Biogeography at the limits of life: Do extremophilic microbial communities show biogeographical regionalization?

Dave Clarke, Mégane Mathieu, Léonie Mourot, Laurent Dufossé, Graham J. C. Underwood, Alex J. Dumbrell, Terry J. Mcgenity

▶ **To cite this version:**

Dave Clarke, Mégane Mathieu, Léonie Mourot, Laurent Dufossé, Graham J. C. Underwood, et al.. Biogeography at the limits of life: Do extremophilic microbial communities show biogeographical regionalization?. Global Ecology and Biogeography, 2017, 26 (12), pp.1435-1446. 10.1111/geb.12670 . hal-01657083

HAL Id: hal-01657083

https://hal.univ-reunion.fr/hal-01657083

Submitted on 6 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Received: 22 May 2017 | Revised: 5 September 2017 | Accepted: 13 September 2017

DOI: 10.1111/geb.12670

RESEARCH PAPER

WILEY Global Ecology and Biogeography A Journal of Macroecology

Biogeography at the limits of life: Do extremophilic microbial communities show biogeographical regionalization?

Dave R. Clark[1] | Mégane Mathieu[2] | Léonie Mourot[2] | Laurent Dufossé[3] | Graham J. C. Underwood[1] | Alex J. Dumbrell[1] | Terry J. McGenity[1]

[1]School of Biological Sciences, University of Essex, Colchester, Essex, United Kingdom

[2]Ecole Nationale d'Industrie Laitière et des Biotechnologies, Mamirolle, France

[3]Laboratoire de Chimie des Substances Naturelles et des Sciences des Aliments, Université de La Réunion, École supérieure d'ingénieurs Réunion océan Indien Agroalimentaire, Sainte-Clotilde, Reunion Island, France

Correspondence
Terry J. McGenity and Alex J. Dumbrell, School of Biological Sciences, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, United Kingdom.
Emails: tjmcgen@essex.ac.uk; adumbrell@essex.ac.uk

Funding information
Natural Environment Research Council, Grant/Award Number: 471757

Editor: Jonathan Davies

Abstract

Aim: Biogeographical regions are the fundamental geographical units for grouping Earth's biodiversity. Biogeographical regionalization has been demonstrated for many higher taxa, such as terrestrial plants and vertebrates, but not in microbial communities. Therefore, we sought to test empirically whether microbial communities, or taxa, show patterns consistent with biogeographical regionalization.

Location: Within halite (NaCl) crystals from coastal solar salterns of western Europe, the Mediterranean and east Africa.

Time period: Modern (2006–2013).

Major taxa studied: Archaea.

Methods: Using high-throughput Illumina amplicon sequencing, we generated the most high-resolution characterization of halite-associated archaeal communities to date, using samples from 17 locations. We grouped communities into biogeographical clusters based on community turnover to test whether these communities show biogeographical regionalization. To examine whether individual taxa, rather than communities, show biogeographical patterns, we also tested whether the relative abundance of individual genera may be indicative of a community's biogeographical origins using machine learning methods, specifically random forest classification.

Results: We found that the rate of community turnover was greatest over subregional spatial scales (< 500 km), whereas at regional spatial scales the turnover was independent of geographical distance. Biogeographical clusters of communities were either not statistically robust or lacked spatial coherence, inconsistent with biogeographical regionalization. However, we identified several archaeal genera that were good indicators of biogeographical origin, providing classification error rates of < 10%.

Main conclusions: Overall, our results provide little support for the concept of biogeographical regions in these extremophilic microbial communities, despite the fact that some taxa do show biogeographical patterns. We suggest that variable dispersal ability among the halite-associated Archaea may disrupt biogeographical patterns at the community level, preventing the formation of biogeographical regions. This means that the processes that lead to the formation of biogeographical regions operate differentially on individual microbial taxa rather than on entire communities.

KEYWORDS

Archaea, dispersal, halite, halophiles, machine learning, macroecology, next generation sequencing, regionalization

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2017 The Authors. *Global Ecology and Biogeography* Published by John Wiley & Sons Ltd

1 | INTRODUCTION

The classification of Earth's biota into biogeographical regions separated by dispersal barriers has captivated ecologists for centuries (Sclater, 1858; Wallace, 1876). The concept of biogeographical regionalization has yielded insight into the origins of biodiversity and areas of endemism (Lamoreux et al., 2006), informed us of species' conservation status (Buckley & Jetz, 2007) and revealed historical connectivity between communities (Cowen, Paris, & Srinivasan, 2006). However, early attempts to define these regions have been superseded by more quantitative methods, improving the robustness and reproducibility of region delineations (Dapporto, Ciolli, Dennis, Fox, & Shreeve, 2015; Kreft & Jetz, 2010; Vilhena & Antonelli, 2015). Coupled with these new methods, the ever-increasing availability of species distribution data has renewed interest in the concept of biogeographical regionalization. As a result, a far greater range of taxa have been studied than ever before in order to define biogeographical regions (Holt et al., 2013). Yet, our knowledge about how Earth's biota may be divided into biogeographical regions is still overwhelmingly based on multicellular (and usually large) eukaryotes. Many inconspicuous, but functionally critical organisms, such as microorganisms, remain poorly studied. Consequently, it is unknown whether microbial communities may be grouped into biogeographical regions, similar to those observed for higher taxa.

Microorganisms are arguably the most functionally diverse and important organisms on Earth (Dinsdale et al., 2008; Fierer et al., 2012), driving every biogeochemical cycle (Falkowski, Fenchel, & Delong, 2008; Zak, Holmes, White, Peacock, & Tilman, 2003). Originally, microorganisms were assumed to have cosmopolitan distributions, with their small size and high population densities making them effective passive dispersers (Baas Becking, 1934; Finlay, 2002). From this assumption, it follows that biogeographical regionalization may not be possible because dispersal limitation is required for areas of endemism to occur (Ficetola, Mazel, & Thuiller, 2017) and produce regions with distinct communities. In contrast to cosmopolitanism, many recent studies have documented relationships between community turnover (the replacement of species) and geographical distance, indicative of dispersal limitation (e.g., Dumbrell, Nelson, Helgason, Dytham, & Fitter, 2010; Lear, Bellamy, Case, Lee, & Buckley, 2014), hinting that biogeographical regionalization of microbial communities could be possible. However, whilst the composition of microbial communities has been shown to differ over biogeographical regional scales, a formal test of whether microbial communities exhibit biogeographical regionalization is lacking.

In order to test for the presence of biogeographical regionalization in microbial communities, an ideal model community should have relatively low diversity, inhabit isolated environments and show a-priori evidence of dispersal limitation. The halite-associated Archaea fulfil these criteria. These Archaea typically belong to the class Halobacteria (more commonly referred to as haloarchaea) and are a major component of halite endolith communities (Henriet, Fourmentin, Delincé, & Mahillon, 2014). Their entombment into the brine inclusions of halite crystals is believed to be an escape mechanism from desiccation and the increasingly chaotropic conditions present in evaporating brines (Hallsworth et al., 2007). Within these pockets they are able to survive over geological time scales (Gramain, Díaz, Demergasso, Lowenstein, & McGenity, 2011; McGenity, Gemmell, Grant, & Stan-Lotter, 2000). As with many extremophilic microbial communities, the halite-associated Archaea are typically less diverse than other microbial systems, facilitating more exhaustive sampling of the total diversity and improving detection of the less abundant endemic taxa, which are indicative of biogeographical regions. Furthermore, these Archaea occupy isolated 'habitat islands' that are physicochemically distinct from the surrounding environment. Many haloarchaea are obligately halophilic and lyse in less saline conditions (Oren, 1994), such as seawater, rendering the surrounding environment a physiological dispersal barrier. Finally, halite crystals form in highly similar conditions worldwide (i.e., saturated NaCl), thus ensuring that species filtering by the environment is low compared with many other environments. Any physicochemical differences between halite crystals (e.g., caused by underlying geology or climate) should themselves be spatially autocorrelated, meaning that species filtering by the environment should enhance, rather than obscure, biogeographical clustering. Such systems are therefore ideal for studies of community turnover and biogeography (Santos, Field, & Ricklefs, 2016). Previous studies of halophilic microbial communities have found evidence of community turnover at regional scales (Pagaling et al., 2009; Zhaxybayeva, Stepanauskas, Mohan, & Papke, 2013), suggesting the potential for biogeographical regions to form. Overall, these properties render the halite-associated Archaea an ideal system in which to test for biogeographical regionalization of microbial communities.

