Biogeographic patterns in populations of marine oil-degrading *Pseudoalteromonas atlantica*.

Prajakta P. Kokate[1], Erika Youngquist[2], Shane Hagen[2], Dominque Joyner[3], Terry C. Hazen[2,3*], Stephen M. Techtmann[1*]

[1]Department of Biological Sciences, Michigan Technological University, Houghton MI.
[2]Department of Microbiology, University of Tennessee Knoxville, Knoxville, TN
[3]Department of Civil and Environmental Engineering, University of Tennessee Knoxville, Knoxville TN

Co-Corresponding Authors

Terry C. Hazen - tchazen@utk.edu
Stephen M. Techtmann – smtechtm@mtu.edu

Running title: Biogeography of *Pseudoalteromonas atlantica*

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 effect. We isolated twenty-three strains of *P. atlantica* from three geographically distant deep marine basins. Whole-genome sequencing and comparative genomic analysis were performed to study the genomic diversity of populations within these three basins. The 16S rRNA gene showed >99.5% identity. However, average nucleotide identity (ANI) 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. Further genome analysis and phylogenetic tree construction also reflected the biogeographical pattern. Strains from the same geographical location had a shorter evolutionary distance. This concordance was more conspicuous in the phylogenetic tree that reflected horizontal gene transfer showing an unambiguous relation between phylogenetic distance and geographical distance, where strains from the same location were closely placed on the phylogenetic tree. Strains from the same location shared many accessory genes. Phenotypic diversity was studied in ten out of twenty-three strains using phenotype microarrays testing carbon source utilization, nitrogen source utilization, and osmotolerance. A genetic basis for phenotypic diversity could be established in most cases but did not appear to be influenced by local environmental conditions. Our study suggests that neutral evolution may have a substantial role in the biodiversity of *P. atlantica*. This work contributes to the body of literature that shows that biogeographic patterns exist in populations of marine microbes throughout the oceans.

Introduction

Biogeographical patterns of microbial diversity are well documented (Bay et al., 2020; Brown et al., 2012; Ghannam, Schaerer, Butler, & Techtmann, 2020; Hanson, Fuhrman, Horner-Devine, & Martiny, 2012; Malmstrom et al., 2013; Martiny et al., 2006; Techtmann et al., 2016). 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, Bohannan, & Whitaker, 2008). On the other hand, is the concept that neutral evolution and dispersal limitation are potentially major determinants of biodiversity (Dhami, Hartwig, Letten, Banf, & Fukami, 2018). Although these two phenomena are not mutually exclusive, identifying the relative contribution of each phenomenon in explaining the biogeographical patterns is vital from the ecological perspective. 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 & 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 (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, Delgado, Alneberg, & Andersson, 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 (Bottos et al., 2018; L. Chen et al., 2020; W. Chen, Jiao, Li, & Du, 2020; McClain, Stegen, & Hurlbert, 2012). 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; Trousselier, Escalas, Bouvier, & Mouillot, 2017; Ward, Cael, Collins, & Young, 2021). Bacterial cells have a high reproductive output, and very few individuals are required to establish connectivity between two places. Unsurprisingly, (Jönsson & 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, van Sebille, & Fredrick, 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 present 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. Some *Pseudoalteromonas* are oil-degrading microbes found in abundance in the microbial consortia found in oil-polluted water bodies (Chronopoulou et al., 2015; Dubinsky et al., 2013; Gutierrez et al., 2013; Harris et al., 2014; Redmond & Valentine, 2012). One species of interest in this study, *Pseudoalteromonas atlantica,* is a motile bacterium that can produce many biologically active extracellular compounds (Chronopoulou et al., 2015; Nordberg H, 2014). *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 H, 2014). 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 H, 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, the Angola Basin, and the Western Atlantic in the Sargasso Sea. 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

The 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 Marine Broth 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). The quality of DNA was measured at the 260/280 ratio on a Nanodrop spectrophotometer (Nanodrop reference). 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 *Pseudoalteromonas atlantica* strains that had greater than 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), 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, Lohse, & Usadel, 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 DH, 2014.). Based on the quality of the assembly, nine strains that showed contamination greater than five percent or genome completeness less than 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, Rodriguez, Phillippy, Konstantinidis, & Aluru, 2018). 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, Serra, & Bork, 2016). The Phandango web application (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 & Goto, 2000).

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 (**Figure 4**). The Biolog Phenotype Microarray (PM) technology was used to identify carbon (PM01A) and nitrogen (PM03B) sources that each of these ten strains could metabolize. 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, the biology PM09 plates were used. Cultures were grown for osmotolerance studies in an ONR7a medium with one g/L peptone as the carbon source. In the Biolog plates, 1X dye H (Biolog, Hayward, CA) was used

to determine metabolic activity. The results of the microarray PM plates were interpreted using a two-step approach. The absorbance was measured at 600 nm for PM01A at 0 hours, 24 hours, and 72 hours. The absorbance for PM03B was measured at 0 hours and 24 hours at 600 nm. For the PM09 plates, absorbance was measured at 560 nm at 0 and 24s hours. Additionally, the plates were visualized manually after seven days to confirm the development of the purple color indicating that the tetrazolium salt, dye H, was reduced to its formazan. A result was considered positive if both the observation methods were in agreement.

