

SCIENTIFIC REPORTS



OPEN

Comparing Sanger sequencing and high-throughput metabarcoding for inferring photobiont diversity in lichens

Fiona Paul¹, Jürgen Otte¹, Imke Schmitt^{1,2} & Francesco Dal Grande^{1,3}

The implementation of HTS (high-throughput sequencing) approaches is rapidly changing our understanding of the lichen symbiosis, by uncovering high bacterial and fungal diversity, which is often host-specific. Recently, HTS methods revealed the presence of multiple photobionts inside a single thallus in several lichen species. This differs from Sanger technology, which typically yields a single, unambiguous algal sequence per individual. Here we compared HTS and Sanger methods for estimating the diversity of green algal symbionts within lichen thalli using 240 lichen individuals belonging to two species of lichen-forming fungi. According to HTS data, Sanger technology consistently yielded the most abundant photobiont sequence in the sample. However, if the second most abundant photobiont exceeded 30% of the total HTS reads in a sample, Sanger sequencing generally failed. Our results suggest that most lichen individuals in the two analyzed species, *Lasallia hispanica* and *L. pustulata*, indeed contain a single, predominant green algal photobiont. We conclude that Sanger sequencing is a valid approach to detect the dominant photobionts in lichen individuals and populations. We discuss which research areas in lichen ecology and evolution will continue to benefit from Sanger sequencing, and which areas will profit from HTS approaches to assessing symbiont diversity.

High-throughput sequencing (HTS) technologies have revolutionized the way we study microbial diversity and other complex ecological communities^{1,2}. In particular, DNA metabarcoding - the identification of organisms from environmental samples using HTS of standardized DNA barcodes - has allowed rapid and cost-effective taxonomic assessments of a wide range of microbial groups from a variety of habitats in marine and terrestrial ecosystems³. Metabarcoding has also advanced our understanding of the distribution, abundance, and community structure of both symbiotic and free-living microorganisms⁴⁻⁶. Because HTS of DNA barcodes, e.g. on the Illumina MiSeq, Ion Torrent, or PacBio platforms, allows taxonomic assessment at a far greater depth and resolution than conventional Sanger sequencing^{2,7}, it also facilitates discovery of rare taxa and detection of previously unrecognized eukaryotic and prokaryotic microbiomes⁸⁻¹². Studying this vastly unexplored microbial biosphere is one of the new frontiers of microbial ecology, currently changing our understanding of species interactions and their ecological and evolutionary dynamics¹³.

One of the most striking examples of the HTS-driven paradigm shifts comes from lichen symbiosis research¹⁴⁻¹⁶. For more than a century, lichens have been viewed as a symbiotic association between two, maximally three partners, i.e. a fungus (the mycobiont) and a green alga and/or cyanobacterium (the photobiont). DNA sequences supported this notion: in most cases researchers detected a single species of mycobiont and photobiont in a thallus based on unambiguous Sanger sequence electropherograms. In recent years, HTS-based studies have uncovered a tremendous species diversity associated with lichen symbioses (e.g.¹⁷⁻¹⁹). Lichen thalli harbor hyper-diverse bacterial as well as fungal communities (e.g.^{16,20}). This previously unrecognized lichen-associated biosphere may act as a significant functional component of the symbiosis, supporting the growth and regeneration of the thallus, and modulating the holobiont's response to environmental triggers²¹⁻²³.

¹Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Senckenberganlage 25, 60325, Frankfurt am Main, Germany. ²Institute of Ecology, Evolution and Diversity, Goethe University Frankfurt am Main, Max-von-Laue-Str. 9, 60438, Frankfurt am Main, Germany. ³Departamento de Farmacología, Farmacognosia y Botánica, Universidad Complutense de Madrid, 28040, Madrid, Spain. Correspondence and requests for materials should be addressed to I.S. (email: imke.schmitt@senckenberg.de) or F.D.G. (email: francesco.dalgrande@senckenberg.de)