Therefore, we examine the regional turnover (replacement of species over biogeographical regional scales) of halite-associated archaeal communities to test whether communities group together in a manner consistent with biogeographical regionalization. Using high-throughput Illumina HiSeq amplicon sequencing, we characterize the archaeal communities of halite from 17 locations, spanning three geographical regions. We apply robust biogeographical clustering methods to determine the extent to which archaeal communities, and taxa, show spatial patterns consistent with biogeographical regionalization. We propose the following three hypotheses:

1. (a) There will be a significant relationship between community turnover and geographical distance, and (b) the rate of community turnover will be greater at biogeographical regional scales than at within-region scales.
2. Communities will form biogeographical clusters that are statistically well supported and spatially coherent.
3. The presence and abundance of some archaeal taxa can predict the (bio)geographical origin of each community.

2 | MATERIALS AND METHODS

We obtained 27 halite samples (in triplicate) from 17 locations in the years between 2006 and 2013 (Figure 1 and Supporting Information

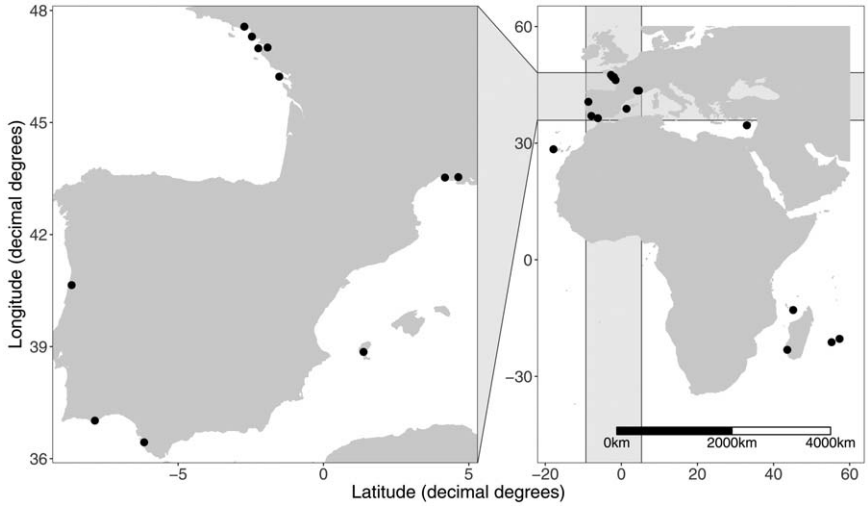

FIGURE 1 Map of sample locations. Further details of samples are available in Supporting Information Appendix S2. The left panel is zoomed in on the grey region in order to distinguish multiple locations along the southwest coast of Europe

Appendix S1). A photographic record of the samples and further details can be found in Supporting Information Appendix S2. We recorded the grain size, which reflects the time taken for the crystals to form, and the impurity colour, which provides a qualitative measure of the types of impurities and physicochemical environment present within the crystal (Sonnenfeld, 1995). Samples were stored in the dark at room temperature.

2.1 | Molecular analyses

DNA was extracted from a 0.25 g aliquot of each sample using MoBio PowerSoil DNA isolation kits, following the manufacturer's instructions (MoBio Laboratories Inc., Carlsbad, CA, U.S.A.). To characterize the archaeal communities, we used a Nextera XT dual indexing strategy, which involves polymerase chain reaction (PCR) amplification of a phylogenetic marker gene, followed by a secondary short-cycle PCR amplification, in which dual Nextera indices are added to the amplicon for multiplexing of samples. We targeted a *c*. 570 bp region of the 16S rRNA gene with the Archaea-specific primers 344F (5′-ACGGGGYG CAGCAGGCGCGA-3′; Raskin, Stromley, Rittmann, & Stahl, 1994) and 915R (5′-GTGCTCCCCCGCCAATTCCT-3′; Stahl & Amann, 1991), both of which were modified to contain Illumina-specific overhang sequences. The 16S rRNA gene was amplified in 25 μl reactions with 12.5 μl of REDTaq® ReadyMixTM (Sigma-Aldrich Co.), 5 μl of each primer (1 μM) and 2.5 μl of template DNA. The PCR protocol included an initial denaturation step at 95 °C for 5 min, followed by 32 cycles of 95 °C for 45 s, 60 °C for 45 s and 72 °C for 1 min. After a final extension step of 72 °C for 5 min, PCR products were held at 4 °C. We purified PCR products using Agencourt AMPure XP PCR Purification beads (Beckman Coulter Ltd, High Wycombe, U.K.) following Illumina's '16S Metagenomic Sequencing Library Preparation' document (https://goo.gl/3Y7oY4).

The index PCR was carried out in 50 μl reactions with 25 μl of KAPA HiFi HotStart ReadyMix (KAPA Biosystems, Wilmington, MA, U.S.A.), 5 μl each of sample specific i5 and i7 Nextera XT index (Illumina), 10 μl of PCR-water (Bioline Reagents Ltd, London, U.K.) and 5 μl of purified PCR product. The PCR was conducted with an initial denaturation at 95 °C for 3 min, followed by eight cycles of 95 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s. Again, a final extension step was included at 72 °C for 5 min, after which PCR products were held at 4 °C. The PCR products were purified using Agencourt AMPure XP PCR Purification beads (Beckman Coulter Ltd) and quantified on a POLARstar Omega (BMG LABTECH GmbH, Ortenburg, Germany) plate reader using the PicoGreen® dsDNA assay. The PCR products were then pooled in equimolar concentrations. The size and concentration of the resulting pool was checked using an Agilent 2100 Bio-analyser (Agilent, Palo Alto, CA, USA). Sequencing was carried out on an Illumina HiSeq 2500 on rapid-run mode, producing two 300 bp sequences, at The Earlham Institute (formerly The Genome Analysis Centre, Norwich Research Park, Norfolk, UK).

2.2 | Bioinformatic analyses

Owing to the length of the amplicon, it was not possible to pair-end align the forward and reverse sequences; therefore, all analyses were based on forward sequences only. This approach has been shown to have little effect on ecological conclusions (Werner, Zhou, Caporaso, Knight, & Angenent, 2012), and in our case, the forward sequence spans the V3 region of the 16S rRNA gene, which has been shown to perform well for profiling archaeal communities (Yu, García-González, Schanbacher, & Morrison, 2008). Sequences were processed according to guidelines outlined by Dumbrell, Ferguson, and Clark (2016). Briefly, we quality trimmed sequences using Sickle (Joshi & Fass, 2011) at a threshold of Q20, trimming only the 3′ end of the sequence and discarding sequences with ambiguous nucleotides. Quality-trimmed sequences were error corrected using the BayesHammer algorithm implemented in SPAdes version 3.10.1, with default parameters (Bankevich et al., 2012; Nikolenko, Korobeynikov, & Alekseyev, 2013). We removed primer sequences, calculated library sizes for each sample, and discarded sequences $<$ 230 nucleotides in length using Linux shell

commands. Samples with excessively small library sizes (< 20,000 sequences) were excluded from further analyses.

We used VSEARCH (Rognes, Flouri, Nichols, Quince, & Mahé, 2016) to cluster sequences into operational taxonomic units (OTUs). First, sequences were de-replicated and singleton sequences discarded, as they are more likely to be artefacts (Flynn, Brown, Chain, MacIsaac, & Cristescu, 2015). We then clustered sequences into OTUs at 97 and 99% sequence similarity (referred to as 97% dataset and 99% dataset). The 97% similarity threshold is the most frequently used, corresponding approximately to intragenus-level similarity (Yarza et al., 2014). The 99% threshold approximates to species-level similarity. We screened OTUs for chimeras against the RDP database (Cole et al., 2009) using VSEARCH, and discarded putative chimeras.

Taxonomy was assigned to OTUs using the RDP classifier (Wang, Garrity, Tiedje, & Cole, 2007), with a minimum confidence threshold of 0.7. We discarded all non-archaeal OTUs. Specific OTUs of interest were identified using BLAST searches against NCBI's 16S ribosomal RNA sequence database (Altschul, Gish, Miller, Myers, & Lipman, 1990).

2.3 | Statistical analyses

We rarefied OTU tables to the smallest library size in each dataset (97% dataset, 27,554 sequences; 99% dataset, 26,578). We checked whether sample age was influencing the OTU richness or community composition using a negative binomial generalised linear model (GLM) and permutation-based multivariate analysis of variance (PERMANOVA), respectively.

In order to address our first hypothesis, we quantified community turnover using the β_{sim} index, which purely quantifies community turnover, the process relevant to biogeographical regionalization (Baselga, 2010), and not nestedness, whereby communities are subsets of each other. Geographical distances between sampled communities were calculated as geodesic distances (Hijmans, 2016). We then tested for correlation between community turnover and geographical distance using Mantel tests, with Spearman's correlation coefficient and 10,000 permutations. We fitted piece-wise regressions to determine breakpoints in the relationship, showing the geographical distance at which the slope of the relationship changes (Castro-Insua, Gómez-Rodríguez, & Baselga, 2016).

