Lawrence Berkeley National Laboratory
LBL Publications

Title

Biogeographic patterns in populations of marine Pseudoalteromonas atlantica isolates

Permalink

https://escholarship.org/uc/item/2xd0n3jd

Authors

Kokate, Prajakta P
Bales, Erika
Joyner, Dominique
et al.

Publication Date

2023-01-17

DOI

10.1093/femsle/fnad081

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

OXFORD FEMS

FEMS Microbiology Letters, 2023, **370**, 1–9

DOI: 10.1093/femsle/fnad081

Advance access publication date: 12 August 2023

Research Letter – Environmental Microbiology and Microbial Ecology

Biogeographic patterns in populations of marine *Pseudoalteromonas atlantica* isolates

Prajakta P. Kokate[1], Erika Bales[2], Dominique Joyner[3], Terry C. Hazen[2,3,*], Stephen M. Techtmann[1,*]

[1]Department of Biological Sciences, Michigan Technological University, Houghton, MI 49931, United States
[2]Department of Microbiology, University of Tennessee Knoxville, Knoxville, TN 37996, United States
[3]Department of Civil and Environmental Engineering, University of Tennessee Knoxville, Knoxville, TN 37996, United States
*Corresponding authors. Terry Hazen University of Tennessee, Knoxville 325 John D. Tickle Engineering Building Knoxville, Tennessee 37996-2313. E-mail: tchazen@utk.edu; Stephen Techtmann Michigan Technological University 740 Dow Environmental Science and Engineering Building 1400 Townsend Drive Houghton, MI 49931. E-mail: smtechtm@mtu.edu

Editor: [Yu-Zhong Zhang]

Abstract

Intra-specific genomic diversity is well documented in microbes. The question, however, remains whether natural selection or neutral evolution is the major contributor to this diversity. We undertook this study to estimate genomic diversity in *Pseudoalteromonas atlantica* populations and whether the diversity, if present, could be attributed to environmental factors or distance effects. We isolated and sequenced twenty-three strains of *P. atlantica* from three geographically distant deep marine basins and performed comparative genomic analyses to study the genomic diversity of populations among these basins. Average nucleotide identity followed a strictly geographical pattern. In two out of three locations, the strains within the location exhibited >99.5% identity, whereas, among locations, the strains showed <98.11% identity. Phylogenetic and pan-genome analysis also reflected the biogeographical separation of the strains. Strains from the same location shared many accessory genes and clustered closely on the phylogenetic tree. Phenotypic diversity between populations was studied in ten out of twenty-three strains testing carbon and nitrogen source utilization and osmotolerance. A genetic basis for phenotypic diversity could be established in most cases but was apparently not influenced by local environmental conditions. Our study suggests that neutral evolution may have a substantial role in the biodiversity of *P. atlantica*.

Keywords: biogeography, neutral evolution, dispersal limitation, genomic diversity, intra-specific, distance effect

Introduction

Biogeographical patterns of microbial diversity are well documented (Martiny et al. 2006, Brown et al. 2012, Hanson et al. 2012, Malmstrom et al. 2013, Techtmann et al. 2016, Bay et al. 2020, Ghannam et al. 2020). However, the causes of this diversity are not clearly defined. On the one hand, there is the concept that "everything is everywhere" and the environment selects (Becking 1934). In other words, microorganisms are dispersed across the globe and can survive and propagate wherever the habitat conditions are conducive (Green et al. 2008). On the other hand, there is the concept that neutral evolution and dispersal limitation are potentially major determinants of biodiversity (Dhami et al. 2018). Although these two phenomena are not mutually exclusive, identifying the relative contribution of each phenomenon in explaining the biogeographical patterns is important to better to understand the ecology and evolution of marine microorganism. For example, where and why are certain organisms found in a particular space, how are these organisms interlinked, and how would they be affected due to deviations in the environmental conditions are pressing questions in microbial ecology (Shirani and Hellweger 2017). Previous work has investigated intra-specific genomic variation as a lens through which to understand the interplay between these two frameworks.

Intra-specific genomic diversity has been explored in *Prochlorococcus* spp. (Johnson et al. 2006, Kashtan et al. 2017). *Prochlorococcus* is one of the most abundant photosynthetic cells in the upper ocean with a high effective population size. Johnson et al. (2006) demonstrated that *Prochlorococcus* shows genomic diversity associated with changes in temperature and intensity of light. Previous work on *Prochlorococcus* compared the genomes from two basins, one from the Pacific Ocean and the other from the Atlantic Ocean. They have shown that there are distinct ecotypes in these two basins that are not randomly distributed (Kashtan et al. 2017). They showed that out of the two basins, lower diversity was observed in the basin that experienced pronounced seasonal changes. Another peculiar observation while comparing genomes from these two basins was that certain genes were ocean-specific but were unrelated phylogenetically, highly suggestive of selective forces at play to retain the nitrate acquisition genes in the Pacific Ocean and the phosphorus acquisition genes in the Atlantic Ocean (Kashtan et al. 2017). In another example, the SAR11 clade has been shown to have undergone adaptive radiation in response to temperature, reiterating the importance of environmental selection on genomic diversity (Brown et al. 2012). Interestingly, intraspecific niche differentiation was observed in highly abundant species (Sjöqvist et al. 2021).

Conversely, a specific physiological example in the cellular slime mold *Dictyostelium discoideum* was reviewed to suggest that stochastic rather than selective forces may be important factors contributing to the patterns of geographical diversity (Nanjundiah 2019). Stochastic factors include dispersal limitation and drift. Dispersal limitation, in particular, is an important influencer of microbial biodiversity and is reported extensively in soil microbiomes as well as water bodies (McClain et al. 2012, Bottos et al.

Received 19 April 2023; **revised** 14 July 2023; **accepted** 11 August 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of FEMS. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/femsle/article/doi/10.1093/femsle/fnad081/7241809 by University of Tennessee Libraries user on 26 August 2023

2018, Chen et al. 2020a, b). The oceans, in particular, represent an interesting system to study biogeography. The oceans are interconnected, and previous studies have suggested that the ocean microbial assemblage represents a persistent seed bank of microorganisms (Gibbons et al. 2013, Troussellier et al. 2017, Ward et al. 2021). Bacterial cells have a high reproductive output, and very few individuals are required to establish connectivity between two places. Unsurprisingly, Jönsson and Watson (2016) observed that distinct regions in the global surface ocean are connected on very short timescales. Conversely, another school of thought argues that the rates of passive dispersal may be inadequate to overcome the biogeographic distinction generated due to chance mutations occurring in geographically isolated regions of the ocean (Ward et al. 2021). Hellweger et al. (2014) used a modeling approach to establish that neutral evolution with dispersal limitation can produce considerable biogeographic patterns in the ocean microbe population. These contradicting reports, each contributing substantially to the understanding of microbial evolution, encouraged us to take up this study. In this study, we attempt to determine the intra-specific genomic diversity and associated biogeographic patterns among strains of *Pseudoalteromonas atlantica* from distinct locations and estimate the contribution of neutral evolution and dispersal limitation versus natural selection due to environmental factors on the biodiversity.

