroscopy of pigments and oxalates in situ within epilithic lichens: *Acarospora* from the Antarctic and Mediterranean. *New Phytol*. 2000:145:271–280.

Honegger R. Lichen-forming fungi and their photobionts. In: Deising HB, editor. *Plant relationships*. *The Mycota*, vol 5. Berlin, Heidelberg: Springer; 2009. p. 307–333.

Huang X, Madan A. CAP3: a DNA sequence assembly program. *Genome Res*. 1999:9:868–877.

Inglis PW, Pappas MCR, Resende LV, Grattapaglia D. Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLoS One*. 2018:13:e0206085.

Jovan S. 2008. *Lichen bioindication of biodiversity, air quality, and climate: baseline results from monitoring in Washington, Oregon, and California*. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Jovan S, Mccune B. Using epiphytic macrolichen communities for biomonitoring ammonia in forests of the greater Sierra Nevada, California. *Water Air Soil Pollut*. 2006:170:69–93.

Kappen L. Some aspects of the great success of lichens in Antarctica. *Antarct Sci*. 2000:12:314–324.

Kerpedjiev P, Abdennur N, Lekschas F, McCallum C, Dinkla K, Strobelt H, Luber JM, Ouellette SB, Azhir A, Kumar N, et al. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol*. 2018:19:125.

Knudsen K. *Acarospora*. In: Nash TH, Gries C, Bungartz F, editors. *Lichen flora of the greater Sonoran Desert region, vol 3, (balance of the microlichens, and the lichenicolous fungi)*. Tempe: Lichens Unlimited, Arizona State University; 2007. p. 1–38.

Korlach J, Gedman G, Kingan SB, Chin CS, Howard JT, Audet JN, Cantin L, Jarvis ED. De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. *GigaScience*. 2017:6:1–16.

Kranner I, Lutzoni F. Evolutionary consequences of transition to a lichen symbiotic state and physiological adaptation to oxidative damage associated with poikilohydry. In: Lerner HR, editor. *Plant responses to environmental stresses*. New York: Routledge; 1999. p. 591–628. https://www.taylorfrancis.com/books/edit/10.1201/9780203743157/plant-responses-environmental-stresses-lerner?refId=4703843a-b5e5-4361-bc29-7340dea4bacb&context=ubx

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, arXiv [q-bio.GN]. http://arxiv.org/abs/1303.3997, preprint: not peer reviewed.

Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018:34:3094–3100.

Llewellyn T, Nowell RW, Aproot A, Temina M, Prescott TAK, Barraclough TG, Gaya E. Metagenomics shines light on the evolution of "sunscreen" pigment metabolism in the Teloschistales (Lichen-forming Ascomycota). *Genome Biol Evol*. 2023:15:evad002.

Lücking R, Hodkinson BP, Leavitt SD. Corrections and amendments to the 2016 classification of lichenized fungi in the Ascomycota and Basidiomycota. *Bryologist*. 2017:120:58–69.

Lutzoni F, Miadlikowska J. Lichens. *Curr Biol*. 2009:19:R502–R503.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021:38:4647–4654.

McDonald TR, Mueller O, Dietrich FS, Lutzoni F. High-throughput genome sequencing of lichenizing fungi to assess gene loss in the ammonium transporter/ammonia permease gene family. *BMC Genom*. 2013:14:225. https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-14-225#citeas

Merges D, Dal Grande F, Greve C, Otte J, Schmitt I. Virus diversity in metagenomes of a lichen symbiosis (*Umbilicaria phaea*): complete viral genomes, putative hosts and elevational distributions. *Environ Microbiol*. 2021:23:6637–6650.

Miller JED, Weill AM, Villella J. Epiphytic macrolichen communities take decades to recover after high severity wildfire in chaparral shrublands. *Divers Distrib*. 2022:28:454–462.

Nash TH III. 2008. *Lichen biology*. Cambridge: Cambridge University Press.

Nelson MP, Leavitt SD, Heller K, Muggia L, Lumbsch HT. Contrasting patterns of climatic niche divergence in *Trebouxia*—a clade of lichen-forming algae. *Front Microbiol*. 2022:13:791546.

Nimis PL, Scheidegger C, Wolseley PA. *Monitoring with lichens—monitoring lichens*. Dordrecht: Springer Netherlands; 2002. p. 1–4.

Pflug JM, Holmes VR, Burrus C, Johnston JS, Maddison DR. Measuring genome sizes using read-depth, k-mers, and flow cytometry: methodological comparisons in beetles (Coleoptera). *G3 (Bethesda)*. 2020:10:3047–3060.

Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A, Manke T. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun*. 2018:9:189.

Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020:11:1432.

Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J, et al. Towards complete and error-free genome assemblies of all vertebrate species. *Nature*. 2021:592:737–746.

Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol*. 2020:21:245.

Romagni J, Gries C. Assessment of fire damage to epiphytic lichens in southeastern Arizona. *Bryologist*. 1997:100:102–108.

Sancho LG, Pintado A. Evidence of high annual growth rate for lichens in the maritime Antarctic. *Polar Biol*. 2004:27:312–319.

Sim SB, Corpuz RL, Simmonds TJ, Geib SM. HiFiAdapterFilt, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly. *BMC Genom*. 2022:23:157.

Sipman HJM, Harris RC. Lichens. In: Lieth H, Werger MJA, editors. *Tropical rain forest ecosystems*. Amsterdam: Elsevier; 1989. p. 303–309.

Spribille T, Tuovinen V, Resl P, Vanderpool D, Wolinski H, Aime MC, Schneider K, Stabentheiner E, Toome-Heller M, Thor G, et al. Basidiomycete yeasts in the cortex of ascomycete macrolichens. *Science*. 2016:353:488–492.

Stajich JE, Palmer JM. *AAFTF—automatic assembly for the fungi*. 2019. https://zenodo.org/record/7439973 (2022 December 15)

Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol*. 2021:38:3022–3027.

Tschermak-Woess E. The algal partner. In: Galun M, editor. *Handbook of Lichenology*. Boca Raton: CRC Press; 1988. p. 39–92.

U'Ren JM, Lutzoni F, Miadlikowska J, Laetsch AD, Arnold AE. Host and geographic structure of endophytic and endolichenic fungi at a continental scale. *Am J Bot*. 2012:99:898–914.

Workman R, Timp W, Fedak R, Kilburn D, Hao S, Liu K. 2018. High molecular weight DNA extraction from recalcitrant plant species for third generation sequencing. Protocols.io. doi:10.17504/protocols.io.4vbgw2n