To investigate our second hypothesis, we adopted a clustering approach as described by Kreft and Jetz (2010). Briefly, this approach involves clustering communities based on the β_{sim} turnover matrix, creating a dendrogram. This dendrogram can be split into *k* clusters representing bioregions. The quality and biological interpretability of the resulting clusters are then checked via statistical metrics and mapping. Biogeographical regionalization may be inferred when clustering solutions are both statistically robust and spatially coherent.

To cluster communities, we used three different clustering algorithms to ensure our conclusions were robust. The unweighted pair-group method using arithmetic averages (UPGMA) defines the distance between clusters as the average distance between all the communities within each cluster. Kreft and Jetz (2010) found that UPGMA best preserved information present in the original distance matrix. Dapporto, Ciolli et al. (2015) also compared clustering algorithms on datasets of varying completeness. They found that for less well-sampled datasets, the Ward method clustered communities most accurately, whereas for intensely sampled datasets, PAM produced the most accurate clusters. To cluster the communities, we used the methodology described by Dapporto, Ramazzotti et al. (2015). This approach overcomes the biases introduced by having zero similarity or tied values in the dissimilarity matrix (Bloomfield, Knerr, & Encinas-Viso, 2017), by repeatedly reshuffling the matrix and reclustering communities. The final clustering solution is then determined by the frequency at which pairs of communities are clustered together in the randomly generated cluster solutions, allowing a more robust final clustering solution. We set the number of matrix randomizations to 50 and the number of clusters (*k*) from 2 to 16. For each value of *k*, we assessed the statistical support of the cluster solution with two metrics, 'mean silhouette width' and 'explained dissimilarity'. The first metric, 'mean silhouette width', is a commonly used metric to evaluate clustering solutions and ranges from minus one, indicating that most communities have been incorrectly clustered, to one, indicating that most communities are correctly clustered. Values < .25 are qualitatively considered to show little evidence of true clustering between the communities (Kaufman & Rousseeuw, 1990). Our second metric, 'explained dissimilarity' (Holt et al., 2013), is a ratio of sums of mean dissimilarity within regions to total dissimilarity over the entire dissimilarity matrix. Explained dissimilarity tends towards one as *k* tends towards the number of communities. We follow the approach of Holt et al. (2013), who indicated that a threshold of .9 provides sufficient support to infer regionalization. However, we also examined the cluster solution that produced the greatest incremental increase in explained dissimilarity, which we refer to as the 'knee solution', because this has been proposed to be a more suitable indicator of optimal cluster number (Kreft & Jetz, 2013). After identifying statistically supported clustering solutions, we determined the spatial coherence of clusters by mapping them. To check whether the measured physicochemical parameters (grain size or impurity) explained any clustering patterns observed, we used PERMANOVA. We included location as the first variable in the model to account for confounding spatial effects. Statistical significance of physicochemical variables was then assessed based on the 'marginal' effects (e.g., after controlling for spatial location), with 999 random permutations. We conducted non-metric multidimensional scaling (NMDS) analysis as a means of visualizing these results.

To test our third hypothesis, we investigated whether the relative abundance of halite-associated archaeal genera could predict the biogeographical origin of a given community using the machine learning method of random forest classification (RFC). Random forests provide an effective method for classification in ecology (Cutler et al., 2007) and are built from an ensemble of classification trees, in which observations of the dependent variable form the leaves and independent variables form the branches. Each tree is trained on a subset of observations and independent variables, and the overall classifier is built by combining predictions from these trees to obtain a more robust classification. We summed the abundances of all OTUs identified to

the genus level and converted these abundances to relative abundances. OTUs not identified to genus were excluded from this analysis. We classified communities (see Supporting Information Appendix S1) based on their biogeographical region (classes: Palearctic, Saharo-Arabian, Madagascan as defined by Holt et al., 2013), geographical region (classes: eastern Europe, western Europe, Mediterranean or west African) and nearest ocean (classes: Atlantic or Indian). We initialized 10,000 trees, and each tree was trained on six archaeal genera. We normalized the sample size from each class to the size of the smallest class to minimize the effects of class size imbalance (e.g., more observations of European communities than African communities). Additionally, for the biogeographical and geographical classifiers, we dropped excessively small classes (Saharo-Arabian; $n = 4$ and west African; $n = 3$), to reduce the imbalance between classes further. We evaluated the overall accuracy of each classifier using the out-of-bag error rate, which quantifies the classifier's ability to classify a given community correctly when it is excluded from the training set. We determined which archaeal genera were the best predictors of biogeographical origin by quantifying variable importance, using the mean decrease in accuracy (MDA) and mean decrease in Gini index (MDGI). The MDA shows the change in accuracy of the classifier with and without a given variable. Important variables will result in a large decrease in accuracy when they are excluded from the classifier, resulting in large MDA values. The MDGI shows the purity of the groups created when the classifier splits the dataset using a given predictor. A good predictor will create homogeneous groups, in which all data points belong to the same class, resulting in a large decrease in MDGI. We also examined partial dependence plots (Hastie, Tibshirani, & Friedman, 2009). In the context of our study, these plots show how the probability of a community being classified into a given biogeographical region changes in relationship to the relative abundance of a given archaeal genus.

All analyses were conducted in R (R Developement Core Team, 2016), using the *vegan* (Oksanen et al., 2015), *recluster* (Dapporto, Ramazzotti et al., 2015) and *randomForest* (Liaw & Wiener, 2002) packages.

3 | RESULTS

3.1 | Diversity of halite-associated Archaea

An initial 17.8 million sequences were reduced to 10.33 million after quality trimming. Error correction, length filtering, and removal of small samples further decreased this total to 10.29 million sequences. These sequences clustered into 1,581 and 10,346 OTUs at the 97 and 99% similarity thresholds, respectively. Sixteen non-archaeal OTUs (12 Bacteria, four unclassified) were removed from each dataset, comprising a total of 294 sequences (< 0.0001% of total sequences). Of the archaeal OTUs, 45.2% were identified to genus level from the 97% dataset, and 59.5% from the 99% dataset (Supporting Information Appendix S4). At the 99% similarity level, these OTUs represented 40 genera from five families (Supporting Information Appendix S3) as identified by the RDP taxonomy; the Halobacteriaceae, Haloferacaceae, Natrialbaceae, Methanosarcinaceae and Nitrososphaeraceae. Most OTUs (58% from the 97% dataset, 79.9% from the 99% dataset) were restricted to 20 or fewer samples, but five OTUs in the 97% dataset and three OTUs in the 99% dataset were detected in every sample (Supporting Information Appendix S6). BLAST analysis of these OTUs revealed their most closely related species as *Halobacterium noricense* (OTU1), *Halorubrum orientale* (OTU2), *Halorubrum sodomense* (OTU21), *Halolamina sediminis* (OTU5) and *Halolamina salina* (OTU92510).

Sample age did not significantly affect OTU richness (97% dataset: slope = −0.01, z-statistic = −0.30, $p = .77$; 99% dataset: slope = −0.02, z-statistic = −0.83, $p = .41$), whereas PERMANOVAs showed that age had a small, but significant, effect on turnover (97% dataset: pseudo-$F_{1,74} = 2.50$, $R^2 = .03$, $p = .03$; 99% dataset: pseudo-$F_{1,74} = 3.13$, $R^2 = .04$, $p = .003$).

3.2 | How is community turnover related to geographical distance?

Mantel tests, used to investigate the relationship between community turnover and geographical distance, showed significant and positive relationships for both datasets (97% dataset: $r_{Mantel} = .26$, $p < .0001$; 99% dataset: $r_{Mantel} = .31$, $p < .0001$), which supports hypothesis 1a. However, piece-wise regressions between geographical distance and community turnover suggested that this correlation was largely driven by high turnover at small spatial scales (Figure 2). For both (97 and 99%) datasets, a steep positive relationship was found at smaller spatial scales, with breakpoints estimated at 420.5 km (standard error = 46.9 km) and 334.6 km (standard error = 23.7 km), respectively. After these breakpoints, community turnover was independent of geographical distance (Figure 2). Davies tests confirmed that the pre-breakpoint slope was significantly greater than the post-breakpoint

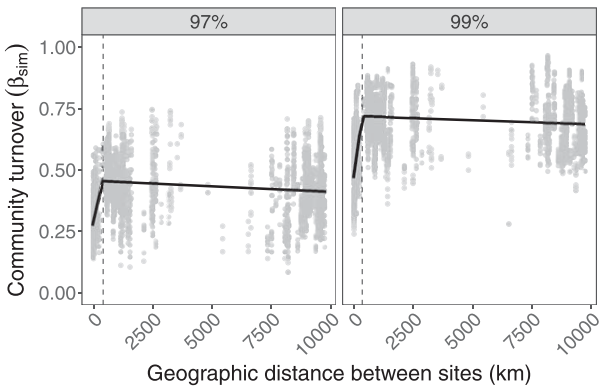

FIGURE 2 The relationship between community turnover and geographical distance, for 97 and 99% similarity operational taxonomic unit (OTU) tables. Values close to one indicate pairs of communities highly similar in composition, whereas values close to zero indicate communities with few OTUs in common. Dashed lines indicate breakpoints (distance in kilometres at which slope changes), which were estimated as 420.5 km (standard error = 46.9 km) and 334.6 km (standard error = 23.7 km). Mantel tests showed statistically significant correlation in both cases ($p < .0001$ in both cases)