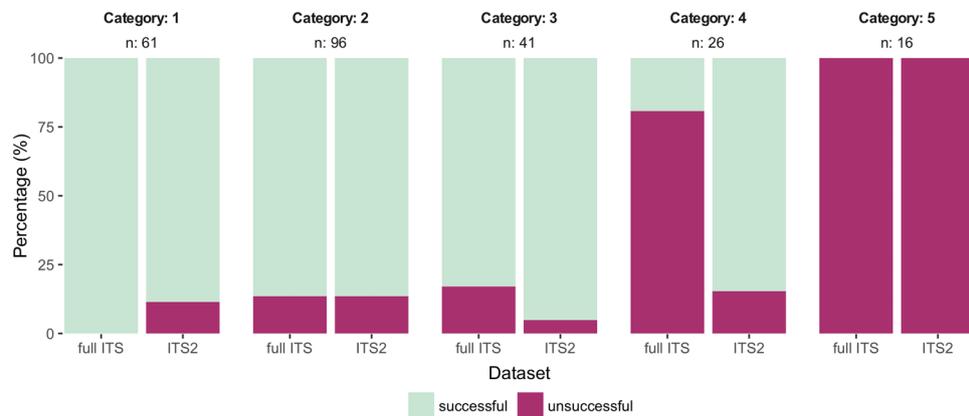


Figure 1. Sanger sequencing success depending on the abundance of the second most abundant photobiont for the full ITS (left) and ITS2 (right) data sets. Abundance categories based on Illumina ITS2 metabarcoding²⁴ are defined as follows: *cat. 1*: only a single photobiont present; *cat. 2*: secondary photobiont accounts for up to 10% of sequence reads; *cat. 3*: secondary photobiont accounts for 10–20% of reads; *cat. 4*: secondary photobiont accounts for 20–30% of reads; *cat. 5*: secondary photobiont represented by more than 30% of sequence reads.

Recent HTS studies on lichen photobiont diversity suggest that the presence of multiple, genetically differentiated algae within a single thallus is a common phenomenon in lichens^{24–26}. This raises the question to what extent Sanger-based studies underestimate within-thallus green algal diversity, and whether the single Sanger sequence obtained from an individual corresponds to the most abundant photobiont in that individual²⁷. These assumptions, however, have never been formally tested.

In this study, we compare HTS and Sanger sequencing methods for estimating the diversity of lichen-associated green algal photobionts. In particular, we address the following questions: (i) Are the photobionts identified via Sanger sequencing always the most abundant taxa in the thallus as inferred by HTS? and (ii) If more than one photobiont is present in a lichen individual: Is there an abundance threshold that precludes the generation of an unambiguous Sanger sequence?

Results

Overall sequencing success. From the Sanger sequencing of the 240 samples, we obtained 183 (full length ITS) and 198 (ITS2) high quality sequences. This corresponds to a sequencing success rate of 76.25% (full length ITS) and 82.5% (ITS2) (Supplementary Table S1). The excluded, low quality sequences were partially sequences containing many ambiguous bases, which reflected variable positions distinguishing the different OTU representative sequences obtained from the Illumina metabarcoding study. We got blast hits of >99% identity to one single OTU representative sequence for all of the Sanger sequences of the full ITS region. In the set of Sanger sequences amplified with the metabarcoding primers covering only ITS2, 99% of sequences were also clearly assignable to a single Illumina metabarcoding OTU. However, one Sanger sequence matched two OTUs at 100% sequence similarity and one Sanger sequence had a highest hit of 98.27% sequence similarity. Both sequences could not be assigned to an OTU, since they contained ambiguous bases at several variable sites that are used to determine OTU membership. We therefore excluded them from all downstream analyses.

Sanger sequencing retrieves most abundant Illumina OTU. Each Sanger sequence (full length ITS, and ITS2-only) that passed the quality filtering corresponded to the most abundant OTU in the HTS metabarcoding dataset (Supplementary Table S2). There was not a single exception to this pattern. This also held true for samples, which had OTU6 as the most abundant algal partner. OTU6 was excluded in the Illumina metabarcoding study²⁵ because it was present in less than 5% ($n = 12$) of the total number of samples. In all of the five cases, in which OTU6 was found to be the predominant photobiont, the Sanger sequences of both datasets (full length ITS and ITS2) were assigned to OTU6.

Sanger sequencing fails when the second most abundant alga reaches 30% abundance. For each of the datasets, we investigated the influence of the frequency distribution of the Illumina metabarcoding OTUs of each sample on the quality of the obtained Sanger sequence. When the predominant photobiont represented 100% of Illumina reads in a sample, 100% of the full length ITS (61 sequences) and 88.52% of the ITS2 (54 sequences) Sanger sequences were of high quality (Fig. 1). The proportion of obtained high quality Sanger sequences decreased with an abundance increase of the second most abundant photobiont. This trend held true for the full length ITS and ITS2-only datasets. Once more than 30% of Illumina reads were assigned to the second most abundant photobiont OTU, 0% (0 sequences) of Sanger sequences were of high quality, while 100% (16 sequences) of Sanger sequences failed to pass previous quality control measures (Fig. 1, Supplementary Table S1).

