George et al. Environmental Microbiome (2025) 20:50
https://doi.org/10.1186/s40793-025-00711-8

Environmental Microbiome

RESEARCH Open Access

Spatial scale modulates stochastic and deterministic influence on biogeography of photosynthetic biofilms in Southeast Asian hot springs

Christaline George[1], Chananwat Kortheerakul[2], Nitthiya Khunthong[2], Chitrabhanu Sharma[3], Danli Luo[4], Kok-Gan Chan[5], Maurycy Daroch[6], Kevin D. Hyde[3], Patrick K. H. Lee[4], Kian Mau Goh[7*], Rungaroon Waditee-Sirisattha[2*] and Stephen B. Pointing[1*]

Abstract

Hot springs, with their well-characterized major abiotic variables and island-like habitats, are ideal systems for studying microbial biogeography. Photosynthetic biofilms are a major biological feature of hot springs but despite this large-scale studies are scarce, leaving critical questions about the drivers of spatial turnover unanswered. Here, we analysed 395 photosynthetic biofilms from neutral-alkaline hot springs (39–66 °C, pH 6.4–9.0) across a 2100 km latitudinal gradient in Southeast Asia. The Cyanobacteria-dominated communities were categorized into six biogeographic regions, each characterized by a distinct core microbiome and biotic interactions. We observed a significant decline in the explanatory power of major abiotic variables with increasing spatial scale, from 62.6% locally, 55% regionally, to 26.8% for the inter-regional meta-community. Statistical null models revealed that deterministic environmental filtering predominated at local and regional scales, whereas stochastic ecological drift was more influential at the inter-regional scale. These findings enhance our understanding of the differential contribution of ecological drivers and highlight the importance of spatial scale in shaping biogeographic distributions for microorganisms.

Keywords Biogeography, Cyanobacteria, Photosynthetic biofilms, Stochasticity, Thermophiles

---

*Correspondence:
Kian Mau Goh
gohkianmau@utm.my
Rungaroon Waditee-Sirisattha
rungaroon.w@chula.ac.th
Stephen B. Pointing
stephen.pointing@nus.edu.sg
[1] Department of Biological Sciences, National University of Singapore, Singapore 117557, Singapore
[2] Department of Microbiology, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand
[3] Centre of Excellence in Fungal Research & School of Science, Mae Fah Luang University, Chiang Rai 57100, Thailand
[4] School of Energy and Environment & State Key Laboratory of Marine Pollution, City University of Hong Kong, Kowloon, Hong Kong SAR, China
[5] Institute of Biological Sciences, University of Malaya, 50603 Kuala Lumpur, Malaysia
[6] School of Environment and Energy, Peking University Shenzhen Graduate School, Shenzhen 518055, China
[7] Department of Biosciences, Universiti Teknologi Malaysia, 81310 Bahru, Johor, Malaysia

Background

The biogeography of microorganisms describes their demographic patterns of distribution across spatial and temporal scales. Resolving such patterns and their underlying cause is important to understanding microbial contribution to ecosystem functionality and resilience [1, 2]. Ecological theory identifies that biogeographic patterns emerge as the result of stochastic and deterministic influences. Neutral theories predict a tendency towards random patterns in species co-occurrence and environmentally independent spatial autocorrelation [3], whilst deterministic processes due to niche partitioning (i.e. environmental variables) and species interactions lead to segregation in species co-occurrence [4]. Large-scale biogeographic patterns have been demonstrated for microbial distribution in terrestrial [5, 6] and marine [7, 8] microbial ecosystems, although resolving evidence for the underlying drivers can be challenging in part due to the complexity of these systems [9]. Extreme microbial habitats such as hot springs offer the potential to interrogate hypotheses in biogeography within relatively well-constrained natural systems. This is because they occur as spatially discrete island-like habitats with readily defined major abiotic variables, and communities develop with reduced trophic complexity relative to other soil and aquatic habitats.

Hot springs can support distinct planktonic, sediment, and biofilm communities [10], thus making them versatile for addressing different questions in microbial ecology. Multiple studies of hot spring sediment and planktonic habitats have shown that local community assembly is strongly correlated with abiotic stressors such as temperature and pH [11–16], suggesting that deterministic processes are dominant at local scales. At landscape scales the relative influence of abiotic variables appears to decline [17], and potential endemism for some taxa may occur at an inter-continental scale [18]. An important hot spring microbial niche is occupied by photosynthetic biofilms that comprise the dominant biomass in neutral-alkaline springs at temperatures from the onset of thermophily at 40–45 °C to the upper limits for photosynthesis at 73 °C [19]. The neutral-alkaline hot springs are widely distributed globally in tectonic landscapes [20, 21]. These photosynthetic communities provide a contrast to hot spring communities at higher temperatures and acidic springs that can be dominated by chemoautotrophic hydrogen-oxidizing communities [22]. Previous studies of hot spring photosynthetic biofilms have demonstrated that the dominant photosynthetic Cyanobacteria [23–26] and Chloroflexota [27–29] taxa display patterns of occurrence linked to temperature at local scales from ambient temperatures up to the thermal limit for photosynthesis at approximately 73 °C [19]. The observation of distinct phylotypes for certain photosynthetic groups including Cyanobacteria [23, 30, 31] and Chloroflexota [24, 32] from different distant hot spring locations has led to the proposal that this reflects allopatric speciation due to geographic isolation. The extent to which such patterns hold for communities across continuous broad environmental and spatial gradients to create clearly defined biogeographic regionalization, and the identification of ecological drivers that underly such patterns remain largely unresolved. Prompted by recent studies of planktonic hot spring microbiota where quantitative estimates have revealed that diverse abiotic variables are only able to explain a small fraction of observed microbial distribution at both local [13] and landscape [17] scales; this motivated us to examine the potential contribution of multiple stochastic and deterministic influences on biofilm community assembly.

We hypothesized that an approach which encompassed a large continuous regional geographic scale with concurrent estimation of the influence from abiotic and biotic, stochastic and deterministic drivers would yield valuable novel insight on the ecological drivers of biogeography for photosynthetic biofilms at different spatial scales in hot springs. The hot springs of Southeast Asia support abundant neutral-alkaline hot springs and this region is also under-represented in the global dataset for hot spring microbiota, and so we rationalized that our focus would also substantially enhance the utility of hot springs for future global-scale comparisons. Here we report the interactions that explain microbial biogeography of photosynthetic biofilms in 395 hot springs from Southeast Asia. Hot springs in this region support prolific photosynthetic biofilms as their major biotic component [33–35], they are culturally important and have economic value, and yet their diversity and biogeography has not been systematically studied. We resolved the phylogenetic structure and biogeography of communities using 16S rRNA gene sequencing, with additional shotgun metagenomics to validate our diversity estimation approach. We then identified the interaction of biotic, geographic, and major geochemical factors on the observed patterns and applied statistical null models to quantify the relative influence of selection, dispersal limitation, and genetic drift in shaping community structure at different spatial scales. The insight generated will be essential for future development and testing of hypotheses that further address large scale spatial distributions of microorganisms in hot springs and other systems.

Materials and methods

Sample collection and environmental metadata

Detailed sampling protocols are described in the Supplementary Methods. Briefly, photosynthetic microbial biofilms (*N*= 395) were sampled aseptically into 500 µl DNA/RNA-later preservative solution at 40 thermally defined hot spring sites from 15 geothermal locations in Southeast Asia along a ~ 2,100 km north–south transect in Southeast Asia during March-October 2022. Abiotic variables (temperature, pH, conductivity, nitrate, nitrite, phosphate, hydrogen sulfide) were measured on-site for each location using hand-held probes and colorimetric test kits, or visual observation (pools v flowing water, human usage of hot springs) (Supplementary Table S1).

DNA sequencing

Detailed sequencing and quality control protocols are described in the Supplementary Methods. Briefly, DNA extraction from biofilm samples was carried out using the Powerlyzer Powersoil Kit (Qiagen) optimized for photosynthetic biofilms [35]. Taxonomic diversity was estimated by 16S rRNA gene amplicon sequencing of the V4 region (Illumina Novaseq 6000, PE250 kits), using universal primers 515 F (GTGCCAGCMGCCGCGGTAA) and 806R (GGACTACHVGGGTWTCTAAT) [36] and the Earth Microbiome Project PCR workflow [37], with appropriate negative controls and sample randomization. To better understand and quantify any potential bias in our dataset due to primer selection we also performed shotgun metagenomic sequencing (Illumina Novaseq 6000, PE150 kits) for a subset of samples (*N*= 12) which avoided the biases of a PCR-based approach.

Bioinformatic analysis

Detailed bioinformatic protocols are described in the Supplementary Methods. Briefly, 16S rRNA gene sequences were processed in R (version 4.2.2) [38] using 'DADA2' [39] before taxonomic assignment of amplicon sequence variants (ASVs) using the SILVA 16S database [40, 41] (v138.1) with the RDP naive Bayesian classifier. The data was rarified to the minimum sequencing depth (63 k reads per sample) and further filtered to contain ASVs with > 1% relative abundance. For metagenomes, 'Trimmomatic' v0.39 [42] was employed to eliminate adapters from raw sequencing data and remove low quality reads that were shorter than 50 bp. Processed reads were subjected to taxonomic assignment using the Kraken2 software suite with NCBI nt database (v2023-11–29) [43] and the default parameters. Comparison of metagenomic and 16S rRNA taxonomic classifications were performed using SingleM (v0.18.2) (Github: https://github.com/wwood/singlem).

Statistical analysis

Detailed statistical analysis protocols are described in the Supplementary Methods. All analyses were performed using R (v4.2.2) [38]. Briefly, alpha diversity indices were estimated using 'vegan' [44]. Distance decay was calculated using 'phyloseq' [45] from unweighted UniFrac phylogenetic distances versus pairwise geographical distances in kilometres calculated using 'geodist' [46]. Beta diversity ordinations were determined using Principal Coordinate Analysis (PCoA) of unweighted UniFRAC distances (without transformation) using 'phyloseq' [45]. Heat trees and differential heat trees displaying both quantitative (ASV numbers, read abundances and median proportion of reads) and hierarchal taxonomic classifications were generated using 'metacoder' [47]. Classification of ASVs into ecological guilds was achieved using FAPROTAX [48], with additional manual curation from searches using NCBI databases and the scientific literature. Correlations (Pearson's [*r*]) between ASV predicted function and abiotic variables was achieved using the 'plot_cor' function in the 'microeco' package [49]. Core microbiomes (> 1% relative abundance and ≥ 95% prevalence) were determined using 'microbiome' [50]. Screening of human-associated taxa was achieved by identifying ASVs associated with twenty indicator genera [51]. Co-occurrence network analysis was performed using 'Spiec-Easi' with the'mb'(Meinshausen-buhlmann's neighborhood selection) method using 'netcomi' [52]. Modules were defined with the 'cluster_walktrap' method using 'microeco' [49]. The importance and classification of nodes within and amongst modules in the network was determined by estimating Pi and Zi scores using the 'plot_taxa_roles' function. Mantel's test was performed using Pearson's correlation coefficient and default 999 permutations with 'microeco' [49]. Variance partitioning was performed using 'vegan' [44]. Random forest machine learning algorithms [53] were used to fit ensemble models to the association between biotic and abiotic data with 'rfpermute' v2.5.2 [54] using default settings. Statistical null models [55] were generated using 'microeco' [49]. The beta nearest taxon index (betaNTI) was calculated to infer contribution to ecological convergence or divergence using the 'frequency' model. Significance testing was performed using 'vegan' [44] for: Analysis of Variance (ANOVA), permutational Multivariate Analysis of Variance (PERMANOVA), permutation-based tests of betadispersion, paired T-tests, and post-hoc Tukey's Honest Significant Difference (HSD).