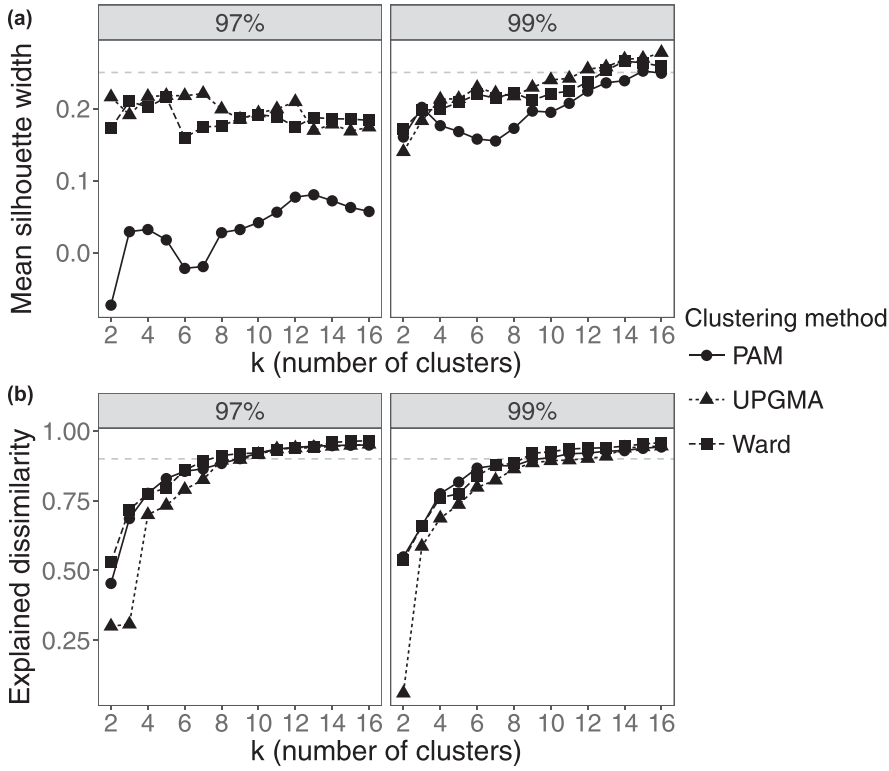

FIGURE 3 The statistical support, quantified as (a) mean silhouette width and (b) explained dissimilarity, of cluster solutions from 2 to 16 clusters, for both 97 and 99% operational taxonomic unit (OTU) datasets. Lines represent three different clustering algorithms used; partitioning around medoids (PAM), unweighted pair group method (UPGMA) and Ward clustering (Ward). In (a), silhouette widths $< .25$ (grey dotted line) are interpreted as showing poor clustering in the data, and in (b) explained dissimilarity of $> .9$ indicates a good cluster solution

slope ($p < .0001$ in both cases), showing that the greatest rate of community turnover was at small, subregional scales, and thus rejecting hypothesis 1b.

3.3 | Do microbial communities cluster into biogeographical regions?

We determined whether archaeal communities group into biogeographical regions by applying three different clustering algorithms (UPGMA, Ward and PAM). To assess the degree of biogeographical clustering within these communities, we first determined the appropriate number of clusters (k) into which our communities should be grouped by examining the cluster quality (using mean silhouette width and explained dissimilarity) for values of k from 2 to 16. For the 97% dataset, statistical support for cluster solutions was poor, as the mean silhouette width never exceeded .25 for any value of k (Figure 3a). In contrast, for the 99% dataset, all three clustering algorithms exceeded .25 for values of $k > 12$, showing that reasonable statistical support was gained when communities were grouped into > 12 regions. All three clustering algorithms yielded similar results when assessed by the explained dissimilarity metric (Figure 3b). Explained dissimilarity values $> .9$ were considered to provide good support for a given cluster solution. To satisfy this threshold, communities were grouped into 8–10 (97% dataset) or 9–12 (99% dataset) clusters, depending on the cluster algorithm used. For both 97 and 99% datasets, the Ward algorithm required the fewest clusters to reach this threshold, and UPGMA the most. We also identified the number of clusters (k) that resulted in the greatest increase in explained dissimilarity (knee solutions). For the 97% dataset, this occurred when communities were clustered into three (PAM and Ward) or four (UPGMA) clusters, whereas for the 99% dataset, the greatest increase in explained dissimilarity was found when communities grouped into three (UPGMA and Ward) or four (PAM) clusters.

We examined the spatial coherence of cluster solutions for the minimal number of clusters (k) required to exceed the explained dissimilarity threshold of .9, as well as solutions that yielded the greatest increase in explained dissimilarity. For both 97 and 99% datasets and all three clustering algorithms, mapping revealed poor spatial coherence (Figure 4), in disagreement with hypothesis 2, suggesting little support for biogeographical regionalization. There was a large degree of mixing between communities on the west European coastline, Mediterranean and Madagascar, counter to our expectation that communities in these regions would cluster separately. Mapping of the knee solutions again revealed clusters with poor spatial coherence, with many European communities clustering together with Madagascan communities (Supporting Information Appendix S7). NMDS and PERMANOVA showed that archaeal communities clustered only weakly by impurity, but not by grain size (Figure 5).

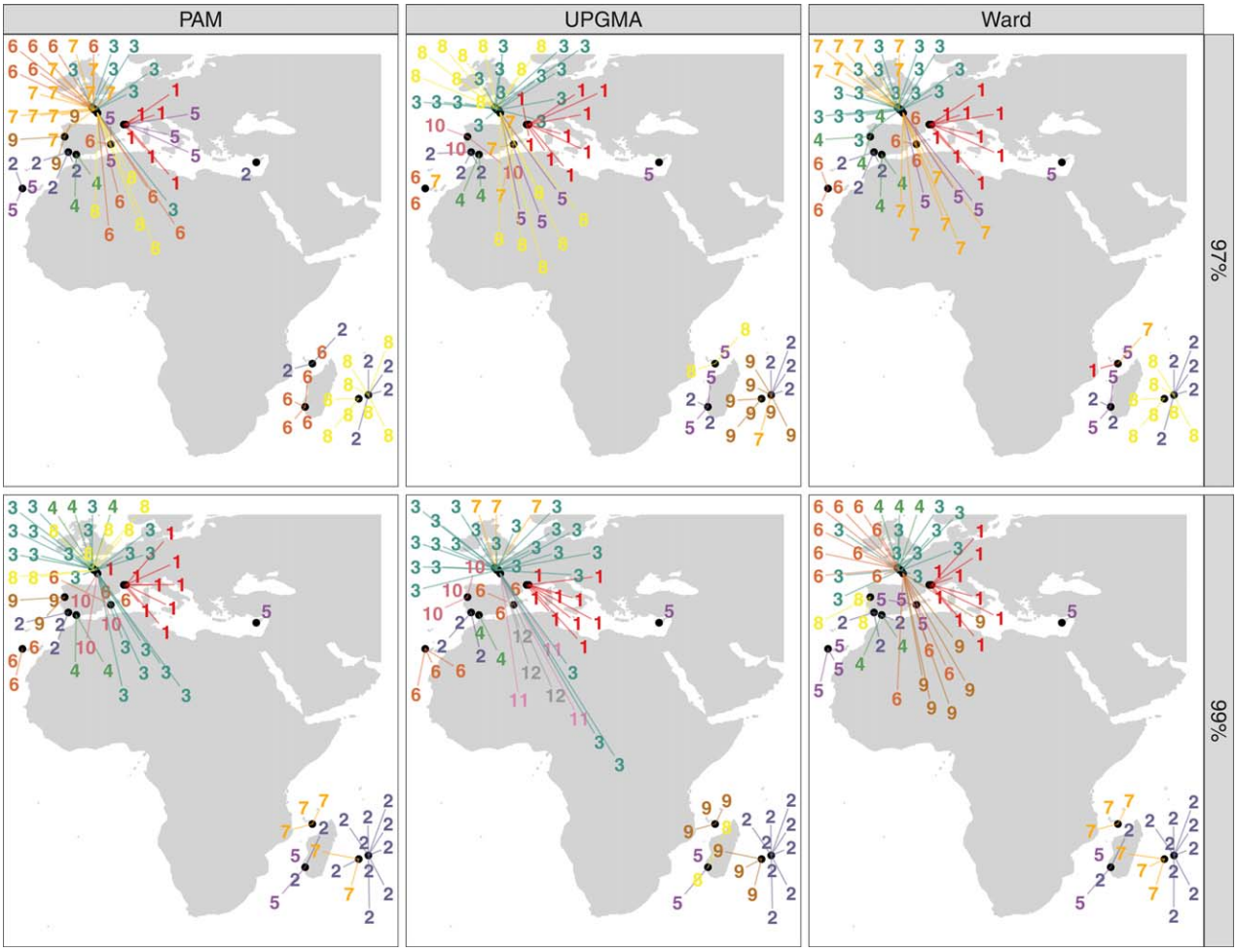

FIGURE 4 The cluster memberships (indicated by colour and number) of communities for each clustering algorithm, for both 97 and 99% operational taxonomic unit (OTU) datasets. For each algorithm, the cluster solution shown is for the minimal value of *k* (number of clusters) that exceeded the explained dissimilarity threshold of .9

3.4 | Can certain haloarchaeal genera be used as indicators of a community's biogeographical origin?

