Genomic diversity and adaptation in Arctic marine bacteria

Michael C. Sadler,[1] Matthias Wietz,[2,3,4] Sayaka Mino,[1,5] Robert M. Morris[1]

AUTHOR AFFILIATIONS See affiliation list on p. 13.

ABSTRACT Arctic marine bacteria experience seasonal changes in temperature, salinity, light, and sea ice cover. Time-series and metagenomic studies have identified spatiotemporal patterns in Arctic microbial communities, but a lack of complete genomes has limited efforts to identify the extent of genomic diversity in Arctic populations. We cultured and sequenced the complete genomes of 34 Arctic marine bacteria to identify patterns of gene gain, loss, and rearrangement that structure genomes and underlie adaptations to Arctic conditions. We found that the most abundant lineage in the Arctic (SAR11) is comprised of diverse species and subspecies, each encoding 50–150 unique genes. Half of the 16 SAR11 genomes harbor a genomic island with the potential to enhance survival in the Arctic by utilizing the osmoprotectant and potential methyl donor glycine betaine. We also cultured and sequenced four species representing an uncultured family of *Pseudomonadales*, four subspecies of *Pseudothioglobus* (SUP05), a genus of high GC *Puniceispirillales* (SAR116), and a family of low GC SAR116. Time-series 16S rRNA amplicon data indicate that this culture collection represents up to 60% of the marine bacterial community in Arctic waters. Their genomes provide insights into the evolutionary processes that underlie bacterial diversity and adaptation to Arctic waters.

IMPORTANCE Genetic diversity has limited efforts to assemble and compare whole genomes from natural populations of marine bacteria. We developed a cultivation-based population genomics approach to culture and sequence the complete genomes of bacteria from the Arctic Ocean. Cultures and closed genomes obtained in this study represent previously uncultured families, genera, and species from the most abundant lineages of bacteria in the Arctic. We report patterns of gene gain, loss, rearrangement, and adaptation in the dominant lineage (SAR11), as well as the size, composition, and structure of genomes from several other groups of marine bacteria. This work demonstrates the potential for cultivation-based high-throughput genomics to enhance understanding of the processes underlying genomic diversity and adaptation.

KEYWORDS SAR11, SUP05, horizontal gene transfer, pangenomics, nanopore, Arctic, cultivation

The oceans are dominated by relatively few but highly diverse lineages of marine bacteria (1, 2). Representatives for some of these lineages have been cultured from the Atlantic and Pacific Oceans, including *Pelagibacterales* (SAR11) (3, 4), *Puniceispirillales* (SAR116) (5), and *Thioglobaceae* (SUP05) (6, 7), but there is significant and unexplored genetic diversity within populations and across oceans. Metagenome-assembled genomes (MAGs) (8–10) and single-cell amplified genomes (SAGs) (11–13) have advanced the understanding of this diversity; however, MAGs and SAGs are typically incomplete and do not provide a full picture of the genetic variation within a highly diverse population of species and strains (14). Advances in cultivation have resulted in the recovery of abundant and novel marine bacteria (15), and modern sequencing techniques produce highly accurate closed bacterial genomes (16, 17). Combined, these

approaches can provide more complete information about the genetic variation within bacterial populations.

The Arctic Ocean is a highly dynamic system characterized by strong spatial and temporal variation, including sea ice cover, daylight, stratification, and freshwater input (18–20). Microbes have evolved specific adaptations to survive in such extremes, such as the ability to use the compatible solute glycine betaine (21), which can enhance osmoregulation and survival in sea ice (22, 23) or can serve as a methyl donor and source of glycine (24). Some Arctic bacteria have been isolated in culture, including *Colwellia* and *Polaribacter* (25–27). However, cultivation-independent methods suggest that current Arctic culture collections represent a relatively small fraction of the community (28). Understanding the genetic variation within populations of Arctic bacteria is particularly important because the Arctic is warming at four times the global rate (29), rapidly losing sea ice (30), and encountering Atlantic water intrusion (31). These changes have the potential to impact bacterial and eukaryotic microbial diversity (32, 33), ultimately changing ecosystem functioning and biogeochemical cycles.

We conducted a high-throughput cultivation-based genomics study to gain insights into the diversity within Arctic bacterial populations. We focused on highly abundant and diverse taxa with few cultured and sequenced representatives, such as SAR11 (3, 34). SAR11 accounts for approximately 25% of bacteria in the Arctic Ocean (33, 35, 36) and is also present in sea ice (37). Like many other marine bacterial lineages, SAR11 has been classified into subclades based on 16S rRNA and internal transcribed spacer sequence analysis (38–40). These classifications are often used to identify patterns of diversity, including in the Arctic (41–43). SAR11 genomes are among the smallest for free-living bacteria (44) and have a high proportion of core genes (39), with most unique genes co-located in a ~50 Kb hypervariable region (HVR) termed HVR2 (45). The evolutionary mechanisms that maintain diversity in HVRs are poorly understood, though homologous recombination is widespread in SAR11 (46) and has been proposed as a driver of diversity in marine bacteria (47). These features have hindered efforts to resolve genetic variation and recover unique genes within populations of SAR11 (14).

This study used comparative genomics to identify patterns of gene gain, loss, and rearrangement in 16 Arctic SAR11. Several other important strains of marine bacteria representing previously uncultured species, genera, and families were also cultured and sequenced (named herein). Rationale for names assigned to previously uncultured families and genera is summarized below and detailed in the protologue. Briefly, *Candidatus* Njordibacter, the genus name derived from "Njord" (the Norse god of wind and seas), and the family *Njordibacteraceae* to encompass this genus. *Candidatus* Levibacter, the genus name derived from the Latin "levis" (lightweight), and the family *Levibacteraceae* to encompass this low GC genus of *Puniceispirillales*. *Candidatus* Ponderosibacter, the genus name derived from the Latin "ponderosus" (heavy), referring to this genus of the high GC *Puniceispirillales*. *Candidatus* Marifrigoribacter, the genus name derived from the Latin "mare" (of the sea) and "frigus" (cold). All 34 new species names correspond to their cultivation ID (e.g., sp. uisw_002).