Results

Southeast Asian hot springs are a tractable system for interrogating photosynthetic biofilm biogeography

We sampled 395 photosynthetic biofilms from 40 neutral-alkaline hot springs (39—66 °C, pH 6.4—9, Supplementary Table S1) along a 2100 km transect of geothermal activity in Southeast Asia (Fig. 1A). Our approach to estimating biofilm community composition employed

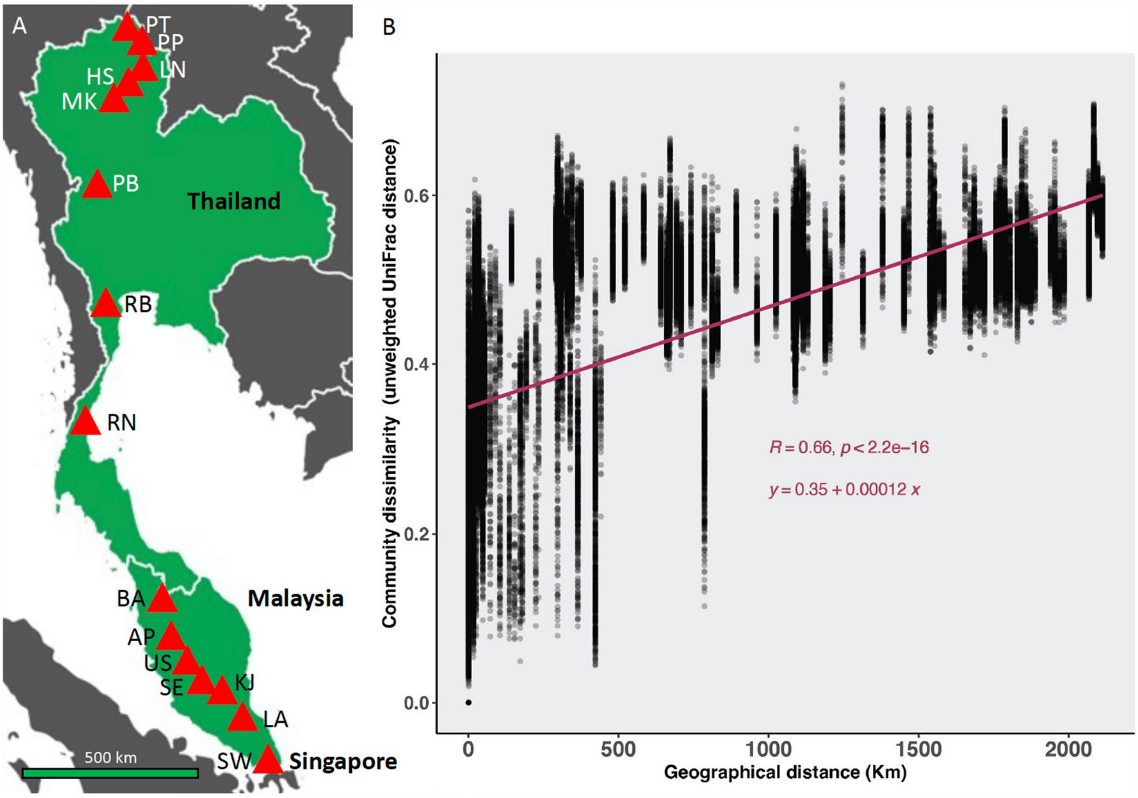

Fig. 1 **A** Map of Southeast Asian geothermal sites from which the 40 hot spring locations (red triangles) were sampled for photosynthetic microbial biofilms (*N* = 395). Alphabetic codes denote sampling location (see Supplementary Table S1). Taxonomic composition and Alpha diversity estimates are shown in Supplementary Figs. 3–4. **B** Distance decay plot illustrating the strong correlation between geographic distance and community dissimilarity for hot spring biofilm communities (*N* = 395). Geographic distance was calculated using pairwise geographic distances. Community dissimilarity was calculated using unweighted UniFRAC distances. Individual distance decay plots for different ecological groups of bacteria are shown in Supplementary Fig. S5

sequencing of 16S rRNA genes from environmental samples (Supplementary Fig. S1). We acknowledge that the Earth Microbiome Project primers we used have recognized limitations in detecting archaeal and eukaryal taxa. We therefore performed shotgun metagenome sequencing for a subset of samples (*N* = 12) to assess whether primer-bias may have affected our dataset (Supplementary Fig. S2). None of the Archaea or microbial Eukarya taxa reached the > 1% relative abundance required for inclusion in our ecological analysis, and rank abundance of bacterial phyla between metagenomes and 16S rRNA genes were largely congruent (Supplementary Fig. S2). Biofilms supported a diverse taxonomic composition comprising 28 bacterial phyla and 2 archaeal phyla, and communities were characterized by four dominant phyla comprising the Bacteroidota, Chloroflexota, Cyanobacteria, and Pseudomonadota (Supplementary Fig. S3). The most abundant and prevalent class in biofilms overall were the oxygenic photoautotrophic Cyanobacteria (Class Cyanophyceae) followed by the anoxygenic photoautotrophic/photoheterotrophic Chloroflexota (Class Chloroflexia) thus underscoring the importance of photosynthesis to community assembly. Many of the geothermal locations in our study are heavily utilized nearby for varied human activities including bathing, laundry, and food preparation (Supplementary Table S1). Our sampling included only undisturbed springs however, to be cautious we performed an estimate of human-associated bacteria present in the biofilms. This revealed that the occurrence of 20 common human-associated genera contributed to only 0.145% of the total relative abundance of ASVs.

Alpha diversity estimates of the rarefied 16S rRNA gene dataset revealed species richness, Pielou's eveness, Shannon's diversity index, and Gini-Simpson's index all displayed a weak but significant negative correlation with temperature (Supplementary Fig. S4). This indicated an overall trend towards lower alpha diversity with increased abundance of fewer taxa as temperature increased. To better focus our subsequent ecological analyses, the rarified dataset was further stringently filtered to contain taxa with relative abundance > 1% (572

ASVs). This final working dataset (rarified and filtered) was subjected to statistical analysis of beta diversity distribution patterns in relation to abiotic variables and biotic interactions, and ecological modelling to resolve the underlying mechanistic ecological drivers influencing the observed distributions.

Hot spring photosynthetic biofilms displayed putative biogeographic regionalization

Resolving large-scale biogeographic patterns for hot spring photosynthetic biofilms requires expanding scale beyond local systems so that regional biogeographic species pools can be determined as a step towards establishing global biogeographic patterns. Our dataset covered a large latitudinal gradient spanning 2100 km that elicited a strongly significant positive distance-decay correlation between community phylogenetic distance and geographical distance of the hot spring communities (Fig. 1B; linear regression: R= 0.66, $P<$ 2.2e-16). We also estimated distance decay patterns for the different ecological guilds (identified using FAPROTAX) within the biofilms and this unveiled differential responses by certain groups. The strongest distance-decay relationships were apparent for photosynthetic (R =0.63, $P<$ 2.2e-16) and heterotrophic (R =0.64, $P<$ 2.2e-16) guilds compared to a relatively weaker pattern for chemoautotrophs (R =0.45, $P<$ 2.2e-16) (Supplementary Fig. S5). Among the photosynthetic bacterial ASVs the Cyanobacteria accounted for the strongest distance decay signal (R =0.61, $P<$ 2.2e-16) whilst the Chloroflexia displayed the weakest pattern (R =0.34, $P<$ 2.2e-16) (Supplementary Fig. S5).

Visualization of beta diversity using PCoA ordination of unweighted UniFrac distances illustrated that communities clustered into six statistically supported putative biogeographic regions that we named North Thailand, Central Thailand, South Thailand, North Malaysia, South Malaysia, and Singapore (99% confidence intervals, one-way PERMANOVA $P<$ 0.001; Fig. 2A). The clustering of regions broadly matched latitudinal distance and was not autocorrelated to abiotic variables or sampling dates. None of the regions could be delineated by their suite of abiotic characteristics alone. Our study suggests that each region extended up to approximately 300 km. Taxonomic patterns that defined the observed biogeographic patterns were visualized using a differential heat tree showing pairwise comparisons of median proportion of reads between the six identified biogeographic regions (Fig. 2B, Supplementary Fig. S6). This illustrated that whilst thermophily appears to be a widespread trait among diverse phyla, some taxonomic groups and notably the Bacteroidota, Chloroflexota, Cyanobacteria, and Proteobacteria (Pseudomonadota) were differentially abundant in pairwise comparisons between regions. Despite these differences in beta diversity the communities displayed consistent predicted functional composition (identified using FAPROTAX) that was dominated by oxygenic photautotrophy and aerobic chemoheterotrophy (Fig. 2C). As the grouping of biogeographic regions was also accompanied by significant betadispersion ($P<$ 2.2e-16, post hoc Tukey $P<$ 0.001) we then set-out to provide further evidence to support or refute the validity of biogeographic regions through identification of core microbiomes and biotic interactions.

Distinct core microbiomes and biotic interactions further support delineation of biogeographic regions

We identified specific taxa that were most abundant in the biogeographic regions (Fig. 3A, Supplementary Fig. S7). Eighteen of the 25 most abundant genera were photosynthetic bacteria and trends for various ecological guilds highlighted that the Cyanobacteria (Cyanophyceae) displayed the most region-specificity/dominance within the Leptolyngbyaceae, Phormidiaceae, and Pseudanabaenaceae. This emphasizes the importance of cyanobacteria as potential keystone taxa in biofilm communities and reinforces the view that they can be used as a key descriptor for specific biofilms. Core microbiome analysis revealed that whilst there was no pan-Southeast Asian core microbiome, each biogeographic region displayed a distinct core comprising 49–99% of ASVs (Fig. 3B). The Cyanobacteria, through various photosynthetic lineages, were part of every core microbiome

(See figure on next page.)

