



OPEN ACCESS

EDITED BY

José A. Siles,
Center for Edaphology and Applied Biology of
Segura (CSIC),
Spain

REVIEWED BY

Yuriy L. Orlov,
I.M. Sechenov First Moscow State Medical
University,
Russia

Adriana Lucia Romero-Olivares,
New Mexico State University,
United States

*CORRESPONDENCE

Dominik Merges
✉ mergesd01@gmail.com

SPECIALTY SECTION

This article was submitted to
Terrestrial Microbiology,
a section of the journal
Frontiers in Microbiology

RECEIVED 14 November 2022

ACCEPTED 23 February 2023

PUBLISHED 22 March 2023

CITATION

Merges D, Dal Grande F, Valim H, Singh G and
Schmitt I (2023) Gene abundance linked to
climate zone: Parallel evolution of gene
content along elevation gradients in lichenized
fungi.

Front. Microbiol. 14:1097787.

doi: 10.3389/fmicb.2023.1097787

COPYRIGHT

© 2023 Merges, Dal Grande, Valim, Singh and
Schmitt. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Gene abundance linked to climate zone: Parallel evolution of gene content along elevation gradients in lichenized fungi

Dominik Merges^{1,2,3*}, Francesco Dal Grande^{1,2,4,5},
Henrique Valim^{1,2}, Garima Singh^{1,2,4} and Imke Schmitt^{1,2,6}

¹Senckenberg Biodiversity and Climate Research Centre, Frankfurt am Main, Germany, ²LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Frankfurt am Main, Germany, ³Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, Uppsala, Sweden, ⁴Department of Biology, University of Padova, Padua, Italy, ⁵National Biodiversity Future Center (NBFC), Palermo, Italy, ⁶Goethe University Frankfurt, Institute of Ecology, Evolution and Diversity, Frankfurt am Main, Germany

Introduction: Intraspecific genomic variability affects a species' adaptive potential toward climatic conditions. Variation in gene content across populations and environments may point at genomic adaptations to specific environments. The lichen symbiosis, a stable association of fungal and photobiont partners, offers an excellent system to study environmentally driven gene content variation. Many of these species have remarkable environmental tolerances, and often form populations across different climate zones. Here, we combine comparative and population genomics to assess the presence and absence of genes in high and low elevation genomes of two lichenized fungi of the genus *Umbilicaria*.

Methods: The two species have non-overlapping ranges, but occupy similar climatic niches in North America (*U. phaea*) and Europe (*U. pustulata*): high elevation populations are located in the cold temperate zone and low elevation populations in the Mediterranean zone. We assessed gene content variation along replicated elevation gradients in each of the two species, based on a total of 2050 individuals across 26 populations. Specifically, we assessed shared orthologs across species within the same climate zone, and tracked, which genes increase or decrease in abundance within populations along elevation.

Results: In total, we found 16 orthogroups with shared orthologous genes in genomes at low elevation and 13 at high elevation. Coverage analysis revealed one ortholog that is exclusive to genomes at low elevation. Conserved domain search revealed domains common to the protein kinase superfamily. We traced the discovered ortholog in populations along five replicated elevation gradients on both continents and found that the number of this protein kinase gene linearly declined in abundance with increasing elevation, and was absent in the highest populations.

Discussion: We consider the parallel loss of an ortholog in two species and in two geographic settings a rare find, and a step forward in understanding the genomic underpinnings of climatic tolerances in lichenized fungi. In addition, the tracking of gene content variation provides a widely applicable framework for retrieving biogeographical determinants of gene presence/absence patterns. Our work provides insights into gene content variation of lichenized fungi in relation to climatic gradients, suggesting a new research direction with implications for understanding evolutionary trajectories of complex symbioses in relation to climatic change.

KEYWORDS

environmental change, gene presences/absence, gene loss/gain, metagenomics, convergent evolution, population genomics, PacBio, protein kinase

Introduction

Intraspecific genomic variability substantially affects a species' adaptive potential to ecological interactions and climatic conditions (Hirsch et al., 2014; Badet et al., 2020; Resl et al., 2022). Variation in genomic content at the gene-level, i.e., presence/absence of genes, within species are regularly found in bacteria, and are often associated with adaptations to specific environments and antibiotic resistances (Masignani et al., 2005; Heuer et al., 2011). Recent studies suggest that eukaryotic species, like bacterial ones, can have intraspecific variation in genomic content at the gene-level (Hirsch et al., 2014; Badet et al., 2020; Gerdol et al., 2020). To this date, it is largely unknown how relevant such variation in gene content is for eukaryotes, and if it is ecologically important for adaptations to specific environments. Intraspecific variation in gene content contributes to the formation of genetically diverse populations, and therefore its characterization is vital to understand the mechanisms of adaptation to specific environments (Plissonneau et al., 2018; Drott et al., 2021; Merges et al., 2022).

