J. Phycol. **54,** 720–733 (2018)

DOI: 10.1111/jpy.12771

THE EFFECTS OF CONTEMPORARY SELECTION AND DISPERSAL LIMITATION ON THE COMMUNITY ASSEMBLY OF ACIDOPHILIC MICROALGAE[1]

Chia-Jung Hsieh[3]

Department of Life Science, Tunghai University, Taichung 40704, Taiwan

Shing Hei Zhan[3]

Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

Chen-Pan Liao[3]

Department of Life Science, Tunghai University, Taichung 40704, Taiwan

Sen-Lin Tang

Biodiversity Research Center, Academia Sinica, Taipei 11529, Taiwan

Liang-Chi Wang

Department of Earth and Environmental Sciences, National Chung Cheng University, Chiayi 621, Taiwan

Tsuyoshi Watanabe

Tohoku National Fisheries Research Institute, Fisheries Research Agency, Miyagi 985-0001, Japan

Paul John L. Geraldino

Department of Biology, University of San Carlos, Cebu 6000, Philippines

and Shao-Lun Liu[2]

Department of Life Science, Tunghai University, Taichung 40704, Taiwan

Extremophilic microalgae are primary producers in acidic habitats, such as volcanic sites and acid mine drainages, and play a central role in biogeochemical cycles. Yet, basic knowledge about their species composition and community assembly is lacking. Here, we begin to fill this knowledge gap by performing the first large-scale survey of microalgal diversity in acidic geothermal sites across the West Pacific Island Chain. We collected 72 environmental samples in 12 geothermal sites, measured temperature and pH, and performed *rbc*L amplicon-based 454 pyrosequencing. Using these data, we estimated the diversity of microalgal species, and then examined the relative contribution of contemporary selection (i.e., local environmental variables) and dispersal limitation on the assembly of these communities. A species delimitation analysis uncovered seven major microalgae (four red, two green, and one diatom) and higher species diversity than previously appreciated. A distance-based redundancy analysis with variation partitioning revealed that dispersal limitation has a greater influence on the community assembly of microalgae than contemporary selection. Consistent with this finding, community similarity among the sampled sites decayed more quickly over geographical distance than differences in environmental factors. Our work paves the way for future studies to understand the ecology and biogeography of microalgae in extreme habitats.

Key index words: **acidophilic microalgae; community assembly; contemporary selection; dispersal limitation; pyrosequencing; *rbc*L**

***Abbreviations*: ESU, evolutionary significant unit; GMYC, generalized mixed Yule-coalescent model; ML, maximum likelihood; OTU, operational taxonomic unit**

The effect of abiotic factors and geographical separation on the structure of microbial communities is an essential topic in ecology and biogeography (Hanson et al. 2012). Contemporary selection along environmental gradients (i.e., local environmental variables such as temperature and pH) and dispersal limitation by historical ecological and evolutionary drivers influence the composition and relative

[1]Received 10 October 2017. Accepted 7 June 2018. First Published Online 28 July 2018. Published Online 12 September 2018, Wiley Online Library (wileyonlinelibrary.com).
[2]Author for correspondence: e-mail shaolunliu@thu.edu.tw.
[3]These authors contributed equally to this work.
Editorial Responsibility: M. Vis (Associate Editor)

abundance of species in microbial communities (Nemergut et al. 2013). The effect of contemporary selection and dispersal limitation varies depending on the ecosystem (reviewed in Hanson et al. 2012). In nonextreme freshwater ecosystems (e.g., rivers, lakes, and reservoirs), contemporary selection has been shown to affect the similarity among microbial communities more than dispersal limitation (e.g., Fagervold et al. 2014, Liu et al. 2015a,b, Watson and Kling 2017). In extreme ecosystems, however, it is unclear whether contemporary selection or dispersal limitation is more important.

Microbial mats found in extreme habitats have simple species composition and low trophic complexity. Thus, extremophilic microbial mats provide an uncomplicated system to study the influence of contemporary selection and dispersal limitation on community assembly (Caruso et al. 2011). It is well-established that dispersal limitation affects community assembly in extremophilic microbes (e.g., Papke et al. 2003, Whitaker 2003, Takacs-Vesbach et al. 2008, Jones et al. 2016). However, Caruso et al. (2011) showed that there is a complex balance between contemporary selection and dispersal limitation, suggesting that both these forces play equally important roles. It is yet unsettled whether dispersal limitation (e.g., Papke et al. 2003, Whitaker 2003, Takacs-Vesbach et al. 2008, Jones et al. 2016) or contemporary selection (e.g., Ciniglia et al. 2004, Aguilera et al. 2006, Kuang et al. 2013, Chen et al. 2016, Huang et al. 2016) has a greater effect on community assembly in extremophilic microbes.

Among the extremophiles in microbial mats are acidophilic microalgae (pH ≤4; Gimmler 2001), which are crucial in arsenic biogeochemical cycles (Qin et al. 2009) and primary production (Boyd et al. 2009, Ňancucheo and Johnson 2012). Acidophilic microalgae exhibit high affinity for CO_2 fixation and high tolerance to low pH and high concentrations of heavy metals and salts. Due to this unique combination of features, there is a growing interest in exploiting acidophilic microalgae in biotechnological applications—for example, biofuel feedstocks (Hirooka et al. 2014), wastewater treatment (Selvaratnam et al. 2014), rare element recovery (Minoda et al. 2015), and health foods (Graziani et al. 2013). Research into the biotechnological applications of acidophilic microalgae would benefit from improving our understanding of their ecology and biogeography (Varshney et al. 2015, Chen et al. 2016, Dhakar and Pandey 2016).

Few biogeographical and ecological studies have focused on acidophilic microalgae. A growing collection of studies suggests that contemporary selection is the primary force shaping their community assembly (Ciniglia et al. 2004, Novis and Harding 2007, Pinto et al. 2007, Toplin et al. 2008, Aguilera 2013, Skorupa et al. 2013, Hsieh et al. 2015). The studies published to date used low-resolution genetic markers such as 18S ribosomal RNA (e.g., Gross et al. 2001, Amaral-Zettler 2013), examined a limited number of samples from restricted geographical regions (e.g., Pinto 1993, Aguilera et al. 2010), or took morphology-based approaches that often fail to capture cryptic species (e.g., Pinto 1993, Gimmler 2001). No study has taken a next-generation sequencing (NGS)-based approach using a high-resolution genetic marker (e.g., *rbc*L) to explore the biodiversity of acidophilic microalgae at a large geographical scale.

The aims of this study are to provide a more complete picture of acidophilic microalgal biodiversity and to understand how environmental and geographical factors may influence their community assembly. We performed 454 pyrosequencing on Form I *rbc*L gene amplicons on 72 environmental samples from 12 acidic geothermal sites in three countries (Japan, Taiwan, and the Philippines) across the West Pacific Island Chain (Fig. 1). Using distance-based redundancy analyses (dbRDA) with variation partitioning and distance-decay analyses, we examined the effects of environmental variables (temperature, pH, light intensity, humidity, and microhabitat type) and geographical separation (sampling site and distance) on the community assembly of acidophilic microalgae. More specifically, we asked: What are the relative roles of contemporary selection and dispersal limitation on the community assembly of acidophilic microalgae? How do these two external factors affect the distribution of different microalgal species?

MATERIALS AND METHODS

Sample collection. Seventy-two samples that were blue-green or brown in appearance were collected from 12 sites across four geothermal regions (Akita and Ōita Prefectures in Japan, Yangmingshan National Park in Taiwan, and Negros Oriental Province in the Philippines) in the West Pacific Island Chain (Fig. 1; see Table S1 in the Supporting Information). The samples were collected from seven microhabitats in three habitats: aquatic at high light (pool and stream) and nonaquatic at high (epilith and soil) or low light (interlith, endolith, and sulfur fume; see examples in Fig. 2). The samples were classified into high and low light conditions; in high light conditions, the samples were directly exposed to sunlight without vertical shading by other rocks (i.e., pool, stream, epilith, and soil), whereas in low light conditions, the samples were shaded by other rocks or inside the rocks (i.e., interlith, endolith, and sulfur fume; Fig. 2). The nonaquatic samples were collected using a hammer and chisel (endolithic, interlithic, or epilithic samples) or a shovel (soil or sulfur fume samples), and the aquatic samples were collected using a toothbrush. Temperature was measured using a digital micro-sensor Type-K thermometer (PTM-806; Lutron Electronic, Taiwan, Taipei.). The pH of the aquatic habitats was measured using a portable waterproof pH meter (PH30; CLEAN Instruments, Taiwan, Taipei.), and the pH of the non-aquatic ones using pH test papers that range from 0.4 to 5.6 (CR, TB, BPB, and BCG in ADVANTEC®, Japan, Tokyo.). Within 1–5 d of collection, the samples were preserved in a portable fridge with ice under dark condition during the transportation. After transportation, the samples were immediately stored at −80°C.