Fig. 2 **A** Beta diversity patterns for hot spring photosynthetic biofilms ($N=$ 395) indicated statistically supported biogeographic regions. Beta diversity was estimated using Principal Coordinate Analysis of unweighted UniFRAC distances (Confidence interval 99%), where NT = North Thailand, CT = Central Thailand, ST = South Thailand, NM = North Malaysia, SM = South Malaysia, SG = Singapore. Symbol shapes denote hot spring location codes (see Supplementary Table S1). **B** A Differential heat tree matrix showing: hierarchal taxonomic coverage represented as nodes, number of ASVs assigned to each taxon as node size/width, and significant differences in pairwise comparisons of median proportions of reads between biogeographic regions as colored nodes determined using a Wilcox rank-sum test. Bacterial and archaeal lineages that were relatively enriched in biofilms are depicted as colored regions in the rows (green branches and nodes) and columns (brown branches and nodes). The labelled gray tree on the lower left is a key for the smaller unlabeled trees in the matrix. Additionally, a heat tree showing all levels of taxonomic coverage with all nodes labeled is shown in Supplementary Fig. S6. C) Functional profiling of hot spring photosynthetic biofilms using FAPROTAX revealed all biofilm samples were dominated by oxygenic photoautotrophy and aerobic chemoheterotrophy

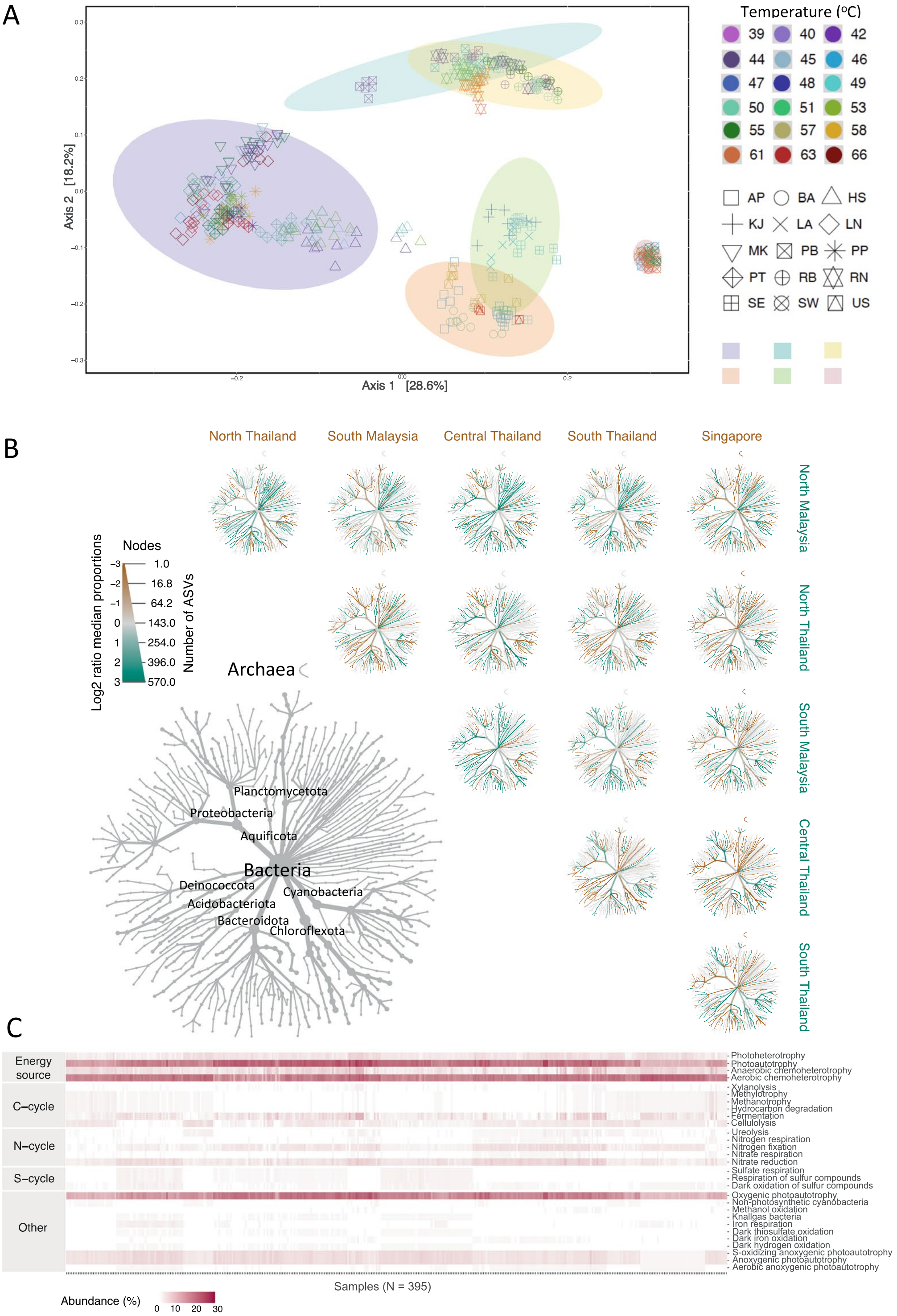

Fig. 2 (See legend on previous page.)

underscoring biofilm dependence upon photoautotrophy for energetic and carbon input to the system. Whilst the Chloroflexota were also core to all regional microbiomes they variously comprised anoxygenic photosynthetic (e.g. *Chloroflexus, Chloroploca, Roseiflexus*) or non-photosynthetic (e.g. RBG-13–54-9, Anaerolineae) taxa. Each

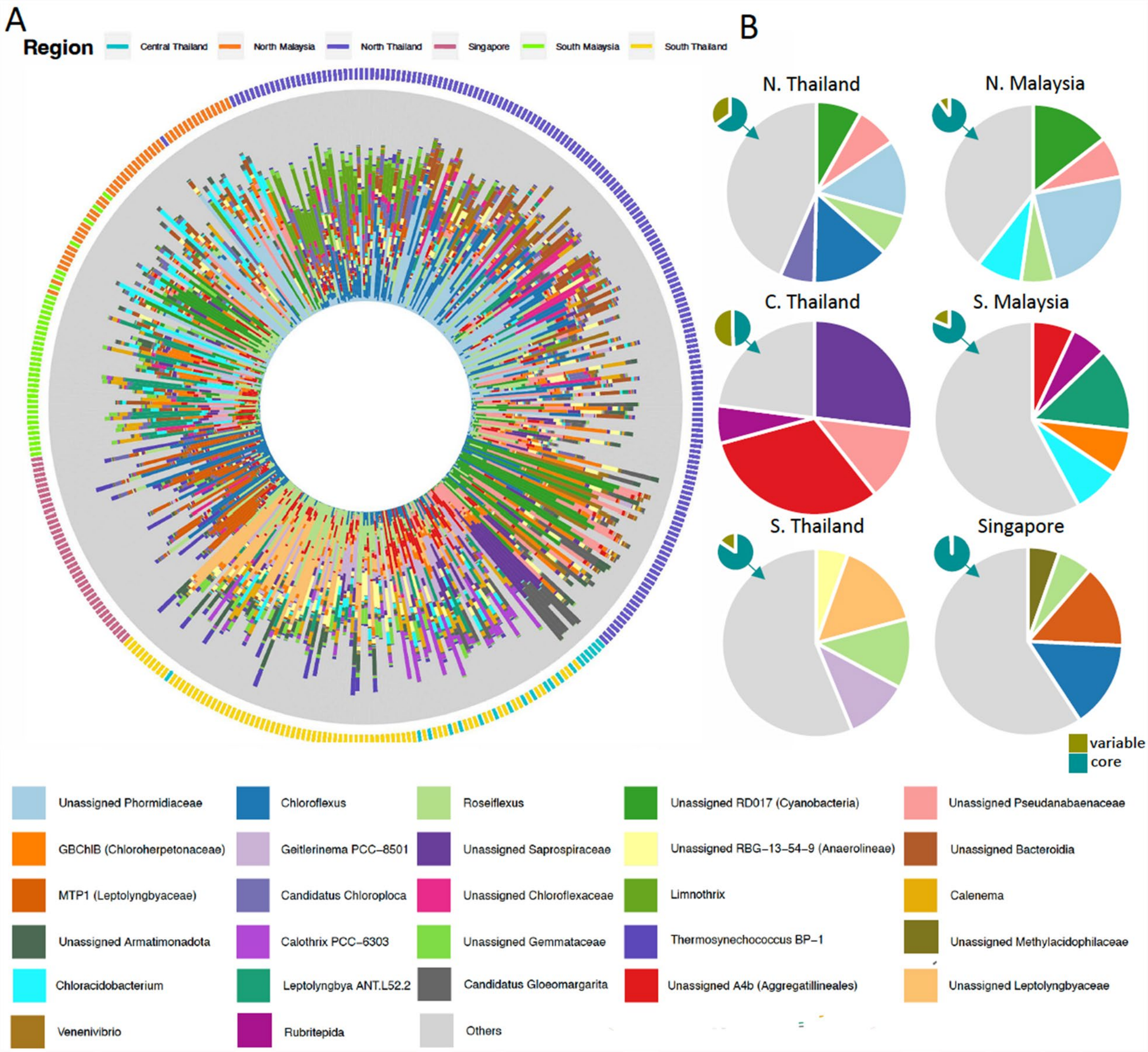

Fig. 3 **A** Relative abundance of the 25 most abundant ASVs in each sample ($N = 395$) shown clustered at genus level (bar plots). The coloured bars at the circumference of the iris plot denote biogeographic region for each biofilm sample. Iris plots for individual ecological groups of taxa are shown in the Supplementary Fig. S7. **B** Composition of the core microbiome for biofilms in each biogeographic region ($N = 395$). The core microbiome was defined as comprising ASVs with ≥ 95% prevalence and > 1% relative abundance. Small pie charts indicate the core versus variable microbiome, and large pie charts show the taxonomic composition of ASVs in each core microbiome clustered at genus level

of the core regional microbiomes supported multiple ASVs associated with nitrogen-fixing taxa and this likely reflected the low combined nitrogen levels in hot spring water (Supplementary Table S1).

Recognizing the potential influence of biotic interactions in community assembly of biofilms, we performed a co-occurrence network analysis to identify putative relationships between taxa occurrence in the communities (Fig. 4, Supplementary Fig. S8). This revealed region-specific modules of interaction and inter-module interaction that spanned multiple trophic levels and were dominated by Alphaproteobacteria, Anaerolineae, Bacteroidia, Chloroflexia, Cyanophyceae, and Gammaproteobacteria (Fig. 4, Supplementary Fig. S8). Hub taxa within and among modules comprised photosynthetic Cyanobacteria and Chloroflexia although some photosynthetic taxa that were part of core microbiomes were not identified as hub taxa. Interactions between taxa were further visualized using chord diagrams to highlight associations (Fig. 4). This revealed a particularly strong putative