The lichen symbiosis, a stable association of mainly fungal and photobiont partners as well as an associated microbiome, lends itself to the study of environmentally-driven gene content, because many species have remarkable environmental tolerances and maintain populations in different climate zones (Kappen, 2000; Werth and Sork, 2014; Singh et al., 2017; Grimm et al., 2021; Jung et al., 2021; Tanunchai et al., 2022). Previous studies showed that environmental differentiation in lichens can be found at the level of single nucleotide polymorphisms (SNPs), which often significantly correlate with differences in geography and ecology and may thus be involved in environmental specialization (Peksa and Skaloud, 2011; Castillo et al., 2012; Hodkinson et al., 2012). Population genomic analyses based on SNPs suggest the presence of genome-wide differentiation between populations in different climate zones (Dal Grande et al., 2017; Rolshausen et al., 2022). To date, the only study which assesses variation in gene content associated with environmental adaptation in lichens is limited to a single species: Singh et al. (2021) show that some gene clusters associated with natural product biosynthesis in *Umbilicaria pustulata* have elevation-specific distributions. However, we currently lack information on the extent of parallel evolution in populations of different species that have independently evolved under similar environmental conditions. We do not know, e.g., (1) whether different species of lichen-forming fungi maintain homologous population-specific genes along comparable environmental gradients and (2) whether intraspecific variation in gene content (e.g., abundance patterns of genes) has independently converged on the same altitudinal patterns in different species, and can thus be linked to environmental preferences of lichens.

Modern evolutionary approaches leverage DNA sequencing to infer ecological and evolutionary processes that occur at the population level. Here we combine comparative genomics and population genomics to assess the presence/absence of genes in high

elevation and low elevation genomes of two lichenized fungi of the genus *Umbilicaria* (*Umbilicaria phaea* and *U. pustulata*). The two species have evolved on different continents under similar climatic selective pressures: *U. pustulata* in Europe and *U. phaea* in North America each occupy the cold temperate as well as the Mediterranean climate zone. We tracked gene content variation along replicated elevation gradients in both species based on a total of 2050 individuals in 26 populations. Specifically, we addressed the following research questions: (a) Which genes are linked to environmental conditions at high elevation (cold temperate climate) and low elevation (Mediterranean climate) across species? To address this question, we assessed which genes are exclusive to climate zones. (b) Do abundances of genes specific to a climate zone co-vary with elevation in populations of *U. phaea* and *U. pustulata*? To address the second question, we assessed which genes increase or decline in abundance with increasing elevation.

Materials and methods

Study site and sample collection

We sampled 15 *U. pustulata* populations along three elevational gradients in Spain and Italy and 11 *U. phaea* populations along two elevational gradients in California, United States (Merges et al., 2021; Singh et al., 2022). Our choice of gradients was based on the high abundance of *U. pustulata* and *U. phaea* in these respective areas, and because the gradients span two contrasting bioclimates, the Mediterranean and the cold temperate zones. A population of lichens is here defined as a group of individuals collected on rocks within an area of approximately 10 × 10 m. Detailed sampling procedures are described in Dal Grande et al. (2017) and Rolshausen et al. (2020). The altitudinal spacing of populations along the gradients can be seen in Supplementary Table S1. The studied *Umbilicaria* species entail foliose, monophyllous lichens, which are attached to rock surfaces with a central holdfast. This growth form facilitates the recognition of individuals. Details of the sampling of European material are described in Singh et al. (2022) and the details of sampling of the North American material in Dal Grande et al. (2017) and Merges et al. (2021). Briefly, two of the European gradients are located in Central Spain, Sierra de Gredos (40.2028, -5.2334 and 39.9946, -4.8679) and one on the island of Sardinia (40.7577, 9.0794). We collected fragments of 100 individuals each, at Mount Limbara (Sardinia, Italy; 6 populations), Sierra de Gredos (Sistema Central, Spain; 6 populations) and Talavera-Puerto de Pico (Sistema Central, Spain; 3 populations), as described in Dal Grande et al. (2017). The Californian gradients are spatially separated by approx. 700 km. We collected fragments of 50 individuals each, at four populations along the Sierra Nevada gradient (38.084, -120.484) and at seven population along the Mt. Jacinto gradient (33.435, -116.484). Schemes indicating the geographic location and sites of the sample collection are given in

Supplementary Table S1. We additionally collected four whole lichen thalli, one low-altitude individual from the Sierra Nevada population and one from the high-altitude population, as well as a low-altitude and a high-altitude individual from populations of the Sierra des Gredos gradient for the reconstruction of reference genomes using PacBio Sequel II data (Merges et al., 2021; Singh et al., 2022; Supplementary Table S2).

DNA extraction for population pooled sequencing

Genomic DNA was extracted separately from each fragment from all populations using a cetyltrimethylammonium bromide-based (CTAB) method (Cubero and Crespo, 2002; Dal Grande et al., 2017). Further, we created a pooled sample for each population containing equal amounts of DNA from each sample and Novogene Co., Ltd. (Cambridge, United Kingdom) performed the library preparation (200–300 bp insert size; Dal Grande et al., 2017). Libraries were sequenced on an Illumina HiSeq2000 with 150 bp paired-end chemistry at ~90× coverage per population (i.e., Pool-seq; Dal Grande et al., 2017).

DNA extraction for genomic sequencing

Genomic DNA for genome sequencing was extracted from dry thallus material of two samples of the same species (i.e., *U. phaea* or *U. pustulata*) collected in different climatic zones (i.e., low elevation/Mediterranean climate zone and high elevation/temperate climate zone). Lichen thalli were thoroughly washed with sterile water and checked under the stereomicroscope for the presence of possible contamination or other lichen thalli. DNA was extracted from all of the samples using CTAB-based method (Mayjonade et al., 2016) as presented in Merges et al. (2021).