## RESULTS

High-throughput cultivation-based genomics

We cultured and sequenced the complete genomes of 34 Arctic marine bacteria to identify differences in genomic diversity and adaptations to Arctic conditions (Fig. 1). Cultures were selected from 106 bacteria obtained by high-throughput dilution-to-extinction cultivation and sequenced using the Oxford Nanopore Technologies (ONT) platform (Table S1). All genomes are single circular contigs, range in size from 1.29 to 3.78 Mbp, and have GC contents between 29% and 52% (Table S2). They include representatives from globally distributed lineages, such as *Pelagibacter* (SAR11, $n = 16$), *Pseudothioglobus* (SUP05, $n = 4$), *Puniceispirillales* (SAR116, $n = 2$), *Methylophilaceae* (OM43, $n = 2$), and *Haliaceae* (OM60, $n = 1$), as well as uncultured and understudied lineages, including a low GC family of SAR116 (48) and a previously unrecognized
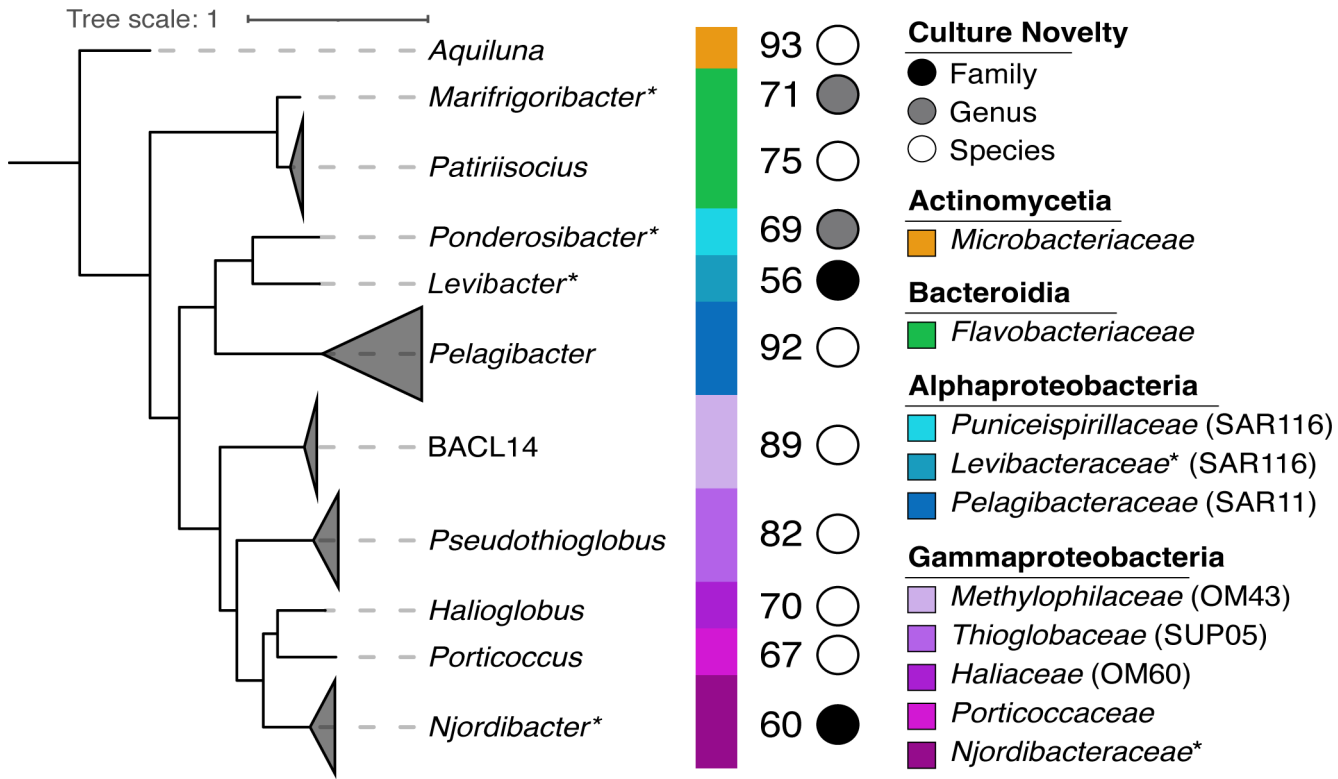
FIG 1 Diversity of Arctic bacteria cultured and sequenced herein. Phylogenomic analysis of Arctic marine bacteria sequenced in this study, with genus affiliation (italics), class and family taxonomic grouping (color bars), maximum amino acid identity (AAI) of all branch members to the closest cultured relative (number), and culture novelty in circles (family, genus, and species). Wedges are proportional to the number of cultures sequenced for each lineage. Family and genus names were assigned using the Genome Taxonomy Database (GTDB)-Tk placement tool, except for those proposed in this study, marked with an asterisk.

family of *Pseudomonadales* that is common in the Arctic (*Njordibacteraceae* herein). Other cultured strains belonging to genera and species without previous representation were also sequenced, including *Flavobacteriaceae* and *Actinomycetes*.

All genomes represent newly cultured species (Fig. 1, Table S1). The four SUP05 genomes have 95% average nucleotide identity (ANI) when compared to each other and less than 95% ANI when compared to described species. The SUP05 genomes encode key proteins needed to fix inorganic carbon using the Calvin-Benson-Bassham cycle, including phosphoribulokinase and the large and small subunits of form Ia Ribulose Bisphosphate Carboxylase. *Levibacter* and *Ponderosibacter,* belonging to the SAR116 clade, have GC contents of 31.4% and 49.8%, coding densities of 95.9% and 88.5%, and amino acid identity (AAI) of 56% and 69% when compared to their nearest cultured relatives, respectively (Table S2). Phylogenetic analysis indicates these cultures represent distinct and diverse SAR116 lineages (Fig. S1). Both SAR116 genomes encode genes for bacteriorhodopsin, carotenoid biosynthesis, carbon monoxide dehydrogenase, and key proteins for sulfur metabolism. The *Porticoccus, Patiriisocius,* and *Halioglobus* genomes have low AAI values when compared to their nearest cultured relatives (67%, 75%, and 70%, respectively), indicating substantial diversity in these lineages. The *Marifrigoribacter* genome has a similarly low AAI value when compared to its nearest cultured relative (71%) and represents the first genome of an uncultured genus of *Flavobacteriaceae*. Notably, the *Porticoccus* genome has 67% AAI and is 40% (1.4 Mbp) smaller than its nearest cultured relative.
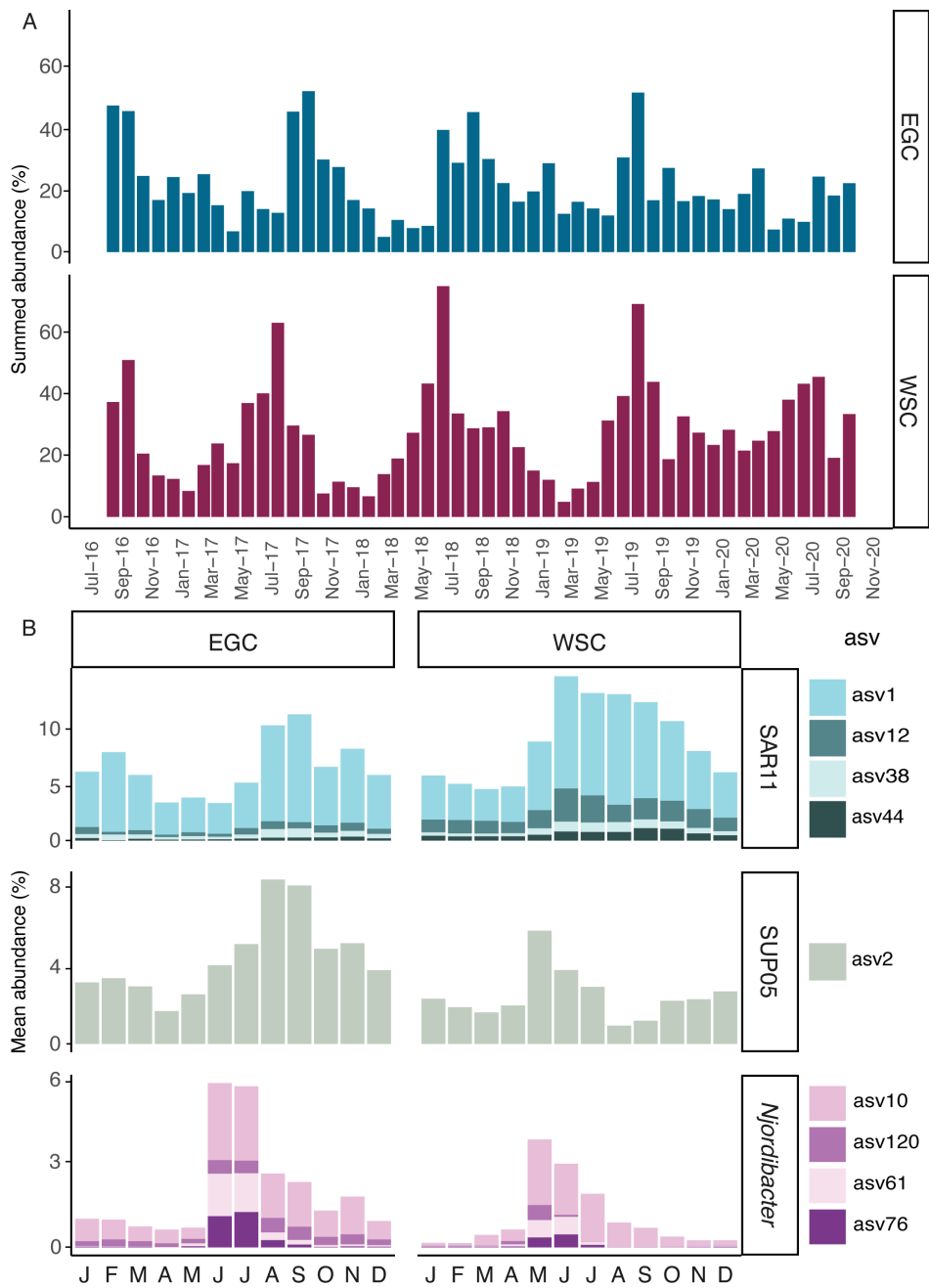
FIG 2 Relative abundance of bacterial ASVs from the FRAM Observatory with 100% identity to 16S rRNA gene sequences from cultures obtained in this study. (A) Summed relative abundance of all matching ASVs from August 2016 to September 2020 in the EGC and WSC. (B) Mean relative abundance values calculated for ASVs matching sequenced *Pelagibacter* (SAR11), *Pseudothioglobus* (SUP05), and *Njordibacter* cultures in the EGC (left) and WSC (right). Bars in panel B represent monthly averages. Individual ASVs for each lineage in panel B are shown on the right.