Results

Genome relatedness

Each strain's 16S ribosomal RNA sequence was compared to the partial sequence of Pseudoalteromonas atlantica strain NBRC 103033 16S ribosomal RNA. Each strain showed > 99.5% sequence identity (Table 2), and these results corroborated with the ANI values, a robust index to evaluate genome relatedness. The percent homology using ANI values was > 97% (Figure 1), above the threshold range (95-96%) used to demarcate species (Figueras, Beaz-Hidalgo, Hossain, & Liles, 2014; Goris et al., 2007). 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 97.6-98.11%, between strains from GAB and EAB ranged from 97.4-97.7%, and between strains from

GAB and WAB ranged between 97.6-97.8%. Strains from WAB and EAB were more homologous with each other than with strains from GAB.

Core and pan genomes

A total of 7841 genes were annotated from the 23 genomes, and the Roary package segregated them into core and accessory genes (Figure 2). The Roary package identified 3074 core genes (present in at least 22 of the 23 genomes), 149 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. Figure 3 provides a phylogenetic representation of the strains and groups the genes as core genome and pan-genome in each strain. Notably, at least for strains from GAB and WAB, most accessory genes were present in all the strains collected from that location. The phylogenetic tree was generated using two different methods. 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, represented in the phylogenetic tree in Figure 4. In contrast, the gene-presence-absence output file from Roary was used to generate another phylogenetic tree (Figure 5), which represents local horizontal gene transfer as a critical factor of diversity among strains. Strains from the same location clustered closer on the phylogenetic tree, directly correlating geographical distance and phylogenetic distance. Figure 6 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 S2-S9.

Phenotype analyses

The phenotype microarray results are shown in supplemental Tables S10, S11, and S12 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 S12). 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 S11). 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 S10). For example, Angola- and Angola-22 could metabolize N-acetyl-D-Glucosamine, succinic acid, fumaric acid, and malic acid, unlike the other eight isolates. 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 S11). 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 S12).

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. 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) 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 acid 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 S13) 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.

Discussion

Decades since the (Becking, 1934)Bass-Becking hypothesis (1934) was first published, a multitude of scientists have contributed either supporting or refuting the hypothesis. A myriad of reports supports that natural selection contributes to biogeographic patterns (M.-Y. Chen et al., 2021; Kivlin, Winston, Goulden, & Treseder, 2014; Techtmann et al., 2016). However, 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 & 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 the oil-degrading marine microbe *P. atlantica*, 2) We found an association between the genotype and the phenotype in some instances, and 3) We observed that neutral evolution and dispersal limitation might be a major contributor of the biogeographic patterning.

The 16S ribosomal RNA sequencing confirms that all strains in this study were from the same species; ANI analysis demonstrates a biogeographical pattern. Strains from the same location showed higher percent homology in most cases. 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 92.6-93.3% between species from different geographical locations (Choudoir & Buckley, 2018).

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. The evolutionary distance was also in concordance with geographical distance, 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 (Bottos et al., 2018; L. Chen et al., 2020; W. Chen et al., 2020; Chust et al., 2016; Sun et al., 2021). (L. Chen et al., 2020)

289 studied anammox bacterial communities across the Yangtze river to demonstrate how dispersal
290 limitation drives biogeographical patterns. (Sun et al., 2021) demonstrated that dispersal limitation
291 accounted for 23.4% of the driving force affecting plastisphere bacterial communities. (Techtman
292 et al., 2017) compared Thaumarchaeotal populations in four deep-sea basins and observed
293 substantial diversity in geochemically similar but geographically distant basins, suggesting that
294 dispersal limitation may contribute considerably to biogeography.

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

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

307 We could not identify accessory genes specific to specific environmental factors; we could
308 associate certain genes to specific phenotypes. Three out of the four strains from EAB showed
309 differential phenotypic activity. Angola-9 and Angola-22 were genetically equipped to metabolize
310 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 m 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 carried 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 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 are critical from an ecological perspective. Primarily, this knowledge will help mitigate the effects of climate change

on biodiversity (Driscoll 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 2014). This work adds to the body of literature establishing how dispersal limitation generates biogeographic patterns in populations of marine microbes throughout the oceans.

Data Availability

Genomic data from these isolates has been deposited at in the SRA under bioproject number XXXXXXXX and the assembled genomes have been deposited in GenBank under accession numbers XXXXXXX – XXXXXXXXX

Acknowledgements

Author Contributions

PPK analyzed genomic and phenotypic data and wrote the manuscript. EY isolated the strains, assisted in the genome sequencing, and performed the phenotypic analysis. SH assisted in isolation and performed the genome sequencing, DCJ provided oversight of the phenotypic data collection, TCH oversaw the work and provided direction for experimental set up. SMT oversaw the work, directed the genomic and phenotypic data collection and participated in writing the manuscript.