PacBio library preparation and sequencing

For PacBio single-molecule real-time (SMRT) sequencing, SMRTbell libraries were constructed according to the manufacturer's instructions of the SMRTbell Express Prep kit v2.0 following the Low DNA Input Protocol (Pacific Biosciences, Menlo Park, CA, United States). Total input DNA was approximately 140 and 800 ng, respectively. Ligation with T-overhang SMRTbell adapters was performed at 20°C overnight. Following ligation, the SMRTbell library was purified with an AMPure PB bead clean up step with 0.45X volume of AMPure PB beads. Subsequently a size-selection step with AMPure PB Beads was performed to remove short SMRTbell templates <3 kb. For this purpose, the AMPure PB beads stock solution was diluted with elution buffer (40% volume/volume) and then added to the DNA sample with 2.2X volume. The size and concentration of the final libraries were assessed using the TapeStation (Agilent Technologies) and the Qubit Fluorometer with Qubit dsDNA HS reagents Assay kit (Thermo Fisher Scientific, Waltham, MA, United States). Sequencing primer v4 and Sequel® II Polymerase 2.0 were annealed and bound, respectively, to each

SMRTbell library. SMRT sequencing was performed on the Sequel System II with Sequel II Sequencing Kit 2.0 in “continuous long read” (i.e., CLR) mode, 30 h movie time with no pre-extension and Software SMRTLINK 8.0 (Pacific Biosciences of California, 2022). One SMRT Cell was run for each sample. A SMRT cell contains millions of wells called zero-mode waveguides (ZMWs; Pacific Biosciences of California, 2022). Within each ZMWs single molecules of DNA are immobilized and as the polymerase incorporates each nucleotide, light is emitted, and nucleotide incorporation is measured in real time (Pacific Biosciences of California, 2022).

De novo assembly of PacBio metagenomic sequence reads

We largely followed the pipeline described in Merges et al. (2021). In summary, we generated HiFi reads from the PacBio Sequel II run using the PacBio tool CCS v5.0.0 with default parameters, i.e., --min-passes 3, remove subreads with lengths <50% or >200% of the median subread length, --max-insertion-size to 30 bp (Pacific Biosciences of California, 2022).¹ Metagenomic sequence reads were assembled into contigs using the long-read based assembler metaFlye v2.7 (Kolmogorov et al., 2019). The assembled contigs were scaffolded with LRScaf v1.1.12 (Qin et al., 2019).² To retrieve the mycobiont genome, the received scaffolds were taxonomically binned *via* blastx using DIAMOND (--more-sensitive --frameshift 15 --range-culling) on a custom database (Singh et al., 2022) and the Metagenome Analyzer MEGAN6 Community Edition pipeline (Huson et al., 2007; Buchfink et al., 2014). The completeness of the genomes represented by the binned Ascomycota scaffolds was estimated using Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis in BUSCO v4 using the Ascomycota dataset (Simão et al., 2015).

Pool-seq data processing

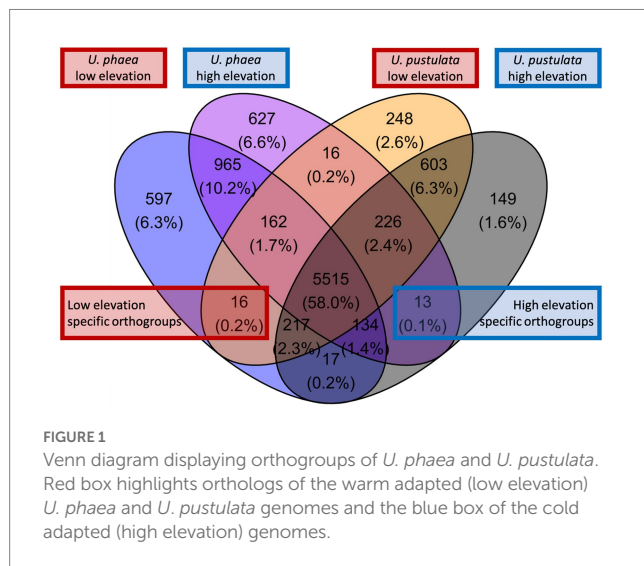
We filtered the pool-seq data for reads shorter than 80 bp, reads with N's, and reads with average base quality scores less than 26 along with their pairs, and discarded them. We mapped the trimmed paired-end reads of each pool to the database of the identified genes using bowtie2 v2.4.1 (Langmead and Salzberg, 2012), using the flags: --very-sensitive-local, --no-mixed, --no-unal, --no-discordant.