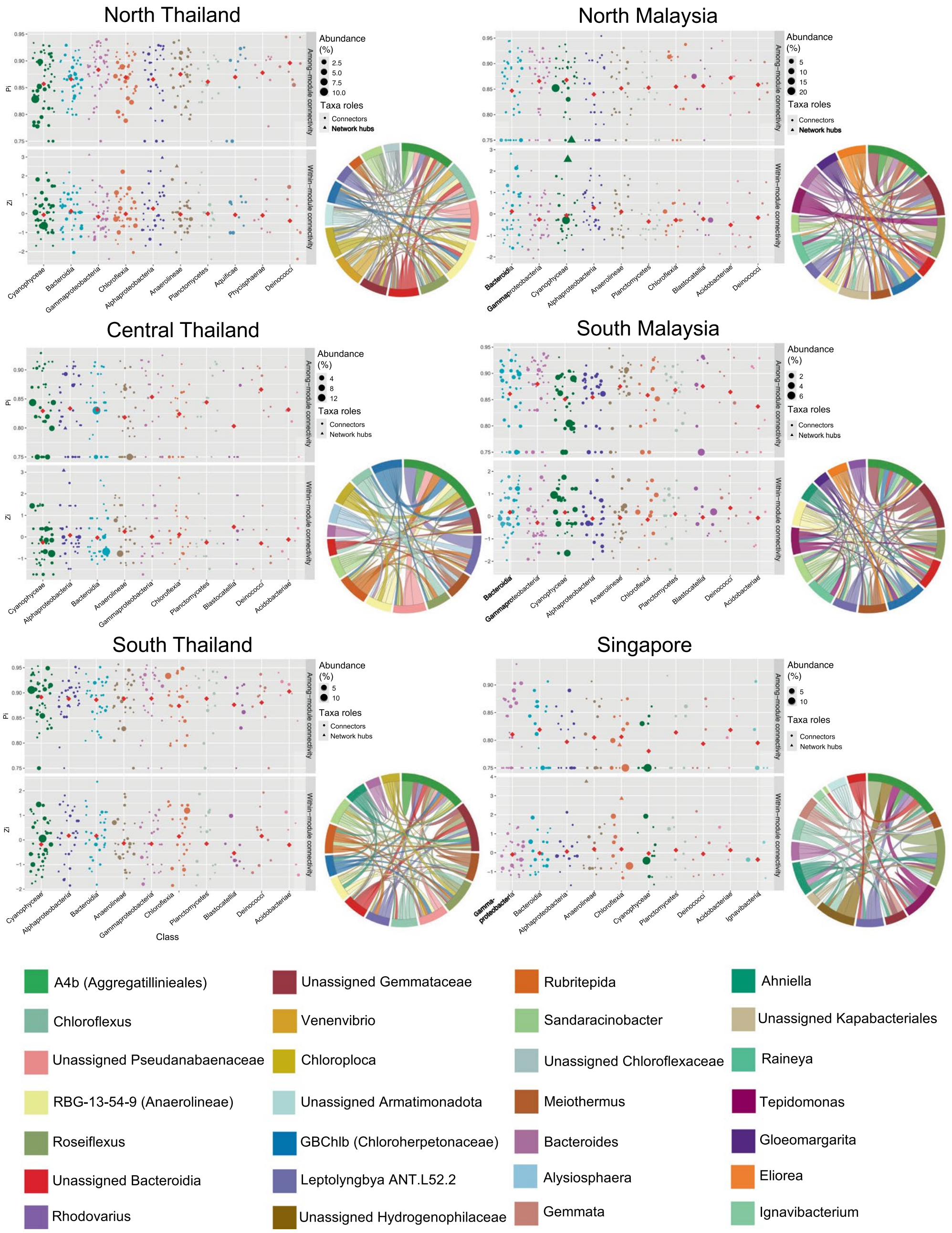

Fig. 4 Putative biotic interactions for bioregions estimated using co-occurrence network analysis. Bubble plots show the patterns of interaction among and within modules of interaction (i.e. putative biotic interactions) at Class level, ranked in order of contribution to interactions for the top ten interacting Classes. Module connectors are shown by circles and module hubs are shown by triangles. Symbol size denotes relative abundance. Red symbols denote mean level of interaction. Interactions for the whole community in each region are shown in Supplementary Fig. S8. Chord diagrams illustrate major associations between abundant ASVs clustered at genus level. Chord thickness denotes relative strength of associations

relationship between Aggregatlineales A4b and several other photosynthetic and heterotrophic taxa. Overall, the most prolific interactions occurred between Anaerolineae ASVs and other phototrophic and heterotrophic taxa suggesting the Anaerolineae occupy a key role in facilitating metabolic cooperation within the biofilms.

Influence of major environmental variables on community assembly declined with increasing spatial scale

A core assumption in extreme environments is that strong selection pressure arises due to the influence of harsh environmental/geochemical conditions. To determine potential deterministic influence of major environmental variables on community assembly we first selected variables that did not display collinearity with each other or with geographic distance (collinearity threshold = Pearson's [*r*] > 0.7, Supplementary Fig. S9). We then employed the widely used Mantel's test to test for linear correlations of biotic data against major geographic and environmental variables relevant to biofilm communities (Fig. 5Ai, Supplementary Fig. S10). Significant positive correlations were observed for pH, temperature, carbonate, phosphate, and hydrogen sulfide; whilst a significant negative correlation occurred for conductivity used as a general proxy for the concentration and valence of soluble ions (overall Mantel's *r* = 0.12). The relative strength of association varied with biogeographic regions (Supplementary Fig. S10) and ecological guilds, with photosynthetic taxa displaying the strongest correlations (Fig. 5Ai). Multivariate variance partitioning analysis revealed that overall, the variables with the strongest combined explanatory power comprised pH, hydrogen sulfide, conductivity, and carbonate (Fig. 5Aii, Supplementary Fig. S10). For individual locations the mean overall variation explained by these abiotic variables was 62.6%, and this declined to 55% at the region scale, and 26.8% at the inter-regional province scale (Fig. 5Aiii, Supplementary Fig. S10).

Recognizing that environmental gradients are seldom truly linear, we also employed machine learning Random Forest analysis with ensemble models of fit to abiotic variables against the biotic dataset (Fig. 5B). This corroborated our findings from linear correlations and identified that overall conductivity, hydrogen sulfide, and pH were the most influential variables on large scale biogeography of biofilm communities. For some regions conductivity and carbonate were relatively more important for the photosynthetic guild than for the overall community, whilst for chemoautotrophs and chemoheterotrophs temperature was relatively more important (Fig. 5B). To further identify the impact of this deterministic influence on putative biofilm community function, we correlated predicted functional traits with abiotic variables (Fig. 5C). This indicated that oxygenic and anoxygenic photosynthesis were positively correlated with carbonate, hydrogen sulfide, and pH; and negatively correlated with conductivity and phosphate. Conversely, chemoautotrophic metabolism was most positively correlated with temperature and heterotrophy was negatively correlated with most abiotic variables.

The contribution of stochastic and deterministic processes varied with spatial scale

Having identified the potential influence of major abiotic variables on community assembly, we repeated our linear and ensemble correlation analysis with inclusion of geographic distance (latitude) as a variable (Supplementary Fig. S11). This suggested that distance was also influential to the observed community distribution and so we then set out to quantify the broader relative influence of niche and neutral ecological drivers on the observed community structure using statistical ecological modelling. By interrogating the phylogenetic data using a net relatedness index against that expected under a purely random community assembly using null models we were able to discern the influence of ecological drivers on spatial assembly of communities whilst avoiding the limitations and bias associated with enforcing correlations against abiotic variables. Net relatedness Index (betaNRI) and Raup-Crick distances were calculated as a standardized measure of mean phylogenetic distance to the nearest taxon in the community (Fig. 6A). The beta nearest taxon index (betaNTI) was then calculated to infer contribution to ecological convergence or divergence using null models for different spatial scales: province (inter-region meta-community), region, location, site; and for the different ecological guilds: photosynthetic, chemoautotrophic, and chemoheterotrophic. This revealed that whilst deterministic selection (homogenous selection) was the dominant process for individual sites, locations, and regions; at the inter-regional province scale the dominant process for turnover of the metacommunity was stochastic ecological drift (Fig. 6B). The dominant ecological drivers also varied among the different ecological guilds for the metacommunity and indicated

(See figure on next page.)

Fig. 5 **A** The influence of non-colinear abiotic variables (Supplementary Fig. S9) on observed taxonomic composition for all taxa plus photosynthetic, chemoautotrophic, and chemoheterotrophic fractions of the community, (i) Mantel's Test and Pearson's correlation between unweighted UniFRAC phylogenetic distances and abiotic variables indicating strength of correlations for the overall meta-community, (ii) Variance Partitioning using transformed Hellinger distances versus abiotic variables indicating the relative contribution from the four most influential variables, (iii) relative contribution of the four topmost variables to observed variation in community distribution at different spatial scales. Individual plots for each region are presented in Supplementary Fig. S10. **B** Random Forest ensemble modelling of the most influential variables on observed community distribution for the overall meta-community, each region, and ecological guilds. N/A denotes no correlation due to insufficient taxa. The more influential variables occupy locations further to the left of each bubble plot. **C** Influence of abiotic variables on predicted metabolic function of biofilms was estimated using Pearson's Correlation [*r*]. Red colour denotes positive correlations, and blue colour denotes negative correlations. * denotes $P = < 0.05$, ** denotes $P = 0.01$, and *** denotes $P = < 0.001$

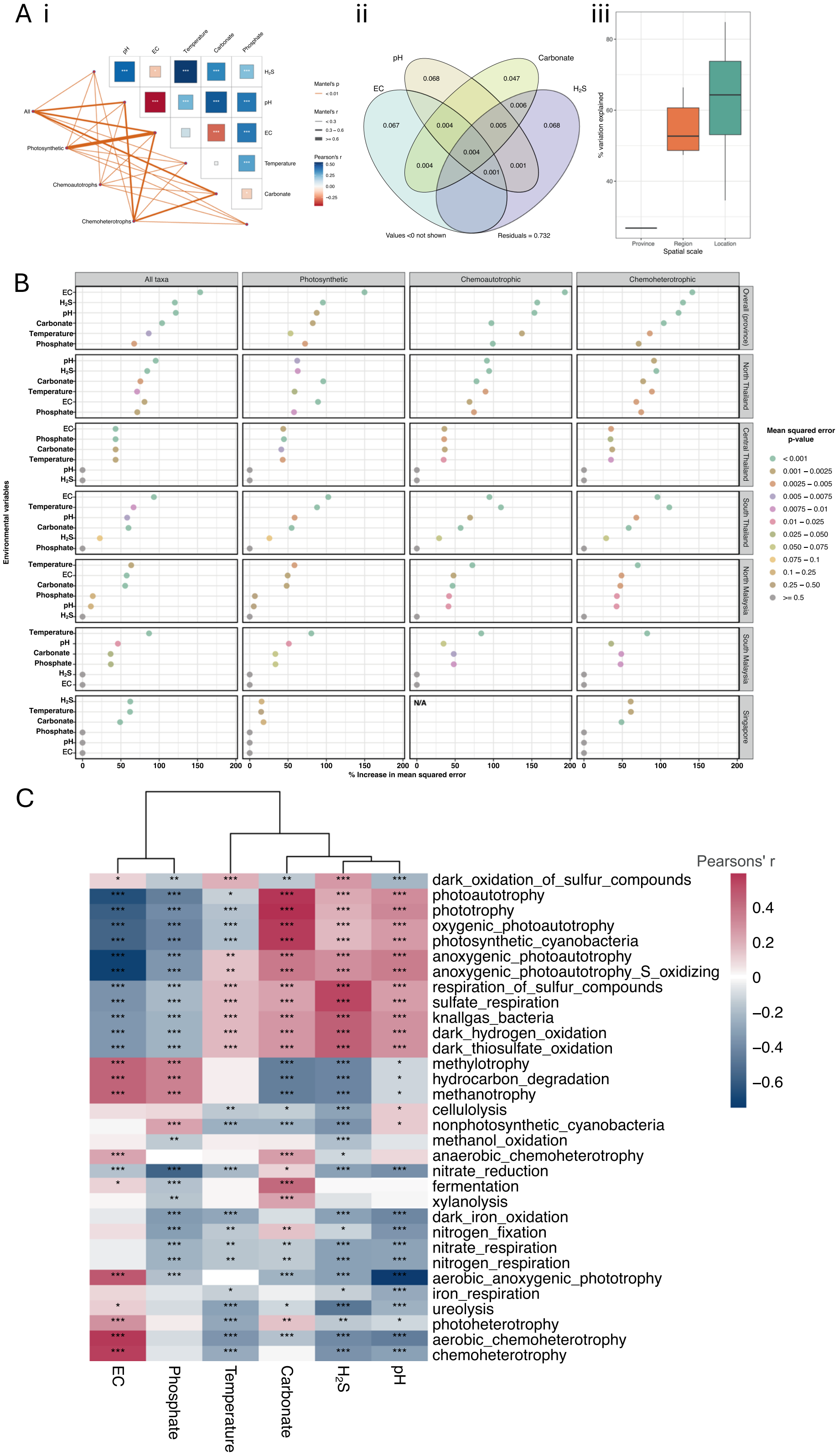

Fig. 5 (See legend on previous page.)