References

Asnicar, F., Thomas, A. M., Beghini, F., Mengoni, C., Manara, S., Manghi, P., . . . Segata, N. (2020). Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nat Commun, 11*(1), 2500. doi:10.1038/s41467-020-16366-7

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 : a journal of computational molecular cell biology, 19*(5), 455-477. doi:10.1089/cmb.2012.0021

Bay, S. K., McGeoch, M. A., Gillor, O., Wieler, N., Palmer, D. J., Baker, D. J., . . . Greening, C. (2020). Soil Bacterial Communities Exhibit Strong Biogeographic Patterns at Fine Taxonomic Resolution. *mSystems, 5*(4). doi:10.1128/mSystems.00540-20

Becking, L. G. M. B. (1934). *Geobiologie of inleiding tot de milieukunde*: WP Van Stockum & Zoon.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. doi:10.1093/bioinformatics/btu170

Bottos, E. M., Kennedy, D. W., Romero, E. B., Fansler, S. J., Brown, J. M., Bramer, L. M., . . . Stegen, J. C. (2018). Dispersal limitation and thermodynamic constraints govern spatial

structure of permafrost microbial communities. *FEMS Microbiol Ecol, 94*(8). doi:10.1093/femsec/fiy110

Brown, M. V., Lauro, F. M., DeMaere, M. Z., Muir, L., Wilkins, D., Thomas, T., . . . Cavicchioli, R. (2012). Global biogeography of SAR11 marine bacteria. *Molecular systems biology, 8*, 595-595. doi:10.1038/msb.2012.28

Chen, L., Liu, S., Chen, Q., Zhu, G., Wu, X., Wang, J., . . . Ni, J. (2020). Dispersal limitation drives biogeographical patterns of anammox bacterial communities across the Yangtze River. *Applied Microbiology and Biotechnology, 104*(12), 5535-5546. doi:10.1007/s00253-020-10511-4

Chen, M.-Y., Teng, W.-K., Zhao, L., Hu, C.-X., Zhou, Y.-K., Han, B.-P., . . . Shu, W.-S. (2021). Comparative genomics reveals insights into cyanobacterial evolution and habitat adaptation. *The ISME journal, 15*(1), 211-227. doi:10.1038/s41396-020-00775-z

Chen, W., Jiao, S., Li, Q., & Du, N. (2020). Dispersal limitation relative to environmental filtering governs the vertical small-scale assembly of soil microbiomes during restoration. *Journal of Applied Ecology, 57*(2), 402-412. doi:https://doi.org/10.1111/1365-2664.13533

Choudoir, M. J., & Buckley, D. H. (2018). Phylogenetic conservatism of thermal traits explains dispersal limitation and genomic differentiation of Streptomyces sister-taxa. *The ISME journal, 12*(9), 2176-2186. doi:10.1038/s41396-018-0180-3

Chronopoulou, P. M., Sanni, G. O., Silas-Olu, D. I., van der Meer, J. R., Timmis, K. N., Brussaard, C. P., & McGenity, T. J. (2015). Generalist hydrocarbon-degrading bacterial communities in the oil-polluted water column of the North Sea. *Microb Biotechnol, 8*(3), 434-447. doi:10.1111/1751-7915.12176

Chust, G., Villarino, E., Chenuil, A., Irigoien, X., Bizsel, N., Bode, A., . . . Borja, A. (2016). Dispersal similarly shapes both population genetics and community patterns in the marine realm. *Scientific Reports, 6*(1), 28730. doi:10.1038/srep28730

Dhami, M. K., Hartwig, T., Letten, A. D., Banf, M., & Fukami, T. (2018). Genomic diversity of a nectar yeast clusters into metabolically, but not geographically, distinct lineages. *Mol Ecol, 27*(8), 2067-2076. doi:10.1111/mec.14535

Dubinsky, E. A., Conrad, M. E., Chakraborty, R., Bill, M., Borglin, S. E., Hollibaugh, J. T., . . . Andersen, G. L. (2013). Succession of Hydrocarbon-Degrading Bacteria in the Aftermath of the Deepwater Horizon Oil Spill in the Gulf of Mexico. *Environmental Science & Technology, 47*(19), 10860-10867. doi:10.1021/es401676y

Figueras, M. J., Beaz-Hidalgo, R., Hossain, M. J., & Liles, M. R. (2014). Taxonomic affiliation of new genomes should be verified using average nucleotide identity and multilocus phylogenetic analysis. *Genome Announc, 2*(6). doi:10.1128/genomeA.00927-14