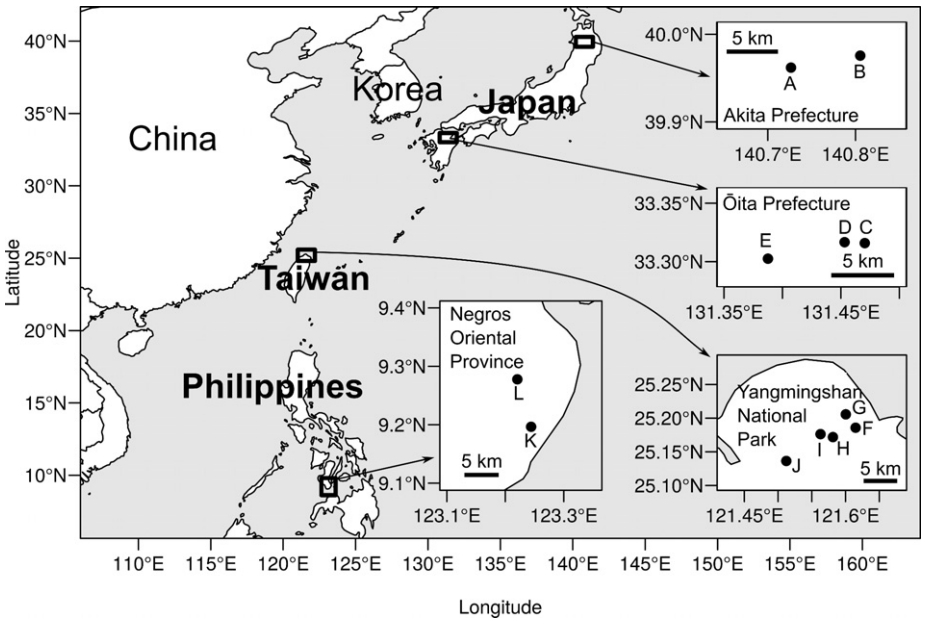

FIG. 1. Sampling sites from three countries (Japan, Taiwan, and the Philippines) in the West Pacific. Twelve geothermal sites are labeled as follows: A = Tamagawa; B = Fukenoyu; C = Umi Jigoku; D = Myoban Onsen; E = Tsukahara Onsen; F = GengZiPing; G = SiHuangPing; H = DaYouKeng; I = MaoCao; J = DiReGu; K = Dauwin; and L = Valencia. Additional information is provided in Table S1. The map was generated using the software R.

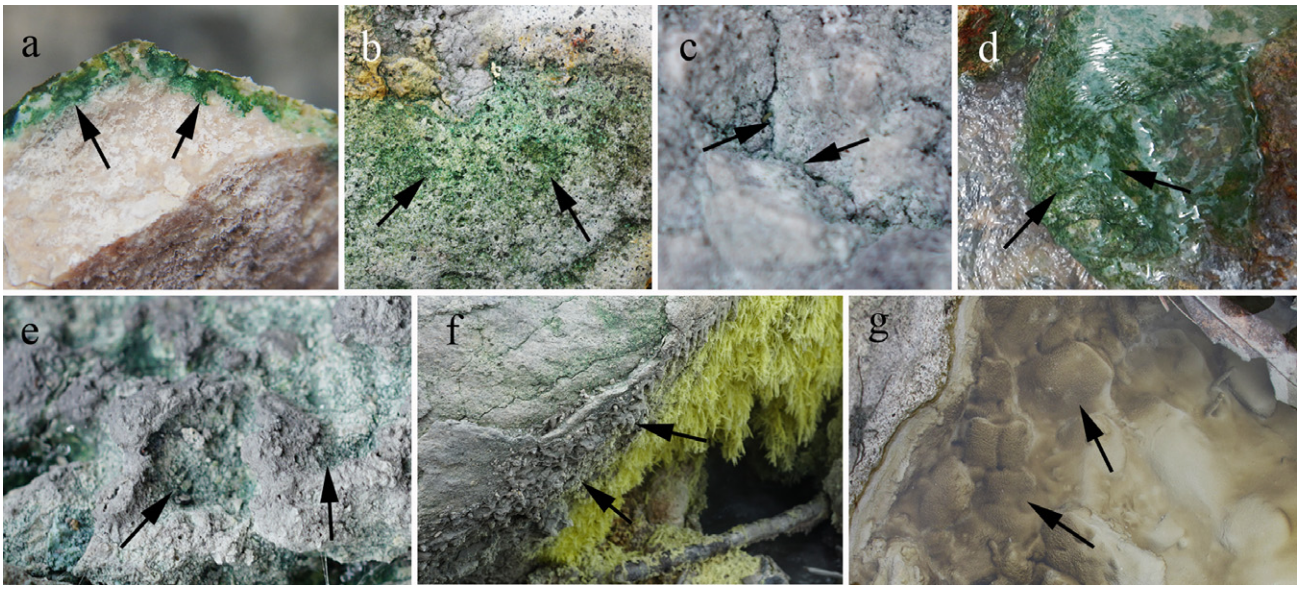

FIG. 2. Examples of the seven microhabitats in geothermal sites in Yangmingshan National Park (Taiwan) where acidophilic microalgae were collected. The arrows point to the microalgae that are blue-green or brown in appearance. (a) Endolith; (b) epilith; (c) interlith, within the rock crevice; (d) stream; (e) soil; (f) sulfur fume; and (g) pool.

During sampling, we deliberately designed a transect scheme to examine the effects of temperature gradient on the community turnover of acidophilic microalgae. We collected six samples along a temperature gradient (35.5°C–63.4°C) from six different small pools (diameter ranging between 20 and 40 cm) along a 10 m transect line at DaYouKeng, Taiwan (sample ID: TWDYK1-6; Table S1). Since these samples were physically close to each other, they presumably shared the same source of geothermal water with a low fluctuation of acidity (pH ranging between 2.18 and 2.22; Table S1). Thus, the sampling design allowed us to examine the effects of temperature on the community turnover of acidophilic microalgae, with minimal effects by other environmental factors (e.g., pH, light condition, and water chemistry).

Nucleic acid extraction and gene amplification. We extracted total DNA from the samples using the PowerSoil DNA Isolation Kit (Mo Bio, Watson Biotechnology Co. Ltd., Taiwan, Taipei.). The DNA was resuspended in 50 μL ddH_2O and then stored at −20°C. For *rbc*L gene amplification, we used

the primers RbcL1F and RbcL1R designed by Toplin et al. (2008). This primer set targets a 480 bp homologous fragment of the Form I *rbc*L gene and works effectively on microalgae but not prokaryotes (Figs. S1 and S2 in the Supporting Information). After 1.5% agarose gel electrophoresis, the expected band from the PCR was extracted using QIAEX II Gel Extraction Kit (Qiagen, USA Germantown, Maryland). The quality of the extracted product was checked using a Maestro Nanodrop Spectrophotometer (Green BioResearch Corp., Taiwan, Miaoli.). To reduce the risk of contamination, molecular analyses were conducted with sterile filtered tips using 70% ethanol-treated, autoclaved, and UV-irradiated pipettes within a laminar-flow hood.

Barcode-tagging PCR, 454 pyrosequencing, and pyrotag denoising. Barcode-tagging PCR (BT-PCR) was performed to add a unique DNA tag primer (four overhanging nucleotides) to the 5′end of primers for the amplicon of each sample (full description in Appendix S1 in the Supporting Information). Then, 200 ng of BT-PCR amplicons from each of the 72 samples were run on a Roche 454 Genome Sequencer GS Junior System at Mission Biotech Company (Taiwan, Taipei.). The raw reads and quality reports were deposited in NCBI SRA (accession SRP090343).

We preprocessed the 454 reads as follows. First, we trimmed problematic reads using a custom scheme similar to that described in Massana et al. (2015; Figs. S3 and S4 in the Supporting Information). Then, we discarded reads that were shorter than 300 bp or had an average base quality ≤30. The remaining reads were demultiplexed into 72 different groups; a read is assigned to a group if its ends are perfectly matched to the tagged barcode and primer sequence (Fig. S3). Reads having more than six gaps were discarded. The remaining reads were further cleaned using our bespoke gap-check pipeline that corrects sequencing errors and discards reads which had more than six gaps, contained premature stop codon(s), or were potentially chimeric (Figs. S3 and S4).

To assess the degree of sequencing saturation in the samples, we perform a rarefaction analysis using iNEXT. Briefly, we down-sampled the sequencing reads of each sample to various fractions, and then computed the Shannon and Simpson diversity indices for each fraction. Plateauing of the diversity indices as the sequencing depth increases indicates that a sufficient amount of sequencing was achieved to capture the sequence diversity in a given sample. The full description is provided in Appendix S1.

Algorithm-based species delimitation. We estimated the number of operational taxonomic units (OTUs) using the UCLUST tool in USEARCH v8.1.1756 (Edgar 2010). Prior to clustering, the cleaned reads were aligned using MUSCLE v3.4 (Edgar 2004) and then clustered using the UCLUST centroid-based algorithm, which assigns a read to a group based on its sequence similarity to the group's centroid (meeting a minimum threshold). Since the number of clusters could be sensitive to our choice of the similarity threshold, we first determined OTUs under various thresholds (from 70% to 97%) before deciding which threshold to ultimately use.

UCLUST groups sequences based on a simple distance score. The method does not account for among-lineage heterogeneity in the rate of sequence evolution. Failing to correct for rate heterogeneity can lead to over- or underestimation of the number of OTUs (Zhang et al. 2013). Hence, methods that model the underlying sequence evolution process have been recommended (Zhang et al. 2013). Another popular method to delimit OTUs is the generalized mixed Yule Coalescent approach (GMYC; Pons et al. 2006), which incorporates speciation and coalescent processes. GMYC determines species boundaries by inferring shifts in the branching rate in a given ultrametric phylogenetic tree. The nodes at which the branching rate transitions (i.e., from a between-species rate to a within-species rate) correspond to evolutionary significant units, or ESUs (which are called as such in order to differentiate them from the OTUs inferred using UCLUST). Here, we applied the single threshold GMYC model, which is implemented in SPLITS v1.0-18 (Monaghan et al. 2009). Unlike UCLUST that takes sequence data as input, GMYC requires an ultrametric phylogenetic tree. Therefore, we conducted an MrBayes analysis (described below) to obtain an ultrametric maximum clade credibility (MCC) tree to use as input to GMYC.

Phylogenetic analyses. To obtain a MCC tree, we reconstructed a phylogeny from longest sequences of each cluster which resulted from the UCLUST analysis that used a cut-off of 97% (denoted as $UCLUST_{97}$). We obtained a MCC tree from a Markov Chain Monte Carlo (MCMC) run with 50 million generations and a thinning frequency of 5,000 generations using MrBayes v.3.2.1 (Ronquist et al. 2012), assuming the GTR+I+G model and the independent gamma rates (relaxed clock) model. The MCC tree was built from the MCMC trees using TreeAnnotator v.1.7.5 (Drummond and Rambaut 2007), after discarding the first 90% of the trees as burn-in.