Spatial and temporal abundance of Arctic bacteria

Spatiotemporal abundance estimates for bacteria cultured in this study were determined by aligning the 16S rRNA gene sequences from cultures with 16S rRNA amplicon sequence variants (ASVs) from two continuous time-series (Fig. 2). These 4-year Arctic datasets were obtained from the West Spitsbergen Current (WSC) and the East Greenland Current (EGC) (49, 50). The 34 cultures matched 18 ASVs with 100% sequence identity (Table S2) and constituted over 60% of the ASV abundance in the WSC and

50% of the ASV abundance in the EGC (Fig. 2A). The most abundant members in the culture collection included SAR11, SUP05, and *Njordibacter*, accounting for 4%–15% of the community in the WSC and 6%–10% of the community in the EGC at peak abundances, respectively (Fig. 2B). ASVs matching SUP05, *Njordibacter*, and *Marifrigoribacter* cultures were more abundant in the EGC, while those matching SAR11, SAR116, and OM43 cultures were more abundant in the WSC (Fig. 2; Fig. S2). Spearman correlation analyses indicate that many of the ASVs matching cultured taxa covary with environmental parameters measured over the same 4 years (Fig. S3). ASVs matching SAR11, SAR116, and OM43 cultures correlate positively with water temperature and negatively with polar water fraction, while ASVs matching *Njordibacter* and SUP05 cultures correlate negatively with water temperature and positively with polar water fraction.

A relatively small fraction of the total SAR11, SUP05, and *Njordibacter* ASVs matched sequences in cultures (4/116, 1/55, and 4/24, respectively). However, these nine ASVs accounted for a substantial percentage of total SAR11, SUP05, and *Njordibacter* ASV abundance detected in the Arctic time-series (70%, 50%, and 90%, respectively) (Fig. S4). Differences in genome and ASV diversity are notable between the groups (Fig. S5; Tables S2 and S3). Many SAR11 genomes are unique species (<95% ANI) but match the same ASV (e.g., uisw_099_02 and uisw_114 matching ASV38), while SAR11 uisw_121 and uisw_116 are the same species (>95% ANI) but match different ASVs. Similarly, *Njordibacter* uisw_002 and uisw_058 have 80% ANI to each other and match the same ASV, while different copies of the 16S rRNA gene in *Njordibacter* uisw_056 match different ASVs. In contrast, all SUP05 genomes represent the same species (>95% ANI) and match the same ASV.

Gene gain, loss, and rearrangement in Arctic SAR11

We evaluated the diversity of SAR11 genomes, which are abundant in the Arctic and well represented in our culture collection, to identify population-wide patterns of genomic diversity. All 16 Arctic SAR11 cultures obtained in this study are members of subclade Ia.1 (Fig. S6). Twelve have ANI values that are <95% to each other, and two pairs have ANI values >95% to each other, indicating that they represent 14 new species and two new subspecies of SAR11 (Fig. 3A; Table S3). Population genomic analysis identified 3,596 total gene clusters, 1,661 singleton gene clusters, and 1,037 core gene clusters (Fig. 3B). Each genome encodes between 1,385 and 1,477 genes. The number of shared genes
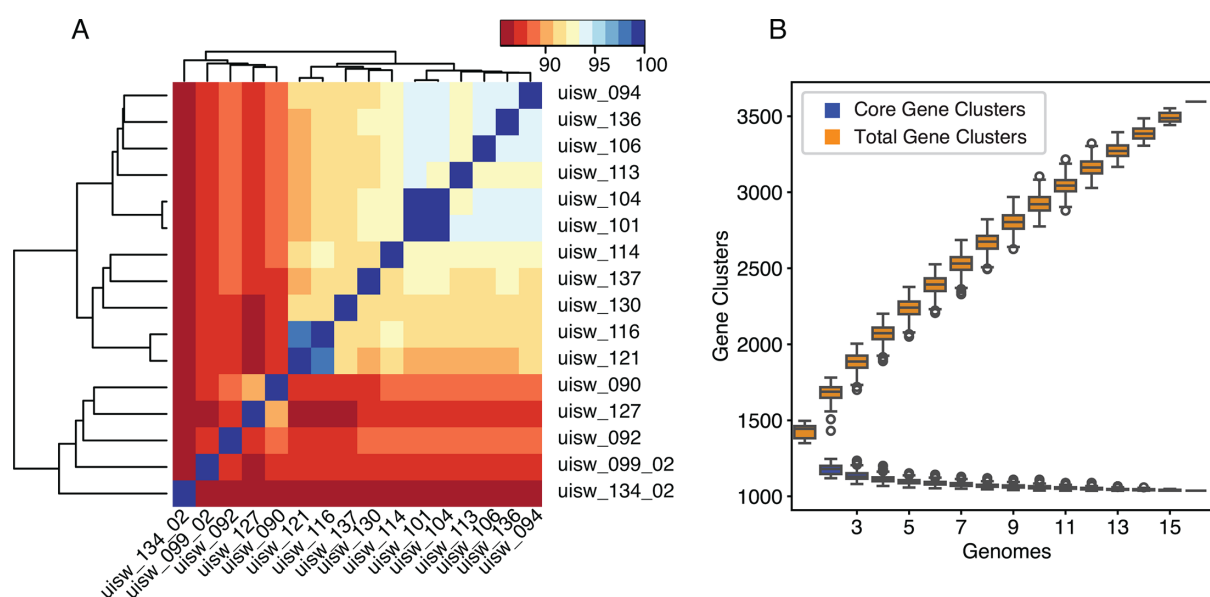
FIG 3 Genomic diversity of Arctic SAR11 cultures. (A) Clustered ANI for each pairwise genome comparison. (B) Gene cluster accumulation curves for the Arctic SAR11 population, total gene clusters (orange), and core gene clusters (blue). Accumulation curves were calculated using 1,000 iterations.