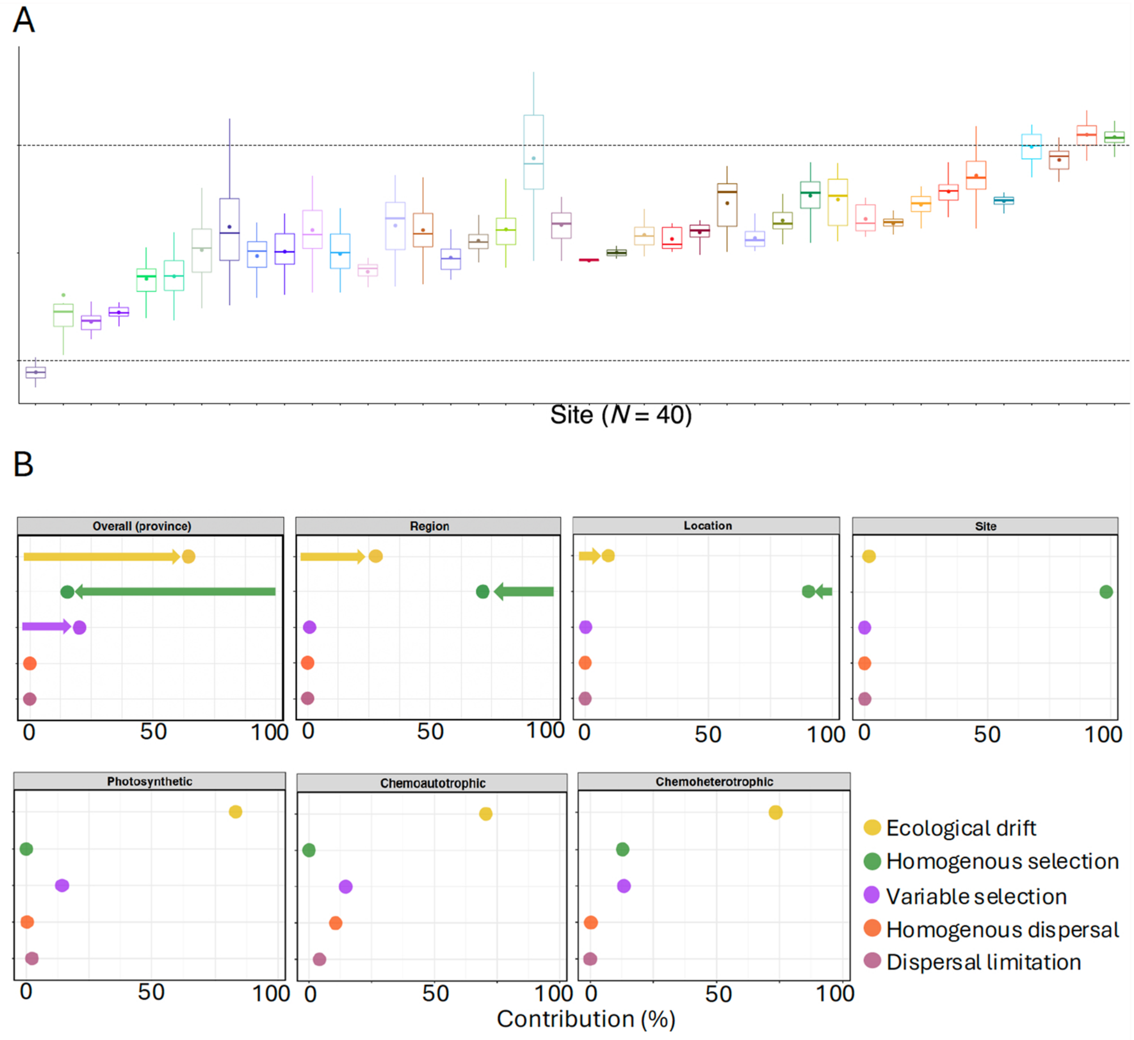

Fig. 6 Quantitative estimation of the contribution of ecological drivers to community assembly. **A** Net relatedness Index was (betaNRI) was calculated as a standardized measure of mean phylogenetic distance to the nearest taxon in the community at each site. Letters denote significant between associations ($P < 0.001$). **B** The beta nearest taxon index (betaNTI) was calculated to infer contribution to ecological convergence or divergence using Raup-Crick null models (RCbray) for different spatial scales: overall (province), region, location, site (arrows denote major shifts due to spatial scale); and for the different ecological guilds (photosynthetic, chemoautotrophic, and chemoheterotrophic) overall

that chemoheterotrophs were under greater influence from homogenous selection compared with the photosynthetic and chemoautotrophic guilds (Fig. 6B). These findings corroborated our other lines of evidence that the hot spring photosynthetic biofilm community is subject to differential influence from ecological drivers at different spatial scales.

Discussion

Our findings expand the global inventory of taxa associated with photosynthetic biofilms in hot springs and addressed an important biogeographic data gap for Southeast Asia where hot springs are common and yet have not been previously studied in a systematic manner. We resolved the bounds of six putative biogeographic regions and provide evidence that the differential contribution of stochastic and deterministic ecological drivers to community assembly at different spatial scales explains the mechanistic basis for their observed biogeography. Identifying this phenomenon across a broad geographic scale among neutral-alkaline hot springs with differing physicochemical properties is an important step towards the delineation and practical use of these habitats as tractable model systems for photosynthetic biofilms in microbial ecology.

Southeast Asian hot springs displayed observable biogeographic patterns

The phenomenon of distance decay may reflect a range of influences on community diversity including environmental filtering, ecological drift, diversification, and dispersal limitation [56, 57]. The strong distance decay signal for the overall community in our study indicated that one or more of these factors was influencing the biogeography of biofilms. We also showed that distance decay patterns were maintained regardless of a negative correlation between alpha diversity and temperature when plotted for individual ecological guilds and this allowed us to be confident that the signal was not biased by deeply-branching lineages that may skew UniFrac estimates, such as for the Aquificota that were associated with higher temperature sites. Of note was the difference in strength for distance decay signal between photosynthetic groups, with photoautotrophic Cyanobacteria (Cyanophyceae) displaying a significantly stronger signal than for the photosynthetic Chloroflexota (Chloroflexia) and other photosynthetic bacteria. This suggested that the Cyanobacteria were the most important phototrophs for delineating biogeographic patterns in hot spring photosynthetic biofilms, and validates descriptive studies that have generally relied on cyanobacterial taxa to describe different biofilms [19].

The beta diversity clustering of communities indicated that biofilm communities were resolved into six statistically supported putative regions. Each region extended up to ~ 300 km and so this may represent the extent of abiotic influence and/or dispersal. Each region was further characterised by a distinct core microbiome where each core comprised Cyanobacteria, Chloroflexota, and other variable taxa. Core microbiomes in tropical soils have been linked to functional stability of communities [58], and we speculate that they perform a similar role in hot spring biofilms. Additional evidence for distinct patterns in putative biotic interactions added further support that the bioregions reflected genuinely different communities. Microbial interactions are thought to present a useful metric of putative contribution to community function [59], and thus our findings highlight that the distinct interaction profiles also reflect functionally distinct communities. A notable difference between core microbiome and biotic interaction estimates was that the A4b (Aggregatilineales) ASVs were part of only one core microbiome and yet emerged as the most influential taxa to biotic interactions for all hot spring regions. This may reflect the growing awareness that less prevalent (conditionally rare) taxa may have important roles in biogeochemical cycling and community function [60]. Overall, the high degree of taxonomic substitution between taxa in the various ecological guilds (photosynthetic, chemoautotrophic, and chemoheterotrophic) among the different biofilms did not significantly affect the predicted dominant functional role of biofilms. We were therefore confident that comparisons across the wide geographic scale were made for similar biofilm communities that were dominated functionally by photoautotrophy and aerobic chemoheterotrophy. Whilst there are some limitations to inferring metabolic function based upon 16S rRNA gene sequences, the approach is widely used in ecological studies and has strong fidelity to broad metabolic categories [48, 61].

The findings also expand the described biogeographic range into Southeast Asia for several taxa within groups including the Aquificota, Bacteroidota, Chloroflexota, Cyanobacteria, and Proteobacteria (Pseudomonadota). This will assist with future global studies to resolve potential endemism in hot spring taxa that are currently hampered by a patchy global dataset. In addition, nearby human usage that was a potential source of transient taxa did not significantly impact our estimates of biofilm composition and so we anticipate a wide variety of locations may be amenable for use in future studies regardless of their usage provided that appropriate quality control measures are employed. Previous studies have recorded distinct phylogenetic lineages of cyanobacteria across inter-continental distances in hot springs [30], and this mirrors observations for other ecological groups [18, 32, 62]. Taken together with our findings this suggests that regions defining similar communities occur in hot spring photosynthetic biofilms at the scale of < ~ 300 kms but that observations at broader spatial scales (e.g. intercontinental) may reflect greater differences and potential endemism arising from yet unresolved contribution from stochastic and deterministic drivers.

Cyanobacteria and Chloroflexota interactions may drive biofilm community assembly and functionality

We demonstrated that, as for globally distributed soil samples [6], a relatively small number of the overall taxa dominated biofilms and overall most taxa were relatively rare. Different photosynthetic biofilms were readily characterized by their dominant cyanobacterial component, and also supported photosynthetic or non-photosynthetic Chloroflexota. A number of studies at a well-studied location, Yellowstone National Park, have highlighted that the *Synechococcus* biofilms (which are not encountered in Southeast Asian hot springs) typical of higher temperatures within the photic gradient invariably co-occur with anoxygenic photosynthetic Chloroflexia such as *Chloroflexus* and *Roseiflexus* [63]. Our study corroborates this co-occurrence for *Thermosynechococcus* biofilms in Southeast Asia but also highlights that photosynthetic Chloroflexota may not be essential taxa for biofilms that are dominated by other cyanobacterial groups. Instead, our core microbiome analysis points to a common association for all biofilm types that includes a Cyanobacteria and Chloroflexota component where the latter may or may not be photosynthetic. This was further supported by our co-occurrence network analysis and suggests that interaction may lie with a metabolic syntrophy that does not require anoxygenic photosynthesis. Therefore, we speculate that non-photosynthetic processes such as sulfur cycling may be a key contribution from this phylum in biofilm ecology. This is supported by the demonstration of physiological linkages between Cyanobacteria and sulfur-dependent metabolism in the Chloroflexota [64–66]. A further observation revealed that the photoautotrophic Cyanobacteria involved in the most significant correlations belonged to genera with known nitrogen-fixing taxa suggests that this group are important for both carbon and nitrogen input to the community. This is consistent with physiological observations that cyanobacteria are capable of nitrogen fixation in hot springs [67, 68]. The FAPROTAX functional prediction data in our study supported these observations although we caution against over-interpreting findings based on inference of function from taxonomic marker genes.