We tested whether the abundance of certain archaeal genera could predict any of three classifiers (biogeographical region, geographical region and nearest ocean) of a community, using random forest classification. All three classifiers performed well, with comparable accuracies (ocean, error rate = 9.33%; biogeographical region, error rate = 8.45%; geographical region, error rate = 8.33%), showing that the biogeographical origin of a community can be predicted accurately from the relative abundance of individual genera. Each classifier was able to predict communities from different biogeographical origins with similar accuracy, suggesting that archaeal relative abundances were equally useful predictors for all classes. The oceanic RFC classified with similar accuracy those communities nearest to the Atlantic or Indian Ocean, with class errors of 8.9 and 10.5%, respectively. The biogeographical region RFC identified communities from the Palearctic region with a 7.7% class error rate, and those from the Madagascan region with a 10.5% class error rate, whereas the geographical region RFC more accurately classified west European (class error = 7.5%) and Mediterranean (class error = 7.7%) communities than east African communities (class error = 10.5%).

To determine which archaeal genera were the best predictors of a community's oceanic, biogeographical or geographical origin, we quantified the importance of each variable (genus) to each RFC (Supporting Information Appendix S4). *Haloquadratum* was the best genus for classifying geographical region, followed by *Halapricum* and *Halobaculum*. Partial dependence plots revealed that, as the relative abundance of *Haloquadratum* exceeded .01, the probability of the community being classified as Mediterranean increased greatly (Figure 6b), reflecting its higher relative abundance in the region (Figure 6a). In contrast, the genera *Halarchaeum* and *Halohasta* were the best for classifying a community's nearest oceanic or biogeographical region, according to both metrics of variable importance (MDA and MDGI). When the relative abundance of *Halarchaeum* exceeded .02, a classification of the community's nearest ocean and biogeographical region as the Indian Ocean and Madagascan biogeographical region, respectively, was most likely (Supporting Information Appendix S8). The finding that certain archaeal genera are good predictors of a community's (bio)geographical origin supports hypothesis 3.

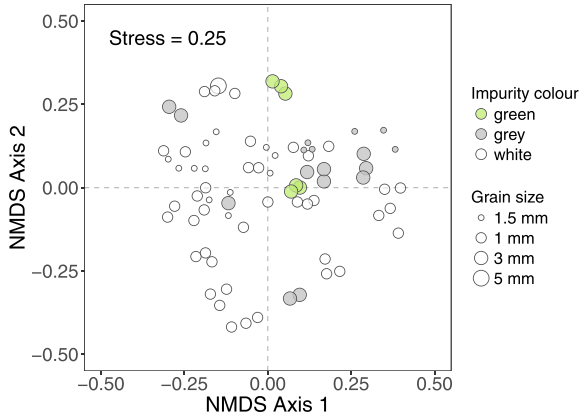

FIGURE 5 Non-metric multidimensional scaling (NMDS) analysis of halite-associated archaeal communities. Each point represents a single community, and points closer together represent compositionally more similar communities. Communities do not appear to cluster by halite properties. Permutation-based multivariate analysis of variance (PERMANOVA) revealed that, after accounting for spatial location, grain size had no significant effect on community composition (pseudo-$F_{1,55} = 2.09$, $R^2 = .01$, $p = .06$), whereas impurity had a significant, but negligible, effect (pseudo-$F_{1,55} = 3.15$, $R^2 = .02$, $p < .05$)

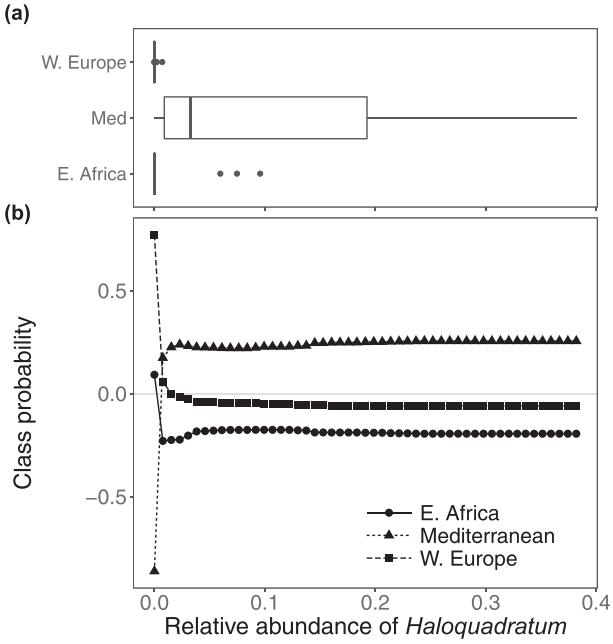

FIGURE 6 (a) The relative abundance of the genus *Haloquadratum* in samples of different geographical origins (E. Afr = East Africa; Med = Mediterranean; W. Eur = West Europe). (b) A partial dependence plot based on a random forest classification. Class probability shows the probability that the random forest classifies a sample to each class (denoted by different line and point styles). As the relative abundance of *Haloquadratum* was notably higher in the Mediterranean, the Mediterranean class probability increased rapidly

4 | DISCUSSION

We studied halite-associated Archaea to determine whether archaeal communities can be clustered into biogeographical regions comparable to those observed for most higher organisms. Our results show that, despite community turnover correlating with geographical distance over small spatial scales (< 500 km), communities do not cluster into spatially coherent biogeographical regions. We found little statistical support for clustering communities into few (two or three) biogeographical regions, which would be the number of regions expected for higher organisms, such as terrestrial vertebrates (Holt et al., 2013) or plants (Takhtajan, 1986). Furthermore, when we clustered communities into a greater number of regions, the spatial configuration of these regions was not consistent with biogeographical regionalization. Lastly, we demonstrated that although communities may not show the expected biogeographical patterns, some individual genera do, as their abundances were found to be good predictors of the biogeographical origin of the community.

Numerous studies have demonstrated that microbial communities differ over continental to regional scales (Lauber, Hamady, Knight, & Fierer, 2009; Papke, Ramsing, Bateson, & Ward, 2003; Whitaker, Grogan, & Taylor, 2003), including studies on halophilic microbes (Pagaling et al., 2009; Zhaxybayeva et al., 2013). However, to our knowledge, no studies have tested quantitatively whether such differences are consistent with the concept of biogeographical regionalization, thus it remains unknown whether the processes that shape microbial communities are capable of forming biogeographical patterns over the spatial scales relevant to other organisms. Glassman et al. (2015) examined fungal spore banks of soils across North America, showing that community turnover was significantly related to geographical distance and, using ordination techniques, that fungal communities appeared to group in a regional manner. Consistent with our study, they found that the highest rate of community turnover occurred over subregional scales, as evidenced by their Mantel correlogram, which shows change from positive to negative correlation over spatial scales of *c.* 500 km. Initially, this might indicate that microbial biogeographical regions are smaller than those defined for higher taxa and more comparable to subregions. However, in our study, this idea is poorly supported by the fact that even for larger values of *k* (indicating more and smaller regions), the spatial coherence of these clusters was poor. A global study of soil fungi (Tedersoo et al., 2014) revealed communities that did not cluster in a spatially coherent manner, which is in contrast to the findings of Glassman et al. (2015) and in agreement with our results. For instance, fungal communities of Europe clustered with those of North America, and those of Oceania clustered with South America. Furthermore, a study of the bacterial communities on *Tamarix* spp. leaf surfaces showed that communities clustered in a manner at odds with their spatial configuration (Finkel et al., 2012). Specifically, communities from around the Dead Sea (Middle East) clustered more closely with those from the Sonoran Desert (North America) than Mediterranean communities. Combined with our results, these studies provide further evidence that biogeographical regionalization may be unlikely in microbial communities.