between any two genomes ranges from 1,126 to 1,234, and the number of singletons in each genome ranges from 50 to 150.
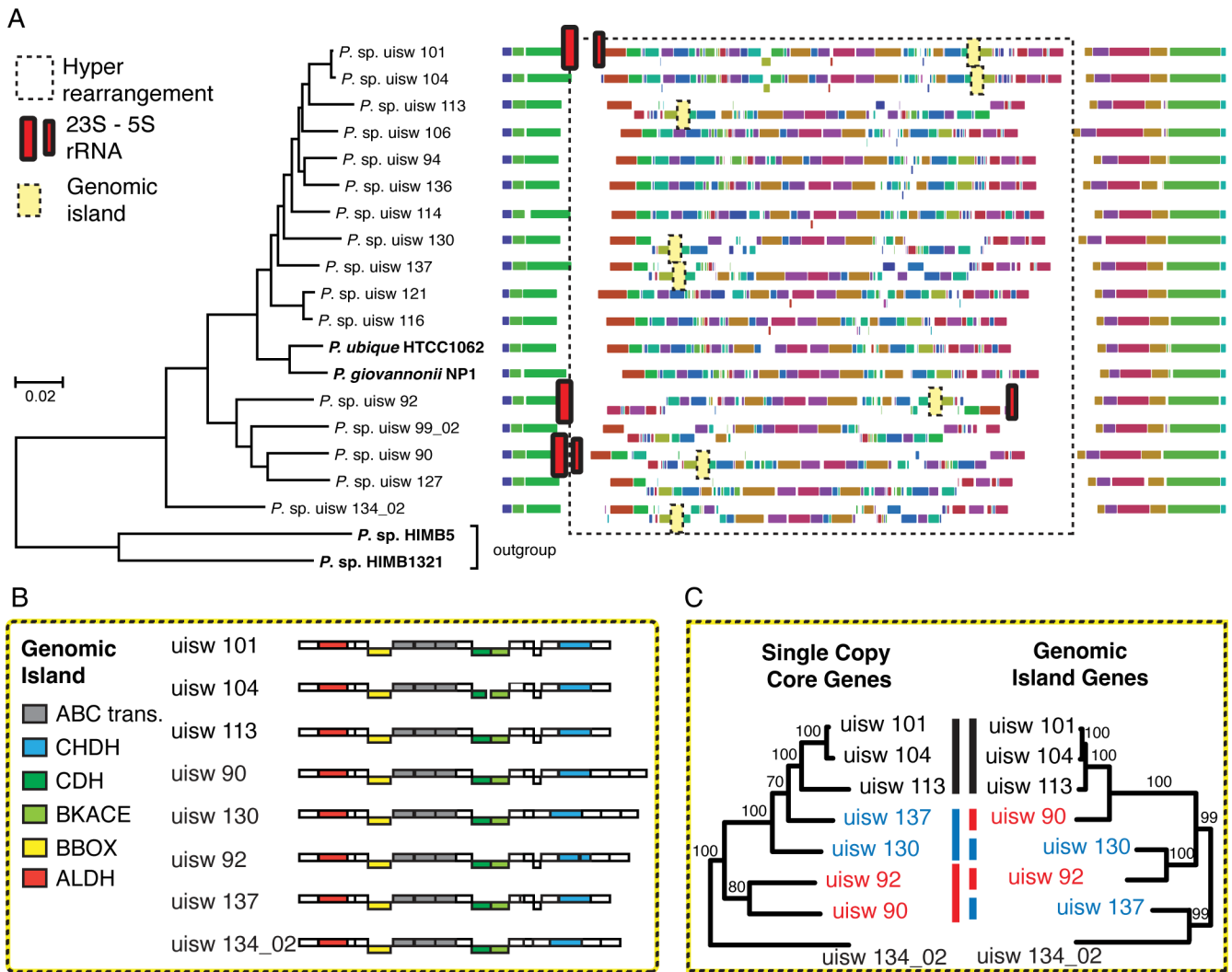
A whole genome alignment with all 16 Arctic SAR11 and two closely related reference genomes (*Pelagibacter ubique and Pelagibacter giovannonii*) was constructed to identify patterns of gene gain, loss, and rearrangement. The alignment revealed notable patterns of rearrangement (Fig. 4). Most notably, there is an ~880 Kb section bound by the 23S and 5S rRNA genes in uisw_092 (Fig. 4A). Patterns of gene gain, loss, and rearrangement in this region were used to identify a unique genomic region that has 17–20 genes and is present in half of the sequenced genomes (Fig. 4B; Table S4). Genes in this region code for a glycine betaine ABC transport system (ABC *trans*.), a choline dehydrogenase (CHDH), an aldehyde dehydrogenase (ALDH), a carnitine dehydrogenase (CDH), a gamma-butyr-obetaine dioxygenase (BBOX), and a beta-keto acid cleavage enzyme (BKACE). Gene order for these genes is preserved and always upstream of genomic betaine/choline/proline ABC transport genes. Phylogenetic trees constructed using concatenated protein sequences from these genes and single-copy core genes are incongruent (Fig. 4C). Gene cluster analysis indicates that of the 17 genes, 15 are unique to the region, with no identifiable homologs occurring elsewhere in the genomes. The remaining two form phylogenetic clusters that are distinct from clusters produced using other copies found in the genomes, such as for aldehyde and choline dehydrogenases (Fig. S7A and B). Most genes in this region (13 out of 17) produced phylogenies that were congruent with each other and with the concatenated set of all 17 genes, including the aldehyde and choline dehydrogenase genes found in this region (Fig. S7C and D).

We searched for this genomic region in public databases to see if it was common but previously unrecognized. There was no evidence of this arrangement in 18 previously sequenced single-contig SAR11 genomes (Table S5). Homologs for the 17 encoded genes were identified in environmental databases. Abundances for 13 have a significant positive correlation with latitude (0.31–0.83, P value < 0.01) (Fig. 5), and seven are rare or absent in samples from lower latitudes, including a lactoylglutathione lyase, a glycerophosphodiester phosphodiesterase, the permease subunit of a proline/glycine betaine ABC transporter, a class II aldolase, a thioesterase, a small multidrug resistance (SMR) transporter, and a drug/metabolite transporter (DMT).