A second phylogenetic tree was reconstructed to determine whether each GMYC ESU belonged to proteobacteria, green algae, or red algae and to explore the evolutionary relationships among the ESUs. We combined the sequences representing the ESUs (again, the longest sequence from each cluster) with 127 additional *rbc*L sequences that consist of (i) the best NCBI BLASTn hit to each of the ESUs and (ii) the updated Form I *rbc*L reference sequences from these sources: proteobacteria (Tabita et al. 2008, Rasigraf et al. 2014), green algae (Lemieux et al. 2014), and red algae (Hsieh et al. 2015; Dryad: https://doi.org/10.5061/dryad.4tc28). We inferred a maximum likelihood (ML) tree using MEGA v.7.0.4 (Kumar et al. 2016), assuming the GTR+I+G model and generating 1,000 bootstrap replicates (Felsenstein 1985). In addition to the bootstrap values, the posterior probability of each node in the consensus ML tree was calculated from the post "burn-in" MCMC trees obtained earlier.

Statistical analysis. We performed three sets of statistical analyses to examine the relationship between community structure (i.e., the composition and relative abundances of ESUs) and abiotic factors and between community structure and geographic components (see the full description in Appendix S1).

To explore the effects of environmental factors and geographical separation on the relative abundances of the ESUs, we performed dbRDA with variation partitioning. We treated the following factors as explanatory variables: temperature, pH, microhabitat types (pool, stream, soil, epilith, endolith, interlith, and sulfur fume), light intensity, humidity, and geographical region (12 geothermal sites; see the contrast matrix in Table S2 in the Supporting Information; Davis 2010). The amount of variation in the relative abundances of the ESUs explained by the abiotic and geographic factors was determined using variation partitioning. We conducted dbRDA on all samples together ("between-region") and also on each region (i.e., Philippines, Japan, and Taiwan) separately ("within-region"). In addition, the decay of community similarity over geographical distance and ecological distance (i.e., with respect to all the environmental variables) among the 72 samples was examined using a Mantel test (Mantel 1967). To check whether the results of the dbRDA and distance decay of similarity (DDS) analyses were robust to the definition of OTUs, we repeated the same analyses on the sets of OTUs delimited using similarity thresholds varying from 70% to 97%.

We examined the distribution of the seven major ESUs with respect to pH, temperature, microhabitat type, and geographic region. We tested whether the relative abundances of the ESUs were significantly different among the four regions or seven microhabitats, using a 10,000-iteration Fisher-Pitman permutation test. Next, we tested whether the relationship between the relative abundance of ESUs and temperature and pH was linear, exponential, or Gaussian. We fitted the data to each of the three models and identified the best-fitting model based on the minimal corrected Akaike Information Criterion (AIC_c; Hurvich and Tsai 1989). We then tested whether there was significant support for the selected model over the null model (i.e., zero correlation). Multiple test correction was performed using false discovery rate (Storey and Tibshirani 2003).

RESULTS

Processed sequence data. Using the 454 pyrosequencing, we obtained 442,721 *rbc*L reads, of which 218,092 remained after filtering. The minimum, median, and maximum read lengths were 329, 470, and 482 bp respectively (Fig. S5 in the Supporting Information). According to a rarefaction analysis, we achieved enough sequencing to capture most of the microalgae diversity in our samples (Table S3 in the Supporting Information). Also, we removed potential contaminant OTUs prior to the statistical analyses (blue/one asterisk in Fig. 3; resulting in filtered GMYC data set, or GMYCft). Even including the contaminant OTUs, our qualitative results did not change (see the GMYC data set in Fig. 5).

The diversity and relative abundances of ESUs. We analyzed the cleaned reads to (i) determine the number of ESUs, (ii) examine the diversity and relative abundances of the ESUs, and (iii) uncover potentially undocumented species.

First, we inferred OTUs using UCLUST under varying sequence similarity-based definitions (Table 1). We discovered 11 OTUs at the most relaxed threshold (i.e., 70%) and 391 OTUs at the strictest (i.e., 97%; Table 1). Distance-based estimates by UCLUST may be inaccurate, because they do not account for speciation and coalescent processes (Fujita et al. 2012). Hence, we also applied the GMYC method, and found at least 51 ESUs (Table 1). In the GMYC analysis that resulted in 51 ESUs, the null model (pure coalescent) was significantly rejected in favor of the alternative (which allows for speciation) ($\ln L_{alternative} = 2{,}848 > \ln L_{null} = 2{,}834$, $P < 0.001$); this result indicates that crude distance-based methods can grossly overestimate the number of OTUs. For subsequent analysis, we took the most conservative estimate (i.e., 51 ESUs)—a practice recommended by Carstens et al. (2013).

Second, we inferred the phylogenetic tree of the 51 ESUs in the reference background of 127 Form I *rbc*L sequences gathered from the literature and the best BLAST hits of the ESUs (Fig. 3; Fig. S6 in the Supporting Information). There were many more phylogenetically informative sites in the *rbc*L fragment in red algae (138–187) than in green algae (49–98; Table S4 in the Supporting Information). We found four well-supported groups of Form I *rbc*L (Fig. 3; Fig. S6). The first group consists of 28 ESUs including four taxa belonging to Streptophyta, two to Euglenozoa, and 21 to Chlorophyta, as well as one cyanobacterium. The second group consists of four proteobacteria. The third comprises 10 ESUs with the most diverse subgroup in Cyanidiales (ESU33-38), followed by two diatoms (ESU40-41), one mesophilic red alga (ESU39) and one Eustigmatophyceae (ESU42). The fourth group contains nine proteobacteria. Among all microbes, microalgae are the most diverse group (35 ESUs), followed by prokaryotes (13 ESUs).

Among the microalgae detected, Rhodophyta was the most dominant group in terms of the total number of reads (84%), followed by Chlorophyta (13%) and diatoms (2%). When classifying into major (>1% relative abundance) or rare ESUs (Fig. 3; Fig. S7 in the Supporting Information), only seven microalgae (four red algae [ESU34-35 and ESU37-38], two green algae [ESU22 and ESU26], and one diatom [ESU40]) were considered major ESUs. Most of the major ESUs were shared among different geothermal regions to different degrees (Table S5 in the Supporting Information; Fig. S8 in the Supporting Information). The overall trend was that the relative abundance of rare microalgal ESUs (i.e., eight green algae [ESU5, 8, 9, 11, 17, 20, 24, 31] and one red alga [ESU36]) was higher in soil and epilithic samples at lower temperatures (18°C–33°C; Table S5). In contrast, two rare autotrophic proteobacteria (ESU44 and ESU50) were more abundant at higher temperatures, in a stream (39°C) and an endolithic (60°C) sample respectively (Table S5). These data suggest that the species richness of microalgae was higher in the nonaquatic environments at lower temperatures, but the diversity pattern of autotrophic prokaryotes was unclear due to insufficient data.

Finally, our analysis revealed undocumented acidophilic autotrophic microbes. The *rbc*L sequences that exhibit low similarity to the sequences in GenBank may belong to novel microalgal species (similarity <90%) or proteobacterial species (similarity ~81%, ranging from 75% to 96%; Table 2).

Effects of contemporary selection and geographical separation on community assembly. The dbRDA with variation partitioning showed the first two axes to explain 35% of the variance in the GMYCft data set with the first axis strongly associated with geographical distance and the second one with the environmental factors (Fig. 4a). Between different geothermal regions, geographical separation explained more variance than the environmental factors (19.63% vs. 14.85%; Fig. 4b); however, within each geothermal region, more variance was explained by environmental factors than geographical separation (Philippines: 12 vs. 0%; Taiwan: 29 vs. 0%; Japan: 24 vs. 10%). The