The *Pseudoalteromonas* genus belongs to the Gammaproteobacteria class. *Pseudoalteromonas* is a diverse genus of bacteria. One species of interest in this study, *P. atlantica*, is a motile bacterium that can produce many biologically active extracellular compounds (Nordberg et al. 2014, Chronopoulou et al. 2015). *Pseudoalteromonas atlantica* is a primary biofilm-forming bacterium, where the enzymes agarases, proteases, etc., facilitate colonizing solid surfaces and using them as substrates. They produce extracellular polysaccharides to form biofilms that allow the concentration of nutrients to support the growth of other marine microorganisms (Nordberg et al. 2014). Some *Pseudoalteromonas* are oil-degrading microbes found in abundance in the microbial consortia found in oil-polluted water bodies (Redmond and Valentine 2012, Dubinsky et al. 2013, Gutierrez et al. 2013, Harris et al. 2014, Chronopoulou et al. 2015). In addition to being relevant in oil degradation and crucially placed ecologically, this species may also be involved in controlling toxic metal concentration in the marine environment (Nordberg et al. 2014).

Here we use genomic and phenotypic tools to study populations of this ubiquitous marine bacterium. We isolated multiple strains of *P. atlantica* from three distinct deep-sea basins—The Great Australian Bight (GAB), the Eastern Atlantic in the Angola Basin (EAB), and the Western Atlantic in the Sargasso Sea (WAB). We addressed three questions: (1) Is there genomic and phenotypic diversity in strains of *P. atlantica* from geographically distant locations? (2) Do the phenotypes observed correlate with the respective genomes? And (3) If diversity is observed, which phenomena contribute substantially to the diversity?

Materials and methods

Sampling

Water samples were taken from south of Australia in the Great Australian Bight (referred to as GAB), The Sargasso Sea in the Western Atlantic basin (referred to as WAB), and off the coast of Angola from the Eastern Atlantic basin (referred to as EAB). Table 1 records the geographical location (latitude and longitude), temperature, salinity, and depth at each sample collection site using the conductivity temperature depth device.

Bacterial isolation and growth

The raw water samples were plated on ONR7a agar medium supplemented with peptone (1 g/l) and 100 ppm of local crude oil. Isolated colonies with similar morphologies were restreaked on the same medium, then transferred into liquid ONR7A supplemented with peptone (1 g/l) and 100 ppm of local crude oil and incubated at 4°C.

DNA extraction and 16S rRNA sequencing

DNA was extracted from each strain when the liquid cultures were in the log phase using the UltraClean Microbial DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA, USA). The quality of DNA was measured at the 260/280 ratio on a Nanodrop spectrophotometer (ThermoFisher Scientific). Bacterial primers 27F and 1492R were used to amplify the 16S rRNA gene. The resulting amplicon was cleaned using the Zymo Research DNA Clean & Concentrator-25 kit (Zymo Research, Irvine, CA, USA) and sequenced using an ABI 3730 sequencer at the University of Tennessee, Knoxville Walter's Life Science sequencing facility to obtain the 16S rRNA gene forward and reverse nucleotide sequences. The 16S rRNA sequences obtained were compared to the NCBI non-redundant database using BLASTn to confirm the taxonomy of the strains.

Whole-genome sequencing

Thirty-two *P. atlantica* strains that had >99.5% 16S rRNA identity to one another were further processed for whole-genome sequencing. Genome libraries were prepared using the Nextera XT Library Preparation Kit (Illumina, San Diego, CA, USA), following the manufacturer's instructions. Paired-end sequencing was done on an Illumina MiSeq platform using a MiSeq Reagent Kit V3 600 cycles. The raw sequenced reads were assessed for quality using FastQC v.0.11.5 (Nurk et al. 2013) and filtered for low-quality reads and adapter regions using Trimmomatic v.0.36 with the following parameters: SLIDINGWINDOW:4:15 MINLEN:36 (Bolger et al. 2014). The *de novo* genome assembly was made with SPAdes v.3.13 (Bankevich et al. 2012). The quality of the genome assembly was checked using CheckM v. 1.1.3 (Parks et al. 2014). Based on the quality of the assembly, nine strains that showed contamination >5% or genome completeness <90% were excluded from further analysis. Gene annotation was achieved using Prokka v.2.1.1 annotation pipeline (Seemann 2014) for the twenty-three genomes that passed the quality check. Supplemental Table S1 tabulates the metrics of these genome assemblies.

Comparative genomics

The average nucleotide identity (ANI) of the annotated DNA sequences of each draft genome was calculated using fastANI (Jain et al. 2018). The pairwise ANI was not normally distributed. A Wilcoxon Rank Sum Test was used to test the hypothesis that the ANI between isolates from the same basin was higher than those from other basins. This test was chosen because the ANI data were not normally distributed. The GenBank output files from Prokka were used to estimate the pan-genome using Roary (Page et al. 2015), which also identified core and accessory genes. Roary was further used to identify genes present exclusively at each sample collection site. The presence of genes in each of the isolates was used to generate the phylogenetic tree using the phylogenetic tree viewer (Huerta-Cepas et al. 2016). The Phandango web application

Downloaded from https://academic.oup.com/femsle/article/doi/10.1093/femsle/fnad081/7241809 by University of Tennessee Libraries user on 26 August 2023

Table 1. Environmental conditions recorded for twenty-three strains used for genome analysis.