A new family of *Pseudomonadales*

A group of *Pseudomonadales* (named *Njordibacter* herein) was frequently identified in culture ($n = 10$). Four cultures were selected for whole genome sequencing and used to identify a new family in the order *Pseudomonadales* (Fig. 6). The genomes range from 2.4 to 2.9 Mbp in length, have a GC content of 43.8%–45.1%, and each contains four copies of the 16S rRNA gene (Table S2). Pairwise ANI values range from 79% to 82% when compared to each other. One genome, uisw_056, and a previously characterized MAG from the Arctic (49) have 99.6% ANI. Phylogenomic analysis using the Genome Taxonomy Database (GTDB) (51) placed this group in a deeply branching and uncultivated genus (ASP10-02a) of *Nitrincolaceae*, but with low AAI (60%) and 16S rRNA sequence similarity (91.7%) to the most closely related isolate. Additionally, the mean relative evolutionary divergence (RED) of this group was 0.875, falling within both the genus and family ranges (52). We therefore compared the genomes of 114 isolates in the order *Pseudomonadales* (Table S6) to each other and to the four *Njordibacter* genomes using AAI and percent of conserved protein (POCP). This yielded over 10,000 pairwise comparisons between 34 genera from seven families, from which family delineations could be distinguished (Fig. 6). Values of 60% AAI and 40% POCP with their closest *Nitrincolaceae* relatives are below the family delineation line, indicating that these *Njordibacter* genomes represent a new family in the order *Pseudomonadales*.

FIG 4 Structure of Arctic SAR11 genomes. (A) Phylogenomic analysis and whole-genome alignment of Arctic SAR11. (B) Alignment and gene content of a genomic island identified in eight SAR11 genomes. (C) Comparison of phylogenetic trees constructed from concatenated protein sequences of single-copy core genes (left) and genomic island genes (right). Phylogenetic trees were constructed with translated and concatenated sequences. Whole genome alignments start with DnaA. In all genomes, the relative position of the 23S and 5S rRNA genes is as depicted for uisw_101, unless otherwise indicated. Previously sequenced isolates are in bold. Enzyme names in panel B are abbreviated as follows: ABC *Trans,* ABC transporter; CHDH, choline dehydrogenase; CDH, carnitine dehydrogenase; BKACE, beta-keto acid cleavage enzyme; BBOX, gamma-butyrobetaine dioxygenase; ALDH, aldehyde dehydrogenase. Bootstrap values in panel C are displayed at each node. Black, red, and blue color blocks highlight node rearrangements.

DISCUSSION

Extensive genomic diversity in Arctic SAR11

Our analysis identified 1,037 core genes in 16 SAR11 subclade Ia.1 genomes obtained from Arctic seawater (Fig. 3B). A similar estimate of 1,047 core genes was obtained in a SAR11 pangenome analysis of five complete genomes from subclade Ia, isolated from diverse locations in the Atlantic and Pacific Oceans (39). Similarity between the core genome estimates using 16 genomes from the Arctic and five genomes from different oceans highlights the extent of core genome conservation within the SAR11 clade (34, 39). The SAR11 pangenome analysis also identified 1,962 total unique gene clusters in the five complete genomes (39). The mean number of unique gene clusters in five Arctic SAR11 genomes was 2,241 ± 117 (95% CI), suggesting that there is greater gene content diversity among Arctic populations of SAR11. This could be due to the mixing of different
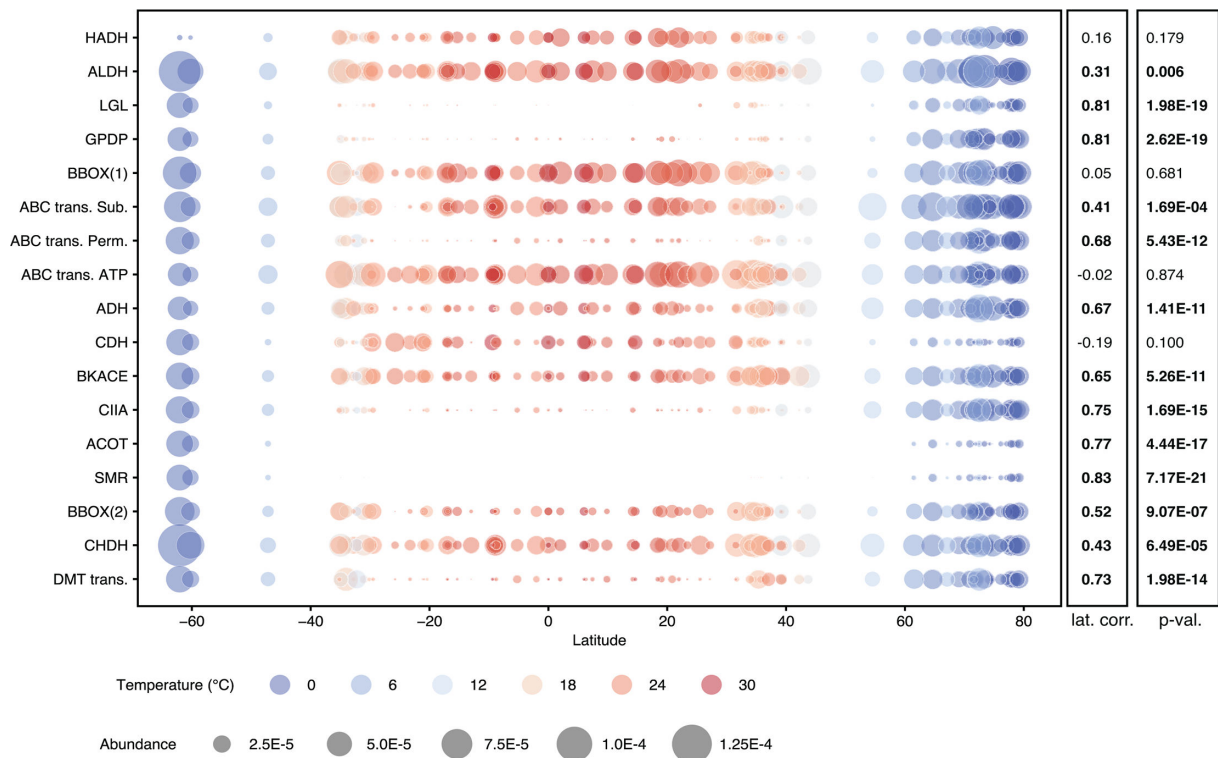
FIG 5 Latitudinal distribution of genes identified in an Arctic SAR11 genomic island. Representative nucleotide sequences from uisw_090 were used to identify homologs in the Ocean Gene Atlas OMRGC v2 metaG data set for all 17 island genes (Table S4). The Spearman's correlation (lat. corr.) and *P* value (p-val.) between abundance and latitude are displayed to the right for each gene. Enzyme names are abbreviated as follows: HADH, 3-hydroxyisobutyrate dehydrogenase; ALDH, aldehyde dehydrogenase; LGL, lactoylglutathione lyase; GPDP, glycerophosphoryl diester phosphodiesterase; BBOX(1), gamma-butyrobetaine dioxygenase; ABC *trans*. Sub, ABC transporter substrate binding; ABC *trans*. Perm., ABC transporter permease; ABC *trans*. ATP, ABC transporter ATP binding; ADH, alcohol dehydrogenase; CDH, carnitine dehydrogenase; BKACE, beta-keto acid cleavage enzyme; CIIA, Class II aldolase; ACOT, acyl-CoA thioesterase; SMR, small multidrug resistance transporter; BBOX(2), gamma-butyrobetaine dioxygenase; CHDH, choline dehydrogenase; DMT trans., drug/metabolite transporter.

populations from Arctic and Atlantic waters (19, 20) or due to enhanced genetic diversity within Arctic populations, as a previous study found that microbial diversity increased with latitude and decreased with temperature (53). Regardless, our analyses suggest that substantial sequencing effort is needed to estimate the gene content and total number of SAR11 species and subspecies in a SAR11 population.