Gene prediction and genome annotation

Functional annotation of genomes, including genes and proteins (antiSMASH; antibiotics and SM Analysis Shell, v5.0) was performed with scripts implemented in the funannotate pipeline (Blin et al., 2017; Palmer and Stajich, 2019). First, the genomes were masked for

1 <https://ccs.how>

2 <https://github.com/shingocat/lrscaf>



repetitive elements, and then the gene prediction was performed using BUSCO2 to train Augustus and self-training GeneMark-ES (Borodovsky and Lomsadze, 2011; Simão et al., 2015). Functional annotation was done with InterProScan (Quevillon et al., 2005), egg-NOG-mapper (Huerta-Cepas et al., 2017, 2019), and evolutionarily-informed expectations of gene content of near-universal single-copy orthologs using BUSCO v 5.1.2 with included Ascomycota dataset (Simão et al., 2015). Secreted proteins were predicted using SignalP (Armenteros et al., 2019) as implemented in the funannotate “annotate” command. Proteins were further characterized by NCBI conserved domain search (Lu et al., 2020).³

Assessing gene content variation in the assembled fungal genomes

To identify presence/absence patterns of genes, we identified orthologs using orthoFinder (Emms and Kelly, 2015, 2019). OrthoFinder provides the most accurate ortholog inference method on the Quest for Orthologs benchmark test (Emms and Kelly, 2015, 2019). In orthoFinder (v.2.5.4), we assigned all genes to orthogroups using protein homology and constructed a pangenome of all four complete genomes (Badet et al., 2020). Shared orthologs (i.e., members of the some orthogroup) of low elevation (warm adapted) and high elevational (cold adapted) genomes were extracted using R v3.6.1 (R Core Team, 2019).

Validating presence/absence of genes at population level

To validate population-level gene presence or absence, we estimated the abundance of each ortholog in the low elevation (warm adapted) and high elevation (cold adapted) population based

on the median coverage of pool-seq reads associated to each ortholog contig. Specifically, we used samtools (v1.15) depth to estimate the coverage of each basepair within the contig (Danecek et al., 2021). We assessed and visualized the data in R v3.6.1 (R Core Team, 2019).

Gene distribution across *Umbilicaria* populations

Bowtie2 (v2.2.2) was used to map pool-seq reads to all ortholog contigs (using default settings). The number of mapped reads was counted per sample and normalized by dividing the number of mapped reads by the total read number of the respective sample to account for differences in sequencing depth. We modeled gene abundance (i.e., normalized read count) as a function of elevation using linear models. Linear models were fitted and plotted in R v3.6.1 (R Core Team, 2019).

Results

HiFi metagenomic sequencing reads of mycobiont

We reconstructed metagenomic sequences from a low-elevation and a high-elevation specimen of *U. pustulata* and *U. phaea*. Sequence output and quality for *U. pustulata* were summarized in Singh et al. (2022), for *U. phaea* in Merges et al. (2021) and in Supplementary Table S2.

Altitude-specific genes in the *de novo* assembled genomes

We screened the *de-novo* assembled genomes of the low- and high-altitude samples for altitude-specific genes (Figure 1). Orthofinder revealed 16 orthogroups with shared orthologous genes (0.2% of the total orthogroups) in low-elevation genomes and 13 in high-elevational genomes (0.1% of the total orthogroups).

Presence/absence of genes at population level

To verify the presence/absence of detected orthologs, the coverage of each ortholog was calculated for the respective population at low and high elevation. The coverage analysis revealed one ortholog present in the genomes at low elevation to be consistently absent in populations at high elevation (Figure 2). The amino acid sequences of the ortholog could not be functionally annotated using the funannotate pipeline and was classified as “hypothetical protein.” NCBI’s conserved domain search revealed an alignment with the catalytic domain of protein kinase superfamily member PKc cd00180 (Position-specific scoring matrix (PSSM) accession cl214531, NCBI Conserved Domain Database) as well as seven Tetratricopeptide repeats (Tetratricopeptide-like helical domain superfamily, InterPro entry IPR011990), indicating putative protein binding surfaces (Supplementary Figure S1).

³ <https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>

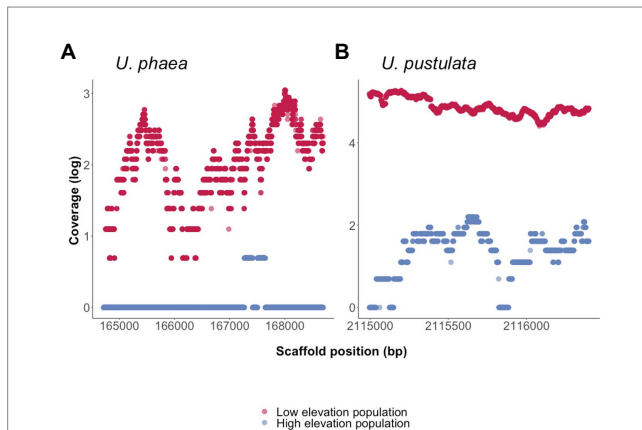


FIGURE 2

Presence/absences of orthologs in four *de novo* sequenced genomes of *U. phaea* and *U. pustulata* were verified by assessing the scaffold coverage in the warm adapted (low elevation) and the cold adapted population (high elevation) respectively. (A) Coverage of ortholog in *U. phaea*: High coverage in warm adapted (low elevation) population and no coverage in high elevational population. (B) Coverage of ortholog in *U. pustulata*: High coverage in warm adapted (low elevation) population.