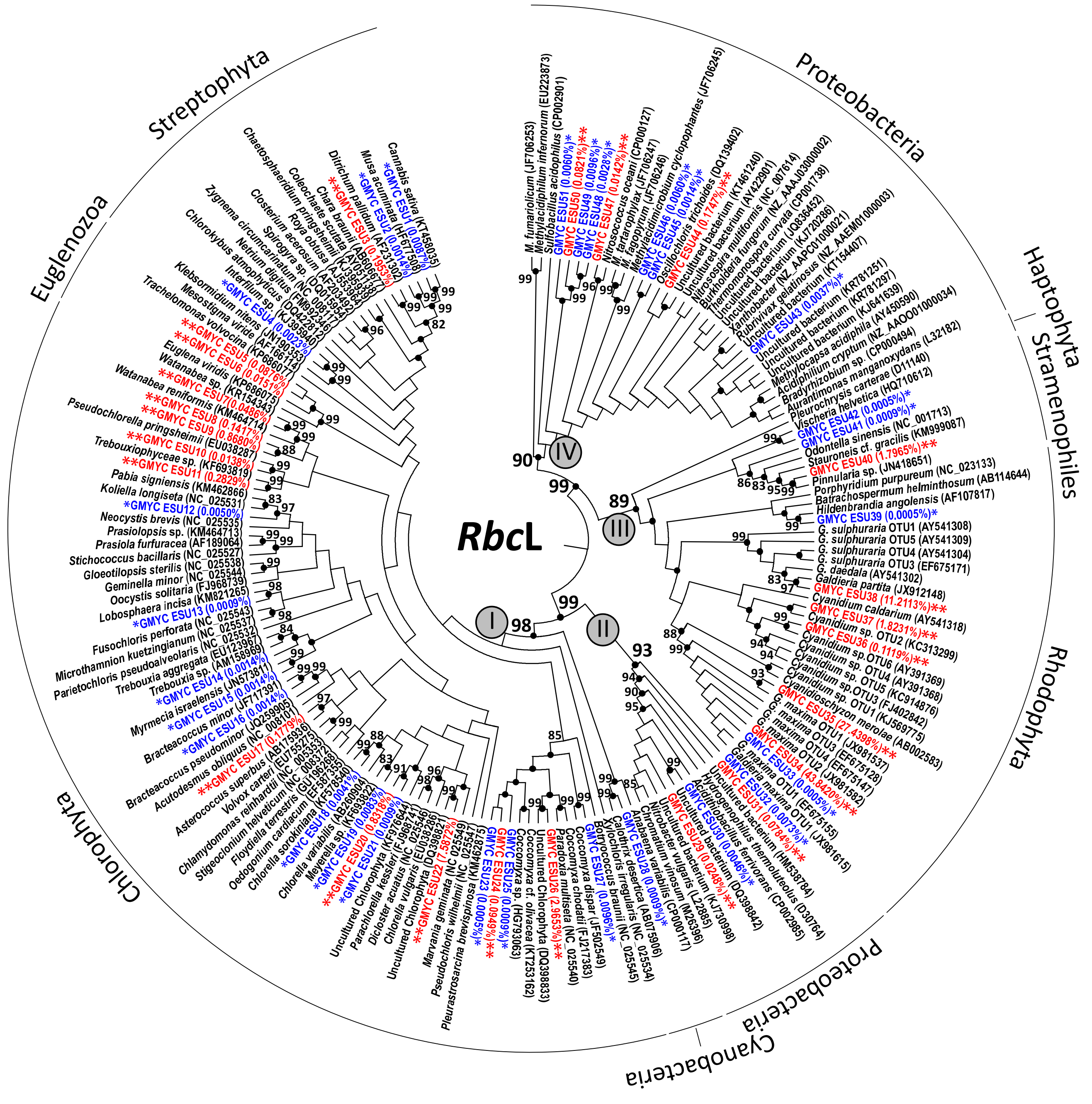

FIG. 3. Maximum likelihood *rbc*L tree of the 51 GMYC ESUs. Bootstrap support (>50%) is shown on the branches, and posterior probability (>80%) shown as dots. Relative abundances and GenBank accessions are indicated in parentheses. The ESUs used for subsequent analyses are denoted by (**), whereas the potential nonacidophilic ESUs are denoted by (*). The phylogram with branch lengths is provided in Figure S5. [Color figure can be viewed at wileyonlinelibrary.com]

similarity of communities among the samples was significantly influenced by geographical separation (across the 12 geothermal sites) as well as pH and temperature ($P < 0.05$; Table S6 in the Supporting Information); no significant effect of microhabitat type, light, or humidity was observed ($P > 0.05$; Table S6). The first and second axes of dbRDA were attributed primarily to the seven major ESUs (Table S7 in the Supporting Information). Between regions, we found that community similarity showed a stronger negative correlation with geographical distance (Pearson's $r = -0.320$, $P < 0.001$; Fig. S9a in the Supporting Information) than ecological distance (Pearson's $r = -0.137$, $P = 0.002$; Fig. S9b).

The patterns observed in GMYCft were tested under different OTU definitions. The variance of community structure explained by the environmental factors increased as the sequence similarity cutoff

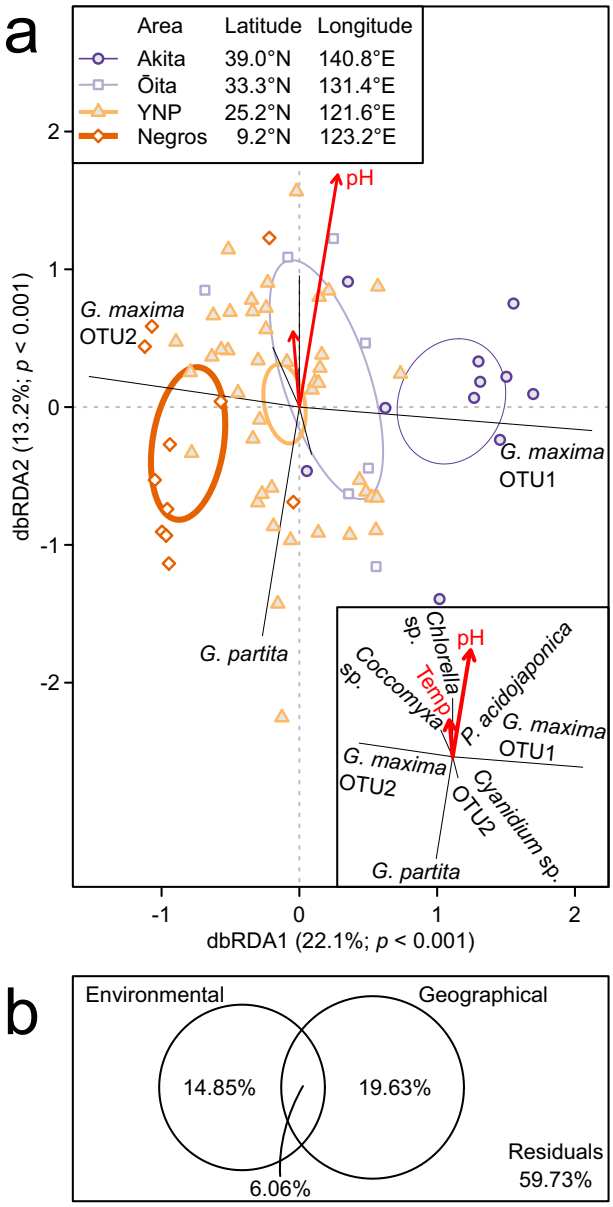

FIG. 4. (a) Results of dbRDA. Lines indicate the major ESUs, and the arrows indicate significant environmental variables. The ellipses (or the 95% confidence limits around the enclosed centroid, with radiuses as the standard error multiplied by 2.4477) indicate different geographical regions. (b) Explained variance of the environmental and geographical components using dbRDA with variation partitioning. [Color figure can be viewed at wileyonlinelibrary.com]

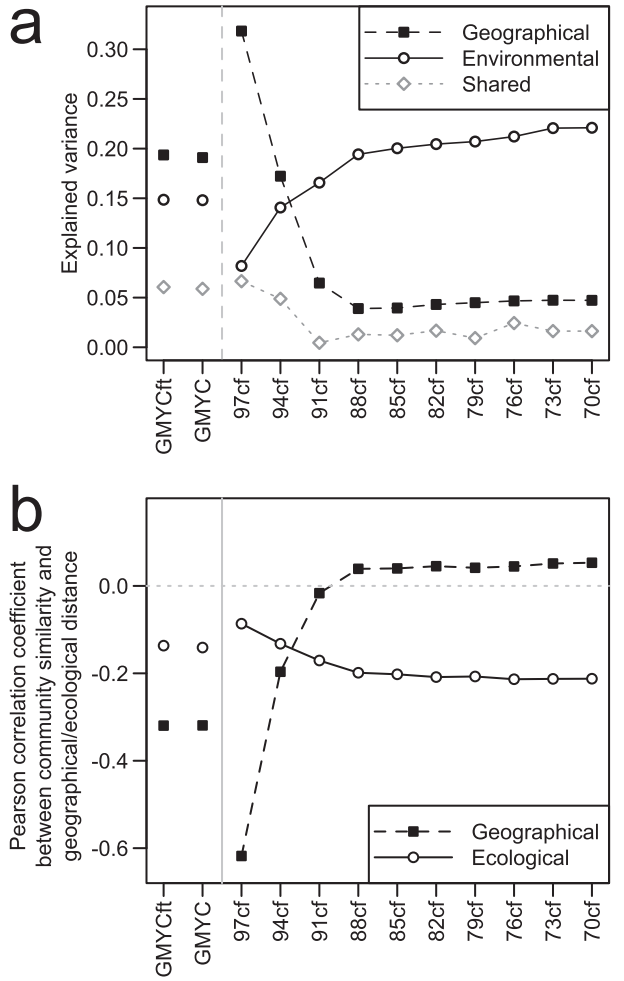

FIG. 5. dbRDA variation (a) explained by environmental factors and geographical separation, and the correlation coefficients of the DDS analyses (b) at various UCLUST sequence similarity cut-offs.

was lowered, but the variance explained by the geographical component decreased dramatically (Fig. 5a). Also, geographical DDS became much more apparent at high sequence similarity cut-offs (≥94%), but ecological DDS did not seem to be affected (Fig. 5b).

Ecological and biogeographical distribution of the seven major ESUs. We examined the distribution of the seven major ESUs by associating their relative abundance with four geographical regions and different environmental variables. The two red algae, ESU34 (*Galdieria maxima* OTU1) and ESU35 (*G. maxima* OTU2), showed different geographical preferences. ESU34 (*G. maxima* OTU1) is more dominant in the northern geothermal regions, whereas ESU35 (*G. maxima* OTU2) is more dominant in the southern regions (Fig. 6). ESU35 (*G. maxima* OTU2) also seems to prefer higher temperatures (M_{linear}: $AIC_c = 16.3$, $m = 0.007$, $P = 0.039$; Fig. 6).

The distribution of three other ESUs (ESU37, ESU38, and ESU22) was explained by the environmental factors. ESU37 (*Cyanidium* sp. OTU2) and ESU38 (*Galdieria partita*) are more dominant at lower temperatures ($M_{exponential}$: $AIC_c = -105$ and -0.694, $r = -0.118$ and -0.024, $P < 0.001$ and $=0.039$; Fig. 6). Additionally, *G. partita* is more dominant under more acidic environments ($M_{exponential}$: $AIC_c = -39.8$, $r = -0.516$, $P < 0.001$; Fig. 6). The

TABLE 1. Number of OTUs inferred using distance-based and coalescent-based approaches. OTUs were inferred using UCLUST, and, equivalently, ESUs from GMYC.