Among the heterotrophic taxa those associated with cellulolytic activity, e.g., *Cytophaga* and *Ignavibacterium*, were notably abundant and this likely reflects the large allochthonous input of plant debris to hot springs erupting in forested areas within tropical Southeast Asia [35]. Other heterotrophic taxa indicated hot springs have potential in bioprospecting for industrial use, e.g., *Ideonella*, a genus associated with breakdown of plastic polymers [69]. The temperature range for the hot springs in our study falls within the lower range for many chemoautotrophic thermophiles [21], and their distribution was patchier compared to other ecological groups. The chemoautotrophic taxa were recovered at very low abundance in biofilms but within the chemoautotrophic group the hydrogen-oxidising *Venenvibrio* (Aquificales), anaerobic sulfur-reducing *Thermodesulfovibrio* (Thermodesulfovibrionales), and putatively facultatively autotrophic Leptospiraceae ASVs were most abundant and displayed some regionalism (Fig. S7). These alternate autotrophic pathways may be important supplements to the largely photoautotrophic metabolism in biofilms, and the sulfur-reducing taxa indicate that biofilms likely support steep micro-environmental gradients within the biofilm structure that might allow anaerobic metabolism within a largely aerobic biofilm, and this is consistent with recent observations in a Chilean hot spring that chemoautotrophic sulfate reduction was supported in photosynthetic biofilms [65]. Future focus on functional meta-omics in biofilms will be important to gain improved insight on how metabolic functionality may influence their ecological distributions.

Deterministic influences only partially explained observed biogeographic distributions

Correlations indicative of deterministic environmental filtering were demonstrated for pH, carbonate, conductivity, temperature, and hydrogen sulfide, although no single abiotic variable emerged as the most explanatory across all sites. The highly significant association with pH aligns with its perceived dominant role in determining biogeographic patterns in other geothermal habitats [13, 17, 70, 71] and the soil biome [5, 6]. Similarly temperature and sulfide have known effects on microbial distributions in hot springs [27, 29, 65, 72]. It remains to be resolved how other factors such as conductivity and carbonate may exert deterministic influence on thermophilic microorganisms at the levels encountered in hot springs. Our findings underscore the complex interplay of major abiotic stressors on biofilms and is corroborated by several location-specific studies identifying these variables as important to biofilm community assembly [12, 23, 31]. Different ecological guilds and their predicted functional traits responded uniquely to abiotic variables, revealing that broad community analyses might obscure specific taxon responses. Oxygenic and anoxygenic photosynthetic traits positively correlated with pH, hydrogen sulfide, and carbonate, aligning with their adaptation to these variables in hot springs and other systems [35, 73–75]. Similarly, chemoautotrophic traits were positively correlated with hydrogen sulfide and temperature, consistent with their adaptation to extreme environments [21]. Conversely, the chemoheterotrophs displayed negative correlations with most abiotic variables, indicating these were likely thermotolerant taxa potentially near their tolerance limits under poly-extreme conditions within the photosynthetic biofilms.

A notable observation was the effect of spatial scale on the influence of abiotic variables toward diversity patterns. The study showed a significant decrease in the explanatory power of major abiotic variables with increasing spatial scale (local: 62.6% vs. province: 26.8%). This pattern reflects how habitat heterogeneity increases with spatial scale [76], emphasizing the need to also quantitatively consider spatial scale when examining abiotic influences on biogeographic patterns. We acknowledge that our study did not measure all potential abiotic variables due to logistical constraints of sampling remote locations. We focused on major variables known to significantly correlate with distribution of hot spring photosynthetic biofilms as general indicators of abiotic influence [12, 23, 25, 31, 77, 78]. However, the potential for bias due to under-estimation of deterministic influence from unmeasured variables such as trace metals, dissolved oxygen or carbon cannot be excluded, although they have typically shown very weak correlation with hot spring community assembly relative to measured major abiotic variables such as temperature, pH, conductivity, and sulfide [17, 23]. The weak explanatory power of environmental variables for microbial distribution patterns is emerging as common in hot springs and other habitats [13, 17, 23, 71, 79], with additional factors such as geological setting, tectonic history, and biotic interactions potentially playing roles [15, 80, 81]. The biogeographic regions identified in our study, spanning hundreds of kilometres, only partially aligned with underlying geological faults. This suggests that a complex interaction of deterministic and stochastic drivers may explain the observed biogeographic patterns.

Spatial scale modulates the relative influence of deterministic and stochastic drivers on biogeography

Having demonstrated the potential deterministic influence of major abiotic variables as well as evidence for distance-related patterns and biotic interactions between ecological guilds within biofilms, we applied statistical models to further resolve the broader interaction of these multiple ecological drivers on community. Statistical approaches using null models are a useful tool to gain insight on the influence of stochastic and deterministic ecological drivers such as selection, dispersal, ecological drift, and speciation that may be shaping microbial biogeography [82]. The approach avoids the biases that are associated with correlation-based analyses of community distribution with regard to abiotic variables [83], because it does not enforce an a priori relationship with potential drivers. Our study provides the first evidence for the relative influence of these processes on photosynthetic microbial biofilms in hot springs at different spatial scales. The fidelity of null model outputs is strongly dependent on the scale and depth of the input matrix, and so large-scale studies such as ours offer important novel insight on large-scale processes. A notable observation was the lack of influence from dispersal limitation and despite earlier hypotheses that this may be a major explanatory variable our data is consistent with recent studies that atmospheric microbial dispersal occurs across large inter-continental distances [84, 85]. It is also worthy of mention that dispersal mechanisms specific to hot springs such as via subsurface aquifers are as yet uncharacterized but may emerge as influential.

The most influential process contributing to observed turnover at individual sites was deterministic homogenous selection, and this gradually declined with increasing spatial scale until at the inter-regional province scale stochastic ecological drift was identified as the major driver, concurrent with variable selection surpassing homogenous selection as a deterministic driver. This novel finding suggests a meta-community where stochastic changes in regional population size are the major driver [82], and was testable by our application of null models to the large phylogenetic dataset [83]. This finding is congruent with our other multiple lines of evidence and analyses demonstrating the effects of environmental variables and biotic interactions on the meta-community. Ecological drift is a stochastic process characterized by weak selection and rare movement of taxa between communities. We envisage these evolutionary forces to operate as continuous variables in hot springs. Thus, whilst moderate levels of both movement (dispersal) and birth–death (selection) are expected, neither process is dominant. Hence, this has also been termed an 'undominated' process [83]. This contrasts with the more traditional view that spatial turnover in hot spring biofilms might arise largely from spatial differences in abiotic conditions (variable selection) and may in part reflect the narrow range of habitat conditions for photosynthetic biofilms in hot springs. Our findings highlight that the 73% variance in communities at the inter-regional province scale that was unexplained by abiotic variables was most likely due to the stochastic influence of ecological drift. Our observations also mirror a recent study that revealed the same pattern occurs for large-scale biogeographic distributions of animal communities where homogenous selection and stochasticity overruled heterogenous selection [86]. This may emerge as a universal rule in terms of the drivers of large scale biogeographic distributions, and help to resolve the large 'unexplained' variance in microbial community structure recorded in other biogeographic studies of hot springs [17, 23] and for other environments such as soils [87].

We further demonstrated a differential contribution of ecological processes acting on the functional guilds within communities. Whilst overall and for photosynthetic and heterotrophic taxa ecological drift was the major factor, autotrophic taxa were subject to somewhat greater influence from variable selection compared to homogenizing selection than for chemoheterotrophs. This points to the autotrophs exhibiting distributions more consistent with niche specialists, compared to heterotrophic taxa tending towards being generalists with a wider niche breadth [88]. This may emerge as a common feature of photosynthetic biofilms in diverse extreme environments given our findings also mirror the relationship observed for cyanobacteria-dominated biofilms in extreme deserts [89]. Possible explanations for the observed stochastic influence in our study may include extinction events due to tectonic variability e.g. thermal surges [90], or thermal variability due to monsoon flooding in the tropics [91]. We envisage these may result in genetic bottlenecks for some populations in hot springs and thus contribute to genetic drift. In addition, ancient historical legacies may contribute to stochasticity, as observed for persistent photosynthetic microbial communities in deserts and lakes [92, 93], and for chemoautotrophic bacteria in volcanic calderas where geothermal history explained more variation than extant geochemical or geographic variables [81]. Other non-extreme aquatic freshwater and marine photosynthetic microbial populations have also been shown to exhibit a lack of dispersal limitation and yet exhibit high levels of stochastic assembly in the absence of satisfactory explanation by environmental variables [92, 94], and this highlights a general similarity that may emerge as typical in diverse microbial systems.

Conclusion

Our study employed multiple lines of evidence to delineate six biogeographic regions for photosynthetic biofilms in Southeast Asian hot springs. Addressing a major regional biodiversity knowledge gap, our research paves the way for enhanced global comparisons. The findings offer new insight on the mechanistic processes that underpin the biogeographic patterns for photosynthetic hot spring biofilms. We elucidated the influence of spatial scale on the interplay between stochastic and deterministic ecological drivers within this system, and showed that homogenizing selection which dominates at local scales gives way to ecological drift with increasing distance and this aligned with delineation of biogeographic regions. Photosynthetic biofilms are also widely distributed in diverse marine and terrestrial environments and so our findings using the hot springs model system have broad applicability in the ongoing effort to identify potential universal constraints on the drivers of global microbial biogeography.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40793-025-00711-8.

Additional file 1
Additional file 2
Additional file 3

Acknowledgements

The authors acknowledge support provided by the Ranong Municipality and Kamphaengphet Provincial Administrative Organization (Thailand) and National Parks Board (Singapore) for facilitating access to field locations. We thank Tancredi Caruso (University College, Dublin) for advice on null model analysis, Lynn Drescher (National University of Singapore) and Chayanard Phukhamsakda (Mae Fah Luang University) for research support, and Poh Moi Goh (National University of Singapore) for technical support.

Author contributions

K.G.C., K.M.G., S.B.P., and R.W.S. designed the study and secured funding; C.G., C.K., N.K., A.B., K.M.G., K.D.H., S.B.P., and R.W.S. conducted field sampling. C.G. and N.K. performed laboratory experiments; C.G. and S.B.P. performed data analysis; K.G.C., M.D., C.G., K.M.G., P.K.H.L., D.L., S.B.P., and R.W.S. interpreted the findings; C.G. and S.B.P. wrote the paper with input from all authors.

Funding

The research was funded by a Singapore Ministry of Education AcRF Tier 2 grant awarded to K.G.C., K.M.G., S.B.P., and R.W.S. (award number MOE-T2EP30123-0007). K.M.G. acknowledges additional support from the Malaysia Ministry of Higher Education Fundamental Research Grant Scheme (award number FRGS/1/2023/STG02/UTM/02/1).

Data availability

All DNA sequence data generated from this study has been deposited in the NCBI Sequence Read Archive under BioProject accession numbers PRJNA1052408 for 16S rRNA gene sequences and PRJNA1106483 for metagenomes. All R codes used in data processing for this study are available at https://github.com/PointingLab.

Declarations

Ethics approval and consent to participate

All prevailing local, national and international regulations and conventions, and normal scientific ethical practices, have been respected.

Competing interests

The authors declare no competing interests.