Basin	Location ID	Sample ID	Depth (m)	Temperature°C	Salinity	Dissolved oxygen	Latitude	Longitude
GAB	**GAB-SS-09-NB**	**GABNB9D**	**1901**	**2.398**	**34.691**	**5.49**	**–35.2879**	**132.0581**
GAB	**GAB-SS-16-NS**	**GABNS16A**	**200**	**12.574**	**35.123**	**7.98**	**–34.6192**	**130.2657**
GAB	**GAB-SS-16-NS**	**GABNS16C**	**200**	**12.574**	**35.123**	**7.98**	**–34.6192**	**130.2657**
GAB	GAB-SS-16-NS	GABNS16E	200	12.574	35.123	7.98	–34.**6192**	130.2657
GAB	GAB-SS-16-NS	GABNS16G	200	12.574	35.123	7.98	–34.**6192**	130.2657
GAB	GAB-SS-16-NS	GABNS16H	200	12.574	35.123	7.98	–34.6192	130.2657
GAB	**GAB-SS-16-2/32 316**	**GAB2316C**	**1650**	**2.621**	**34.624**	**5.45**	**–34.6192**	**130.2657**
WAB	**CST-05-1200**	**CST1**	**1197.36**	**5.6213**	**35.0644**	**6.81**	**25.5009**	**–62.6082**
WAB	CST-05-1200	CST3	1197.36	5.6213	35.0644	6.81	25.5009	–62.6082
WAB	CST-05-1200	CST4	1197.36	5.6213	35.0644	6.81	25.5009	–62.6082
WAB	CST-05-1200	CST5	1197.36	5.6213	35.0644	6.81	25.5009	–62.6082
WAB	CST-05-1200	CST6	1197.36	5.6213	35.0644	6.81	25.5009	–62.6082
WAB	CST-05-1200	CST7	1197.36	5.6213	35.0644	6.81	25.5009	–62.6082
WAB	CST-05-1200	CST9	1197.36	5.6213	35.0644	6.81	25.5009	–62.6082
WAB	**CST-03-5000**	**CST2**	**5005.05**	**2.1253**	**34.8582**	**8.18**	**16.4717**	**–60.0388**
EAB	24 b 3 II iv	Angola-4	1450	3.925	34.93	4.48	–11.9099	6.8424
EAB	**24 b 3 I iii**	**Angola-7**	**1450**	**3.925**	**34.93**	**4.48**	**–11.9099**	**6.8424**
EAB	**19 b 4 III i**	**Angola-9**	**1250**	**4.2**	**34.8**	**2.89**	**–9.1274**	**6.3615**
EAB	19 b 3 III ii	Angola-18	1250	4.2	34.8	2.89	–9.1274	6.3615
EAB	19 b 3 III iv	Angola-20	1250	4.2	34.8	2.89	–9.1274	6.3615
EAB	**24 2/3 2 I ii**	**Angola-22**	**960**	**4.366**	**34.618**	**3.06**	**–11.9099**	**6.8424**
EAB	**19 2/3 2 III ii**	**Angola-30**	**850**	**4.9**	**34.6**	**2.18**	**–9.1274**	**6.3615**
EAB	19 2/3 2 III iii	Angola-31	850	4.9	34.6	2.18	–9.1274	6.3615

(Hadfield et al. 2017) generated a pan-genome tree. Phylophlan 3.0 (Asnicar et al. 2020) was used to construct a strain-level phylogenetic tree based on the alignment of four hundred universal marker genes for prokaryotes from the twenty-three isolates. DuctApe (Galardini et al. 2014) was used to map the genomic data to KEGG pathways (Kanehisa and Goto 2000). To better clarify the impact of distance and geochemistry on the phylogenetic diversity of these populations, we compared the geochemical factors and geographic distance to the phylogenetic distance observed. Geochemical data (temperature, salinity, dissolved oxygen, and depth) were center scaled and then used to construct a Euclidian distance matrix. Geographic distance was calculated from longitude and latitude values for sampling sites using geodesic distance in the distm function from the geosphere package in R (Hijmans et al. 2017). Pairwise phylogenetic distance between tips of the phylogenetic trees was extracted from the phylogenetic trees using the cophenetic.phylo command in the ape package in R (Paradis and Schliep 2019). A mantel test was used to identify significant correlations between geographic distance, geochemical factors, and phylogenetic distance. The mantel test was used as implemented in the vegan package in R (Oksanen et al. 2007). Kendall's rank correlation tau was used with 999 permutations. Phlylogentic trees were plotted using the ggtree package in R (Yu et al. 2017).

Phenotyping using metabolic profiling

The phenotypic characterization was done for ten strains (denoted in Table 1 in bold) representing each basin from different phylogenetic tree branches (Fig. 3). The Biolog phenotype microarray (PM) technology was used to identify carbon (PM01A) and nitrogen (PM03B) sources that each of these ten strains could metabolize. The PM09 panel was used to characterize salt tolerance. After growing to reach the late log phase, the bacterial cells were resuspended at a 10% inoculum concentration in a minimal medium (ONR7a) without any additional carbon source for the PM01A, which tests carbon source utilization profiles. For the determination of nitrogen source utilization profiles using PM03B plates, the cultures were grown in an ONR7a medium without nitrogen and the addition of 1 g/l of peptone as a carbon source. Standard ONR7a medium contains 0.27 g/l of NH_4Cl. For the study of nitrogen sources, this NH_4Cl was omitted from the medium. For the determination of osmotolerance, cultures were grown in an ONR7a medium with 1 g/l peptone as the carbon source. Biolog Redox Dye H (Biolog, Hayward, CA, USA) was added with the inoculum at a 1X concentration to determine the metabolic activity of the test substrates. The results of the microarray PM plates were interpreted using a two-step approach. Absorbance was measured at 600 nm for PM01A and PM03B at 0, 24, and 72 h. For the PM09 plates, absorbance was measured at 560 nm at 0 and 24 h. Additionally, the plates were visualized manually after 7 days to confirm the development of the purple color indicating that the tetrazolium salt, Dye H, was reduced to formazan. A result was considered positive if both the observation methods were in agreement.

Results

Genome relatedness

Each strain's 16S rRNA sequence was compared to the partial sequence of *P. atlantica* strain NBRC 103033 16S rRNA. Each strain showed >99.5% sequence identity (Supplemental Table S2), and these results corroborated with the ANI values (Fig. 1), a robust index to evaluate genome relatedness. The percent homology using ANI values was >97% (Fig. 1b),which is above the threshold range (95%–96%) used to demarcate species (Goris et al. 2007, Figueras et al. 2014). The all-versus-all comparison for strains collected from two locations (GAB and WAB) showed >99.5% homology among strains within the location. At the location EAB, the percent homology ranged from 97.61% (between Angola-7 and Angola-31) to 99.9992% (between Angola-9 and Angola-22). Among locations, the percent homology varied slightly. The percent homology between strains from EAB and WAB ranged from 97.6% to 98.11%, be-

Downloaded from https://academic.oup.com/femsle/article/doi/10.1093/femsle/fnad081/7241809 by University of Tennessee Libraries user on 26 August 2023