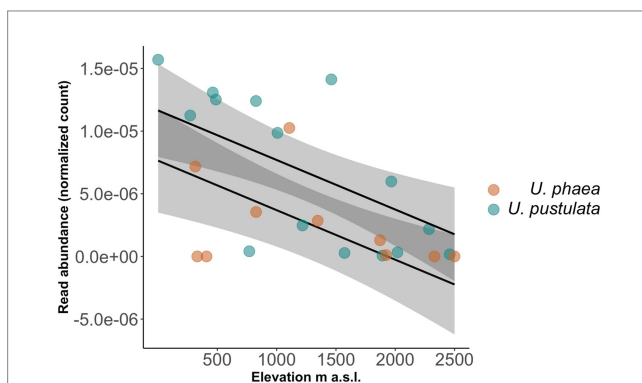


FIGURE 3

Read abundance of identified ortholog decreases significantly with increasing elevation across *U. phaea* (brown circles, lower regression line) and *U. pustulata* (blue circles, upper regression line) populations.

Gene abundance distributions along gradients

The normalized read number of the identified orthologs, annotated as members of the Protein Kinases superfamily, showed a decline with increasing elevation across all populations in both species (value of $p = 0.00373$, Figure 3).

Discussion

Although adaptations to environmental gradients may lead to variation in gene content, assessments of gene presence/absence patterns across populations and species of lichenized fungi are still missing. Here we assess signatures of parallel evolution at the level of gene presence and absence in lichenized fungi of the genus

Umbilicaria, and trace the discovered genes in lichen populations along five replicated elevation gradients across two continents. While our whole genome comparison based on four *de novo* sequenced specimen (two per species) suggested up to 29 elevation-specific orthogroups (i.e., 16 low elevation-specific and 13 high elevation-specific orthogroups, Figure 1), the population-level verification approach showed only one gene, putatively encoding a protein kinase (PK), which linearly declined in abundance with increasing elevation, and was truly absent in the highest population. This suggests either high strain-specificity of certain genes, or high false positive recovery of gene presence/absence patterns when relying on comparative genomics approaches based on only a few individuals. Thereby extrapolating the significance of gene content variation assessed only with comparative genomics approaches on a few individuals can be potentially misleading when interpreting the evolutionary significance of the variation at population level. Regarding the PK gene consistently absent in high elevation genomes and populations, we found that the discovered gene declines linearly across all populations, suggesting an evolutionary benefit only at lower altitudes. Alternatively, the loss of the gene at higher elevations might benefit individuals in cold climates. To our knowledge, we report for the first time parallel gene presence and absence patterns correlating with climatic niches in different species of lichenized fungi. However, it remains to be analyzed, if the identified molecular trait is associated to a particular phenotype that can be associated with an adaptive function.

In bacteria variation in gene content is assumed to be driven by selection for environmental conditions that are relatively rare across the entire range of a species (Qi et al., 2017). Recent evidence suggests that specific populations of lichenized fungi may contain unique biosynthetic gene clusters (Singh et al., 2021), and our current findings show that also other genes can be elevation-specific. The gradual gene loss across populations with increasing elevation may suggest a decline of selective benefit and may indicate that certain variations in gene content could be of functional importance for local adaptation and climatic tolerances in lichenized fungi. The conserved domain search revealed a catalytic domain of a PK, a common eukaryotic protein superfamily. PKs selectively modify other proteins by phosphorylation, changing their enzymatic activity, cellular location and association with other proteins (Cheng et al., 2002; Hanks, 2003; Asano et al., 2005; Heinisch and Rodicio, 2018). Within a genome, PKs are encoded by a large multigene family with genes being distributed among multiple chromosomes. Putatively, the high number of PK genes has arisen by genome segmental duplication events (Asano et al., 2005; Heinisch and Rodicio, 2018). In our study, the presence/absence of a single PK gene may suggest a climate-specific ancestral genome segmental duplication event. Across the tree of life, e.g., in bacteria (Christ and Chin, 2008), fungi (James et al., 2008) and plants (DeBolt, 2010; Prunier et al., 2019), organisms subjected to selection under high temperatures show higher probability for genome segmental duplications (Christ and Chin, 2008; James et al., 2008; DeBolt, 2010; Kondrashov, 2012). For example, adaptation to heat stress through gene duplication of stress-related genes has been shown in *Escherichia coli* (Riehle et al., 2001; Kondrashov, 2012), where an upregulation through gene duplication of genes may play a role in adaptation (Kondrashov, 2012; Kuzmin et al., 2022). However, due to the scarcity of functional annotation

of non-model organisms and the resulting lack of in-depth functional annotation of the gene in question, the mechanisms generating such population level diversification and the underlying molecular mechanisms behind such adaptations are yet to be understood.

While environmental adaptations are commonly highly polygenic (Rivas et al., 2018; Barghi et al., 2019; Hartke et al., 2021; Pfenninger et al., 2021), there is increasing evidence of the effect of single gene content variation (Liu et al., 2021). For example, as has been recently shown in agave, where a single gene encoding a phosphoenolpyruvate carboxylase enhances the plant's climate resilience (Liu et al., 2021). Not only the gain of genes, but also the loss of genes has been associated with adaptive traits, such as the evolution of particular diets in bats (Blumer et al., 2022). Therefore, we consider the parallel loss of a homologous gene in two species and two geographic settings a rare find, and a step forward in understanding the genomic underpinnings of climatic tolerances in lichenized fungi. Future research should address the functional importance of the gene present at low altitude in the Mediterranean climate zone, and specifically explore the effects of variation in gene abundances across populations. Additionally, future research should consider using heterologous expression approaches to reveal whether the gene presence could induce tolerances to warm conditions.