Diversity found in ITS1 compared to ITS2. The OTUs assigned to each sample based on the ITS1 sequence retained from the full ITS Sanger sequences always matched the OTU assignment based on the ITS2 sample. We found no differences in diversity between ITS1 and ITS2 for the *Trebouxia* algae associated with the

two lichen species, *Lasallia pustulata* and *L. hispanica*. For each sample, the OTU assigned based on the ITS1 region also coincided with the predominant photobiont found in the HTS data (Supplementary Table S3).

Discussion

While the implementation of HTS technologies has greatly advanced our understanding of the hyper-diverse bacterial and fungal microbiomes integral to the lichen symbiosis²⁰, most of the taxonomy and diversity estimates of lichen photobionts are still based on conventional Sanger sequencing of a single DNA barcode, the internal transcribed spacer (ITS) rRNA. This is evident from the vast number of ITS sequences of *Trebouxia* spp., the most common green algal genus of lichenized algae, deposited in NCBI GenBank (>6,000 algal sequences from more than 100 lichen genera; accessed in January 2018). Sanger sequences of lichenized algae obtained via amplification and sequencing of mixed symbiont DNA from lichen thalli have been used not only for molecular identification and phylogenetic analysis of photobionts^{28–30}, but also for inferring fungal-algal association patterns^{31–33}, and for reconstructing photobiont phylogeography and ecological niches^{34,35}. The presence of multiple photobionts within a single thallus^{24,25,27,36–40}, however, raises the question whether Sanger-based datasets are reproducible and biologically meaningful. In this study, we compared Sanger-based identification and diversity estimates of photobionts with photobiont abundances observed in the same DNA extracts inferred via Illumina MiSeq HTS of the PCR products using the same primers.

Sanger-based and HTS datasets inform each other. Our results clearly indicate that the photobiont identified by Sanger sequencing always corresponds to the most abundant photobiont in a sample. Furthermore, our results help to verify the accuracy of the HTS sequence filtering steps utilized in²⁵. One of the six photobionts (OTU6) identified via Sanger sequencing was excluded from the HTS photobiont dataset in²⁵, because it was present in less than 5% of the samples (although it was the predominant alga in some samples). The consistent recovery of this OTU using Sanger sequencing suggests that, in order to avoid excessive filtering and keep biologically relevant units in the analysis, the filtering of the HTS OTU table should retain OTUs if they represent the predominant alga in at least one sample.

One limitation that has been attributed to HTS platforms is the limited sequence length (<500 bp)⁷. In our study the addition of the ITS1 portion did not improve taxonomic resolution: the OTUs defined via the full ITS barcode were always concordant with the OTUs based on ITS2. Thus, in the case of the *Trebouxia* photobionts analyzed in this study, all belonging to the *Trebouxia* clade S (*simplex/letharii/jamesii* sensu⁴¹), the implementation of long read sequencing platforms, such as PacBio or Ion Torrent, are not expected to improve species discrimination and phylogenetic information.

Most lichen individuals contain a single algal population. The performance of Sanger sequencing degraded considerably when the second most abundant photobiont in a sample increased in relative abundance, and Sanger sequencing failed entirely when the second most abundant photobiont exceeded 30% of the total sample reads. This is particularly relevant if we consider that almost all algal ITS sequences deposited in public databases were obtained from direct PCR and sequencing of DNA mixtures from lichen thalli without any cloning steps. The inherent limitations of Sanger sequencing in obtaining readable electropherograms from samples with two abundant photobionts thus suggest that the vast majority of lichen thalli indeed host a single predominant algal population. If ITS Sanger sequencing fails because multiple algal strains are present in high abundances (>30% of total reads in sample), it may still be possible to obtain readable Sanger sequences from the sample by using lineage specific algal primers²⁷. Our analysis suggests that each lichen individual contains maximally two, or very rarely three, abundant *Trebouxia* strains. Thus, Sanger ITS sequencing can be a viable approach to comprehensively sample green algal diversity, even if there is more than one abundant photobiont in a thallus.