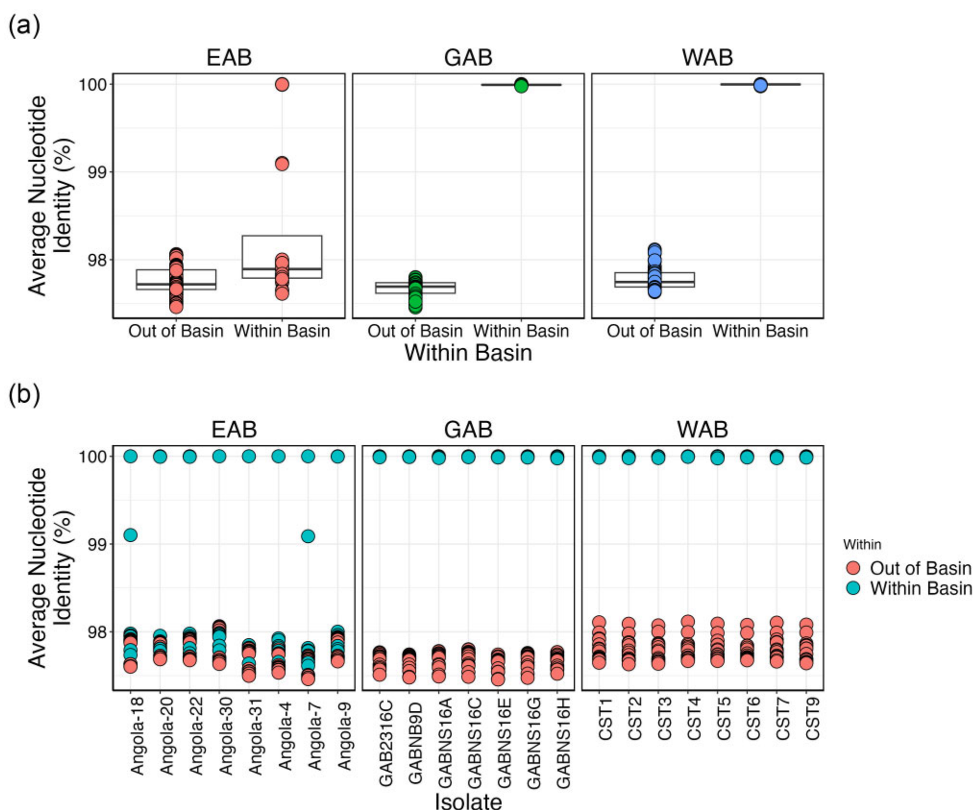

Figure 1. Percent homology using ANI values of all-versus-all comparison of genomes. (a) Comparison of ANI for genomes for isolates compared to other isolates within basin and outside of the basin of isolation. (b) ANI comparison for each isolate.

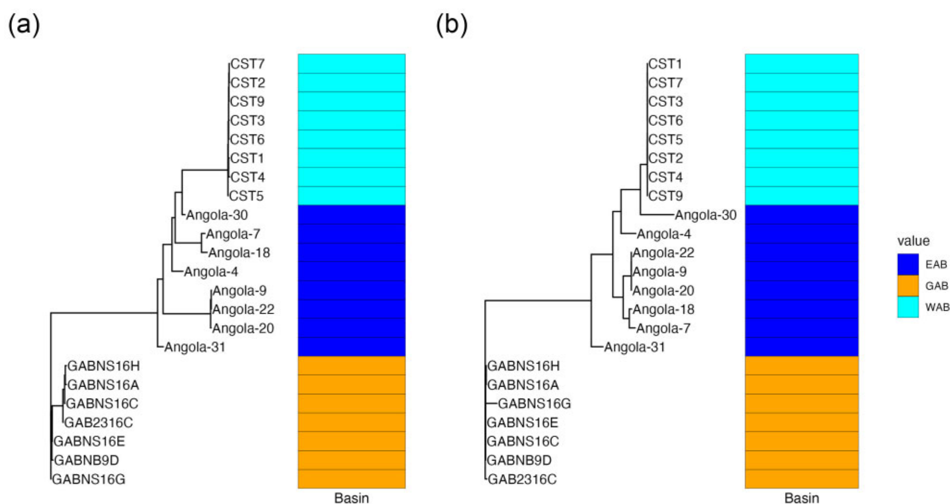

Figure 2. (a) Phylogenetic tree based on the presence and absence of genes isolated studied obtained using the Roary pipeline. Basin of isolation is shown next to the tip labels. (b) Strain-level phylogenetic tree based on four hundred universal marker genes using Phylophlan 3.0. Basin of isolation is shown next to the tip labels.

tween strains from GAB and EAB, ranged from 97.4% to 97.7%, and between strains from GAB and WAB ranged between 97.6% and 97.8%. Strains from WAB and EAB were more homologous with each other than with strains from GAB. The ANI between isolates from the same basin was significantly higher for all three basins (Fig. 1a) (Wilcoxon Rank Sum Test, GAB: *P*-value $<2.2 \times 10^{-16}$, EAB: *P*-value 6×10^{-10}, WAB: *P*-value $<2.2 \times 10^{-16}$). There was a larger range of within-basin ANI for isolates from the Eastern Atlantic Basin.

Core- and pan-genomes

A total of 7852 genes were annotated from the 23 genomes, and the Roary package segregated them into core and accessory genes. The Roary package identified 3073 core genes (present in at least 22 of the 23 genomes), 141 genes were found in at least 21 out of 23 genomes, 1807 genes were categorized as shell genes, and 2831 genes were present in less than three strains. Phylogenetic representation based on core- and pan-genome cluster strains based on basin of isolation (Fig. 2a). Notably, at least for strains

Downloaded from https://academic.oup.com/femsle/article/doi/10.1093/femsle/fnad081/7241809 by University of Tennessee Libraries user on 26 August 2023

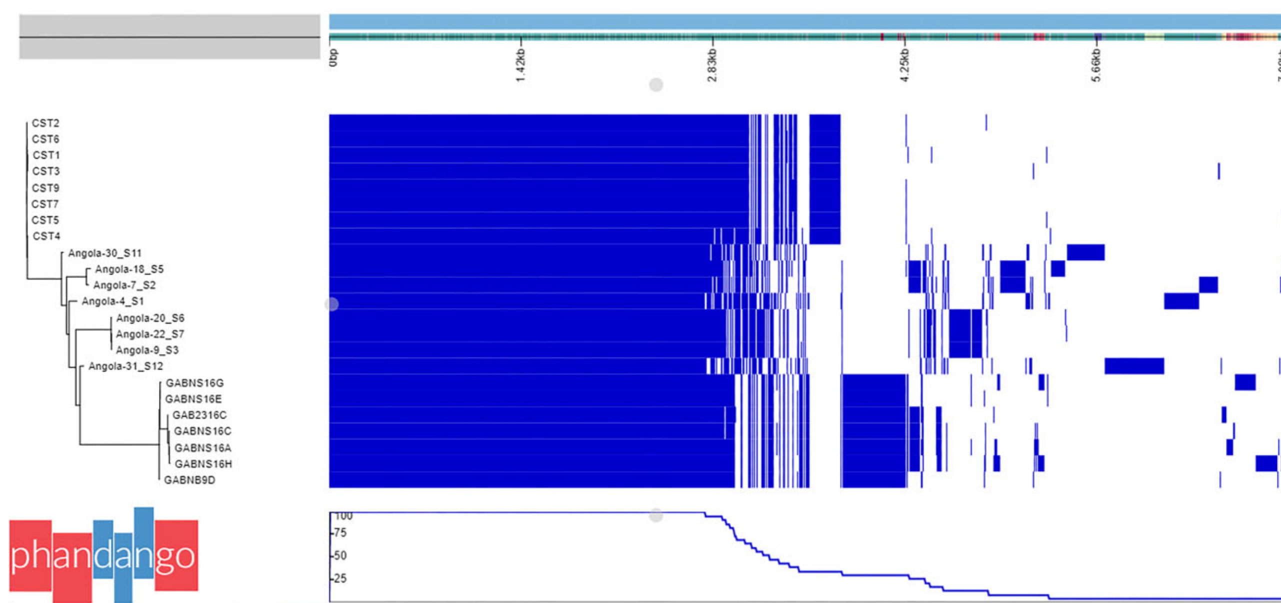

Figure 3. Use of gene presence in each strain to generate the phylogenetic tree, core- and pan-genomes.

from GAB and WAB, most accessory genes were present in all the strains collected from that location (Fig. 3). The core- and pan-genome analysis indicating distinct accessory genes in each basin suggests local horizontal gene transfer as a critical factor of diversity among strains. There was a significantly lower phylogenetic distance between strains from the same basin relative to strains from other basins (Fig. 4a) (Wilcoxon Rank Sum Test, GAB: *P*-value $<2.2 \times 10^{-16}$, EAB: *P*-value $<2.2 \times 10^{-16}$, WAB: *P*-value $<2.2 \times 10^{-16}$).