Received: 1 November 2024 Accepted: 18 April 2025

References

1. Zhou J, Ning D. Stochastic community assembly: does it matter in microbial ecology? Microbiol Mol Biol Rev. 2017;81(4):e00002-17. https://doi.org/10.1128/mmbr.00002-17.
2. Dickey JR, Swenie RA, Turner SC, Winfrey CC, Yaffar D, Padukone A, et al. The utility of macroecological rules for microbial biogeography. Front Ecol Evol. 2021. https://doi.org/10.3389/fevo.2021.633155.
3. Hubbell SP. The Unified neutral theory of biodiversity and biogeography. Princeton: Princeton University Press; 2001.
4. Chave J. Neutral theory and community ecology. Ecol Lett. 2004;7(3):241–53. https://doi.org/10.1111/j.1461-0248.2003.00566.x.
5. Bahram M, Hildebrand F, Forslund SK, Anderson JL, Soudzilovskaia NA, Bodegom PM, et al. Structure and function of the global topsoil microbiome. Nature. 2018;560(7717):233–7. https://doi.org/10.1038/s41586-018-0386-6.
6. Delgado-Baquerizo M, Oliverio AM, Brewer TE, Benavent-González A, Eldridge DJ, Bardgett RD, et al. A global atlas of the dominant bacteria found in soil. Science. 2018;359(6373):320–5. https://doi.org/10.1126/science.aap9516.
7. Sommeria-Klein G, Watteaux R, Ibarbalz FM, Pierella Karlusich JJ, Iudicone D, Bowler C, et al. Global drivers of eukaryotic plankton biogeography in the sunlit ocean. Science. 2021;374(6567):594–9. https://doi.org/10.1126/science.abb3717.
8. Righetti D, Vogt M, Gruber N, Psomas A, Zimmermann NE. Global pattern of phytoplankton diversity driven by temperature and environmental variability. Sci Adv. 2019;5(5):6253. https://doi.org/10.1126/sciadv.aau6253.
9. Martiny JBH, Bohannan BJMM, Brown JH, Colwell RK, Fuhrman JA, Green JL, et al. Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol. 2006;4(2):102–12. https://doi.org/10.1038/nrmicro1341.
10. Chen J-S, Hussain B, Tsai H-C, Nagarajan V, Koner S, Hsu B-M. Analysis and interpretation of hot springs water, biofilms, and sediment bacterial

community profiling and their metabolic potential in the area of Taiwan geothermal ecosystem. Sci Total Environ. 2023;856: 159115. https://doi.org/10.1016/j.scitotenv.2022.159115.

11. Cole JK, Peacock JP, Dodsworth JA, Williams AJ, Thompson DB, Dong H, et al. Sediment microbial communities in Great Boiling Spring are controlled by temperature and distinct from water communities. ISME J. 2013;7(4):718–29. https://doi.org/10.1038/ismej.2012.157.
12. Miller SR, Strong AL, Jones KL, Ungerer MC. Bar-coded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in yellowstone national park. Appl Environ Microbiol. 2009;75(13):4565–72. https://doi.org/10.1128/AEM.02792-08.
13. He Q, Wang S, Hou W, Feng K, Li F, Hai W, et al. Temperature and microbial interactions drive the deterministic assembly processes in sediments of hot springs. Sci Total Environ. 2021;772: 145465. https://doi.org/10.1016/j.scitotenv.2021.145465.
14. Sharp CE, Brady AL, Sharp GH, Grasby SE, Stott MB, Dunfield PF. Humboldt's spa: microbial diversity is controlled by temperature in geothermal environments. ISME J. 2014;8(6):1166–74. https://doi.org/10.1038/ismej.2013.237.
15. Colman DR, Veach A, Stefánsson A, Wurch L, Belisle BS, Podar PT, et al. Tectonic and geological setting influence hot spring microbiology. Environ Microbiol. 2023;25(11):2481–97. https://doi.org/10.1111/1462-2920.16472.
16. Qi Y-L, Chen Y-T, Xie Y-G, Li Y-X, Rao Y-Z, Li M-M, et al. Analysis of nearly 3000 archaeal genomes from terrestrial geothermal springs sheds light on interconnected biogeochemical processes. Nat Commun. 2024;15(1):4066. https://doi.org/10.1038/s41467-024-48498-5.
17. Power JF, Carere CR, Lee CK, Wakerley GLJ, Evans DW, Button M, et al. Microbial biogeography of 925 geothermal springs in New Zealand. Nat Commun. 2018;9(1):2876. https://doi.org/10.1038/s41467-018-05020-y.
18. Whitaker RJ, Grogan DW, Taylor JW. Geographic barriers isolate endemic populations of Hyperthermophilic archaea. Science. 2003;301(5635):976–8. https://doi.org/10.1126/science.1086909.
19. Ward DM, Castenholz RW, Miller SR. Cyanobacteria in geothermal habitats. Dordrecht: Springer; 2012.
20. Tamburello G, Chiodini G, Ciotoli G, Procesi M, Rouwet D, Sandri L, et al. Global thermal spring distribution and relationship to endogenous and exogenous factors. Nat Commun. 2022;13(1):6378. https://doi.org/10.1038/s41467-022-34115-w.
21. Shu W-S, Huang L-N. Microbial diversity in extreme environments. Nat Rev Microbiol. 2022;20(4):219–35. https://doi.org/10.1038/s41579-021-00648-y.
22. Spear JR, Walker JJ, McCollom TM, Pace NR. Hydrogen and bioenergetics in the Yellowstone geothermal ecosystem. Proc Natl Acad Sci U S A. 2005;102(7):2555–60. https://doi.org/10.1073/pnas.0409574102.
23. Becraft ED, Wood JM, Cohan FM, Ward DM. Biogeography of American Northwest Hot Spring A/B'-Lineage Synechococcus Populations. Front Microbiol. 2020;11:77. https://doi.org/10.3389/fmicb.2020.00077.
24. Lau CY, Jing H, Aitchison JC, Pointing SB. Highly diverse community structure in a remote central Tibetan geothermal spring does not display monotonic variation to thermal stress. FEMS Microbiol Ecol. 2006;57(1):80–91. https://doi.org/10.1111/j.1574-6941.2006.00104.x.
25. Kees D, Murugapiran SK, Bennet AC, Hamilton TL. Distribution and genomic variation of thermophilic cyanobacteria in diverse microbial mats at the upper temperature limits of photosynthesis. mSystems. 2022. https://doi.org/10.1128/msystems.00317-22.
26. Alcorta J, Espinoza S, Viver T, Alcaman-Arias ME, Trefault N, Rossello-Mora R, et al. Temperature modulates Fischerella thermalis ecotypes in Porcelana Hot Spring. Syst Appl Microbiol. 2018;41(6):531–43. https://doi.org/10.1016/j.syapm.2018.05.006.
27. Bennett AC, Murugapiran SK, Hamilton TL. Temperature impacts community structure and function of phototrophic Chloroflexi and Cyanobacteria in two alkaline hot springs in Yellowstone National Park. Environ Microbiol Rep. 2020;12(5):503–13. https://doi.org/10.1111/1758-2229.12863.
28. Lau MC, Aitchison JC, Pointing SB. Bacterial community composition in thermophilic microbial mats from five hot springs in central Tibet. Extremophiles. 2009;13(1):139–49. https://doi.org/10.1007/s00792-008-0205-3.
29. Hamilton TL, Bennett AC, Murugapiran SK, Havig JR. Anoxygenic phototrophs span geochemical gradients and diverse morphologies in terrestrial geothermal springs. MSystems. 2019. https://doi.org/10.1128/mSystems.00498-19.
30. Papke RT, Ramsing NB, Bateson MM, Ward DM. Geographical isolation in hot spring cyanobacteria. Environ Microbiol. 2003;5(8):650–9. https://doi.org/10.1046/j.1462-2920.2003.00460.x.
31. Podar PT, Yang Z, Bjornsdottir SH, Podar M. Comparative analysis of microbial diversity across temperature gradients in hot springs from Yellowstone and Iceland. Front Microbiol. 2020;11:1625. https://doi.org/10.3389/fmicb.2020.01625.
32. Gaisin VA, Grouzdev DS, Namsaraev ZB, Sukhacheva MV, Gorlenko VM, Kuznetsov BB. Biogeography of thermophilic phototrophic bacteria belonging to Roseiflexus genus. FEMS Microbiol Ecol. 2016. https://doi.org/10.1093/femsec/fiw012.
33. Sompong U, Hawkins PR, Besley C, Peerapornpisal Y. The distribution of cyanobacteria across physical and chemical gradients in hot springs in northern Thailand. FEMS Microbiol Ecol. 2005;52(3):365–76. https://doi.org/10.1016/j.femsec.2004.12.007.
34. Chan CS, Chan KG, Ee R, Hong KW, Urbieta MS, Donati ER, et al. Effects of physiochemical factors on prokaryotic biodiversity in Malaysian Circumneutral hot springs. Front Microbiol. 2017;8:1252. https://doi.org/10.3389/fmicb.2017.01252.
35. George C, Lim CX, Pointing SB. Community assembly of thermophilic photosynthetic microbial mats along a polyextreme environmental gradient at Sembawang Hot Spring Singapore. Front Microbiol. 2023;14:1189468. https://doi.org/10.3389/fmicb.2023.1189468.
36. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012;6(8):1621–4.
37. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. Nature. 2017;551(7681):457–63. https://doi.org/10.1038/nature24621.
38. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing,Vienna, Austria. https://www.R-project.org/ (2022).
39. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. R package version 1.26.0. Nat Methods. 2016;13(7):581. https://doi.org/10.1038/nmeth.3869.
40. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, et al. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res. 2014;42:D643–8.
41. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(D1):D590–6.
42. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170.
43. Lu J, Rincon N, Wood DE, Breitwieser FP, Pockrandt C, Langmead B, et al. Metagenome analysis using the Kraken software suite. Nat Protoc. 2022;17(12):2815–39. https://doi.org/10.1038/s41596-022-00738-y.
44. Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P, et al. Vegan: Community Ecology Package. R package version 2.6–4. https://CRAN.R-project.org/package=vegan (2023).
45. McMurdie PJ, Holmes S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. R package version 1.4.20. PLoS ONE. 2013. https://doi.org/10.1371/journal.pone.0061217.
46. Padgham M. geodist: Fast, Dependency-Free Geodesic Distance Calculations. R package version 0.0.7. https://github.com/hypertidy/geodist (2021).
47. Foster Z, Sharpton T, Grünwald N. Metacoder: an R package for visualization and manipulation of community taxonomic diversity data. R package version 0.3.6. PLOS Comput Biol. 2017;13(2):1–15. https://doi.org/10.1371/journal.pcbi.1005404.
48. Louca S, Parfrey LW, Doebeli M. Decoupling function and taxonomy in the global ocean microbiome. Science. 2016;353(6305):1272–7. https://doi.org/10.1126/SCIENCE.AAF4507.
49. Liu C, Cui Y, Li X, Yao M. microeco: an R package for data mining in microbial community ecology. R package version 1.1.1. FEMS Microbiol Ecol. 2021. https://doi.org/10.1093/femsec/fiaa255.
50. Leo Lahti, Shetty S. Tools for microbiome analysis in R. R package version 1.20.0. http://microbiome.github.com/microbiome (2017).