A translocation of the 5S rRNA gene in SAR11 uisw_092 was used to define a large (~880 kbp) genomic region with a higher frequency of inversions and gene indels relative to regions flanking the origin of replication, suggesting that there is more selective pressure to conserve gene order near the origin of replication (Fig. 4A). We also identified a genomic region that encodes genes for the uptake and production of glycine betaine (Fig. 4; Table S4). Evidence that phylogenies for genes in this region are congruent with each other and incongruent with the phylogeny of single-copy core genes suggests that this region is a genomic island acquired by horizontal gene transfer (HGT) (Fig. 4C; Fig. S7). Although the insertion site appears to be conserved, there are no obvious genes that facilitate HGT in the region. However, homologous recombination is widespread in SAR11 (46) and has been proposed as a mechanism for the transfer of genomic islands (47). It has also been proposed as a mechanism for genomic diversification in other marine bacteria, including HIMB59 (a sister clade to SAR11) and OM43 (54, 55). In the case of HIMB59, more phosphorus-related genes were identified on a genomic island under phosphorus-limiting conditions (54). Similarly, the genomic island encoding genes for the uptake of glycine betaine in Arctic SAR11 could enhance the survival and metabolic activities of cells experiencing low temperatures and

high salinities during freezing, as reported for other bacteria (21–23). The island may also serve as an important methyl donor and source of glycine (24). Our evidence that genes in the Arctic SAR11 genomic island are more abundant at polar latitudes and less abundant or absent in temperate regions suggests that it is an adaptation to Arctic conditions, maintained in a large fraction of the population (50%) by HGT (Fig. 5). A similar polar distribution pattern has been observed in metagenomic read recruitment for SAR11 clade la.1 (56), suggesting that this could give SAR11 clade la.1 populations an advantage during the annual formation and melting of sea ice.

Variable patterns of diversity in abundant Arctic populations

Time-series ASV data from the Fram Strait suggest that the most frequently cultured bacteria in our culture collection (SAR11, SUP05, and *Njordibacter*) are among the most abundant Arctic taxa (Fig. 2B). There are, however, notable differences in the diversity of ASVs between these lineages and their matches to genomes in our culture collection (Fig. S5). Many unique species of SAR11 and *Njordibacter* match the same ASV, and in some cases, unique ASVs match the same species or different copies of the 16S rRNA gene in the same genome. These observations underline that ASVs cannot delineate patterns of bacterial speciation (57). However, closed genomes can be used to significantly improve estimates of diversity that are based on ASV analyses. Evidence that the Arctic SUP05 population is comprised of subspecies with similar genomes with relatively high ANI values also raises questions about how genomic diversity is maintained, particularly when compared to SAR11. Higher frequencies of genomic rearrangement have been reported as a mechanism for diversification in marine nitrogen fixers (58). A similar mechanism may help maintain diversity in SUP05, although more closed genomes are needed to quantify differences in the frequencies of gene gain, loss, rearrangement, and recombination in both SUP05 and SAR11.
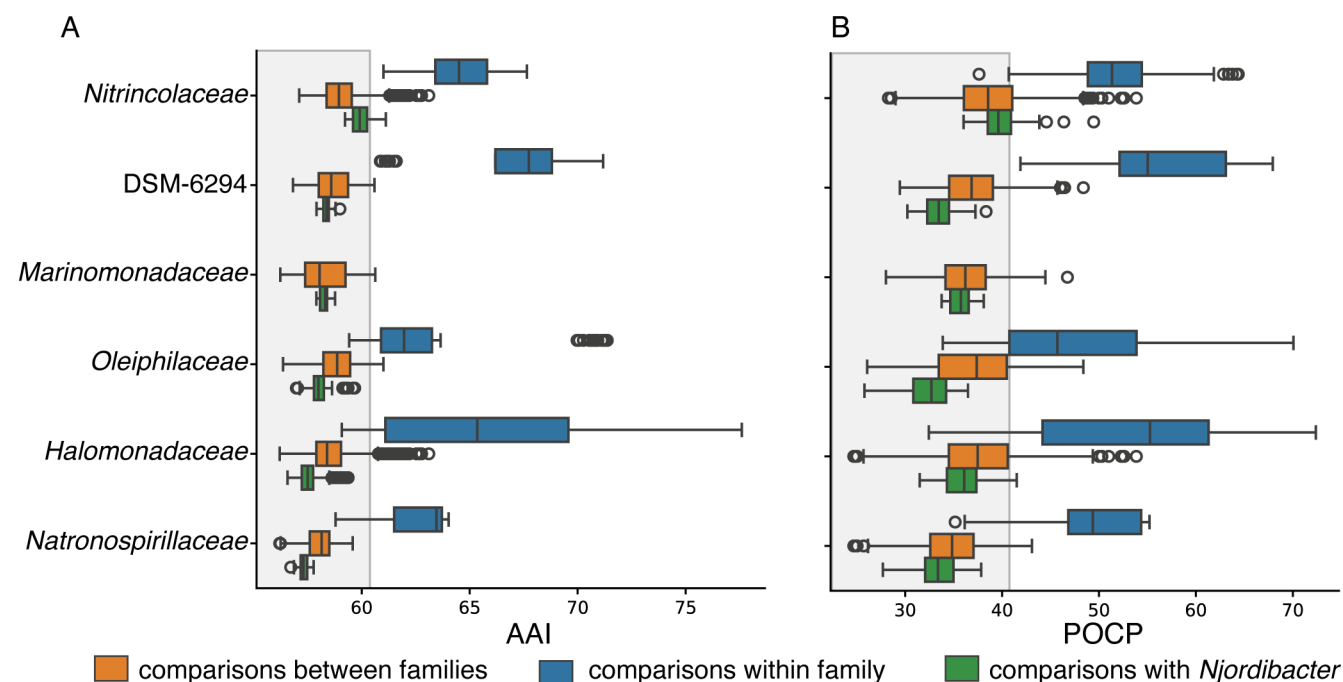
FIG 6 Whole-genome-based family and genus delineations in the order *Pseudomonadales*. (A) All pairwise average AAI comparisons. (B) All pairwise POCP comparisons. Representatives from the six most closely related families to *Njordibacter* were analyzed (Table S6). The shaded box indicates the family-level delineation in the order *Pseudomonadales*. Circles represent values outside the 1.5 interquartile range from the first and third quartiles.

Sulfur metabolism in high and low GC SAR116

The SAR116 clade consists of high GC (50% ± 7%) and low GC (31% ± 1%) lineages (48). Both clades are globally distributed, but metagenomic and single-cell analyses have suggested distinct differences in their ability to metabolize dimethylsulfoniopropionate (DMSP) (48). Specifically, high GC lineages encode DMSP lyases (*dddL* and *dddP*), which produce a volatile compound resulting in a loss of sulfur to the atmosphere as dimethyl sulfide (DMS). In contrast, low GC lineages encode DMSP demethylases (*dmdA*), which retains sulfur as 3-(methylsulfanyl)propanoate. Interestingly, the high GC SAR116 we cultured from the Arctic, *Ponderosibacter*, encodes *dmdA*, indicating that some high GC members also demethylate DMSP. This highlights the importance of closed genome sequences to verify the presence, and more importantly, the absence of key metabolic functions.