Threshold cut-off (similarity level % in subscript)	Number of OTUs
$UCLUST_{70}$	11
$UCLUST_{73}$	12
$UCLUST_{76}$	19
$UCLUST_{79}$	23
$UCLUST_{82}$	30
$UCLUST_{85}$	43
$UCLUST_{88}$	56
$UCLUST_{91}$	65
$UCLUST_{94}$	95
$UCLUST_{97}$	391
$UCLUST_{97}$ + GMYC	51 (ESUs)

Chlorella-like ESU22 (tentatively, *Chlorella* sp.) was more dominant at moderate temperatures (~38°C) ($M_{gaussian}$: $AIC_c = -19.8$, $h = 0.400$, $P = 0.001$; Fig. 6) and less acidic environments (M_{linear}: $AIC_c = -30.3$, $m = 0.087$, $P < 0.001$; Fig. 6).

The last two ESUs, ESU26 (*Coccomyxa* sp.) and ESU40 (*Pinnularia acidojaponica*; see Fig. S10 in the Supporting Information for its morphologies), showed no preference within the environmental ranges examined in this study.

Community turnover of acidophilic microalgae along temperature gradient. In the survey of six pool samples along a temperature gradient (35.5°C–63.4°C) at DaYouKeng, Taiwan, the species diversity of microalgae increased from 4 or 5 ESUs to 7 ESUs ($q0$) as temperature decreased (Fig. 7; Table S5). The diversity indexes were the lowest at the highest temperature (Shannon: $q1 = 1.107$; Simpson: $q2 = 1.035$) and the highest at the lowest temperature (Shannon: $q1 = 3.513$; Simpson: $q2 = 2.778$).

Five ESUs were largely responsible for the community turnover. ESU34 (*Galdieria maxima* OTU1) and ESU35 (*G. maxima* OTU2) were the dominant microbes at 42.7°C and above, indicating that the two *G. maxima* might prefer higher temperatures than the rest of acidophilic microalgae in the regions surveyed. At lower temperatures, ESU22 (*Chlorella* sp.) was the most abundant, followed by ESU40 (*Pinnularia acidojaponica*) and ESU38 (*Galdieria partita*). Furthermore, two *Coccomyxa* green algae (ESU24 and ESU26) were present at low relative abundances at the lower temperatures (Fig. 7).

DISCUSSION

Our ESU delimitation analyses using *rbc*L revealed at least 51 ESUs of autotrophs in acidic geothermal regions. This result should be interpreted primarily with microalgae in mind, because the detection of autotrophic prokaryotes is less effective due to lower primer efficacy. Previous studies found about four ESUs in the same acidic geothermal regions, also using *rbc*L (Toplin et al. 2008, Hsieh et al. 2015). Therefore, we raised the count of microalgal ESUs from four to 34 (discounting the ESUs of autotrophic prokaryotes and land plants), suggesting that the species diversity of microalgae in acidic geothermal habitats is substantially greater than previously appreciated.

There are four different forms of CO_2-fixed RubisCO (Form I, II, II/III, and III; Tabita et al. 2008). Form I *rbc*L sequences across different autotrophic microbes (e.g., photoautotrophic microalgae and chemolithoautotrophic proteobacteria in this study) have been categorized into five types (Tabita et al. 2008, Rasigraf et al. 2014): Forms IA and IB are green plastid-associated and Forms IC to IE are red plastid-associated. In our survey, we detected all five types of Form I *rbc*L. Our data show that while green algae are the most *speciose* group, Cyanidiales is the most *dominant* group. Given that five out of seven major microalgae are red plastid-associated, this result suggests that microalgae with plastids of the red algal type are the dominant ones thriving in acidic geothermal habitats.

Considerably fewer microalgae species have been detected in acidic environments than in neutral environments, in which as many as a few hundred species of microalgae have been reported (e.g., Ferris et al. 2005, Walker et al. 2005, Aguilera et al. 2010). Unsurprisingly, this observation is consistent with the view that high acidity is too harsh for most microalgae (Witkowski et al. 2011). Nonetheless, we detected hidden diversity of acidophilic microalgae with some ESUs exhibiting low sequence identity to the *rbc*L sequences in GenBank; we speculate that these ESUs may represent species yet to be recognized. One example is an acidophilic Cyanidiales species (ESU36; *Cyanidium* sp. OTU6) that appears to have never been reported anywhere in the world (Hsieh et al. 2015).

The discovery of eightfold more microalgal ESUs in this study than in previous studies is not surprising. Sanger sequencing used in previous studies is effective for capturing dominant species in environmental samples as opposed to NGS from this study which can detect far less abundant OTUs. For example, we detected the presence of four dominant red algae (i.e., two *Galdieria maxima* OTUs, *G. partita*, and *Cyanidium* sp. OTU2), which were also detected using a Sanger-cloning method in our previous study (Hsieh et al. 2015). But, we were able to find 12 ESUs of green algae and one ESU of red algae (ESU36) that occur at low relative abundances (<1%). More NGS-based surveys, such as the present one, are important to reveal the hidden biodiversity of acidophilic microalgae and more broadly extremophilic microalgae.

Our dbRDA and DDS analysis revealed that geographical distance predicts the similarity among communities of acidophilic microalgae across different geothermal regions. This result suggests that dispersal limitation is present not only in extremophilic prokaryotes (see examples in Whitaker 2003, Jones

TABLE 2. The top BLASTn hit with the highest bitscore and the lowest E-value of each of the 51 GMYC ESUs.

GMYC unit	Top hit	% ID	Bitscore/E-value	Taxonomic group	*rbc*L type
ESU1	*Cannabis sativa* (KT458035)[a]	100	891/0	Streptophyta	B
ESU2	*Musa acuminate* (HF677508)[a]	100	863/0	Streptophyta	B
ESU3	*Ditrichum pallidum* (AF231302)	97	813/0	Streptophyta	B
ESU4	*Interfilum* sp. (KJ395940)	95	769/0	Streptophyta	B
ESU5	*Trachelomonas volvocina* (KP686077)	88	531/6.00E-147	Euglenozoa	B
ESU6	*Euglena viridis* (KP686075)	87	510/9.00E-141	Euglenozoa	B
ESU7	*Watanabea* sp. (KR154343)	98	828/0	Chlorophyta	B
ESU8	*Watanabea reniformis* (KM464714)	87	525/3.00E-145	Chlorophyta	B
ESU9	*Bracteacoccus pseudominor* (JQ259905)	88	547/7.00E-152	Chlorophyta	B
ESU10	*Pseudochlorella pringsheimii* (EU038287)	99	841/0	Chlorophyta	B
ESU11	Trebouxiophyceae sp. (KF693819)	97	802/0	Chlorophyta	B
ESU12	*Pabia signiensis* (KM462866)	92	669/0	Chlorophyta	B
ESU13	*Lobosphaera incise* (KM821265)	99	837/0	Chlorophyta	B
ESU14	*Trebouxia* sp. (AM158969)	88	553/1.00E-153	Chlorophyta	B
ESU15	*Myrmecia* sp. (JN573811)	97	815/0	Chlorophyta	B
ESU16	*Bracteacoccus minor*(JF717391)	95	747/0	Chlorophyta	B
ESU17	*Asterococcus superbus* (AB175936)	92	675/0	Chlorophyta	B
ESU18	*Chlorella sorokiniana* (KF578540)	99	852/0	Chlorophyta	B
ESU19	*Meyerella* sp. (KF693822)	94	725/0	Chlorophyta	B
ESU20	Uncultured Chlorophyta (KF976644)	91	575/3.00E-160	Chlorophyta	B
ESU21	Uncultured Chlorophyta (KF976644)	90	610/8.00E-171	Chlorophyta	B
ESU22[b]	Uncultured Chlorophyta (DQ398821)[c]	98	665/0	Chlorophyta	B
ESU23	*Coccomyxa* sp. (HG793063)	91	616/2.00E-172	Chlorophyta	B
ESU24	*Coccomyxa* sp. (HG793063)	92	667/0	Chlorophyta	B
ESU25	*Coccomyxa* cf. *olivacea* (KT253162)	89	595/2.00E-166	Chlorophyta	B
ESU26[b]	Uncultured Chlorophyta (DQ398833)[c]	94	736/0	Chlorophyta	B
ESU27	*Coccomyxa dispar* (JF502549)	91	531/7.00E-147	Chlorophyta	B
ESU28	*Calothrix desertica* (AB075906)	94	710/0	Cyanobacteria	B
ESU29	Uncultured bacterium (KJ730998)	83	379/2.00E-101	Proteobacteria	A
ESU30	Uncultured bacterium (DQ398842)[c]	96	667/0	Proteobacteria	A
ESU31	Uncultured bacterium (HM538784)	75	198/7.00E-47	Proteobacteria	A
ESU32	*Acidithiobacillus ferrivorans* (CP002985)	77	198/8.00E-47	Proteobacteria	A
ESU33	*Galdieria maxima* (JX981615)[c]	96	782/0	Rhodophyta	D
ESU34[b]	*Galdieria maxima* (EF675155)[c]	96	752/0	Rhodophyta	D
ESU35[b]	*Galdieria maxima* (JX981537)[c]	99	854/0	Rhodophyta	D
ESU36	*Cyanidium* sp. (KC914876)[c]	85	484/5.00E-133	Rhodophyta	D
ESU37[b]	*Cyanidium* sp. (KC313299)[c]	96	776/0	Rhodophyta	D
ESU38[b]	*Galdieria partita* (JX912148)[c]	99	870/0	Rhodophyta	D
ESU39	*Hildenbrandia angolensis* (AF107817)	93	553/1.00E-153	Rhodophyta	D
ESU40[b]	*Pinnularia* sp. (JN418651)	98	826/0	Bacillariophyta	D
ESU41	*Stauroneis* cf. *gracilis* (KM999087)	88	560/9.00E-156	Bacillariophyta	D
ESU42	*Vischeria helvetica* (HQ710612)	98	819/0	Heterokontophyta	D
ESU43	Uncultured bacterium (KJ641639)	85	364/5.00E-97	Proteobacteria	C&E
ESU44	Uncultured bacterium (KT461240)	86	390/9.00E-105	Proteobacteria	C&E
ESU45	Uncultured bacterium (AY422901)[c]	79	233/2.00E-57	Proteobacteria	C&E
ESU46	Uncultured bacterium (KR781251)	81	261/8.00E-66	Proteobacteria	C&E
ESU47	Uncultured bacterium (KR781297)	81	259/2.00E-65	Proteobacteria	C&E
ESU48	Uncultured bacterium (KJ720286)	75	137/1.00E-28	Proteobacteria	C&E
ESU49	Uncultured bacterium (KT154407)	80	251/5.00E-63	Proteobacteria	C&E
ESU50	Uncultured bacterium (JQ836452)	75	165/6.00E-37	Proteobacteria	C&E
ESU51	*Thermomonospora curvata* (CP001738)	78	248/6.00E-62	Proteobacteria	C&E