One possible reason for no evidence of biogeographical regionalization in these communities is that some halophilic Archaea may be differentially susceptible to long-distance dispersal. Previous studies of halophilic microbial communities have identified several potential

mechanisms for long-distance dispersal of haloarchaea. Despite the hostility of this environment, animal vectors may passively disperse viable Archaea between sites. Organisms such as birds and invasive invertebrates, such as brine shrimp (*Artemia* spp.), have been found to harbour diverse haloarchaea (Brito-Echeverría, López-López, Yarza, Antón, & Rosselló-Móra, 2009; Riddle, Baxter, & Avery, 2013; Yim et al., 2015), which may help them to spread between habitat islands. Furthermore, wind- or human-mediated dispersal of halite crystals may disperse entombed haloarchaea. Wind is known to play a role in dispersing free-living microbes over continental distances (Favet et al., 2013; Kellogg & Griffin, 2006) and is likely to disperse small halite crystals, along with endolithic microbes, over such distances. Human transport of salt as a commercial product and as a de-icing agent on roads may also aid the dispersal of halite endolithic communities. However, such dispersal would select for those Archaea capable of survival in halite crystals, filtering out some taxa, as evidenced by the disparity between brine and halite crystal archaeal communities described previously (Henriet et al., 2014). Finally, dispersal via seawater could be possible for some haloarchaeal taxa, because viable cells have been isolated from seawater and coastal sediments (Purdy et al., 2004; Rodriguez-Valera, Ruiz-Berraquero, & Ramos-Cormenzana, 1979). Seawater may also provide a means of dispersal between ancient and modern halite deposits (McGenity, Hallsworth, & Timmis, 2008). Ancient halite deposits can become exposed in deep water horizons, where they may dissolve, creating stratified deep-sea brines, which are a potential source of extremely halophilic Archaea (Antunes, Ngugi, & Stingl, 2011). However, although short-term (*c*. 24 hr) or partial survival at seawater salinity has been found in a number of haloarchaea (Torreblanca et al., 1986), the majority of genera detected in this study, particularly the most abundantly detected genera, are known exclusively from hypersaline habitats, and their cells lyse at seawater salinity. Therefore, seawater is an unlikely medium for their dispersal. Furthermore, the deposition of cells from ancient halite into modern hypersaline environments would be most likely to occur over regional extents (e.g., owing to oceanic currents), thus increasing the compositional similarity of sites within a region. Finally, even with connectivity between ancient and modern halite, there is no guarantee that those cells will become established and multiply (Jones, Ramoneda, Rivett, & Bell, 2017). Therefore, the influence of ancient haloarchaea on the clustering patterns observed here should be minimal. Even so, the degree to which other potential dispersal vectors contribute to connectivity between sites is unknown and warrants further research, as connectivity between contemporary halite deposits may be a better measure of isolation for these communities than geographical distance alone.

An alternative explanation as to why biogeographical clustering was not observed in these archaeal communities is that environmental filtering, because of physicochemical differences between the halite crystals, could obscure biogeographical clustering. Within hypersaline systems, salinity (concentration of sodium chloride; NaCl) has been shown to be the predominant physicochemical variable causing environmental filtering of microbial communities (Baati et al., 2008; Benlloch et al., 2002; Casamayor et al., 2002; Herlemann et al., 2011). However, the role of physicochemical differences in structuring microbial communities between hypersaline habitats is less well known, as most research has focused on within-site salinity gradients. Despite this, we suggest that physicochemical differences between halite samples are unlikely to explain the clustering patterns observed. Halite is an evaporite mineral, formed by the precipitation of sodium chloride from concentrated brine. Given that halite precipitates only when the concentration of NaCl (sodium chloride) exceeds *c*. 32% w/v (McGenity et al., 2000), it is not possible for large differences in NaCl concentrations to occur between sites. Furthermore, as all the halite samples used here were formed in the same way (i.e. by progressive evaporation of seawater), the precipitation point of halite is most likely to be similar across sites. Other ions are also present in varying concentrations within the source brines that could have an effect on the composition of archaeal communities within the brine. Differences in the concentrations of these ions may be caused by differing underlying geology or by differing climate. However, both geology and climate are themselves spatially autocorrelated. Therefore, if physicochemical differences between habitats dictate differences in the microbial communities, we would expect these effects to enhance any biogeographical clustering, because sites within the same region will have physicochemically similar brines. Nonetheless, we observed little evidence of environmental filtering on the microbial communities, suggesting that the physicochemical environment has a minimal influence on our conclusions.

The fact that these dispersal vectors are likely to disperse haloarchaea with differing physiological capabilities selectively may explain why, despite finding no evidence of biogeographical regionalization at the community level, our population-level analyses revealed several haloarchaeal genera with distinct biogeographical patterns. The square haloarchaeon, *Haloquadratum*, was found to be a good indicator of geographical region, as it was found in abundance in the Mediterranean, yet was scarce in western Europe and eastern Africa. Despite this, *Haloquadratum* has been detected globally in hypersaline brines (Di Meglio et al., 2016; Oh, Porter, Russ, Burns, & Dyall-Smith, 2010; Podell et al., 2014). A previous study of halite-associated Archaea found *Haloquadratum* to be a very small component of the halite-associated community, despite being highly abundant in the hypersaline brine that was the source of the halite (Henriet et al., 2014). Furthermore, Gramain et al. (2011) demonstrated that *Haloquadratum* resumed growth slowly after halite entombment compared with other haloarchaea, inferring that it is a relatively poor survivor in halite. Yet, this fails to parsimoniously explain our finding that *Haloquadratum* was an abundant member of Mediterranean halite samples. Significantly, Gramain et al. (2011) also observed that the recovery time of *Haloquadratum* was dramatically enhanced when co-entombed with the geographically widespread halophilic bacterium, *Salinibacter ruber* (Antón et al., 2008; Di Meglio et al., 2016; Ventosa, de la Haba, Sánchez-Porro, & Papke, 2015). Despite the ubiquity of *S. ruber* in hypersaline environments, metabolomic profiles of geographically distant strains show biogeographical patterns (Rosselló-Mora et al., 2008). We speculate that the presence of a particular *S. ruber* variant or other halophilic organism in this region may facilitate the survival of *Haloquadratum* in halite, perhaps via metabolite transfer (Bolhuis, Te Poele, & Rodríguez-Valera, 2004; Elevi Bardavid & Oren, 2008). We also identified *Halarchaeum* as the best genus in

predicting a sample's oceanic and biogeographical origins, because it was largely restricted to Madagascan samples. Despite this finding, *Halarchaeum* spp. have been isolated previously from globally distributed commercial salt samples (Minegishi, Echigo, Nagaoka, Kamekura, & Usami, 2010; Shimane et al., 2015; Youssef, Ashlock-Savage, & Elshahed, 2012), hinting that, despite its wide distribution, it may be highly abundant only in certain regions.

Overall, we found little evidence to support the existence of biogeographical regions in communities of extremely halophilic Archaea. We demonstrated that, despite finding evidence of a distance-decay relationship in these communities, clustering them into regions did not produce spatially coherent regions. We suggest that the cause of this may be long-distance dispersal of some haloarchaeal taxa, as we identified three particularly abundant and widespread species that were universally detected across all samples. However, certain individual taxa are able to indicate a given community's biogeographical origins accurately, suggesting highly differential dispersal abilities in haloarchaea. Taken together, our results suggest that geographical distance alone may be a poor indicator of isolation in microbial communities and that more work is needed to examine the role of connectivity in microbial biogeography.

ACKNOWLEDGMENTS

We are grateful to Dr Andrew Crombie (University of East Anglia) for providing the Portuguese samples and to Dr Pascal Danthu (Centre de Coopération Internationale en Recherche Agronomique pour le Développement) for providing the Madagascan samples. We also thank Farid Benyahia for assistance with the molecular analyses of these samples. D.R.C. was supported by a Natural Environment Research Council PhD studentship (471757).

AUTHOR CONTRIBUTIONS

D.R.C., G.J.C.U., A.J.D., L.D. and T.J.M. designed the research. L.M., M.M. and D.R.C. performed the research. L.M., M.M., L.D. and D.R.C. contributed reagents and analytical tools. A.J.D., T.J.M. and D.R.C. analysed the data. All authors contributed to the writing of the manuscript.

DATA ACCESSIBILITY

Sequence data from this study are deposited in the European Nucleotide Archive under the project accession number: PRJEB19885. All other data needed to reproduce this study can be found in Supporting Information Appendix S1 (Supplementary file 1).

ORCID

Dave R. Clark http://orcid.org/0000-0002-2306-0220

REFERENCES

Altschul, S. F., Gish, W., Miller, W. T., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*, 403–410.

Antón, J., Peña, A., Santos, F., Martínez-García, M., Schmitt-Kopplin, P., & Rosselló-Mora, R. (2008). Distribution, abundance and diversity of the extremely halophilic bacterium *Salinibacter ruber*. *Saline Systems*, *4*, 15.

Antunes, A., Ngugi, D. K., & Stingl, U. (2011). Microbiology of the Red Sea (and other) deep-sea anoxic brine lakes. *Environmental Microbiology Reports*, *3*, 416–433.