Conclusion

Early efforts to assemble genomes of marine bacteria directly from seawater produced only partial sequences for the most abundant lineages (59). While advances in sequencing have improved our ability to identify genetic variation in nature, the extent of genomic diversity within bacterial populations remains elusive, particularly for abundant lineages such as SAR11 (10, 14). Our high-throughput cultivation-based genomics approach produced 34 closed bacterial genomes, including several for the most abundant lineages in the Arctic (Fig. 2). These cultures and genomes provide insights into the diversity and adaptation of Arctic populations, as well as high-quality reference sequences that will enable population genomic analyses when similar high-quality sequences become available for populations in other oceans.

MATERIALS AND METHODS

Sample collection

Seawater samples were collected in the Arctic Ocean aboard the RV Kronprins Haakon from 18 to 21 May 2023, through an auger hole bored through ~2 m thick sea ice. Samples for cultivation (50 mL) were collected at 81.04° N, 10.62° E, from 25 m below the ice/water interface with a temperature of −1.8°C, salinity of 34.2 PSU, and 0.28 relative flourescence units, using a 2 L Hydro-Bios water sampler (Hydro-Bios, Altenholz, Germany). A 1 mL seawater sample was amended to 10% (vol/vol) glycerol, flash frozen in liquid nitrogen, and stored at −80°C until high-throughput dilution-to-extinction cultivation. Seawater for culture media (10 L) was collected at 80.96° N, 9.66° E from 1 m below the ice/water interface into acid-washed and Milli-Q rinsed cubitainers. Seawater was then filter-sterilized using a tangential-flow-filtration (TFF) system equipped with a 30 kDa Pellicon XL Polyethersulfone Biomax filter (MilliporeSigma, Burlington, MA). The resulting media was collected in 1 L acid-washed and autoclaved polycarbonate bottles, incubated for 2 months at 4°C, and checked for bacterial growth with a flow cytometer to ensure sterility prior to use.

Bacteria cultivation

Cultures were obtained by high-throughput dilution-to-extinction cultivation using cryopreserved Arctic seawater as previously described (60). Briefly, 0.059 mL of cryopreserved seawater with a cell density of 2.4×10^5 cells per mL was diluted to ~33 cells per mL in 30 kDa filter-sterilized seawater media. A 1.5 mL aliquot was then added to each well of three acid-washed and sterile 96-well Teflon plates, incubated at 4°C, and monitored for growth once a week for 13 weeks using a Guava easyCyte flow cytometer (Cytek, Fremont, CA). Wells that were positive for growth ($>2 \times 10^4$ cells per mL) were subjected to 16S rRNA sequence analysis when cell densities reached $>10^5$ cells per mL and preserved in 10% (vol/vol) glycerol that was frozen and stored at −80°C.

Culture identification

DNA for PCR was extracted from 100 µL of cell culture and sequenced using a physical lysis procedure as previously described (61). Briefly, potassium hydroxide–dithiothreitol was added, and samples were subjected to one freeze-thaw cycle. The pH was then adjusted to 8.0 using Tris-HCl. DNA was purified using 2× (vol/vol) DNA mag beads (Sergi Lab Supplies, Seattle, WA) with two 80% ethanol washes and eluted in 20 µL 10 mM Tris-HCl. 16S rRNA gene fragments were then amplified using PCR with primers 27F_B (5′ AGRGTTYGATYMTGGCTCAG 3′) and 926R_B (5′ CCGYCAATTCMTTTRAGTTT 3′), and in some cases using a second semi-nested PCR reaction with primers 519F (5′ CAGCMGCCGCGGTAATWC 3′) and 926R_B. The following PCR conditions were used throughout: 94°C 120 s, 38 cycles of amplificaiton (94°C 20 s, 55°C 45 s, 72°C 120 s). PCR products were sequenced by Genewiz (Genewiz, Seattle, WA). Cultures were putatively identified by aligning sequences in the Silva database v138.1 (62, 63).

Genome sequencing

Cultures selected for whole-genome sequencing were revived from freezer stocks in 1 L acid-washed and autoclaved polycarbonate bottles containing TFF sterilized Puget Sound seawater media. Cells were collected on 47 mm 0.2 µm pore size Isopore membrane filters (MilliporeSigma, Burlington, MA) when cultures reached maximum cell densities (between 10^5 and 10^6 cells per mL). High molecular weight DNA was extracted using the AutoGen QuickGene DNA Tissue Kit (Autogen, Holliston, MA) following the extraction protocol for animal tissue with minor modifications as noted below. Filters were cut into small pieces using sterile forceps and scissors and placed in sterile 2 mL DNA LoBind tubes (Eppendorf, Hamburg, Germany) containing 200 µL of TE buffer. Filters were then frozen at −80°C for 20 minutes and heated until thawed at 95°C. All recommended extraction volumes were doubled, and DNA was eluted in 200 µL of molecular-grade water. DNA was cleaned using 1× (vol/vol) DNA magnetic beads (Sergi Lab Supplies, Seattle, WA) with two 80% ethanol washes, then eluted in 20 µL of molecular-grade water. DNA was quantified using a Qubit dsDNA HS kit (Invitrogen, Waltham, MA) and sequenced using the ONT R10.4.1 Flongle flow cells with the SQK-RAD114 rapid library prep kit (Oxford Nanopore Technologies, Oxford, United Kingdom). Bases were called with Dorado v4.2.0 (github.com/nanoporetech/dorado), using the dna_r10.4.1_e8.2_400bps_hac@v4.2.0 model.

Genome assembly and annotation

Bacterial genomes were assembled with Flye v2.9.1 (64) and polished with Medaka v1.7.2 (github.com/nanoporetech/medaka) using the UseGalaxy web platform (65). Genome annotation was performed by NCBI using the Prokaryotic Genome Annotation Pipeline v6.8 (66, 67). The quality of ONT genomes was evaluated by comparing ONT-only genomes constructed with varying levels of coverage (9–500×) to ONT-Illumina hybrid genome assemblies obtained for two previously sequenced strains, the SAR11 strain NP1 (68) and the SUP05 strain EF3 (69). Coverage for ONT-only genomes was varied by subsampling ONT reads (accessions: SRX26378910 and SRX22361185) with Rasusa v2.0.0 (70). Hybrid genomes were created by polishing ONT genomes with Illumina reads (accessions: SRX26378911 and SRX23025519) using BWA-mem2 v2.2.1 (71) and Pilon v1.2.0 (72). The quality of ONT-only genomes was determined by identifying the number of mismatches, indels, and excess CDSs relative to hybrid assemblies using Quast v5.2.0 (73) (Fig. S7). Only closed genomes with greater than 10× coverage, corresponding to >99.9% accuracy, were used for further analyses.