Which research questions can be addressed using Sanger sequencing? The ability of Sanger sequencing to always, and unfailingly, detect the predominant algal species in a lichen thallus supports the continued implementation of this sequencing technique for addressing a wide variety of questions in lichen ecology and evolution. Specifically, Sanger ITS barcoding will continue to be useful in the following types of studies: (i) Assessing coevolutionary patterns in the lichen symbiosis: abundances and distributions of species exhibit often great stability on ecological timescales. Sanger sequencing of the ITS barcode marker allows for the rapid identification of the most common, i.e. biologically relevant, photobionts associated with a lichen. Sanger-derived photobiont databases can thus be used to study patterns of fungal-algal associations (e.g., generalism vs. specificity) across spatial and temporal scales, and for testing hypotheses of coevolution between fungi and their most common photobionts; (ii) Understanding symbiont-mediated niche expansion and conservation: according to the (bio) mass-ratio hypothesis, common species are disproportionately significant in ecosystem function^{42–44}. Transposing the concept to lichen associations, the most abundant photobionts will numerically dominate lichen communities and thus play a disproportionately large role in the lichen community and ecosystem processes. Sanger-derived ITS photobiont datasets can thus be used to reconstruct ecological niches of the most common photobionts in a lichen community or population, and to study distributional trends of photobionts in response to environmental and/or anthropogenic cues. This is particularly relevant when assessing and predicting frequency and stability of fungal-algal interactions against a background of ongoing global change. Any change in the distribution of the most common species will be immediately significant for the entire community, because the most common species play important roles in engineering environments and supporting a large number of biotic interactions compared to rare species^{45,46}.

Which research questions can be addressed using high-throughput metabarcoding? Because of its ultra-high-throughput and relatively low cost per read, Illumina MiSeq has become the method of choice for highly-multiplexed barcoding sequencing, i.e. the massive sequencing of hundreds, if not thousands, of samples

in a single run. Additionally, and most importantly, HTS platforms allow the profiling of symbionts at the level of the whole community.

Thus, the most relevant application of the Illumina MiSeq (as any other HTS) platform is in the field of community ecology. Especially for species-rich communities, such as the lichen microbiomes, abundances of different taxa can vary considerably among different zones of a lichen thallus²⁶, among individual hosts, as well as among different host taxa or habitats. As recently shown by²⁶, only a combination of ultrastructural and HTS sequencing approaches on different parts of a lichen thallus (especially for species with foliose or fruticose growth form) and on many individuals will allow to fully uncover this potential variation⁵.

Rare species are increasingly recognized as potentially hidden drivers of microbiome function in various ecosystems^{8,47,48}. This is especially true when rare or less abundant species strongly interact with more common species⁴⁶. The role of relatively rare photobionts still needs to be investigated in lichens. HTS metabarcoding is thus the preferred method to obtain taxonomically more comprehensive symbiont datasets, as well as to investigate and model patterns of interactions among the photobiont taxa themselves within and among lichen thalli in various habitats. For instance, aspects of the insurance hypothesis⁴⁹ could be tested using metabarcoding approaches. This hypothesis states that the presence of rare symbionts may buffer against environmental changes in order to maintain ecosystem function (in this case the survival of the lichen thallus) by allowing species turnovers, i.e. flexible abundance fluctuations in response to environmental changes. HTS platforms open up new avenues for investigating lichen symbioses and their complex microbiomes and will shed light on processes that may only be visible when the entire symbiont community is comprehensively considered.

Methods

Sampling. We collected populations of the lichen-forming fungi *Lasallia hispanica* and *L. pustulata* along an altitudinal gradient in the Sierra de Gredos mountain range in Central Spain. We sampled each species at six sites, and 20 individuals per site, resulting in a total collection of 240 individuals. Details on the study organisms, sampling, and sampling sites are given in²⁵.

Datasets. In the present study, we compared three datasets generated for each of the 240 individuals: (1) A metabarcoding dataset based on algal ITS2; (2) A Sanger sequence set based on the full length algal ITS; and (3) A Sanger sequence set based on ITS2 using the same primers as in the metabarcoding study.

Metabarcoding dataset. Thallus pieces of ~8 mm in diameter were collected in proximity to the centre of each thallus⁵⁰. We extracted total DNA using the CTAB method⁵¹, and amplified *Trebouxia* algal communities with algal-specific primers targeting the ITS2 region (FDGITS2-f: AGCGAAATGCGATACGTAGTGT; FDGITS2-r: GGGTGTCTTGCCTGACCTC²⁵). After library preparation the reads were paired-end sequenced (2 × 300 bp) on an Illumina MiSeq sequencer at StarSeq (Mainz, Germany). For a detailed account of Illumina sequencing results and bioinformatics processing of the reads please refer to²⁵. Briefly, we quality-filtered, paired-end assembled, and demultiplexed the reads, before grouping them into operational taxonomic units (OTUs) using the de novo, single-linkage clustering algorithm SWARM v.2.0⁵². This OTU calling approach clusters sequences based on abundance information, as well as global sequence similarity. We excluded chimeric sequences and OTUs not assigned to the algal genus *Trebouxia* and filtered the OTU table for erroneous sequence occurrences as follows: we removed OTUs falling below 0.005% of the total read count in each sample, as well as OTUs present in less than 5% of samples (n = 12)^{25,53–55}.