Galardini, M., Mengoni, A., Biondi, E. G., Semeraro, R., Florio, A., Bazzicalupo, M., . . . Mocali, S. (2014). DuctApe: a suite for the analysis and correlation of genomic and OmniLog™ Phenotype Microarray data. *Genomics, 103*(1), 1-10. doi:10.1016/j.ygeno.2013.11.005

Ghannam, R. B., Schaerer, L. G., Butler, T. M., & Techtmann, S. M. (2020). Biogeographic Patterns in Members of Globally Distributed and Dominant Taxa Found in Port Microbial Communities. *mSphere, 5*(1). doi:10.1128/mSphere.00481-19

Gibbons, S. M., Caporaso, J. G., Pirrung, M., Field, D., Knight, R., & Gilbert, J. A. (2013). Evidence for a persistent microbial seed bank throughout the global ocean. *Proceedings of the National Academy of Sciences, 110*(12), 4651-4655. doi:doi:10.1073/pnas.1217767110

412 Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., & Tiedje, J. M.
413 (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence
414 similarities. *Int J Syst Evol Microbiol, 57*(Pt 1), 81-91. doi:10.1099/ijs.0.64483-0
415 Green, J. L., Bohannan, B. J. M., & Whitaker, R. J. (2008). Microbial Biogeography: From
416 Taxonomy to Traits. *Science, 320*(5879), 1039-1043. doi:doi:10.1126/science.1153475
417 Gutierrez, T., Singleton, D. R., Berry, D., Yang, T., Aitken, M. D., & Teske, A. (2013).
418 Hydrocarbon-degrading bacteria enriched by the Deepwater Horizon oil spill identified by
419 cultivation and DNA-SIP. *The ISME journal, 7*(11), 2091-2104. doi:10.1038/ismej.2013.98
420 Hadfield, J., Croucher, N. J., Goater, R. J., Abudahab, K., Aanensen, D. M., & Harris, S. R. (2017).
421 Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics, 34*(2),
422 292-293. doi:10.1093/bioinformatics/btx610
423 Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C., & Martiny, J. B. (2012). Beyond
424 biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol,*
425 *10*(7), 497-506. doi:10.1038/nrmicro2795
426 Harris, A. P., Techtmann, S. M., Stelling, S. C., Utturkar, S. M., Alshibli, N. K., Brown, S. D., &
427 Hazen, T. C. (2014). Draft Genome Sequence of Pseudoalteromonas sp. Strain ND6B, an
428 Oil-Degrading Isolate from Eastern Mediterranean Sea Water Collected at a Depth of
429 1,210 Meters. *Genome Announc, 2*(6). doi:10.1128/genomeA.01212-14
430 Hellweger, F. L., van Sebille, E., & Fredrick, N. D. (2014). Biogeographic patterns in ocean
431 microbes emerge in a neutral agent-based model. *Science, 345*, 1346-1349.
432 doi:10.1126/science.1254421
433 Huerta-Cepas, J., Serra, F., & Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization
434 of Phylogenomic Data. *Molecular Biology and Evolution, 33*(6), 1635-1638.
435 doi:10.1093/molbev/msw046
436 Jain, C., Rodriguez, R. L., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High
437 throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat*
438 *Commun, 9*(1), 5114. doi:10.1038/s41467-018-07641-9
439 Johnson, Z. I., Zinser, E. R., Coe, A., McNulty, N. P., Woodward, E. M., & Chisholm, S. W.
440 (2006). Niche partitioning among Prochlorococcus ecotypes along ocean-scale
441 environmental gradients. *Science, 311*(5768), 1737-1740. doi:10.1126/science.1118052
442 Jönsson, B. F., & Watson, J. R. (2016). The timescales of global surface-ocean connectivity. *Nat*
443 *Commun, 7*(1), 11239. doi:10.1038/ncomms11239
444 Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids*
445 *Res, 28*(1), 27-30. doi:10.1093/nar/28.1.27
446 Kashtan, N., Roggensack, S. E., Berta-Thompson, J. W., Grinberg, M., Stepanauskas, R., &
447 Chisholm, S. W. (2017). Fundamental differences in diversity and genomic population
448 structure between Atlantic and Pacific Prochlorococcus. *The ISME journal, 11*(9), 1997-
449 2011. doi:10.1038/ismej.2017.64
450 Kivlin, S. N., Winston, G. C., Goulden, M. L., & Treseder, K. K. (2014). Environmental filtering
451 affects soil fungal community composition more than dispersal limitation at regional scales.
452 *Fungal Ecology, 12*, 14-25. doi:https://doi.org/10.1016/j.funeco.2014.04.004
453 Malmstrom, R. R., Rodrigue, S., Huang, K. H., Kelly, L., Kern, S. E., Thompson, A., . . . Chisholm,
454 S. W. (2013). Ecology of uncultured Prochlorococcus clades revealed through single-cell
455 genomics and biogeographic analysis. *The ISME journal, 7*(1), 184-198.
456 doi:10.1038/ismej.2012.89

Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., . . . Staley, J. T. (2006). Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology, 4*(2), 102-112. doi:10.1038/nrmicro1341

McClain, C. R., Stegen, J. C., & Hurlbert, A. H. (2012). Dispersal, environmental niches and oceanic-scale turnover in deep-sea bivalves. *Proceedings of the Royal Society B: Biological Sciences, 279*(1735), 1993-2002. doi:doi:10.1098/rspb.2011.2166

Nanjundiah, V. (2019). Many roads lead to Rome: Neutral phenotypes in microorganisms. *J Exp Zool B Mol Dev Evol, 332*(8), 339-348. doi:10.1002/jez.b.22909

Nordberg H, C. M., Dusheyko S, Hua S, Poliakov A, Shabalov I, Smirnova T, Grigoriev IV, Dubchak I. (2014). The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Res, 42*(1), D26-31.

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A., Korobeynikov, A., Lapidus, A., . . . Pevzner, P. A. (2013). *Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads*, Berlin, Heidelberg.

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., . . . Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics, 31*(22), 3691-3693. doi:10.1093/bioinformatics/btv421

Parks DH, I. M., Skennerton CT, Hugenholtz P, Tyson GW. (2014.). Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. . *Genome research, 25*, 1043-1055.

Redmond, M. C., & Valentine, D. L. (2012). Natural gas and temperature structured a microbial community response to the Deepwater Horizon oil spill. *Proceedings of the National Academy of Sciences, 109*(50), 20292-20297. doi:doi:10.1073/pnas.1108756108

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics, 30*(14), 2068-2069. doi:10.1093/bioinformatics/btu153 %J Bioinformatics

Shirani, S., & Hellweger, F. L. (2017). Neutral Evolution and Dispersal Limitation Produce Biogeographic Patterns in Microcystis aeruginosa Populations of Lake Systems. *Microb Ecol, 74*(2), 416-426. doi:10.1007/s00248-017-0963-5

Sjöqvist, C., Delgado, L. F., Alneberg, J., & Andersson, A. F. (2021). Ecologically coherent population structure of uncultivated bacterioplankton. *The ISME journal, 15*(10), 3034-3049. doi:10.1038/s41396-021-00985-z

Sun, Y., Zhang, M., Duan, C., Cao, N., Jia, W., Zhao, Z., . . . Wang, J. (2021). Contribution of stochastic processes to the microbial community assembly on field-collected microplastics. *Environmental Microbiology, 23*(11), 6707-6720. doi:https://doi.org/10.1111/1462-2920.15713

Techtman, S. M., Mahmoudi, N., Whitt, K. T., Campa, M. F., Fortney, J. L., Joyner, D. C., & Hazen, T. C. (2017). Comparison of Thaumarchaeotal populations from four deep sea basins. *FEMS Microbiol Ecol, 93*(11). doi:10.1093/femsec/fix128

Techtmann, S. M., Fitzgerald, K. S., Stelling, S. C., Joyner, D. C., Uttukar, S. M., Harris, A. P., . . . Hazen, T. C. (2016). Colwellia psychrerythraea Strains from Distant Deep Sea Basins Show Adaptation to Local Conditions. *Frontiers in Environmental Science, 4*. doi:10.3389/fenvs.2016.00033

Troussellier, M., Escalas, A., Bouvier, T., & Mouillot, D. (2017). Sustaining Rare Marine Microorganisms: Macroorganisms As Repositories and Dispersal Agents of Microbial Diversity. *Frontiers in Microbiology, 8*. doi:10.3389/fmicb.2017.00947

Ward, B. A., Cael, B. B., Collins, S., & Young, C. R. (2021). Selective constraints on global plankton dispersal. *Proceedings of the National Academy of Sciences, 118*(10), e2007388118. doi:doi:10.1073/pnas.2007388118

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-16-2/32316**	**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