[a]Those two land plants were used as internal controls for the similarity threshold cut-off (<0.01%) of potential contaminants since both do not grow in acidic geothermal regions.
[b]Italic boldface indicates major ESUs with relative abundance ≥1%.
[c]The best BLASTn hits are associated with autotrophic microbes from acidic geothermal regions.

et al. 2016) but also extremophilic eukaryotes (this study). The result is largely attributed to the distribution of ESU34 (*Galdieria maxima* OTU1) and ESU35 (*G. maxima* OTU2). We found that these two red algae are generally dominant, but ESU34 (*G. maxima* OTU1) is more dominant than ESU35 (*G. maxima* OTU2) in the northern parts of the West Pacific Island Chain, whereas ESU35 (*G. maxima* OTU2) is more dominant in the southern parts. This pattern was not found in previous studies (Toplin et al. 2008, Ciniglia et al. 2014, Hsieh et al. 2015), probably due to low sample sizes and the limitations of Sanger sequencing-based approaches. With an increase of sample size and the benefits of NGS, we were able to observe the biogeographical pattern in this study.

Dispersal limitation among acidophilic microalgae communities has been reported at the inter-continental scale (Gross et al. 2001). Certain species,

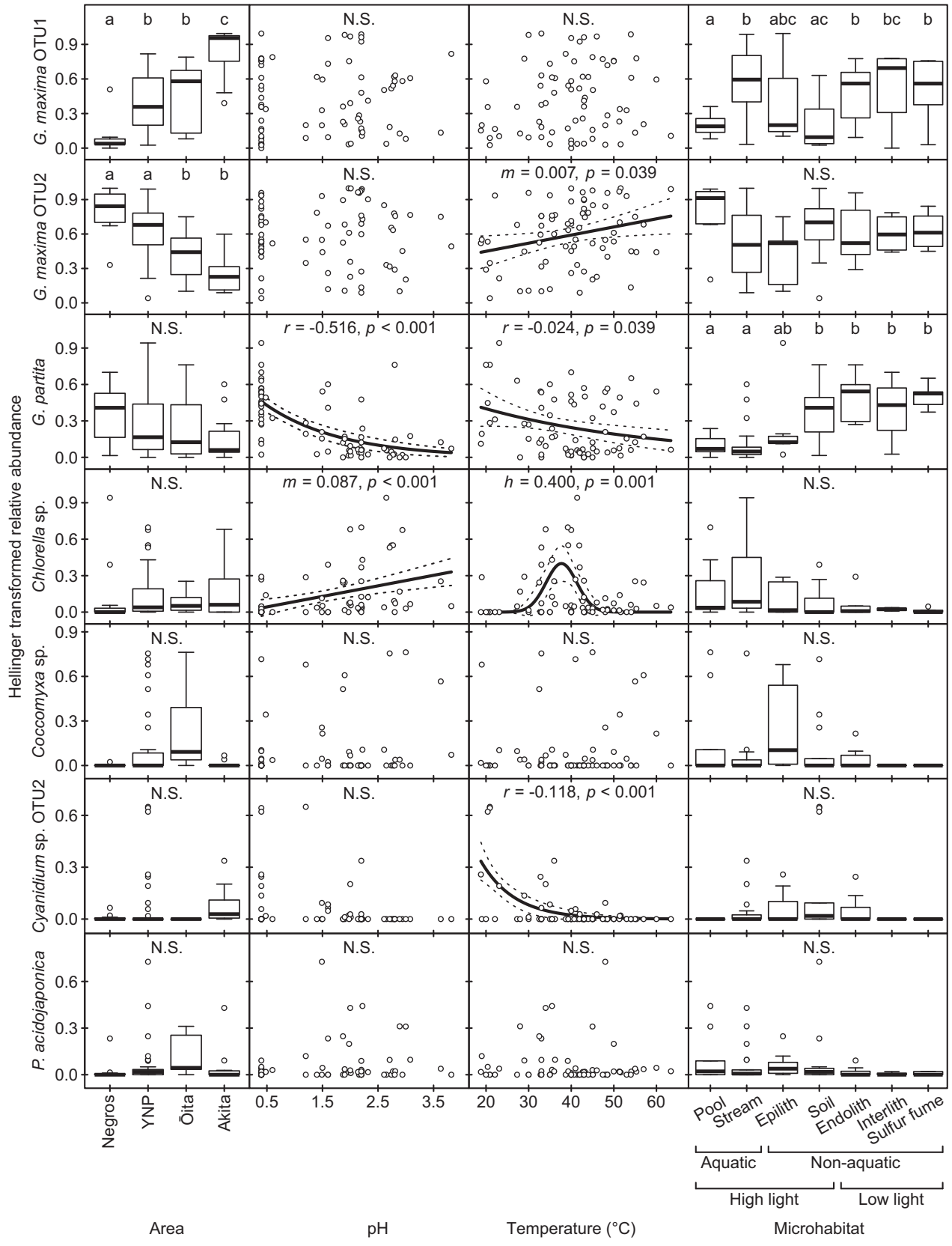

FIG. 6. Relative abundance of each major ESU in relation to geographical (region) and environmental factors (pH, temperature, and the microhabitat type). The letters (a, b, and c) denote different statistically significant groups with $P < 0.05$. "N.S." denotes not statistically significant.

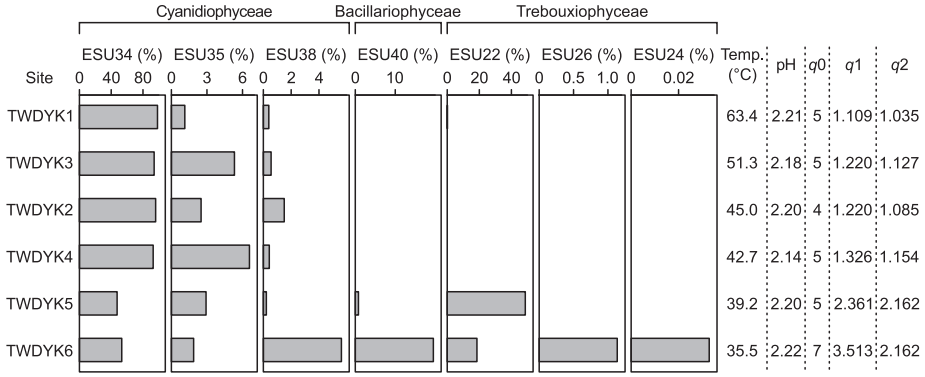

FIG. 7. Community turnover of acidophilic microalgae along a temperature gradient. Three diversity indexes ($q0$, $q1$, and $q2$) were estimated using iNEXT.

however, seem to be dispersed globally (Toplin et al. 2008, Ciniglia et al. 2014, Hsieh et al. 2015). For instance, it has been speculated that *Cyanidioschyzon merolae* and *Galdieria sulphuraria* might be distributed in distantly separated continents via dispersion by volcanic explosions (Reeb and Bhattacharya 2010), the migration of water birds (Castenholz and McDermott 2010), or air currents (Ciniglia et al. 2014). Our data suggest that dispersal limitation can affect the spatial distribution of acidophilic microalgae at a between-region scale (a spatial scale that is much smaller than the continental level). The pattern of geographical distribution observed here is driven mostly by ESU34 (*G. maxima* OTU1) and ESU35 (*G. maxima* OTU2). A probable explanation for dispersal constraint on ESU34 (*G. maxima* OTU1) and ESU35 (*G. maxima* OTU2) may be their lack of tolerance for desiccation, as often seen in Cyanidiales (Toplin et al. 2008); this may also mean that their dispersal by air is limited. Another explanation might be the lack of physiological flexibility, as hypothesized by Hsieh et al. (2015). Moreover, we found no evidence for dispersal limitation in other common acidophilic microalgae besides ESU34 (*G. maxima* OTU1) and ESU35 (*G. maxima* OTU2). Indeed, dispersal by air or birds is reportedly common in different groups of microalgae (Proctor 1959, Sherwood et al. 2017). Consistent with this observation, we found two acidophilic green algae (ESU7 and ESU11) in some air dust samples that we collected as part of our fieldwork (Fig. S1).

In addition to dispersal limitation, contemporary selection influences the community assembly of acidophilic microalgae at a within-region scale (in this study, within an island). The dbRDA results demonstrated that pH and temperature are key predictors of community similarity at a within-region scale. For instance, ESU38 (*Galdieria partita*) prefers extremely acidic environments at lower temperatures, and ESU22 (*Chlorella* sp.) prefers less acidic conditions at moderate temperatures. Interestingly, Doemel and Brock (1970) reported that Cyanidiales can grow at temperatures as high as ~60°C; here, we see that they can grow at even greater temperatures (e.g., ESU35 [*G. maxima* OTU2]). In this study, we have examined just a handful of environmental factors that underlie the biogeographical distribution of microbial organisms. It is quite possible that some of the patterns found here might be correlated to abiotic factors that we have not measured or even considered. One such factor is organic carbon content, which has been suggested to influence the distribution of different species of Cyanidiales (Pinto 1993). Another factor is the intensity of UV light, which changes along latitude. The growth of acidophilic microalgae (e.g., Cyanidiales) is sensitive to UV light in geothermal regions (Lehr et al. 2007). Revisiting the biogeographical questions that we have addressed here with fuller ecological data is an important direction for future studies.