Sanger sequencing (full length ITS rDNA). We used the 240 DNA extracts described above to amplify full length algal ITS using the universal primers T1 and T4⁵⁶. The PCR-reactions contained 0.65 U TaKaRa ExTaq (Clontech Laboratories Inc., Palo Alto, CA, USA), 2.5 µL buffer, 18.5 µL water, 0.5 µL bovine serum albumin (BSA, 10 mg ml⁻¹), 2 µL dNTP mixture (2.5 mM each), about 5 ng total DNA (0.5 µL) and 0.22 µM (0.5 µL) forward and reverse primers. The following PCR conditions were used: initial denaturation at 95 °C for 4 min, followed by 35 cycles of 95 °C for 30 s, 63 °C for 20 s, 72 °C for 20 s and a final elongation for 5 min at 72 °C. We sequenced the amplicons using BigDye v3.1 chemistry (Applied Biosystems, Foster City, CA, USA) under the following cycle sequencing conditions: initial denaturation for 1 min at 95 °C, followed by 30 cycles of 96 °C for 10 s, 50 °C for 10 s, 60 °C for 2 min. The products were purified and sequenced on an ABI 3730 DNA Analyzer (Applied Biosystems).

Sanger sequencing (ITS2). From the 240 DNA extracts we also amplified the ITS2 region, using the same primers we used in the metabarcoding study (FDGITS2-f and FDGITS2-r). The PCR mixture as well as the cycling conditions were the same as described above for the full length ITS sequences. Sanger sequencing conditions were also the same as for the full length ITS fragments.

Editing of Sanger sequences. Sanger sequences were manually evaluated in Geneious v9.1.7 (Biomatters Ltd., Auckland, New Zealand). In a first step, the forward and reverse sequences of each sample were paired-end assembled using the command De Novo Assembler with the Geneious Assembler, Highest Sensitivity and without trimming the sequences. If the assembly was unsuccessful without trimming the sequences, sequences were paired-end assembled allowing the sequences to be trimmed. Secondary peaks were automatically identified as peaks with a secondary peak height above 25% of the maximal peak. The assembled sequences were then manually evaluated and edited based on their electropherogram. We directly excluded sequences, for which the paired-end assembly failed and/or led to significantly shortened sequences. Further, we excluded sequences with ambiguous base calls and/or underlying peaks in the electropherogram that could not be resolved. Since it is impossible to perform robust taxonomic assignment for these sequences, we excluded them from the dataset and from further analyses, and we noted them as unsuccessful Sanger sequences. All sequences passing the quality filter were aligned in a multiple pairwise alignment with MAFFT version 7.308⁵⁷ using default settings to further

edit the sequences where needed. We then extracted the algal ITS1 (for the full length ITS dataset) and ITS2 (for both full length ITS dataset and ITS2 dataset) variable regions from the sequences using ITSx version 1.0.11⁵⁸ with default settings.

Comparison of datasets. To compare the Sanger sequence of each sample to the OTUs in the metabarcoding dataset, we ran a local nucleotide BLAST comparing each Sanger sequence to the representative sequences of the OTUs. We performed a local BLAST run using blastn v 2.2.30+⁵⁹ with default settings, accepting sequence similarities of 95% and higher. We excluded sequences from further analysis, if there was (1) no clear BLAST hit to a single OTU representative sequence, i.e. the second-best BLAST hit was not significantly lower in score and identity than best BLAST hit and (2) the sequence similarity was not 99% or higher. If a Sanger sequence met these criteria, the matching OTU was assigned to the respective sample. We then checked whether the OTU assigned based on the Sanger sequence of each sample matched the most abundant OTU found in the metabarcoding dataset.

We examined whether the quality of the Sanger sequence was influenced by the abundance distribution of the OTUs in the metabarcoding dataset. We first split our dataset into the Sanger sequences of the full ITS sequence (amplified with universal primers T1 and T4) and the Sanger sequences of only ITS2 (amplified with primers FDGITS2-f and FDGITS2-r). Therefore, we assigned each sample to a frequency category based on the frequency of the second most abundant algal OTU in the metabarcoding dataset. We defined frequency categories as follows:

- category 1: most abundant alga accounts for 100% of sequences;
- category 2: secondary photobiont accounts for up to 10% of sequence reads;
- category 3: secondary photobiont accounts for 10–20% of reads;
- category 4: secondary photobiont accounts for 20–30% of reads;
- category 5: secondary photobiont represented by more than 30% of sequence reads.