Baas Becking, L. G. M. (1934). *Geobiologie of inleiding tot de milieukunde*. Hague, The Netherlands: W.P. Van Stockum & Zoon.

Baati, H., Guermazi, S., Amdouni, R., Gharsallah, N., Sghir, A., & Ammar, E. (2008). Prokaryotic diversity of a Tunisian multipond solar saltern. *Extremophiles*, *12*, 505–518.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., … Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*, 455–477.

Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, *19*, 134–143.

Benlloch, S., López-López, A., Casamayor, E. O., Øvreås, L., Goddard, V., Daae, F. L., … Rodríguez-Valera, F. (2002). Prokaryotic genetic diversity throughout the salinity gradient of a coastal solar saltern. *Environmental Microbiology*, *4*, 349–360.

Bloomfield, N. J., Knerr, N., & Encinas-Viso, F. (2017). A comparison of network and clustering methods to detect biogeographical regions. *Ecography*. https://doi.org/10.1111/ecog.02596

Bolhuis, H., Te Poele, E. M., & Rodríguez-Valera, F. (2004). Isolation and cultivation of Walsby's square archaeon. *Environmental Microbiology*, *6*, 1287–1291.

Brito-Echeverría, J., López-López, A., Yarza, P., Antón, J., & Rosselló-Móra, R. (2009). Occurrence of *Halococcus* spp. in the nostrils salt glands of the seabird *Calonectris diomedea*. *Extremophiles*, *13*, 557–565.

Buckley, L. B., & Jetz, W. (2007). Environmental and historical constraints on global patterns of amphibian richness. *Proceedings of the Royal Society B: Biological Sciences*, *274*, 1167–1173.

Casamayor, E. O., Massana, R., Benlloch, S., Øvreås, L., Díez, B., Goddard, V. J., … Pedrós-Alió, C. (2002). Changes in archaeal, bacterial and eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar saltern. *Environmental Microbiology*, *4*, 338–348.

Castro-Insua, A., Gómez-Rodríguez, C., & Baselga, A. (2016). Break the pattern: Breakpoints in beta diversity of vertebrates are general across clades and suggest common historical causes. *Global Ecology and Biogeography*, *25*, 1279–1283.

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., … Tiedje, J. M. (2009). The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. *Nucleic Acids Research*, *37*, D141–D145.

Cowen, R. K., Paris, C. B., & Srinivasan, A. (2006). Scaling of connectivity in marine populations. *Science*, *311*, 522–527.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*, 2783–2792.

Dapporto, L., Ciolli, G., Dennis, R. L. H., Fox, R., & Shreeve, T. G. (2015). A new procedure for extrapolating turnover regionalization at mid-small spatial scales, tested on British butterflies. *Methods in Ecology and Evolution*, *6*, 1287–1297.

Dapporto, L., Ramazzotti, M., Fattorini, S., Vila, R., Talavera, G., & Dennis, R. L. H. (2015). *recluster: Ordination methods for the analysis of beta-diversity indices* (R package version 2.8).

Di Meglio, L., Santos, F., Gomariz, M., Almansa, C., López, C., Antón, J., & Nercessian, D. (2016). Seasonal dynamics of extremely halophilic microbial communities in three Argentinian salterns. *FEMS Microbiology Ecology*, *92*, fiw184.

Dinsdale, E. A., Edwards, R. A., Hall, D., Angly, F., Breitbart, M., Brulc, J. M., … Rohwer, F. (2008). Functional metagenomic profiling of nine biomes. *Nature*, *452*, 629–632.

Dumbrell, A. J., Ferguson, R. M. W., & Clark, D. R. (2016). Microbial community analysis by single-amplicon high-throughput next generation sequencing: Data analysis – from raw output to ecology. In T. J. McGenity, K. N. Timmis, & B. Nogales (Eds.), *Hydrocarbon and lipid microbiology protocols: Microbial quantitation, community profiling and array approaches* (pp. 155–206). Heidelberg, Germany: Springer.

Dumbrell, A. J., Nelson, M., Helgason, T., Dytham, C., & Fitter, A. H. (2010). Relative roles of niche and neutral processes in structuring a soil microbial community. *The ISME Journal*, *4*, 337–345.

Elevi Bardavid, R., & Oren, A. (2008). Dihydroxyacetone metabolism in *Salinibacter ruber* and in *Haloquadratum walsbyi*. *Extremophiles*, *12*, 125–131.

Falkowski, P. G., Fenchel, T., & Delong, E. F. (2008). The microbial engines that drive Earth's biogeochemical cycles. *Science*, *320*, 1034–1039.

Favet, J., Lapanje, A., Giongo, A., Kennedy, S., Aung, Y. Y., Cattaneo, A., … Gorbushina, A. A. (2013). Microbial hitchhikers on intercontinental dust: Catching a lift in Chad. *The ISME Journal*, *7*, 850–867.

Ficetola, G. F., Mazel, F., & Thuiller, W. (2017). Global determinants of zoogeographical boundaries. *Nature Ecology and Evolution*, *1*, 89.

Fierer, N., Leff, J. W., Adams, B. J., Nielsen, U. N., Bates, S. T., Lauber, C. L., … Caporaso, J. G. (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences USA*, *109*, 21390–21395.

Finkel, O. M., Burch, A. Y., Elad, T., Huse, S. M., Lindow, S. E., Post, A. F., & Belkin, S. (2012). Distance-decay relationships partially determine diversity patterns of phyllosphere bacteria on *Tamarix* trees across the Sonoran desert. *Applied and Environmental Microbiology*, *78*, 6187–6193.

Finlay, B. J. (2002). Global dispersal of free-living microbial eukaryote species. *Science*, *296*, 1061–1063.

Flynn, J. M., Brown, E. A., Chain, F. J. J., MacIsaac, H. J., & Cristescu, M. E. (2015). Toward accurate molecular identification of species in complex environmental samples: Testing the performance of sequence filtering and clustering methods. *Ecology and Evolution*, *5*, 2252–2266.

Glassman, S. I., Peay, K. G., Talbot, J. M., Smith, D. P., Chung, J. A., Taylor, J. W., … Bruns, T. D. (2015). A continental view of pine-associated ectomycorrhizal fungal spore banks: A quiescent functional guild with a strong biogeographic pattern. *New Phytologist*, *205*, 1619–1631.

Gramain, A., Díaz, G. C., Demergasso, C., Lowenstein, T. K., & McGenity, T. J. (2011). Archaeal diversity along a subterranean salt core from the Salar Grande (Chile). *Environmental Microbiology*, *13*, 2105–2121.

Hallsworth, J. E., Yakimov, M. M., Golyshin, P. N., Gillion, J. L. M., D'Auria, G., De Lima Alves, F., … McGenity, T. J. (2007). Limits of life in $MgCl_2$-containing environments: Chaotropicity defines the window. *Environmental Microbiology*, *9*, 801–813.

Hastie, T. J., Tibshirani, R. J., & Friedman, J. H. (2009). *Boosting and additive trees*. T. J. Hastie, R. J. Tibshirani, & J. H. Friedman (Eds.), *The elements of statistical learning*. New York, NY: Springer.

Henriet, O., Fourmentin, J., Delincé, B., & Mahillon, J. (2014). Exploring the diversity of extremely halophilic archaea in food-grade salts. *International Journal of Food Microbiology*, *191*, 36–44.

Herlemann, D. P. R., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J. J., & Andersson, A. F. (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *The ISME Journal*, *5*, 1571–1579.

Hijmans, R. J. (2016). *Spherical trigonometry* (R package version 1.5-5).

Holt, B. G., Lessard, J. P., Borregaard, M. K., Fritz, S. A., Araújo, M. B., Dimitrov, D., … Rahbek, C. (2013). An update of Wallace's zoogeographic regions of the world. *Science*, *339*, 74–78.

Jones, M. L., Ramoneda, J., Rivett, D. W., & Bell, T. (2017). Biotic resistance shapes the influence of propagule pressure on invasion success in bacterial communities. *Ecology*, *98*, 1743–1749.

Joshi, N., & Fass, J. (2011). *Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files* (Version 1.33) [Software]. Retrieved from https://github.com/najoshi/sickle

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. New York, NY: Wiley.

Kellogg, C. A., & Griffin, D. W. (2006). Aerobiology and the global transport of desert dust. *Trends in Ecology and Evolution*, *21*, 638–644.

Kreft, H., & Jetz, W. (2010). A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography*, *37*, 2029–2053.

Kreft, H., & Jetz, W. (2013). Comment on "An update of Wallace's zoogeographic regions of the world". *Science*, *341*, 343.

Lamoreux, J. F., Morrison, J. C., Ricketts, T. H., Olson, D. M., Dinerstein, E., McKnight, M. W., & Shugart, H. H. (2006). Global tests of biodiversity concordance and the importance of endemism. *Nature*, *440*, 212–214.