In addition to the core- and pan-genome tree, a phylogenetic tree was constructed using an alignment of conserved marker genes. The Phylophlan pipeline was used to compare the sequences of four hundred universal marker genes in each strain to estimate the evolutionary relationship among the strains (Fig. 2b). Strains from the same location clustered more closely on the phylogenetic tree. Similarly to the pan-genome tree, the marker gene tree demonstrated less phylogenetic distance between isolates from the same basins compared to isolates from other basins (Fig. 4b) (Wilcoxon Rank Sum Test, GAB: *P*-value $<2.2 \times 10^{-16}$, EAB: *P*-value 5.7×10^{-13}, WAB: *P*-value $<2.2 \times 10^{-16}$). Figure 5 is a Venn diagram indicating the number of genes present exclusively in each location and the number of genes shared between locations. The list of genes can be found in Supplemental Tables S3–S10.

A mantel test was used to correlate geographic distance with the observed phylogenetic distance from the pangenome trees as well as the tree from the conserved marker genes. This analysis demonstrated a significant correlation between geographic distance and phylogenetic distance from the conserved maker tree (Mantel statistic *r*: 0.2732, Significance: 9×10^{-4}). Geographic distance was also highly correlated to phylogenetic distance from the gene presence–absence tree (Mantel statistic *r*: 0.1661, Significance: 0.0116). A mantel test was also used to test if the phylogenetic distance was correlated with geochemical factors in the samples from which the isolates were derived. This analysis demonstrated that there was no significant correlation between geochemical factors and the phylogenetic distance from the marker gene tree (Mantel statistic *r*: 0.06571, Significance: 0.0936). Conversely, a significant correlation was observed between the geochemical conditions and the phylogenetic distance of the gene presence-absence tree (Mantel statistic *r*: 0.1669, Significance: 0.005).

Phenotype analyses

The phenotype microarray results are shown in Supplemental Tables S1, S12 and S13 for carbon source utilization, nitrogen source utilization, and osmotolerance. All ten isolates grew well at up to 10% NaCl, 4% KCl, 5% Sodium sulfate, and 20% Ethylene glycol, indicating high osmotolerance. All ten isolates also grew well when stressed with 6% NaCl combined with different osmolytes (Supplemental Table S13). All the isolates could grow using L-leucine, L-Proline, and L-Citrulline as sole nitrogen sources. Also, all isolates showed active metabolism on most dipeptides represented in PM3 (Supplemental Table S12). Many monosaccharides, disaccharides, and a trisaccharide (maltotriose) could be used as the sole source of carbon by all ten isolates. Interestingly, all the isolates could use the L-enantiomers of three amino acids (L-alanine, L-serine, and L-threonine) as a carbon source, but not their D-enantiomers.

Intraspecific phenotypic variation in carbon source utilization was evident (Supplemental Table S11). For example, Angola-9 and Angola-22 could metabolize N-acetyl-D-Glucosamine, succinic acid, fumaric acid, and malic acid, unlike the other eight isolates from the same location. Angola-30 was the only isolate that could use L-rhamnose as a carbon source. Only one isolate, GABNB9D, could metabolize α-ketobutyric acid. All isolates except GABNS16A could metabolize D-aspartic acid as a carbon source. The extent of intraspecific phenotypic variation was lesser for nitrogen source utilization (Supplemental Table S12). Three isolates, Angola-30, CST1, and CST2, could not metabolize L-Alanine, whereas Angola-30, GABNS16C, and GAB2316C were unable to use L-Tryptophan. On the other hand, Salt tolerance was fairly similar in all the isolates studied (Supplemental Table S13).

Genotype-phenotype correlation

Two strains from EAB showed a distinct carbon source utilization profile compared to the other strains. Angola-9 and Angola-22 were the only strains that could use N-Acetyl-D-Glucosamine.

Downloaded from https://academic.oup.com/femsle/article/doi/10.1093/femsle/fnad081/7241809 by University of Tennessee Libraries user on 26 August 2023