Conclusion

Our study demonstrates how comparative genomics in combination with population genomic data can reveal patterns of gene content variation across climatic gradients. In addition, the tracking of gene content variation across populations provides a widely applicable framework for retrieving meaningful biogeographical determinants of gene presence/absence patterns. We contribute to understanding convergence and parallel evolution at the genomic level, by providing insights into gene content variation of lichenized fungi in relation to climatic gradients. This suggests a promising new research direction with implications for understanding evolutionary trajectories in relation to climatic change.

Data availability statement

Raw sequence reads were deposited in the National Centre for Biotechnology Information (NCBI) Sequence Read Archive under the BioProject accession numbers PRJNA693984 and PRJNA820300.

References

- Armenteros, A., Juan, J., Konstantinos, D., Kaae, C., Version, D., and Armenteros, A. (2019). SignalP 5. 0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi: 10.1038/s41587-019-0036-z
- Asano, T., Tanaka, N., Yang, G., Hayashi, N., and Komatsu, S. (2005). Genome-wide identification of the rice calcium-dependent protein kinase and its closely related kinase gene families: Comprehensive analysis of the CDPKs gene family in rice. *Plant Cell Physiol.* 46, 356–366. doi: 10.1093/pcp/pci035
- Badet, T., Oggenfuss, U., Abraham, L., Mcdonald, B. A., and Croll, D. (2020). A 19-isolate reference-quality global pangenome for the fungal wheat pathogen *Zymoseptoria tritici*. *BMC Biol.* 18:12. doi: 10.1186/s12915-020-0744-3
- Barghi, N., Tobler, R., Nolte, V., Jaksic, A. M., Mallard, F., Otte, K. A., et al. (2019). Genetic redundancy fuels polygenic adaptation in *Drosophila*. *PLoS Biol.* 17:e3000128. doi: 10.1371/journal.pbio.3000128
- Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., et al. (2017). antiSMASH 4.0 -- improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* 45, W36–W41. doi: 10.1093/nar/gkx319
- Blumer, M., Brown, T., Freitas, M. B., Destro, A. L., Oliveira, J. A., Morales, A. E., et al. (2022). Gene losses in the common vampire bat illuminate molecular adaptations to blood feeding. *Sci. Adv.* 8:eabm6494. doi: 10.1126/sciadv.abm6494

Author contributions

DM and IS conceived the ideas and wrote the manuscript. IS and FD collected the data. DM, GS, and HV performed genome assembly and annotations. DM analyzed the data. FD provided analytical guidance. All authors contributed to the various drafts and gave final approval for publication.

Funding

This research has been partly funded by the Hesse's "state initiative for the development of scientific and economic excellence" (LOEWE initiative) through the LOEWE Center for Translational Biodiversity Genomics (TBG).

Acknowledgments

We thank Carola Greve and Jürgen Otte for the laboratory procedures, and Christoph Sinai, Tilman Schell, and Anjuli Calchera (all Frankfurt am Main) for support with bioinformatics. We thank Barbara Feldmeyer and Markus Pfenninger (Frankfurt am Main) for their valuable suggestions on analyses.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1097787/full#supplementary-material>