51. Archer SDJ, Lee KC, Caruso T, Alcami A, Araya JG, Cary SC, et al. Contribution of soil bacteria to the atmosphere across biomes. Sci Total Environ. 2023;871: 162137. https://doi.org/10.1016/j.scitotenv.2023.162137.
52. Peschel S, Muller CL, von Mutius E, Boulesteix AL, Depner M. NetCoMi: network construction and comparison for microbiome data in R. R pacakge version 1.1.0. Brief Bioinf. 2021. https://doi.org/10.1093/bib/bbaa290.
53. Cutler DR, Edwards TC Jr, Beard KH, Cutler A, Hess KT, Gibson J, et al. Random forests for classification in ecology. Ecology. 2007;88(11):2783–92. https://doi.org/10.1890/07-0539.1.
54. Archer E. rfPermute: Estimate Permutation p-values for Random Forest Importance Metrics. v2.5.2 edn; 2011. https://cran.r-project.org/web/packages/rfPermute.
55. Chase JM, Kraft NJB, Smith KG, Vellend M, Inouye BD. Using null models to disentangle variation in community dissimilarity from variation in α-diversity. Ecosphere. 2011;2(2):24. https://doi.org/10.1890/ES10-00117.1.
56. Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK, et al. Global patterns in the biogeography of bacterial taxa. Environ Microbiol. 2011;13(1):135–44. https://doi.org/10.1111/j.1462-2920.2010.02315.x.
57. Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. Beyond biogeographic patterns: processes shaping the microbial landscape. Nat Rev Microbiol. 2012;10(7):497–506. https://doi.org/10.1038/nrmicro2795.
58. Jiao S, Chen W, Wei G. Core microbiota drive functional stability of soil microbiome in reforestation ecosystems. Glob Chang Biol. 2022;28(3):1038–47. https://doi.org/10.1111/gcb.16024.
59. Kodera SM, Das P, Gilbert JA, Lutz HL. Conceptual strategies for characterizing interactions in microbial communities. IScience. 2022;25(2):103775. https://doi.org/10.1016/j.isci.2022.103775.
60. Jousset A, Bienhold C, Chatzinotas A, Gallien L, Gobet A, Kurm V, et al. Where less may be more: how the rare biosphere pulls ecosystems strings. ISME J. 2017;11(4):853–62. https://doi.org/10.1038/ismej.2016.174.
61. Djemiel C, Maron P-A, Terrat S, Dequiedt S, Cottin A, Ranjard L. Inferring microbiota functions from taxonomic genes: a review. GigaScience. 2022. https://doi.org/10.1093/gigascience/giab090.
62. Hedlund BP, Reysenbach AL, Huang L, Ong JC, Liu Z, Dodsworth JA, et al. Isolation of diverse members of the Aquificales from geothermal springs in Tengchong. China Front Microbiol. 2015;6:157. https://doi.org/10.3389/fmicb.2015.00157.
63. Hamilton TL, Vogl K, Bryant DA, Boyd ES, Peters JW. Environmental constraints dfining the distribution, composition, and evolution of chlorophototrophs in thermal features of Yellowstone National Park. Geobiology. 2012;10:236–49. https://doi.org/10.1111/j.1472-4669.2011.00296.x.
64. Kawai S, Kamiya N, Matsuura K, Haruta S. Symbiotic growth of a thermophilic Sulfide-oxidizing photoautotroph and an elemental sulfur-Disproportionating Chemolithoautotroph and cooperative dissimilatory oxidation of Sulfide to Sulfate. Front Microbiol. 2019;10:1150. https://doi.org/10.3389/fmicb.2019.01150.
65. Konrad R, Vergara-Barros P, Alcorta J, Alcaman-Arias ME, Levican G, Ridley C, et al. Distribution and activity of sulfur-metabolizing bacteria along the temperature gradient in phototrophic mats of the Chilean hot spring Porcelana. Microorganisms. 2023. https://doi.org/10.3390/microorganisms11071803.
66. Hamilton TL, Klatt JM, de Beer D, Macalady JL. Cyanobacterial photosynthesis under sulfidic conditions: insights from the isolate Leptolyngbya sp. strain hensonii. ISME J. 2018;12(2):568–84. https://doi.org/10.1038/ismej.2017.193.
67. Steunou A-S, Bhaya D, Bateson MM, Melendrez MC, Ward DM, Brecht E, et al. In situ analysis of nitrogen fixation and metabolic switching in unicellular thermophilic cyanobacteria inhabiting hot spring microbial mats. Proc Natl Acad Sci. 2006;103(7):2398–403. https://doi.org/10.1073/pnas.0507513103.
68. Estrella Alcaman M, Fernandez C, Delgado A, Bergman B, Diez B. The cyanobacterium Mastigocladus fulfills the nitrogen demand of a terrestrial hot spring microbial mat. ISME J. 2015;9(10):2290–303. https://doi.org/10.1038/ismej.2015.63.
69. Yoshida S, Hiraga K, Takehana T, Taniguchi I, Yamaji H, Maeda Y, et al. A bacterium that degrades and assimilates poly(ethylene terephthalate). Science. 2016;351(6278):1196–9. https://doi.org/10.1126/science.aad6359.
70. Sriaporn C, Campbell KA, Van Kranendonk MJ, Handley KM. Genomic adaptations enabling Acidithiobacillus distribution across wide-ranging hot spring temperatures and pHs. Microbiome. 2021;9(1):135. https://doi.org/10.1186/s40168-021-01090-1.
71. Barbosa C, Tamayo-Leiva J, Alcorta J, Salgado O, Daniele L, Morata D, et al. Effects of hydrogeochemistry on the microbial ecology of terrestrial hot springs. Microbiol Spectrum. 2023;11(5):e00249-e323. https://doi.org/10.1128/spectrum.00249-23.
72. Skirnisdottir S, Hreggvidsson GO, Hjörleifsdottir S, Marteinsson VT, Petursdottir SK, Holst O, et al. Influence of Sulfide and temperature on species composition and community structure of hot spring microbial mats. Appl Environ Microbiol. 2000;66(7):2835–41. https://doi.org/10.1128/AEM.66.7.2835-2841.2000.
73. Singh Y, Gulati A, Singh DP, Khattar JIS. Cyanobacterial community structure in hot water springs of Indian North-Western Himalayas: a morphological, molecular and ecological approach. Algal Res. 2018;29:179–92. https://doi.org/10.1016/j.algal.2017.11.023.
74. Wu Y. Combined effect of bicarbonate and water in photosynthetic oxygen evolution and carbon neutrality. Acta Geochimica. 2023;42(1):77–88. https://doi.org/10.1007/s11631-022-00580-9.
75. Prairie YT, Cole JJ. The Carbon Cycle in Lakes: A Biogeochemical Perspective. In: Mehner T, Tockner K, editors. Encyclopedia of Inland Waters. 2nd ed. Oxford: Elsevier; 2022. p. 89–101.
76. Loke LHL, Chisholm RA. Measuring habitat complexity and spatial heterogeneity in ecology. Ecol Lett. 2022;25(10):2269–88. https://doi.org/10.1111/ele.14084.
77. Kostesic E, Mitrovic M, Kajan K, Markovic T, Hausmann B, Orlic S, et al. Microbial diversity and activity of biofilms from geothermal springs in Croatia. Microb Ecol. 2023. https://doi.org/10.1007/s00248-023-02239-1.
78. Bennett AC, Murugapiran SK, Kees ED, Sauer HM, Hamilton TL. Temperature and geographic location impact the distribution and diversity of photoautotrophic gene variants in alkaline yellowstone hot springs. Microbiol Spectrum. 2022;10(3):e01465-e1521. https://doi.org/10.1128/spectrum.01465-21.
79. Massello FL, Chan CS, Chan KG, Goh KM, Donati E, Urbieta MS. Meta-analysis of microbial communities in hot springs: recurrent taxa and complex shaping factors beyond pH and temperature. Microorganisms. 2020. https://doi.org/10.3390/microorganisms8060906.
80. Upin HE, Newell DL, Colman DR, Boyd ES. Tectonic settings influence the geochemical and microbial diversity of Peru hot springs. Commun Earth Environ. 2023. https://doi.org/10.1038/s43247-023-00787-5.
81. Takacs-Vesbach C, Mitchell K, Jackson-Weaver O, Reysenbach A-L. Volcanic calderas delineate biogeographic provinces among Yellowstone thermophiles. Environ Microbiol. 2008;10(7):1681–9. https://doi.org/10.1111/j.1462-2920.2008.01584.x.
82. Vellend M. Conceptual synthesis in community ecology. Q Rev Biol. 2010;85(2):183–206. https://doi.org/10.1086/652373.
83. Stegen JC, Lin X, Fredrickson JK, Konopka AE. Estimating and mapping ecological processes influencing microbial community assembly. Front Microbiol. 2015;6:370. https://doi.org/10.3389/fmicb.2015.00370.
84. Mayol E, Arrieta JM, Jiménez MA, Martínez-Asensio A, Garcias-Bonet N, Dachs J, et al. Long-range transport of airborne microbes over the global tropical and subtropical ocean. Nat Commun. 2017;8(1):201. https://doi.org/10.1038/s41467-017-00110-9.
85. Barberán A, Henley J, Fierer N, Casamayor EO. Structure, inter-annual recurrence, and global-scale connectivity of airborne microbial communities. Sci Total Environ. 2014;487:187–95. https://doi.org/10.1016/j.scitotenv.2014.04.030.
86. Soininen J, Graco-Roza C. Homogeneous selection and stochasticity overrule heterogeneous selection across biotic taxa and ecosystems. Oikos. 2024. https://doi.org/10.1111/oik.10517.
87. Liu C, Yao M, Stegen JC, Rui J, Li J, Li X. Long-term nitrogen addition affects the phylogenetic turnover of soil microbial community responding to moisture pulse. Sci Rep. 2017;7(1):17492. https://doi.org/10.1038/s41598-017-17736-w.
88. He Q, Wang S, Feng K, Michaletz ST, Hou W, Zhang W, et al. High speciation rate of niche specialists in hot springs. ISME J. 2023. https://doi.org/10.1038/s41396-023-01447-4.
89. Caruso T, Chan Y, Lacap DC, Lau MCY, McKay CP, Pointing SB. Stochastic and deterministic processes interact in the assembly of desert microbial

communities on a global scale. ISME J. 2011;5(9):1406–13. https://doi.org/10.1038/ismej.2011.21.

90. Hamilton TL, Havig J. Meet Me in the middle: median temperatures impact cyanobacteria and photoautotrophy in eruptive yellowstone hot springs. MSystems. 2022;7(1): e0145021. https://doi.org/10.1128/msystems.01450-21.
91. Lacap DC, Barraquio W, Pointing SB. Thermophilic microbial mats in a tropical geothermal location display pronounced seasonal changes but appear resilient to stochastic disturbance. Environ Microbiol. 2007;9(12):3065–76. https://doi.org/10.1111/j.1462-2920.2007.01417.x.
92. Monchamp M-E, Spaak P, Domaizon I, Dubois N, Bouffard D, Pomati F. Homogenization of lake cyanobacterial communities over a century of climate change and eutrophication. Nature Ecol Evolution. 2018;2(2):317–24. https://doi.org/10.1038/s41559-017-0407-0.
93. Bahl J, Lau MCY, Smith GJD, Vijaykrishna D, Cary SC, Lacap DC, et al. Ancient origins determine global biogeography of hot and cold desert cyanobacteria. Nature Commun. 2011;2(1):163. https://doi.org/10.1038/ncomms1167.
94. Milke F, Wagner-Doebler I, Wienhausen G, Simon M. Selection, drift and community interactions shape microbial biogeographic patterns in the Pacific Ocean. ISME J. 2022;16(12):2653–65. https://doi.org/10.1038/s41396-022-01318-4.

Publisher's Note