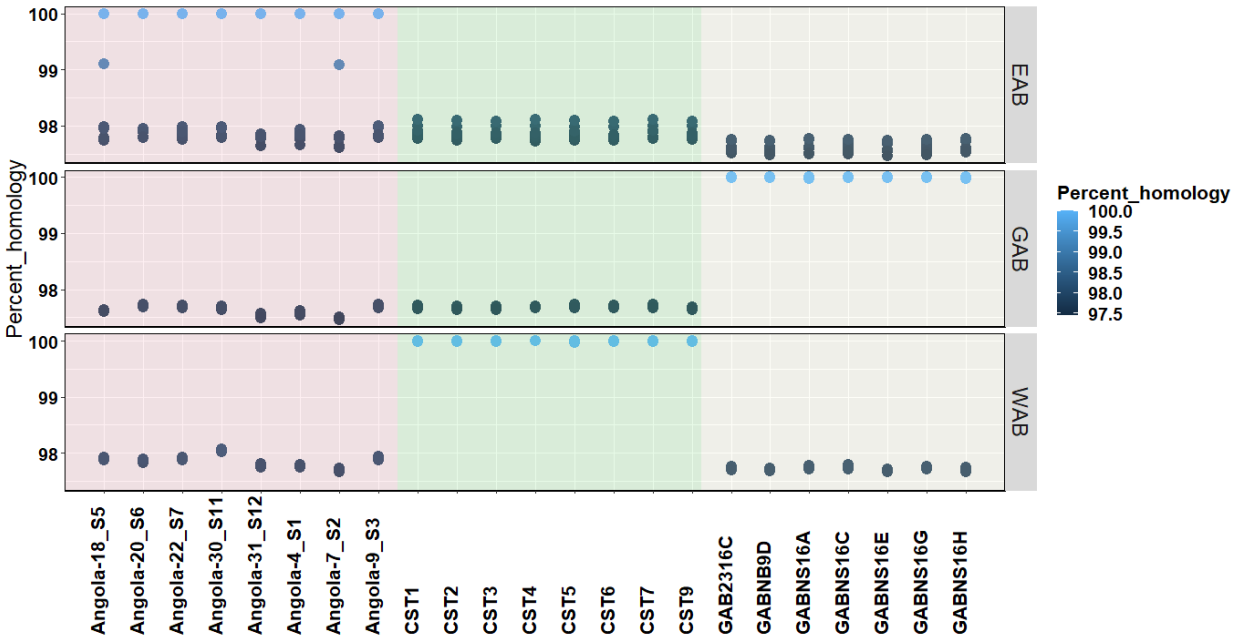
509

510

511

Table 2: Percent Identity of the 16S rRNA with *Pseudoalteromonas atlantica* strain NBRC 103033

Location	Strain	16S ribosomal RNA Percent Identity	Location	Strain	16S ribosomal RNA Percent Identity
EAB	Angola-4	99.658	GAB	GABNS16E	99.589
EAB	Angola-7	100	GAB	GABNS16G	99.589
EAB	Angola-9	99.658	GAB	GABNS16H	99.686
EAB	Angola-18	99.757	WAB	CST-1	99.658
EAB	Angola-20	99.658	WAB	CST-2	99.658
EAB	Angola-22	99.658	WAB	CST-3	99.658
EAB	Angola-30	99.785	WAB	CST-4	99.291
EAB	Angola-31	99.92	WAB	CST-5	99.658
GAB	GAB2316C	99.589	WAB	CST-6	99.658
GAB	GABNB9D	99.589	WAB	CST-7	99.658
GAB	GABNS16A	99.686	WAB	CST-9	99.658
GAB	GABNS16C	99.686			



517

518 Figure 1: Percent homology using ANI values of all-versus-all comparison of genomes.

519

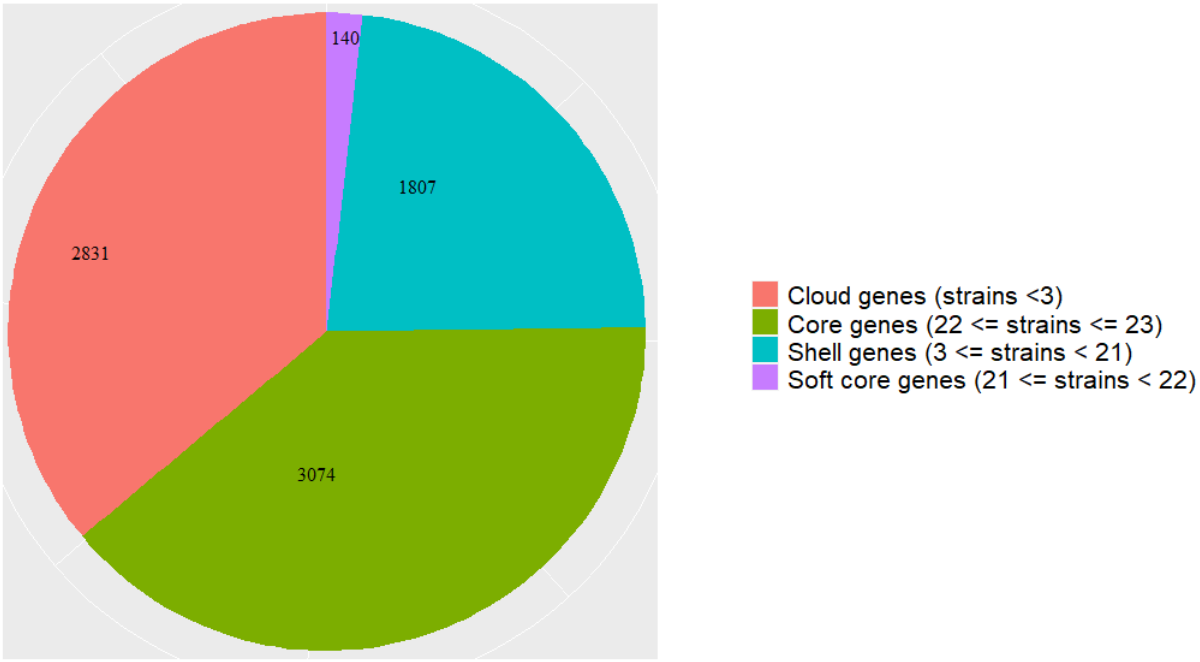


Figure 2: Comparative genomic analysis using Roary to identify genes in the core and pan genomes.