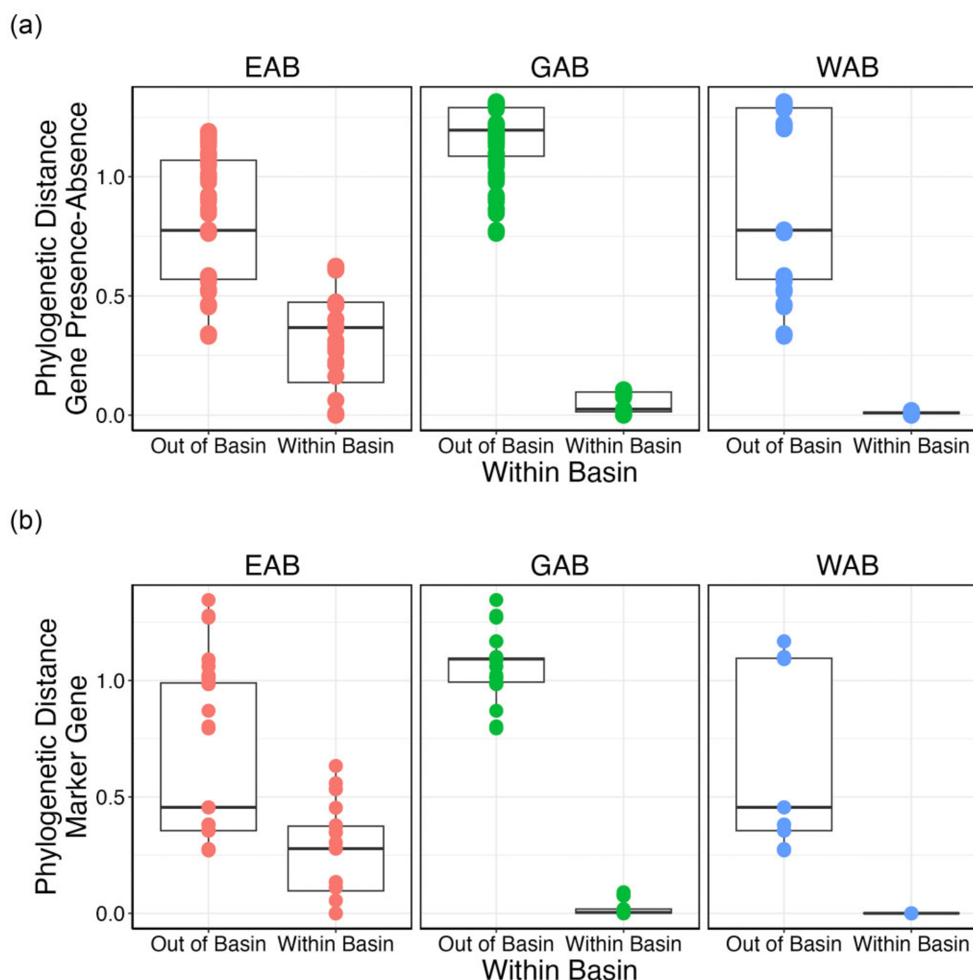

Figure 4. Comparison of phylogenetic distance to isolates within the same basin and to isolates derived from other basins. (a) Phylogenetic distances from the gene presence absence tree from the Roary pipeline. (b) Phylogenetic distances from the marker gene tree. Colors of points represent the basin.

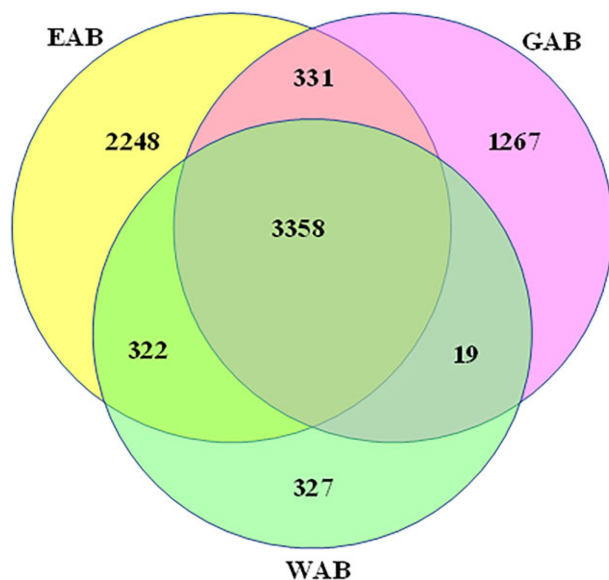

Figure 5. Venn diagram showing genes present in all three locations, the accessory genes present exclusively in strains from each location, and genes shared by strains from two different locations.

When the genomes of these two strains were compared with the other eight strains using Roary, one gene, *wbpA*, encodes UDP-N-acetylglucosamine-6-dehydrogenase (an enzyme involved in the metabolism of aminosugars, including N-Acetyl-D-Glucosamine), which was found only in the genomes of Angola-9 and Angola-22.

In another instance, L-rhamnose was used as a carbon source by only one strain, Angola-30. Again, when the genome of Angola-30 was compared with the other nine strains, Angola-30 was shown to possess a gene *rhaM* that codes for L-rhamnose mutarotase (involved in L-rhamnose metabolism). This gene was not found in any of the other nine strains.

Phenotypic heterogeneity was observed in additional instances. For example, Angola-9 and Angola-22 could use succinic, malic, and fumaric acids as carbon sources. However, the *dctA* gene responsible for transporting these sugars into the bacterial cell was present in all ten strains. The genomes were compared using The KEGG IDs (Supplemental Table S14) obtained as an output using the DucTape package, and K00571, which represents adenine-specific DNA-methyltransferase, was present only in Angola-9 and Angola-22. The authors note that further studies would be required to identify whether this gene was responsible for the differential phenotype in these two strains.

Downloaded from https://academic.oup.com/femsle/article/doi/10.1093/femsle/fnad081/7241809 by University of Tennessee Libraries user on 26 August 2023

Discussion

In the decades since the Bass-Becking hypothesis (Becking 1934) was first published, a multitude of scientists have contributed either to supporting or refuting the hypothesis. A myriad of reports support that natural selection contributes to biogeographic patterns (Kivlin et al. 2014, Techtmann et al. 2016, Chen et al. 2021). On the other hand, a large number of reports suggest that neutral evolution and dispersal limitation potentially play a substantial role in microbial diversity (Hellweger et al. 2014, Shirani and Hellweger 2017). Martiny et al. (2006) argued that by the time a microorganism has traveled a long distance, it may have replicated multiple times and diverged considerably from the source population.

In this study, we have three salient findings: (1) We observed biogeographic patterns in populations of the marine bacterium *P. atlantica*, (2) We found associations between the genotype and the phenotype, and (3) We observed that neutral evolution and dispersal limitation are a major contributor of the biogeographic patterning.

Twenty-three genomes of *P. atlantica* showed a biogeographical pattern when compared for genomic relatedness at the nucleotide level. Generally, strains from the different locations showed a lower percent identity compared to strains from the same basin. A previous study on *Streptomyces griseus* has demonstrated a similar finding where two recently diverged but genetically distinct species show a biogeographical pattern. This pattern was reflected in their pairwise ANI values, which ranged between 92.6% and 93.3% between species from different geographical locations (Choudoir and Buckley 2018). Interestingly, compared to *S. griseus* collected from soil samples that have diverged into separate lineages, the *P. atlantica* strains from the oceans appear more related to each other.

Similar to the ANI results, accessory genes also showed biogeographical patterns. Strains from the same location had many accessory genes in common. A set of genes, albeit smaller in number, is shared between strains from different locations. These accessory genes shared within and among locations suggest that the strains have diverged from their source population and that a limited dispersal between locations can be potentially observed. This finding supports the idea that populations of *P. atlantica* in different oceanic basins have acquired distinct genetic content in each basin contributing to the geographic pattern observed. The evolutionary distance was also in concordance with geographical distance. Furthermore, there was not a significant correlation between phylogenetic distance of a tree constructed from conserved marker genes and environmental conditions, providing additional evidence that dispersal limitation may be a crucial contributor to biodiversity among these strains. The influence of dispersal limitation in biogeographical patterns has been reported previously (Chust et al. 2016, Bottos et al. 2018, Chen et al. 2020a, b, Sun et al. 2021). Chen et al. (2020b) studied anammox bacterial communities across the Yangtze river to demonstrate how dispersal limitation drives biogeographical patterns. (Sun et al. 2021) demonstrated that dispersal limitation accounted for 23.4% of the driving force affecting plastisphere bacterial communities. Techtmann et al. (2017) compared Thaumarchaeotal populations in four deep-sea basins and observed substantial diversity in geochemically similar but geographically distant basins, suggesting that dispersal limitation may contribute considerably to biogeography.