We then checked which frequency categories resulted in a high-quality Sanger sequence. We calculated the proportions of samples with and without Sanger sequences in each of the categories and made stacked bar plots using R v. 3.3.3⁶⁰ and the ggplot2 package⁶¹.

We compared the taxonomic resolution of ITS1 compared to ITS2, since we also had Sanger sequences covering the full ITS, from which we extracted the ITS1 region during the ITS extraction (see above). For this, we first clustered all ITS1 Sanger sequences at 99% sequence identity with vsearch version 2.4.3⁶² using the cluster_size command with default settings. The generated OTU representative sequences were then compared to the OTUs from the metabarcoding dataset. We compared the assigned OTUs based on ITS1 of each sample to the most abundant OTU from the metabarcoding dataset, noting whether they matched or not.

Sequence data are available in NCBI GenBank under accession numbers MH298924 - MH299121 and MH299122 - MH299304 for the ITS2 and the full ITS sequences, respectively.

References

1. Aylagas, E., Borja, Á., Irigoien, X. & Rodríguez-Ezpeleta, N. Benchmarking DNA Metabarcoding for Biodiversity-Based Monitoring and Assessment. *Front Mar Sci* **3** (2016).
2. Kennedy, P. G., Cline, L. C. & Song, Z. Probing promise versus performance in longer read fungal metabarcoding. *New Phytol* **217**, 973–976 (2018).
3. Hajibabaei, M. The golden age of DNA metasytematics. *Trends Genet* **28**, 535–537 (2012).
4. Zhou, J. *et al.* High-throughput metagenomic technologies for complex microbial community analysis: Open and closed formats. *MBio* **6** (2015).
5. Bálint, M. *et al.* Millions of reads, thousands of taxa: Microbial community structure and associations analyzed via marker genes. *FEMS Microbiology Reviews* **40**, 686–700 (2016).
6. Dickie, I. A. & John, M. G. In *Molecular Mycorrhizal Symbiosis* 473–491 <https://doi.org/10.1002/9781118951446.ch26> (2016).
7. Tedersoo, L., Tooming-Klunderud, A. & Anslan, S. PacBio metabarcoding of Fungi and other eukaryotes: Errors, biases and perspectives. *New Phytol* **217**, 1370–1385 (2017).
8. Ainsworth, T. D. *et al.* The coral core microbiome identifies rare bacterial taxa as ubiquitous endosymbionts. *ISME J* **9**, 2261–2274 (2015).
9. Divakar, P. K. *et al.* Evolution of complex symbiotic relationships in a morphologically derived family of lichen-forming fungi. *New Phytol* **208**, 1217–1226 (2015).
10. Wedin, M. *et al.* Microbiome change by symbiotic invasion in lichens. *Environ Microbiol* **18**, 1428–1439 (2016).
11. Baker, C. C. M., Bittleston, L. S., Sanders, J. G. & Pierce, N. E. Dissecting host-associated communities with DNA barcodes. *Philos Trans R Soc B Biol Sci* **371**, 20150328 (2016).
12. Martin, F. M., Uroz, S. & Barker, D. G. Ancestral alliances: Plant mutualistic symbioses with fungi and bacteria. *Science* **356** (2017).
13. Selosse, M.-A., Schneider-Maunoury, L. & Martos, F. Time to re-think fungal ecology? Fungal ecological niches are often prejudged. *New Phytol* **217**, 968–972 (2018).
14. Grube, M. *et al.* In *The Ecological Genomics of Fungi* 191–212 <https://doi.org/10.1002/9781118735893.ch9> (John Wiley & Sons, Inc, 2013).
15. Greshake, B. *et al.* Potential and pitfalls of eukaryotic metagenome skimming: A test case for lichens. *Mol Ecol Resour* **16**, 511–523 (2016).
16. Meiser, A., Otte, J., Schmitt, I. & Grande, F. D. Sequencing genomes from mixed DNA samples - Evaluating the metagenome skimming approach in lichenized fungi. *Sci Rep* **7** (2017).
17. Bates, S. T., Cropsey, G. W. G., Caporaso, J. G., Knight, R. & Fierer, N. Bacterial communities associated with the lichen symbiosis. *Appl Environ Microbiol* **77**, 1309–1314 (2011).
18. Fernández-Mendoza, F., Fleischhacker, A., Kopun, T., Grube, M. & Muggia, L. ITS1 metabarcoding highlights low specificity of lichen mycobiomes at a local scale. *Mol Ecol* **26**, 4811–4830 (2017).
19. Spribille, T. *et al.* Basidiomycete yeasts in the cortex of ascomycete macrolichens. *Science* **353**, 488–492 (2016).
20. Aschenbrenner, I. A., Cernava, T., Berg, G. & Grube, M. Understanding microbial multi-species symbioses. *Front Microbiol* **7**, 1–9 (2016).