We further explored ESU distribution along the temperature gradient at a local scale. We found more complex ESU compositions in cooler environments, consistent with the trend observed in different case studies (e.g., Huss et al. 2002, Walker et al. 2005, Ueda et al. 2009). This observation could be explained by the adaptation of Cyanidiales to extreme habitats. At high temperatures Cyanidiales have a competitive advantage over other algae (e.g., green algae and diatoms) and thus tend to dominate (thereby decreasing species diversity), but at lower temperatures the advantage lessens and consequently other algae can compete with Cyanidiales better (thereby increasing species diversity).

The dbRDA revealed that unlike temperature and pH, the type of microhabitat (humidity and exposure to light) did not seem to affect the community structure of acidophilic microalgae. However, previous work suggested that humidity and exposure to light have an effect, but the data were limited in terms of sample size and the geographical scale of sample collection (Ciniglia et al. 2004, Skorupa et al. 2013, Hsieh et al. 2015). Nevertheless, further work is needed to understand the role of microhabitat type—humidity, exposure to light, water motion,

or another aspect of microhabitats that we have yet to consider.

According to the literature, contemporary selection influences the assembly of communities of microbes found in acidic habitats (e.g., acid-mine drainage and acidic geothermal regions) more strongly than dispersal limitation (e.g., Pinto 1993, Ciniglia et al. 2004, 2014, Toplin et al. 2008, Kuang et al. 2013, Skorupa et al. 2013, Hsieh et al. 2015, Chen et al. 2016, Huang et al. 2016). In this study, we found that geographic separation is a stronger driver of the community assembly of acidophilic microalgae than environmental factors at a between-region scale (i.e., between the islands surveyed here), but at within-region and local scales contemporary selection is the main driver. This result shows that both dispersal limitation and contemporary selection affect community assembly, but their relative contribution varies depending on the geographical scale.

Lu et al. (2016) reported that contemporary selection affects community assembly more strongly at more relaxed definitions of OTUs (i.e., lower similarity thresholds). Under stricter definitions (i.e., higher similarity thresholds) more OTUs are inferred and changes in their relative abundances are more likely to be detected; therefore, there is more statistical power to detect geographical effects on community assembly (Hanson et al. 2012). These observations suggest that there was more statistical power to detect biogeographic patterns using stricter OTU definitions. Alternatively, these observations may also suggest that environmental effects are subtler and therefore more difficult to detect than geographical effects. Consistent with these observations, we found that under more relaxed thresholds for OTU delimitation, the effects of geographical separation on the community assembly can be missed. Additionally, the choice of the genetic marker is important. In previous studies, the 18S ribosomal RNA has been used, but this marker provides low resolution for OTU delimitation in microalgae compared to *rbc*L (e.g., Toplin et al. 2008). Hence, the use of the *rbc*L marker may be another reason that we discerned a stronger influence on community assembly by geographical distance than environmental factors. We recommend the approach taken herein to study the relationship between abiotic and geographical factors and microbial community assembly (i.e., by using a suitable genetic marker and by employing different methods as well as various thresholds for OTU delimitation).

We thank Pin-Chen Chen and Dr. Yin-Ru Chiang for helping with the primer efficacy tests. We also thank Drs. Sarah P. Otto, Chih-Hao Hsieh, and Hsiao-Pei Lu for their constructive comments. SHZ was supported by the Canadian Institutes of Health Research Doctoral Research Award. The work was supported by grants from the Ministry of Science and Technology, Taiwan and the Jiayi Foundation, Taiwan to SLL. The authors have declared no competing interests in this work.

Aguilera, A. 2013. Eukaryotic organisms in extreme acidic environments, the Río Tinto case. *Life* 3:363–74.

Aguilera, A., Manrubia, S. C., Gómez, F., Rodríguez, N. & Amils, R. 2006. Eukaryotic community distribution and its relationship to water physicochemical parameters in an extreme acidic environment, Río Tinto (southwestern Spain). *Appl. Environ. Microbiol.* 72:5325–30.

Aguilera, Á., Souza-Egipsy, V., González-Toril, E., Rendueles, O. & Amils, R. 2010. Eukaryotic microbial diversity of phototrophic microbial mats in two Icelandic geothermal hot springs. *Int. Microbiol.* 13:21–32.

Amaral-Zettler, L. A. 2013. Eukaryotic diversity at pH extremes. *Front. Microbiol.* 3:1–17.

Boyd, E. S., King, S., Tomberlin, J. K., Nordstrom, D. K., Krabbenhoft, D. P., Barkay, T. & Geesey, G. G. 2009. Methylmercury enters an aquatic food web through acidophilic microbial mats in Yellowstone National Park, Wyoming. *Environ. Microbiol.* 11:950–9.

Carstens, B. C., Pelletier, T. A., Reid, N. M. & Satler, J. D. 2013. How to fail at species delimitation. *Mol. Ecol.* 22:4369–83.

Caruso, T., Chan, Y., Lacap, D. C., Lau, M. C. Y., McKay, C. P. & Pointing, S. B. 2011. Stochastic and deterministic processes interact in the assembly of desert microbial communities on a global scale. *ISME J.* 5:1406–13.

Castenholz, R. W. & McDermott, T. R. 2010. The cyanidiales: ecology, biodiversity, and biogeography. *In* Seckbach, J. & Chapman, D. J. [Eds.] *Red Algae in the Genomic Age*. Springer, Dordrecht, The Netherlands, pp. 357–71.

Chen, L. X., Huang, L. N., Méndez-García, C., Kuang, J. L., Hua, Z. S., Liu, J. & Shu, W. S. 2016. Microbial communities, processes and functions in acid mine drainage ecosystem. *Curr. Opin. Biotechnol.* 38:150–8.

Ciniglia, C., Yang, E. C., Pollio, A., Pinto, G., Iovinella, M., Vitale, L. & Yoon, H. S. 2014. Cyanidiophyceae in Iceland: plastid *rbc* L gene elucidates origin and dispersal of extremophilic *Galdieria sulphuraria* and *G. maxima* (Galdieriaceae, Rhodophyta). *Phycologia* 53:542–51.

Ciniglia, C., Yoon, H. S., Pollio, A., Pinto, G. & Bhattacharya, D. 2004. Hidden biodiversity of the extremophilic Cyanidiales red algae. *Mol. Ecol.* 13:1827–38.

Davis, M. J. 2010. Contrast coding in multiple regression analysis: strengths, weaknesses, and utility of popular coding structures. *J. Data Sci.* 8:61–73.

Dhakar, K. & Pandey, A. 2016. Wide pH range tolerance in extremophiles: towards understanding an important phenomenon for future biotechnology. *Appl. Microbiol. Biotechnol.* 100:2499–510.

Doemel, W. N. & Brock, T. D. 1970. The upper temperature limit of *Cyanidium caldarium*. *Arch. Mikrobiol.* 72:326–32.

Drummond, A. J. & Rambaut, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–7.

Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–1.

Fagervold, S. K., Bourgeois, S., Pruski, A. M., Charles, F., Kerhervé, P., Vétion, G. & Galand, P. E. 2014. River organic matter shapes microbial communities in the sediment of the Rhône prodelta. *ISME J.* 8:2327–38.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–91.

Ferris, M., Sheehan, K., Kuhl, M., Cooksey, K., Wigglesworth-Cooksey, B., Harvey, R. & Henson, J. 2005. Algal species and light microenvironment in a low-pH, geothermal microbial mat community. *Appl. Environ. Microbiol.* 71:7164–71.

Fujita, M. K., Leaché, A. D., Burbrink, F. T., McGuire, J. A. & Moritz, C. 2012. Coalescent-based species delimitation in an integrative taxonomy. *Trends Ecol. Evol.* 27:480–8.

Gimmler, H. 2001. Acidophilic and acidotolerant algae. *In* Rai, L. C. & Gaur, J. P. [Eds.] *Algal Adaptation to Environmental Stresses-Physiological, Biochemical, and Molecular Mechanisms*. Springer, Berlin, pp. 259–90.

Graziani, G., Schiavo, S., Nicolai, M. A., Buono, S., Fogliano, V., Pinto, G. & Pollio, A. 2013. Microalgae as human food: chemical and nutritional characteristics of the thermo-acidophilic microalga *Galdieria sulphuraria*. *Food Funct.* 4:144–52.

Gross, W., Heilmann, I., Lenze, D. & Schnarrenberger, C. 2001. Biogeography of the Cyanidiaceae (Rhodophyta) based on 18S ribosomal RNA sequence data. *Eur. J. Phycol.* 36:275–80.

Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C. & Martiny, J. B. H. 2012. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat. Rev. Microbiol.* 10:497–506.

Hirooka, S., Higuchi, S., Uzuka, A., Nozaki, H. & Miyagishima, S. Y. 2014. Acidophilic green alga *Pseudochlorella* sp. YKT1 accumulates high amount of lipid droplets under a nitrogen-depleted condition at a low-pH. *PLoS ONE* 9:e107702.

Hsieh, C. J., Zhan, S. H., Lin, Y., Tang, S. L. & Liu, S. L. 2015. Analysis of *rbc*L sequences reveals the global biodiversity, community structure, and biogeographical pattern of thermoacidophilic red algae (Cyanidiales). *J. Phycol.* 51:682–94.

Huang, L. N., Kuang, J. L. & Shu, W. S. 2016. Microbial ecology and evolution in the acid mine drainage model system. *Trends Microbiol.* 24:581–93.

Hurvich, C. M. & Tsai, C. L. 1989. Regression and time series model selection in small samples. *Biometrika* 76:297–307.

Huss, V. A. R., Ciniglia, C., Cennamo, P., Cozzolino, S., Pinto, G. & Pollio, A. 2002. Phylogenetic relationships and taxonomic position of *Chlorella*-like isolates from low pH environments (pH < 3.0). *BMC Evol. Biol.* 2:13.