Lauber, C. L., Hamady, M., Knight, R., & Fierer, N. (2009). Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Applied and Environmental Microbiology*, *75*, 5111–5120.

Lear, G., Bellamy, J., Case, B. S., Lee, J. E., & Buckley, H. L. (2014). Fine-scale spatial patterns in bacterial community composition and function within freshwater ponds. *The ISME Journal*, *8*, 1715–1726.

Liaw, A., & Wiener, M. (2002). *Breiman and Cutler's random forests for classification and regression* (R package version 4.6–12).

McGenity, T. J., Gemmell, R. T., Grant, W. D., & Stan-Lotter, H. (2000). Origins of halophilic microorganisms in ancient salt deposits. *Environmental Microbiology*, *2*, 243–250.

McGenity, T. J., Hallsworth, J. E., & Timmis, K. N. (2008). Connectivity between 'ancient' and 'modern' hypersaline environments, and the salinity limits of life. In F. Briand (Ed.), *CIESM workshop monographs no. 33: The Messinian salinity crisis from mega-deposits to microbiology – A consensus report* (pp. 115–120). Monaco: CIESM Publisher.

Minegishi, H., Echigo, A., Nagaoka, S., Kamekura, M., & Usami, R. (2010). *Halarchaeum acidiphilum* gen. nov., sp. nov., a moderately acidophilic haloarchaeon isolated from commercial solar salt. *International Journal of Systematic and Evolutionary Microbiology*, *60*, 2513–2516.

Nikolenko, S. I., Korobeynikov, A. I., & Alekseyev, M. A. (2013). BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics*, *14*, S7.

Oh, D., Porter, K., Russ, B., Burns, D., & Dyall-Smith, M. (2010). Diversity of *Haloquadratum* and other haloarchaea in three, geographically distant, Australian saltern crystallizer ponds. *Extremophiles*, *14*, 161–169.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., … Wagner, H. (2015). *vegan: Community ecology package* (R package version 2.3-1).

Oren, A. (1994). The ecology of the extremely halophilic archaea. *FEMS Microbiology Reviews*, *13*, 415–439.

Pagaling, E., Wang, H., Venables, M., Wallace, A., Grant, W. D., Cowan, D. A., ... Heaphy, S. (2009). Microbial biogeography of six salt lakes in Inner Mongolia, China, and a salt lake in Argentina. *Applied and Environmental Microbiology*, *75*, 5750–5760.

Papke, R. T., Ramsing, N. B., Bateson, M. M., & Ward, D. M. (2003). Geographical isolation in hot spring cyanobacteria. *Environmental Microbiology*, *5*, 650–659.

Podell, S., Emerson, J. B., Jones, C. M., Ugalde, J. A., Welch, S., Heidelberg, K. B., ... Allen, E. E. (2014). Seasonal fluctuations in ionic concentrations drive microbial succession in a hypersaline lake community. *The ISME Journal*, *8*, 979–990.

Purdy, K. J., Cresswell-Maynard, T. D., Nedwell, D. B., McGenity, T. J., Grant, W. D., Timmis, K. N., & Embley, T. M. (2004). Isolation of haloarchaea that grow at low salinities. *Environmental Microbiology*, *6*, 591–595.

R Developement Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raskin, L., Stromley, J. M., Rittmann, B. E., & Stahl, D. A. (1994). Group-specific 16S rRNA hybridization probes to describe natural communities of methanogens. *Applied and Environmental Microbiology*, *60*, 1232–1240.

Riddle, M. R., Baxter, B. K., & Avery, B. J. (2013). Molecular identification of microorganisms associated with the brine shrimp *Artemia franciscana*. *Aquatic Biosystems*, *9*, 7.

Rodriguez-Valera, F., Ruiz-Berraquero, F., & Ramos-Cormenzana, A. (1979). Isolation of extreme halophiles from seawater. *Applied and Environmental Microbiology*, *38*, 164–165.

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, *4*, e2409v1.

Rosselló-Mora, R., Lucio, M., Peña, A., Brito-Echeverría, J., Lopez-Lopez, A., Valens-Vadell, M., ... Schmitt-Kopplin, P. (2008). Metabolic evidence for biogeographic isolation of the extremophilic bacterium *Salinibacter ruber*. *The ISME Journal*, *2*, 242–253.

Santos, A. M. C., Field, R., & Ricklefs, R. E. (2016). New directions in island biogeography. *Global Ecology and Biogeography*, *25*, 751–768.

Sclater, P. L. (1858). On the general geographic distribution of the members of the class Aves. *Zoological Journal of the Linnean Society*, *2*, 130–136.

Shimane, Y., Minegishi, H., Echigo, A., Kamekura, M., Itoh, T., Ohkuma, M., ... Hatada, Y. (2015). *Halarchaeum grantii* sp. nov., a moderately acidophilic haloarchaeon isolated from a commercial salt sample. *International Journal of Systematic and Evolutionary Microbiology*, *65*, 3830–3835.

Sonnenfeld, P. (1995). The color of rock salt—A review. *Sedimentary Geology*, *94*, 267–276.

Stahl, A. D., & Amann, R. (1991). Development and application of nucleic acid probes. In E. Stackebrandt & M. Goodfellow (Eds.), *Nucleic acid techniques in bacterial systematics* (pp. 205–248). Chichester, U.K.: Wiley.

Takhtajan, A. L. (1986). *The floristic regions of the world*. Berkeley, CA: University of California Press.

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N. S., Wijesundera, R., ... Abarenkov, K. (2014). Global diversity and geography of soil fungi. *Science*, *346*, 1256688.

Torreblanca, M., Rodriguez-Valera, F., Juez, G., Ventosa, A., Kamekura, M., & Kates, M. (1986). Classification of non-alkaliphilic halobacteria based on numerical taxonomy and polar lipid composition, and description of *Haloarcula* gen. nov. and *Haloferax* gen. nov. *Systematic and Applied Microbiology*, *8*, 89–99.

Ventosa, A., de la Haba, R. R., Sánchez-Porro, C., & Papke, R. T. (2015). Microbial diversity of hypersaline environments: A metagenomic approach. *Current Opinion in Microbiology*, *25*, 80–87.

Vilhena, D. A., & Antonelli, A. (2015). A network approach for identifying and delimiting biogeographical regions. *Nature Communications*, *6*, 6848.

Wallace, A. R. (1876). *The geographical distribution of animals*. Cambridge, U.K.: Cambridge University Press.

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, *73*, 5261–5267.

Werner, J. J., Zhou, D., Caporaso, J. G., Knight, R., & Angenent, L. T. (2012). Comparison of Illumina paired-end and single-direction sequencing for microbial 16S rRNA gene amplicon surveys. *The ISME Journal*, *6*, 1273–1276.

Whitaker, R. J., Grogan, D., & Taylor, J. W. (2003). Geographic barriers isolate endemic populations of hyperthermophilic Archaea. *Science*, *301*, 976–978.

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K. H., ... Rosselló-Móra, R. (2014). Uniting the classification of cultured and uncultured Bacteria and Archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology*, *12*, 635–645.

Yim, K. J., Kwon, J., Cha, I. T., Oh, K. S., Song, H. S., Lee, H. W., ... Roh, S. W. (2015). Occurrence of viable, red-pigmented haloarchaea in the plumage of captive flamingoes. *Scientific Reports*, *5*, 16425.

Youssef, N. H., Ashlock-Savage, K. N., & Elshahed, M. S. (2012). Phylogenetic diversities and community structure of members of the extremely halophilic Archaea (order Halobacteriales) in multiple saline sediment habitats. *Applied and Environmental Microbiology*, *78*, 1332–1344.

Yu, Z., García-González, R., Schanbacher, F. L., & Morrison, M. (2008). Evaluations of different hypervariable regions of archaeal 16S rRNA genes in profiling of methanogens by Archaea-specific PCR and denaturing gradient gel electrophoresis. *Applied and Environmental Microbiology*, *74*, 889–893.

Zak, D. R., Holmes, W. E., White, D. C., Peacock, A. D., & Tilman, D. (2003). Plant diversity, soil microbial communities, and ecosystem function: Are there any links? *Ecology*, *84*, 2042–2050.

Zhaxybayeva, O., Stepanauskas, R., Mohan, N. R., & Papke, R. T. (2013). Cell sorting analysis of geographically separated hypersaline environments. *Extremophiles*, *17*, 265–275.

BIOSKETCH

Dave Clark is a microbial ecologist interested in the processes that determine microbial distributions over macroecological scales. The balance between biotic and abiotic interactions, and dispersal limitation are of particular interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Clark DR, Mathieu M, Mourot L, et al. Biogeography at the limits of life: Do extremophilic microbial communities show biogeographical regionalization? *Global Ecol Biogeogr*. 2017;26:1435–1446. https://doi.org/10.1111/geb.12670