Indeed, Shirani and Hellweger (2017) created a mechanistic model to simulate many lake systems, e.g. Great Lakes, Klamath River, Yahara River, and Chattahoochee River, to demonstrate biogeography in the cyanobacterium *Microcystis aeruginosa*, and showed that strongly connected lakes show lower diversity. They concluded that neutral evolution and dispersal limitation could be essential factors affecting microbial biogeography. These studies agree with our study, where the percent identity based on ANI values reflected the geographical distance between the three locations of our study.

Although we did not find any genes directly associated with an adaptation to specific environmental conditions, we note that such an association can be complicated as multiple environmental factors could act on the different gene(s) or at the genome level. There is no conclusive method to claim that the distance effect observed is not confounded by an unforeseen, undetermined, unmeasured environmental factor.

While we could not identify accessory genes specific to specific environmental factors; we could associate certain genes to specific phenotypes. Three out of the four strains from EAB showed differential phenotypic activity. Angola-9 and Angola-22 were genetically equipped to metabolize N-acetyl glucosamine. Further, only these two strains possessed the adenine-specific DNA-methyltransferase, the phenotypic consequences of which require further studies. Interestingly, however, these strains have been collected from different sites within the same location. The depth of the collection site for Angola-9 and Angola-22 was 1250 and 960 m, respectively, whereas the temperature and salinity are comparable. We cannot determine whether any common environmental factors acted as selective forces on these two genomes.

Another strain, Angola-30, carried the gene *rhaM* required for L-rhamnose metabolism. Neither of the other nine strains could metabolize L-rhamnose or carry this gene. Although there is no evidence to eliminate the possibility that a selective force caused the retention of this accessory gene, it appears unlikely that selective pressure due to any environmental factor, biotic or abiotic, may be causing Angola-30 to maintain the *rhaM* gene. In this case, the most parsimonious explanation would be neutral evolution unless the collection site of Angola-30 has excessive L-rhamnose and reduced availability of other carbon sources.

Two strains from GAB, GABNS16A and GABNS16C, were isolated from the same collection site, indicating identical environmental factors. Nonetheless, four KEGG IDs could be mapped to GABNS16A but not to GABNS16C. These pathways represent dioxin degradation, type IV pilus assembly protein (PilW), FMN-dependent NADH-azoreductase, and transposase. Strains with identical environmental conditions, showing differences in the genome, indirectly suggest that natural selection may not contribute extensively to biodiversity.

In this study, we have demonstrated genomic and phenotypic biogeographical patterns. We also argue that neutral evolution with dispersal limitation may play a substantial role in the biodiversity of marine organisms. Understanding dispersal and dispersal limitation is critical from an ecological perspective. Primarily, this knowledge could help mitigate the effects of climate change on biodiversity (Driscoll et al. 2014). Dispersal estimates are factored in stochastic models that are used to understand population trajectories through time and also predict the risk of extinction of species (Driscoll et al. 2014). This work adds to the body of literature establishing how dispersal limitation generates biogeographic patterns in populations of marine microbes throughout the oceans.

Downloaded from https://academic.oup.com/femsle/article/doi/10.1093/femsle/fnad081/7241809 by University of Tennessee Libraries user on 26 August 2023

Author contributions

P.P.K. analyzed genomic and phenotypic data and wrote the manuscript. E.B. isolated the strains, assisted in the genome sequencing, and performed the phenotypic analysis, D.J. provided oversight of the phenotypic data collection and T.C.H. oversaw the work and provided direction for the experimental setup. S.M.T. oversaw the work, directed the genomic and phenotypic data collection, and participated in writing the manuscript.

Acknowledgments

We would like to acknowledge Shane Hagen, who assisted in isolation and performed the genome sequencing. We thank BP International and its affiliates (BP Exploration & Production, Inc., BP Egypt, BP Azerbaijan, and BP Angola) for enabling this work. This work was supported by BP America Inc. We gratefully acknowledge the captains and crews of all of the research vessels and express our appreciation for ship time and assistance in the collection of samples.

Supplementary data

Supplementary data is available at *FEMSLE Journal* online.

Conflict of interest: None declared.

Funding

This research was supported by contract A13-0119-001 Deep Sea Basin Microbiology between the University of Tennessee and BP America.

Data availability

Genomic data from these isolates have been deposited in the SRA under bioproject number PRJNA929979. The supplemental dataset is available at https://figshare.com/articles/dataset/Supplemental_Core_and_Pan_Genome_Data/23688894

References

Asnicar F, Thomas AM, Beghini F *et al.* Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nat Commun* 2020;**11**: 2500. https://doi.org/10.1038/s41467-020-16366-7.

Bankevich A, Nurk S, Antipov D *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;**19**:455–77.

Bay SK, McGeoch MA, Gillor O *et al.* Soil bacterial communities exhibit strong biogeographic patterns at fine taxonomic resolution. *Msystems* 2020;**5**:e00540–20.

Becking L. *Geobiologie of Inleiding Tot De Milieukunde*. The Hague: WP Van Stockum & Zoon, 1934.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;**30**:2114–20.

Bottos EM, Kennedy DW, Romero EB *et al.* Dispersal limitation and thermodynamic constraints govern spatial structure of permafrost microbial communities. *FEMS Microbiol Ecol* 2018;**94**:1–14.

Brown MV, Lauro FM, DeMaere MZ *et al.* Global biogeography of SAR11 marine bacteria. *Mol Syst Biol* 2012;**8**:1–13.

Chen L, Liu S, Chen Q *et al.* Dispersal limitation drives biogeographical patterns of anammox bacterial communities across the Yangtze River. *Appl Microbiol Biotechnol* 2020a;**104**:5535–46.

Chen M-Y, Teng W-K, Zhao L *et al.* Comparative genomics reveals insights into cyanobacterial evolution and habitat adaptation. *ISME J* 2021;**15**:211–27.

Chen W, Jiao S, Li Q *et al.* Dispersal limitation relative to environmental filtering governs the vertical small-scale assembly of soil microbiomes during restoration. *J Appl Ecol* 2020b;**57**:402–12.

Choudoir MJ, Buckley DH. Phylogenetic conservatism of thermal traits explains dispersal limitation and genomic differentiation of *Streptomyces sister*-taxa. *ISME J* 2018;**12**:2176–86.

Chronopoulou PM, Sanni GO, Silas-Olu DI *et al.* Generalist hydrocarbon-degrading bacterial communities in the oil-polluted water column of the North Sea. *Microb Biotechnol* 2015;**8**:434–47.

Chust G, Villarino E, Chenuil A *et al.* Dispersal similarly shapes both population genetics and community patterns in the marine realm. *Sci Rep* 2016;**6**:28730.

Dhami MK, Hartwig T, Letten AD *et al.* Genomic diversity of a nectar yeast clusters into metabolically, but not geographically, distinct lineages. *Mol Ecol* 2018;**27**:2067–76.

Driscoll DA, Banks SC, Barton PS *et al.* The trajectory of dispersal research in conservation biology. Systematic review. *PloS one* 2014;**9**:e95053–e.