21. Hodkinson, B. P., Gittel, N. R., Schadt, C. W. & Lutzoni, F. Photoautotrophic symbiont and geography are major factors affecting highly structured and diverse bacterial communities in the lichen microbiome. *Environ Microbiol* **14**, 147–161 (2012).
22. Grube, M. *et al.* Exploring functional contexts of symbiotic sustain within lichen-associated bacteria by comparative omics. *ISME J* **9**, 412–24 (2015).
23. Cernava, T. *et al.* Deciphering functional diversification within the lichen microbiota by meta-omics. *Microbiome* **5**, 82 (2017).
24. Moya, P., Molins, A., Martínez-Alberola, F., Muggia, L. & Barreno, E. Unexpected associated microalgal diversity in the lichen *Ramalina farinacea* is uncovered by pyrosequencing analyses. *PLoS One* **12** (2017).
25. Dal Grande, F. *et al.* Environment and host identity structure communities of green algal symbionts in lichens. *New Phytol* **217**, 277–289 (2018).
26. Molins, A., Moya, P., García-Breijo, F. J., Reig-Armiñana, J. & Barreno, E. A multi-tool approach to assess microalgal diversity in lichens: isolation, Sanger sequencing, HTS and ultrastructural correlations. *Lichenol* **50**, 123–138 (2018).
27. del Campo, E. M. *et al.* The genetic structure of the cosmopolitan three-partner lichen *Ramalina farinacea* evidences the concerted diversification of symbionts. *FEMS Microbiol Ecol* **83**, 310–323 (2013).
28. Leliaert, F. *et al.* DNA-based species delimitation in algae. *Eur J Phycol* **49**, 179–196 (2014).
29. Sadowska-Deś, A. D. *et al.* Integrating coalescent and phylogenetic approaches to delimit species in the lichen photobiont *Trebouxia*. *Mol Phylogenet Evol* **76**, 202–10 (2014).
30. Škaloud, P., Friedl, T., Hallmann, C., Beck, A. & Dal Grande, F. Taxonomic revision and species delimitation of coccoid green algae currently assigned to the genus *Dictyochloropsis* (Trebouxiophyceae, Chlorophyta). *J Phycol* **52**, 599–617 (2016).
31. Dal Grande, F. *et al.* Molecular phylogeny and symbiotic selectivity of the green algal genus *Dictyochloropsis* s.l. (Trebouxiophyceae): A polyphyletic and widespread group forming photobiont-mediated guilds in the lichen family Lobariaceae. *New Phytol* **202**, 455–470 (2014).
32. Werth, S. & Sork, V. L. 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 (2014).
33. Singh, G. *et al.* Fungal–algal association patterns in lichen symbiosis linked to macroclimate. *New Phytol* **214**, 317–329 (2017).
34. Fernández-Mendoza, F. *et al.* Population structure of mycobionts and photobionts of the widespread lichen *Cetraria aculeata*. *Mol Ecol* **20**, 1208–1232 (2011).
35. Rolshausen, G., Dal Grande, F., Sadowska-Deś, A. D., Otte, J. & Schmitt, I. Quantifying the climatic niche of symbiont partners in a lichen symbiosis indicates mutualist-mediated niche expansions. *Ecography* <https://doi.org/10.1111/ecog.03457> (2017).
36. Casano, L. M. *et al.* Two *Trebouxia* algae with different physiological performances are ever-present in lichen thalli of *Ramalina farinacea*. Coexistence versus Competition? *Environ Microbiol* **13**, 806–818 (2011).
37. Dal Grande, F. *et al.* Insights into intrathalline genetic diversity of the cosmopolitan lichen symbiotic green alga *Trebouxia decolorans* Ahmadjian using microsatellite markers. *Mol Phylogenet Evol* **72**, 54–60 (2014).
38. Catalá, S. *et al.* Coordinated ultrastructural and phylogenomic analyses shed light on the hidden phycobiont diversity of *Trebouxia* microalgae in *Ramalina fraxinea*. *Mol Phylogenet Evol* **94**, 765–777 (2016).
39. Piercey-Normore, M. Vegetatively reproducing fungi in three genera of the Parmeliaceae share divergent algal partners. *Bryologist* **112**, 773–785 (2009).
40. Nyati, S., Werth, S. & Honegger, R. Genetic diversity of sterile cultured *Trebouxia* photobionts associated with the lichen-forming fungus *Xanthoria parietina* visualized with RAPD-PCR fingerprinting techniques. *Lichenol* **45**, 825–840 (2013).