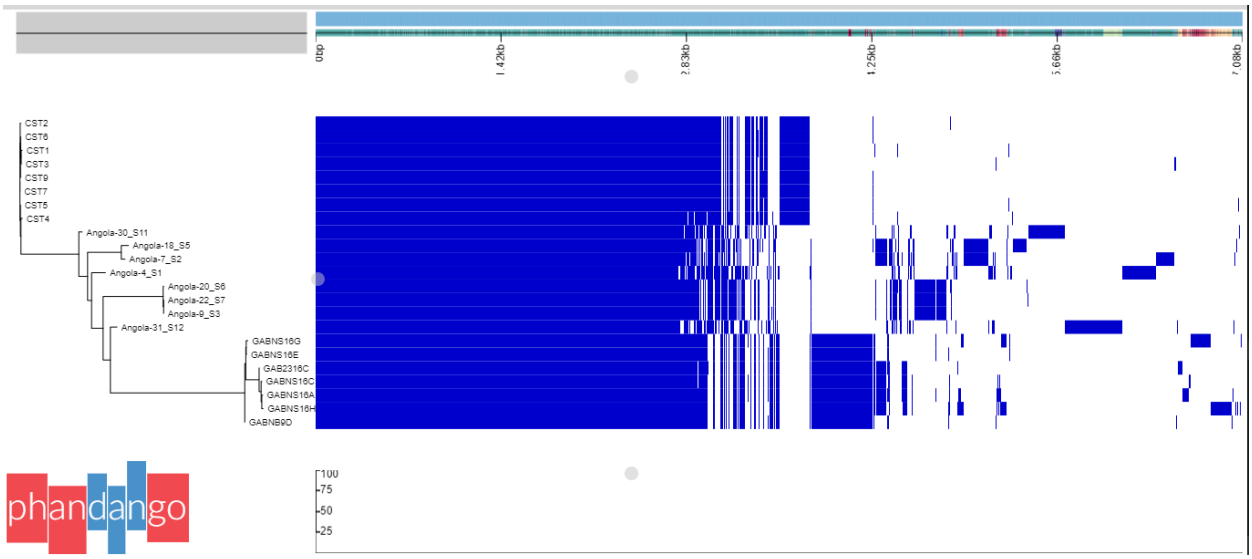
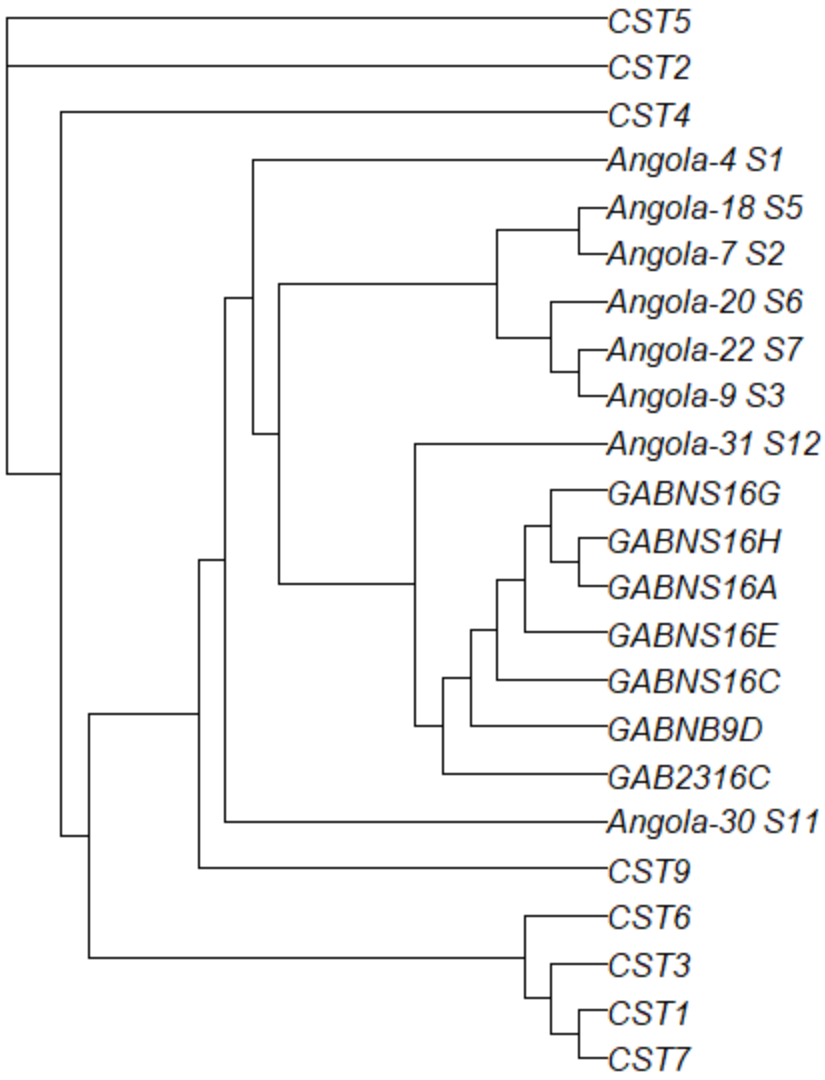
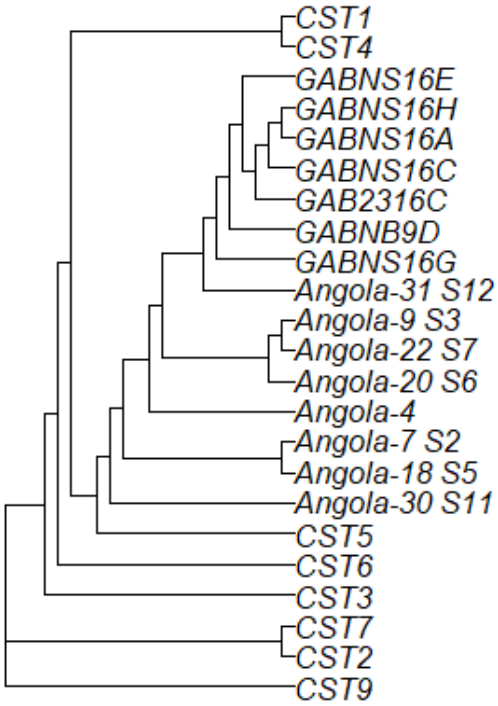