- Borodovsky, M., and Lomsadze, A. (2011). Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. *Curr. Protoc. Bioinformatics* 4, 4.6.1–4.6.10. doi: 10.1002/0471250953.bi0406s35
- Buchfink, B., Xie, C., and Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176
- Castillo, R. V., Beck, A., and Lumbsch, H. T. (2012). Photobiont selectivity and specificity in *Caloplaca* species in a fog-induced community in the Atacama Desert, northern Chile. *Fungal Biol.* 116, 665–676. doi: 10.1016/j.funbio.2012.04.001
- Cheng, S., Willmann, M. R., Chen, H., and Sheen, J. (2002). Calcium signaling through protein kinases. The *Arabidopsis* calcium-dependent protein kinase gene family. *Plant Physiol.* 129, 469–485. doi: 10.1104/pp.005645
- Christ, D., and Chin, J. W. (2008). Engineering *Escherichia coli* heat-resistance by synthetic gene amplification. *Protein Eng. Des. Sel.* 21, 121–125. doi: 10.1093/protein/gzm085
- Cubero, O. F., and Crespo, A. (2002). “Isolation of nucleic acids from lichens,” in *Protocols in Lichenology*. eds. I. Kranner, R. Beckett and A. Varma (Berlin, Heidelberg: Springer), 381–391.
- Dal Grande, F., Sharma, R., Meiser, A., Rolshausen, G., Büdel, B., Mishra, B., et al. (2017). Adaptive differentiation coincides with local bioclimatic conditions along an elevational cline in populations of a lichen-forming fungus. *BMC Evol. Biol.* 17:93. doi: 10.1186/s12862-017-0929-8
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008. doi: 10.1093/gigascience/giab008
- DeBolt, S. (2010). Copy number variation shapes genome diversity in *Arabidopsis* over immediate family generational scales. *Genome Biol. Evol.* 2, 441–453. doi: 10.1093/gbe/evq033
- Drott, M. T., Rush, T. A., Satterlee, T. R., Giannone, R. J., and Abraham, P. E. (2021). Microevolution in the pansecondary metabolome of *Aspergillus flavus* and its potential macroevolutionary implications for filamentous fungi. *PNAS* 118:e2021683118. doi: 10.1073/pnas.2021683118
- Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:157. doi: 10.1186/s13059-015-0721-2
- Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238. doi: 10.1186/s13059-019-1832-y
- Gerdol, M., Moreira, R., Cruz, F., Gómez-garrido, J., Vlasova, A., Rosani, U., et al. (2020). Massive gene presence-absence variation shapes an open pan-genome in the Mediterranean mussel. *Genome Biol.* 21, 1–21. doi: 10.1186/s13059-020-02180-3
- Grimm, M., Grube, M., Schiefelbein, U., Zühlke, D., and Bernhardt, J. (2021). The lichens' microbiota, still a mystery? *Front. Microbiol.* 12:623839. doi: 10.3389/fmicb.2021.623839
- Hanks, S. K. (2003). Genomic analysis of the eukaryotic protein kinase superfamily: a perspective. *Genome Biol.* 4:111. doi: 10.1186/gb-2003-4-5-111
- Hartke, J., Waldvogel, A., Sprenger, P. P., Schmitt, T., Menzel, F., Pfenninger, M., et al. (2021). Little parallelism in genomic signatures of local adaptation in two sympatric, cryptic sister species. *J. Evol. Biol.* 34, 937–952. doi: 10.1111/jeb.13742
- Heinisch, J. J., and Rodicio, R. (2018). Protein kinase C in fungi—more than just cell wall integrity. *FEMS Microbiol. Rev.* 42:fux051. doi: 10.1093/femsre/fux051
- Heuer, H., Schmitt, H., and Smalla, K. (2011). Antibiotic resistance gene spread due to manure application on agricultural fields. *Curr. Opin. Microbiol.* 14, 236–243. doi: 10.1016/j.mib.2011.04.009
- Hirsch, C. N., Foerster, J. M., Johnson, J. M., Sekhon, R. S., Muttoni, G., Vaillancourt, B., et al. (2014). Insights into the maize pan-genome and pan-transcriptome. *Plant Cell* 26, 121–135. doi: 10.1105/tpc.113.119982
- Hodkinson, B. P., Gittel, N. R., Schadt, C. W., and Lutzoni, F. (2012). Photoautotrophic symbiont and geography are major factors affecting highly structured and diverse bacterial communities in the lichen microbiome. *Environ. Microbiol.* 14, 147–161. doi: 10.1111/j.1462-2920.2011.02560.x
- Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., Von Mering, C., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* 34, 2115–2122. doi: 10.1093/molbev/msx148
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hern, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085
- Huson, D. H., Auch, A. F., Qi, J., and Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Res.* 17, 377–386. doi: 10.1101/gr.5969107
- James, T. C., Usher, J., Campbell, S., and Bond, U. (2008). Lager yeasts possess dynamic genomes that undergo rearrangements and gene amplification in response to stress. *Curr. Genet.* 53, 139–152. doi: 10.1007/s00294-007-0172-8
- Jung, P., Brust, K., Schultz, M., Büdel, B., Donner, A., Lakatos, M., et al. (2021). Opening the gap: Rare lichens with rare cyanobionts – Unexplored cyanobiont diversity in cyanobacterial lichens of the order Lichinales. *Front. Microbiol.* 12:728378. doi: 10.3389/fmicb.2021.728378
- Kappen, L. (2000). Some aspects of the great success of lichens in Antarctica. *Antarct. Sci.* 12, 314–324. doi: 10.1017/S095410200000377
- Kolmogorov, M., Rayko, M., Yuan, J., Pevzner, E., and Pevzner, P. (2019). metaFlye: scalable long-read metagenome assembly using repeat graphs. *BioRxiv*. doi: 10.1101/637637
- Kondrashov, F. A. (2012). Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc. Biol. Sci.* 279, 5048–5057. doi: 10.1098/rspb.2012.1108
- Kuzmin, E., Taylor, J. S., and Boone, C. (2022). Retention of duplicated genes in evolution. *Trends Genet.* 38, 59–72. doi: 10.1016/j.tig.2021.06.016
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Liu, D., Hu, R., Zhang, J., Guo, H., Cheng, H., Li, L., et al. (2021). Overexpression of an Agave phosphoenolpyruvate carboxylase improves plant growth and stress tolerance. *Cells* 10:582. doi: 10.3390/cells10030582
- Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, R., et al. (2020). CDD/SPARCLE: the conserved domain database in. *Nucleic Acids Res.* 48, D265–D268. doi: 10.1093/nar/gkz991
- Masignani, V., Medini, D., Donati, C., and Rappuoli, R. (2005). The microbial pan-genome. *Curr. Opin. Genet. Dev.* 15, 589–594. doi: 10.1016/j.gde.2005.09.006
- Mayjonade, B., Gouzy, J., Donnadieu, C., Pouilly, N., Callot, C., Langlade, N., et al. (2016). Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. *BioTechniques* 61, 203–205. doi: 10.2144/000114460
- Merges, D., Grande, F. D., Greve, C., Otte, J., and 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. doi: 10.1111/1462-2920.15802
- Merges, D., Schmidt, A., Schmitt, I., Neuschulz, E. L., Dal Grande, F., and Bálint, M. (2022). Metatranscriptomics reveals contrasting effects of elevation on the activity of bacteria and bacterial viruses in soil. *Mol. Ecol.* 1–12. doi: 10.1111/mec.16756
- Pacific Biosciences of California (2022). Available at: https://downloads.paccloud.com/public/software/installers/smrtlink_11.1.0.166339
- Palmer, J., and Stajich, J. (2019). *Funannotate v1.7.4*. Zenodo. doi: 10.5281/ZENODO.2604804
- Peksa, O., and Skaloud, P. (2011). Do photobionts influence the ecology of lichens? A case study of environmental preferences in symbiotic green alga *Asterochloris* (Trebouxiophyceae). *Mol. Ecol.* 20, 3936–3948. doi: 10.1111/j.1365-294X.2011.05168.x
- Pfenninger, M., Reuss, F., Kiebler, A., Mody, K., Mishra, B., and Thines, M. (2021). Genomic basis for drought resistance in European beech forests threatened by climate change. *eLife* 10:e65532. doi: 10.7554/eLife.65532
- Plissonneau, C., Hartmann, F. E., and Croll, D. (2018). Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biol.* 16:5. doi: 10.1186/s12915-017-0457-4
- Prunier, F., Giguère, I., Ryan, N., Guy, R., Soolanayakanahally, R., Isabel, N., et al. (2019). Gene copy number variations involved in balsam poplar (*Populus balsamifera* L.) adaptive variations. *Mol. Ecol.* 28, 1476–1490. doi: 10.1111/mec.14836
- Qi, Q., Zhao, M., Wang, S., Ma, X., Wang, Y., Gao, Y., et al. (2017). The biogeographic pattern of microbial functional genes along an altitudinal gradient of the Tibetan pasture. *Front. Microbiol.* 8:976. doi: 10.3389/fmicb.2017.00976
- Qin, M., Wu, S., Li, A., Zhao, F., Feng, H., Ding, L., et al. (2019). LRScf: improving draft genomes using long noisy reads. *BMC Genomics* 20:955. doi: 10.1186/s12864-019-6337-2
- Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., et al. (2005). InterProScan: protein domains identifier. *Nucleic Acids Res.* 33, W116–W120. doi: 10.1093/nar/gki442
- R Core Team (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Resl, P., Bujold, A. R., Tagirdzhanova, G., Meidl, P., Rallo, S. F., Kono, M., et al. (2022). Transport potential among lichen fungal symbionts. *Nat. Commun.* 13:2634. doi: 10.1038/s41467-022-30218-6
- Riehle, M. M., Bennett, A. F., and Long, A. D. (2001). Genetic architecture of thermal adaptation in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 98, 525–530. doi: 10.1073/pnas.98.2.525
- Rivas, M. J., Saura, M., Pérez-figueroa, A., Panova, M., Johansson, T., André, C., et al. (2018). Population genomics of parallel evolution in gene expression and gene sequence during ecological adaptation. *Sci. Rep.* 8:16147. doi: 10.1038/s41598-018-33897-8
- Rolshausen, G., Dal, F., Jürgen, G., and Imke, O. (2022). Lichen holobionts show compositional structure along elevation. *Mol. Ecol.* doi: 10.1111/mec.16471. [Epub ahead of print].
- Rolshausen, G., Hallman, U., Grande, F. D., Otte, J., Knudsen, K., and Schmitt, I. (2020). Expanding the mutualistic niche: Parallel symbiont turnover along climatic gradients. *Proc. R. Soc. B* 287:20192311. doi: 10.1098/rspb.2019.2311

- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Singh, G., Calchera, A., Merges, D., Valim, H., Otte, J., and Schmitt, I. (2022). A candidate gene cluster for the bioactive natural product gyrophoric acid in lichen-forming fungi. *Microbiol. Spectr.* 10:e0010922. doi: 10.1128/spectrum.00109-22
- Singh, G., Calchera, A., Schulz, M., Drechsler, M., Bode, H. B., Schmitt, I., et al. (2021). Climate-specific biosynthetic gene clusters in populations of a lichen-forming fungus. *Environ. Microbiol.* 23, 4260–4275. doi: 10.1111/1462-2920.15605
- Singh, G., Grande, F. D., Divakar, P. K., Crespo, A., and Schmitt, I. (2017). Fungal-algal association patterns in lichen symbiosis linked to macroclimate. *New Phytol.* 214, 317–329. doi: 10.1111/nph.14366
- Tanunchai, B., Schroeter, S. A., Ji, L., Fareed, S., Wahdan, M., Hossen, S., et al. (2022). More than you can see: unraveling the ecology and biodiversity of lichenized fungi associated with leaves and needles of temperate tree species using high-throughput sequencing. *Front. Microbiol.* 13:907531. doi: 10.3389/fmicb.2022.907531
- Werth, S., and Sork, V. L. (2014). Ecological specialization in Trebouxia (Trebouxiophyceae) photobionts of *Ramalina menziesii* (Ramalinaceae) across six range-covering ecoregions of Western North America. *Am. J. Bot.* 101, 1127–1140. doi: 10.3732/ajb.1400025