41. Leavitt, S. D. *et al.* Fungal specificity and selectivity for algae play a major role in determining lichen partnerships across diverse ecogeographic regions in the lichen-forming family Parmeliaceae (Ascomycota). *Mol Ecol* **24**, 3779–3797 (2015).
42. Grime, J. P. Benefits of plant diversity to ecosystems: Immediate, filter and founder effects. *J Ecol* **86**, 902–910 (1998).
43. Connolly, S. R. *et al.* Commonness and rarity in the marine biosphere. *Proc Natl Acad Sci* **111**, 8524–8529 (2014).
44. Houadria, M. & Menzel, F. What determines the importance of a species for ecosystem processes? Insights from tropical ant assemblages. *Oecologia* **184**, 885–899 (2017).
45. Jiguet, F. *et al.* Population trends of European common birds are predicted by characteristics of their climatic niche. *Glob Chang Biol* **16**, 497–505 (2010).
46. Gaston, K. J. Common Ecology. *Bioscience* **61**, 354–362 (2011).
47. Marcelino, V. R. & Verbruggen, H. Multi-marker metabarcoding of coral skeletons reveals a rich microbiome and diverse evolutionary origins of endolithic algae. *Sci Rep* **6** (2016).
48. Jousset, A. *et al.* Where less may be more: How the rare biosphere pulls ecosystems strings. *ISME Journal* **11**, 853–862 (2017).
49. Yachi, S. & Loreau, M. Biodiversity and ecosystem productivity in a fluctuating environment: The insurance hypothesis. *Proc Natl Acad Sci* **96**, 1463–1468 (1999).
50. Dal Grande, F. *et al.* Adaptive differentiation coincides with local bioclimatic conditions along an elevational cline in populations of a lichen-forming fungus. *BMC Evol Biol* **17**, 93 (2017).
51. Cubero, O. F. & Crespo, A. In *Protocols in Lichenology SE – 23* (eds Kranner, I., Beckett, R. & Varma, A.) 381–391 https://doi.org/10.1007/978-3-642-56359-1_23 (Springer Berlin Heidelberg, 2002).
52. Mahé, F., Rognes, T., Quince, C., De Vargas, C. & Dunthorn, M. Swarmv2: highly-scalable and high-resolution amplicon clustering. *PeerJ* **3**, e1420 (2015).
53. Bokulich, N. A. *et al.* Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods* **10**, 57–9 (2013).
54. Callahan, B. J. *et al.* DADA2: High resolution sample inference from amplicon data. *Nat Methods* **13**, 581–583 (2016).
55. Krohn, A. *et al.* Optimization of 16S amplicon analysis using mock communities: implications for estimating community diversity. *PeerJ* **219** (2016).
56. Kroken, S. & Taylor, J. W. Phylogenetic species, reproductive mode, and specificity of the green alga *Trebouxia* forming lichens with the fungal genus *Letharia*. *Bryologist* **103**, 645–660 (2000).
57. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* **30**, 772–780 (2013).
58. Bengtsson-Palme, J. *et al.* Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods Ecol Evol* **4**, 914–919 (2013).
59. Altschul, S. F. *et al.* Blast and Psi-Blast: Protein Database Search Programs. *Nucleic Acid Res* **25**, 2289–4402 (1997).
60. R Core Team. R Core Team (2017). R: A language and environment for statistical computing. *R Found Stat Comput Vienna, Austria URL* <http://www.R-project.org/> R Foundation for Statistical Computing (2017).
61. Wickham, H. *ggplot2 Elegant Graphics for Data Analysis*. *Media* **35** (2009).
62. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).

Acknowledgements

We thank Gregor Rolshausen (Frankfurt) for useful comments on the manuscript. F.D.G. acknowledges postdoctoral fellowship support from “Ayudas para la contratación de personal posdoctoral de formación en docencia e investigación - Convocatoria UCM CT65/16” and from the Spanish “Juan de la Cierva-incorporación” program (IJCI-2016-29149).

Author Contributions

F.D.G. and I.S. conceived and designed the study. J.O. and F.D.G. performed the experiments; F.P. and F.D.G. analyzed the data. F.D.G., F.P. and I.S. wrote the paper. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-26947-8>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018