Jones, D. S., Schaperdoth, I. & Macalady, J. L. 2016. Biogeography of sulfur-oxidizing *Acidithiobacillus* populations in extremely acidic cave biofilms. *ISME J.* 10:2879–91.

Kuang, J. L., Huang, L. N., Chen, L. X., Hua, Z. S., Li, S. J., Hu, M., Li, J. T. et al. 2013. Contemporary environmental variation determines microbial diversity patterns in acid mine drainage. *ISME J.* 7:1038–50.

Kumar, S., Stecher, G. & Tamura, K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33:1870–4.

Lehr, C. R., Frank, S. D., Norris, T. B., D'lmperio, S., Kalinin, A. V., Toplin, J. A., Castenholz, R. W. & McDermott, T. R. 2007. Cyanidia (Cyanidiales) population diversity and dynamics in an acid-sulfate-chloride spring in Yellowstone National Park. *J. Phycol.* 43:3–14.

Lemieux, C., Otis, C. & Turmel, M. 2014. Chloroplast phylogenomic analysis resolves deep-level relationships within the green algal class Trebouxiophyceae. *BMC Evol. Biol.* 14:211.

Liu, L., Yang, J., Lv, H., Yu, X., Wilkinson, D. M. & Yang, J. 2015a. Phytoplankton communities exhibit a stronger response to environmental changes than bacterioplankton in three subtropical reserviors. *Envrion. Sci. Technol.* 49:108509–58.

Liu, L., Yang, J., Yu, Z. & Wilkinson, D. M. 2015b. The biogeography of abundant and rare bacterioplankton in the lakes and reservoirs of China. *ISME J.* 9:2068–77.

Lu, H. P., Yeh, Y. C., Sastri, A. R., Shiah, F. K., Gong, G. C. & Hsieh, C. 2016. Evaluating community–environment relationships along fine to broad taxonomic resolutions reveals evolutionary forces underlying community assembly. *ISME J.* 10:2867–78.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27:209–20.

Massana, R., Gobet, A., Audic, S., Bass, D., Bittner, L., Boutte, C., Chambouvet, A. et al. 2015. Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. *Environ. Microbiol.* 17:4035–49.

Minoda, A., Sawada, H., Suzuki, S., Miyashita, S., Inagaki, K., Yamamoto, T. & Tsuzuki, M. 2015. Recovery of rare earth elements from the sulfothermophilic red alga *Galdieria sulphuraria* using aqueous acid. *Appl. Microbiol. Biotechnol.* 99:1513–9.

Monaghan, M. T., Wild, R., Elliot, M., Fujisawa, T., Balke, M., Inward, D. J. G., Lees, D. C. et al. 2009. Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Syst. Biol.* 58:298–311.

Ňancucheo, I. & Johnson, D. B. 2012. Acidophilic algae isolated from mine-impacted environments and their roles in sustaining heterotrophic acidophiles. *Front. Microbiol.* 3:325.

Nemergut, D. R., Schmidt, S. K., Fukami, T., O'Neill, S. P., Bilinski, T. M., Stanish, L. F., Knelman, J. E. et al. 2013. Patterns and processes of microbial community assembly. *Microbiol. Mol. Biol. Rev.* 77:342–56.

Novis, P. & Harding, J. S. 2007. Extreme acidophiles: freshwater algae associated with acid mine drainage. *In* Seckbach, J. [Ed.] *Algae and Cyanobacteria in Extreme Environments*. Springer, Dordrecht, The Netherland, pp. 445–63.

Papke, R. T., Ramsing, N. B., Bateson, M. M. & Ward, D. M. 2003. Geographical isolation in hot spring cyanobacteria. *Environ. Microbiol.* 5:650–9.

Pinto, G. 1993. Acid-tolerant and acidophilic algae from Italian environments. *Giorn. Bot. Ital.* 127:400–6.

Pinto, G., Ciniglia, C., Cascone, C. & Pollio, A. 2007. Species composition of Cyanidiales assemblages in Pisciarelli (Campo Flegrei, Italy) and description of *Galdieria phlegrea* sp. nov. *In* Seckbach, J. [Ed.] *Algae and Cyanobacteria in Extreme Environments*. Springer, Dordrecht, The Netherlands, pp. 489–502.

Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., Kamoun, S. et al. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.* 55:595–609.

Proctor, V. W. 1959. Dispersal of freshwater algae by migratory water birds. *Science* 130:623–4.

Qin, J., Lehr, C. R., Yuan, C., Le, X. C., McDermott, T. R. & Rosen, B. P. 2009. Biotransformation of arsenic by a Yellowstone thermoacidophilic eukaryotic alga. *Proc. Natl. Acad. Sci. USA* 106:5213–7.

Rasigraf, O., Kool, D. M., Jetten, M. S. M., Sinninghe Damsté, J. S. & Ettwig, K. F. 2014. Autotrophic carbon dioxide fixation via the Calvin-Benson-Bassham cycle by the denitrifying methanotroph "Candidatus Methylomirabilis oxyfera". *Appl. Environ. Microbiol.* 80:2451–60.

Reeb, V. & Bhattacharya, D. 2010. The thermo-acidophilic Cyanidiophyceae (Cyanidiales). *In* Seckbach, J. & Chapman, D. J. [Eds.] *Red Algae in the Genomic Age*. Springer, Dordrecht, The Netherlands, pp. 411–26.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B. et al. 2012. Mrbayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61:539–42.

Selvaratnam, T., Pegallapati, A. K., Montelya, F., Rodriguez, G., Nirmalakhandan, N., VanVoorhies, W. & Lammers, P. J. 2014. Evaluation of a thermo-tolerant acidophilic alga, *Galdieria sulphuraria*, for nutrient removal from urban wastewaters. *Bioresour. Technol.* 156:395–9.

Sherwood, A. R., Dittbern, M. N., Johnston, E. T. & Conklin, K. Y. 2017. A metabarcoding comparison of windward and leeward airborne algal diversity across the Ko'olau mountain range on the island of O'ahu, Hawai'i. *J. Phycol.* 53:437–45.

Skorupa, D. J., Reeb, V., Castenholz, R. W., Bhattacharya, D. & Mcdermott, T. R. 2013. Cyanidiales diversity in Yellowstone National Park. *Lett. Appl. Microbiol.* 57:459–66.

Storey, J. D. & Tibshirani, R. 2003. Statistical significance for genome-wide studies. *Proc. Natl. Acad. Sci. USA* 100:9440–5.

Tabita, F. R., Hanson, T. E., Satagopan, S., Witte, B. H. & Kreel, N. E. 2008. Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms. *Philos. Trans. R Soc. L. B Biol. Sci.* 363:2629–40.

Takacs-Vesbach, C., Mitchell, K., Jackson-Weaver, O. & Reysenbach, A. L. 2008. Volcanic calderas delineate biogeographic provinces among Yellowstone thermophiles. *Environ. Microbiol.* 10:1681–9.

Toplin, J. A., Norris, T. B., Lehr, C. R., McDermott, T. R. & Castenholz, R. W. 2008. Biogeographic and phylogenetic diversity of thermoacidophilic cyanidiales in Yellowstone National

Park, Japan, and New Zealand. *Appl. Environ. Microbiol.* 74:2822–33.
Ueda, A., Watanabe, T., Akaneya, K. & Katano, N. 2009. Diatoms in Akita Prefecture, northern part of Japan, part 1 - Diatoms in strongly acidic hot springs. *Diatom* 25:116–9.
Varshney, P., Mikulic, P., Vonshak, A., Beardall, J. & Wangikar, P. P. 2015. Extremophilic micro-algae and their potential contribution in biotechnology. *Bioresour. Technol.* 184:363–72.
Walker, J., Spear, J. & Pace, N. 2005. Geobiology of a microbial endolithic community in the Yellowstone geothermal environment. *Nature* 434:861–4.
Watson, S. B. & Kling, H. 2017. Lake of the Woods phyto- and picoplankton: spatiotemporal patterns in blooms, community composition, and nutrient status. *Lake Reserv. Manag.* 33:415–32.
Whitaker, R. J. 2003. Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* 301:976–8.
Witkowski, A., Radziejewska, T., Wawrzyniak-Wydrowska, B., Lange-Bertalot, H., Bąk, M. & Gelbrecht, J. 2011. Living on the pH edge: diatom assemblages of low-pH lakes in western Pomerania (NW Poland). *In* Kociolek, P. [Ed.] *The Diatom World*. Springer, New York, pp. 365–84.
Zhang, J., Kapli, P., Pavlidis, P. & Stamatakis, A. 2013. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* 29:2869–76.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web site:

Figure S1. Primer efficacy test.

Figure S2. The original full-length gels for the cropped gel images in Figure S1.

Figure S3. A custom pyrotag denoising pipeline.

Figure S4. The *rbc*L UCHIME reference database construction.

Figure S5. Distribution of the read length of the filtered 454 pyrotags.

Figure S6. Maximum likelihood tree of 51 GMYC ESUs.

Figure S7. Relative abundance of 51 GMYC ESUs under the classification into major and rare group.

Figure S8. Venn diagram of 51 ESUs among four geothermal areas.

Figure S9. Relationship between community similarity (1 minus Bray-Curtis dissimilarity) and geographical and ecological distance.

Figure S10. SEM morphology of *Pinnularia acidojaponica*.

Table S1. Sample information.

Table S2. Additional planned contrast sets for parameterizing light intensity and humidity (under seven microhabitat types) in dbRDA.

Table S3. Read number, sampling coverage, and the ESU richness for all samples analyzed in this study using iNEXT.

Table S4. Phylogenetically informative sites of each ESU diagnosed using DIVEIN.

Table S5. List showing the read number of the 51 GMYC ESUs for each sample.

Table S6. Permutation tests of dbRDA variables.

Table S7. dbRDA loading of seven major ESUs

Appendix S1. Supplementary materials and methods.