Dubinsky EA, Conrad ME, Chakraborty R *et al.* Succession of hydrocarbon-degrading bacteria in the aftermath of the deepwater horizon oil spill in the gulf of Mexico. *Environ Sci Technol* 2013;**47**:10860–7.

Figueras MJ, Beaz-Hidalgo R, Hossain MJ *et al.* Taxonomic affiliation of new genomes should be verified using average nucleotide identity and multilocus phylogenetic analysis. *Genome Announc* 2014;**2**:6–7.

Galardini M, Mengoni A, Biondi EG *et al.* DuctApe: a suite for the analysis and correlation of genomic and OmniLog™ Phenotype Microarray data. *Genomics* 2014;**103**:1–10.

Ghannam RB, Schaerer LG, Butler TM *et al.* Biogeographic patterns in members of globally distributed and dominant taxa found in port microbial communities. *mSphere* 2020;**5**:e00481–19.

Gibbons SM, Caporaso JG, Pirrung M *et al.* Evidence for a persistent microbial seed bank throughout the global ocean. *Proc Natl Acad Sci USA* 2013;**110**:4651–5.

Goris J, Konstantinidis KT, Klappenbach JA *et al.* DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 2007;**57**:81–91.

Green JL, Bohannan BJM, Whitaker RJ. Microbial biogeography: from taxonomy to traits. *Science* 2008;**320**:1039–43.

Gutierrez T, Singleton DR, Berry D *et al.* Hydrocarbon-degrading bacteria enriched by the Deepwater Horizon oil spill identified by cultivation and DNA-SIP. *ISME J* 2013;**7**:2091–104.

Hadfield J, Croucher NJ, Goater RJ *et al.* Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics* 2017;**34**:292–3.

Hanson CA, Fuhrman JA, Horner-Devine MC *et al.* Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Micro* 2012;**10**:497–506.

Harris AP, Techtmann SM, Stelling SC *et al.* Draft genome sequence of *Pseudoalteromonas* sp. strain ND6B, an oil-degrading isolate from Eastern Mediterranean Sea water collected at a depth of 1,210 meters. *Genome Announc* 2014;**2**:e01212–14.

Hellweger FL, van Sebille E, Fredrick ND. Biogeographic patterns in ocean microbes emerge in a neutral agent-based model. *Science* 2014;**345**:1346–9.

Hijmans RJ, Williams E, Vennes C *et al.* Package 'geosphere'. *Spherical Trigonometry* 2017;**1**:1–45.

Downloaded from https://academic.oup.com/femsle/article/doi/10.1093/femsle/fnad081/7241809 by University of Tennessee Libraries user on 26 August 2023

Huerta-Cepas J, Serra F, Bork P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol* 2016;**33**:1635–8.

Jain C, Rodriguez RL, Phillippy AM *et al.* High throughput ANI analysis of 90 K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018;**9**:5114.

Johnson ZI, Zinser ER, Coe A *et al.* Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* 2006;**311**:1737–40.

Jönsson BF, Watson JR. The timescales of global surface-ocean connectivity. *Nat Commun* 2016;**7**:11239.

Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;**28**:27–30.

Kashtan N, Roggensack SE, Berta-Thompson JW *et al.* Fundamental differences in diversity and genomic population structure between Atlantic and Pacific Prochlorococcus. *ISME J* 2017;**11**:1997–2011.

Kivlin SN, Winston GC, Goulden ML *et al.* Environmental filtering affects soil fungal community composition more than dispersal limitation at regional scales. *Fungal Ecol* 2014;**12**:14–25.

Malmstrom RR, Rodrigue S, Huang KH *et al.* Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic analysis. *ISME J* 2013;**7**:184–98.

Martiny JBH, Bohannan BJM, Brown JH *et al.* Microbial biogeography: putting microorganisms on the map. *Nat Rev Micro* 2006;**4**:102–12.

McClain CR, Stegen JC, Hurlbert AH. Dispersal, environmental niches and oceanic-scale turnover in deep-sea bivalves. *Proc R Soc B* 2012;**279**:1993–2002.

Nanjundiah V. Many roads lead to Rome: neutral phenotypes in microorganisms. *J Exp Zool B Mol Dev Evol* 2019;**332**:339–48.

Nordberg HCM, Dusheyko S, Hua S *et al.* The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucl Acids Res* 2014;**42**:D26–31.

Nurk S, Bankevich A, Antipov D *et al.* Assembling genomes and mini-metagenomes from highly chimeric reads, *Research in Computational Molecular Biology*. Berlin, Heidelberg: Springer, 2013, 158–70.

Oksanen J, Kindt R, Legendre P *et al.* The vegan package. *Community ecology package*, 2007;**10**:719.

Page AJ, Cummins CA, Hunt M *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;**31**:3691–3.

Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 2019;**35**:526–8.

Parks DHIM, Skennerton CT, Hugenholtz P *et al.* Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2014;**25**:1043–55.

Redmond MC, Valentine DL. Natural gas and temperature structured a microbial community response to the Deepwater Horizon oil spill. *Proc Natl Acad Sci USA* 2012;**109**:20292–7.

Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;**30**:2068–9.

Shirani S, Hellweger FL. Neutral evolution and dispersal limitation produce biogeographic patterns in *Microcystis aeruginosa* populations of Lake systems. *Microb Ecol* 2017;**74**:416–26.

Sjöqvist C, Delgado LF, Alneberg J *et al.* Ecologically coherent population structure of uncultivated bacterioplankton. *ISME J* 2021;**15**:3034–49.

Sun Y, Zhang M, Duan C *et al.* Contribution of stochastic processes to the microbial community assembly on field-collected microplastics. *Environ Microbiol* 2021;**23**:6707–20.

Techtmann SM, Fitzgerald KS, Stelling SC *et al. Colwellia psychrerythraea* strains from distant deep sea basins show adaptation to local conditions. *Front Environ Sci* 2016;**4**:1–10.

Techtmann SM, Mahmoudi N, Whitt KT *et al.* Comparison of Thaumarchaeotal populations from four deep sea basins. *FEMS Microbiol Ecol* 2017;**93**:fix128.

Troussellier M, Escalas A, Bouvier T *et al.* Sustaining rare marine microorganisms: macroorganisms as repositories and dispersal agents of microbial diversity. *Front Microbiol* 2017;**8**:947.

Ward BA, Cael BB, Collins S *et al.* Selective constraints on global plankton dispersal. *Proc Natl Acad Sci USA* 2021;**118**:e2007388118.

Yu G, Smith DK, Zhu H *et al.* ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 2017;**8**:28–36.

Received 19 April 2023; **revised** 14 July 2023; **accepted** 11 August 2023
© The Author(s) 2023. Published by Oxford University Press on behalf of FEMS. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/femsle/article/doi/10.1093/femsle/fnad081/7241809 by University of Tennessee Libraries user on 26 August 2023