Figure 3: Use of gene presence in each strain to generate the phylogenetic tree, core- and pan-genomes.



527
528 Figure 4: Strain-level phylogenetic tree based on four hundred universal marker genes using
529 Phylophlan 3.0.

530



531

532 Figure 5: Phylogenetic tree based on presence and absence of genes in the 23 isolates studied
533 obtained using Roary pipeline

534

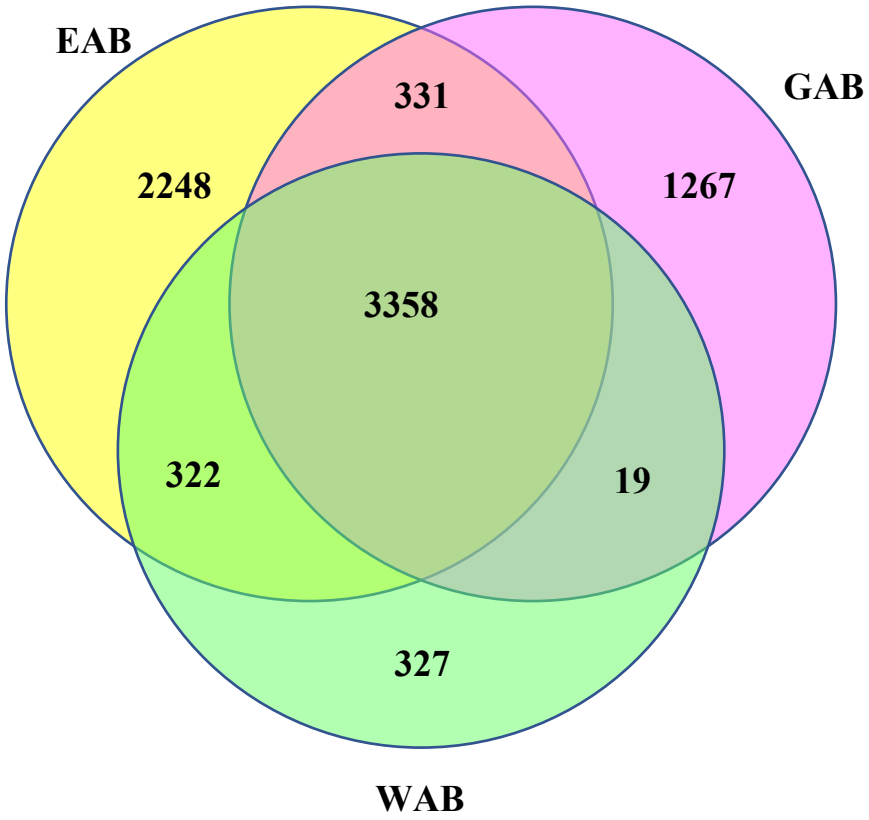


Figure 6: Venn diagram showing genes present in all three locations, the accessory genes present exclusively in strains from each location, and genes shared by strains from two different locations.