Spatial and temporal abundance

16S rRNA sequences of cultures were matched against 5,511 ASVs derived from the HAUSGARTEN/FRAM Observatory in Fram Strait. ASVs originated from year-round autonomous sampling in the polar-influenced East Greenland Current (49) and the

Atlantic-influenced West Spitsbergen Current (50) between 2016 and 2020, with approximately biweekly resolution. Using Geneious Prime v2023.2.1 and R v4.2.2, we investigated their abundances and environmental correlations in Fram Strait over time to establish a broader spatiotemporal and ecological context. The latitudinal distributions of genes within the Arctic SAR11 genomic island were determined by identifying homologs for each gene in the Ocean Gene Atlas OMRGC v2 metaG data set. Representative nucleotide sequences from uisw_090 were used in a BLAST search of the database with an expect threshold of E-10 or less. The abundance of each homolog was normalized to the percent of mapped reads, as previously described (74, 75). Homolog abundance for each gene at each station were summed. Spearman's correlations were calculated for latitudinal distance from the equator using the "spearmanr" function in scipy v1.14.0 (76), and corrected for multiple hypothesis tests using the Benjamini/Hochberg method from statsmodels v0.14.2 (77).

Phylogenomics and population genomics

Taxonomic classifications were assigned to sequenced genomes using the GTDB-Tk v2.3.2, with the *de novo* workflow and reference database v214 (78). New species and strains were given the prefix "uisw" followed by the cultivation number. Most genomes came from pure cultures (n = 27). The suffix "_01" or "_02" was added to the cultivation number if two complete genomes were recovered from a mixed culture. The genomes of these cultures were analyzed using average AAI with ezaai v1.2.3 (79), POCP with POCP v2.3.2 (80), ANI with pyANI v0.2.12 (81), and RED with PhyloRank v0.1.12 (github.com/dparks1134/PhyloRank), using the GTDB v214 database. Genome structure was evaluated through the visualization of linear co-similarity blocks using progressive-Mauve v2015-02-25 (82). All phylogenetic trees were constructed with MUSCLE v3.8.31 (83) and RAxML v8.2.11 with model GTRGAMMA (84) using the ETE3 v3.1.3 phylogenetic analysis pipeline (85). Whole-genome phylogenies were constructed using the bacterial_71 single-copy core gene collection in anvi'o v8 (86). SAR11 gene cluster data were created using the anvi'o pangenomic workflow with the flags –use-ncbi-blast, – minbit 0.5, and –mcl-inflation 10, as previously described (87).

Protologue

Candidatus Njordibacter gen. nov.

Njor.di.bac'ter. N.L. masc. n. *bacter*, rod; N.L. masc. n. *Njordibacter*, a rod named after *Njord,* the Norse god of wind and seas.

Candidatus Njordibacteraceae fam. nov.

Njor.di.bac.te.ra.ce'ae. N.L. masc. n. *Njordibacter*, type genus of the family; suff. -aceae, ending to denote a family; N.L. fem. pl. n. *Njordibacteraceae*, the family of the genus *Njordibacter*. The description of the family *Njordibacteraceae* is the same as for the genus *Njordibacter*.

Candidatus Levibacter gen. nov.

Le.vi.bac'ter. L. masc. adj. *levis*, light in weight; N.L. masc. n *bacter*, rod; N.L. masc. n. *Levibacter,* a light rod alluding to the low GC clade of the order *Puniceispirillales*.

Candidatus Levibacteraceae fam. nov.

Le.vi.bac.ter.a.ce'ae. N.L. masc. n. *Levibacter*, type genus of the family; suff. -aceae, ending to denote a family; N.L. fem. pl. n. *Levibacteraceae*, the family of the genus *Levibacter*. The description of the family *Levibacteraceae* is the same as for the genus *Levibacter*.

Candidatus Ponderosibacter gen. nov.

Pon.de.ro.si.bac'ter L. masc. adj. *ponderosus,* heavy, weighty; N.L. masc. n. *bacter*, rod; N.L. masc. n. *Ponderosibacter*, a heavy rod alluding to the high GC lineage of the order *Puniceispirillales*.

Candidatus Marifrigoribacter gen. nov.

ma.ri.fri.go.ri.bac'ter L. neut. n. *mare,* the sea; L. neut. N. *frigor,* cold; N.L. masc. n. *bacter*, rod; N.L. masc. n. *Marifrigoribacter*, a rod from the cold of the sea.

AUTHOR AFFILIATIONS

[1]School of Oceanography, University of Washington, Seattle, USA
[2]Deep-Sea Ecology and Technology, Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany
[3]Max Planck Institute for Marine Microbiology, Bremen, Germany
[4]Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Oldenburg, Germany
[5]Laboratory of Microbiology, Faculty of Fisheries Science, Hokkaido University, Hakodate, Japan

AUTHOR ORCIDs

Michael C. Sadler http://orcid.org/0000-0002-1818-8473
Matthias Wietz http://orcid.org/0000-0002-9786-3026
Sayaka Mino http://orcid.org/0000-0003-2151-9795
Robert M. Morris http://orcid.org/0000-0001-9030-3790

AUTHOR CONTRIBUTIONS

Michael C. Sadler, Conceptualization, Formal analysis, Funding acquisition, Methodology, Visualization, Writing – original draft, Writing – review and editing | Matthias Wietz, Formal analysis, Visualization, Writing – review and editing | Sayaka Mino, Methodology, Writing – review and editing | Robert M. Morris, Conceptualization, Formal analysis,

Funding acquisition, Project administration, Supervision, Visualization, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

All data are publicly available at the NCBI under Bioproject PRJNA1129510. Individual genome accession and sequence read archive accession numbers are available in Table S2.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Fig. S1 (mBio01555-25-S0001.eps). SAR116 phylogeny.
Fig. S2 (mBio01555-25-S0002.eps). Spatiotemporal ASV distribution.
Fig. S3 (mBio01555-25-S0003.eps). ASV correlation.
Fig. S4 (mBio01555-25-S0004.eps). ASV abundance by lineage.
Fig. S5 (mBio01555-25-S0005.eps). ASV abundance by genome match.
Fig. S6 (mBio01555-25-S0006.eps). Arctic SAR11 ITS phylogeny.
Fig. S7 (mBio01555-25-S0007.eps). Genomic island gene cluster phylogeny.
Fig. S8 (mBio01555-25-S0008.eps). Sequencing accuracy.
Supplemental Tables (mBio01555-25-S0009.xlsx). Tables S1 to S6.

REFERENCES

1. Rappé MS, Giovannoni SJ. 2003. The uncultured microbial majority. Annu Rev Microbiol 57:369–394. https://doi.org/10.1146/annurev.micro.57.030502.090759
2. Lloyd KG, Steen AD, Ladau J, Yin J, Crosby L. 2018. Phylogenetically novel uncultured microbial cells dominate earth microbiomes. mSystems 3:e00055-18. https://doi.org/10.1128/mSystems.00055-18
3. Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, Giovannoni SJ. 2002. SAR11 clade dominates ocean surface bacterioplankton communities. Nature 420:806–810. https://doi.org/10.1038/nature01240
4. Oh HM, Kang I, Lee K, Jang Y, Lim SI, Cho JC. 2011. Complete genome sequence of strain IMCC9063, belonging to SAR11 subgroup 3, isolated from the Arctic Ocean. J Bacteriol 193:3379–3380. https://doi.org/10.1128/JB.05033-11
5. Oh HM, Kwon KK, Kang I, Kang SG, Lee JH, Kim SJ, Cho JC. 2010. Complete genome sequence of "*Candidatus* Puniceispirillum marinum" IMCC1322, a representative of the SAR116 clade in the Alphaproteobacteria. J Bacteriol 192:3240–3241. https://doi.org/10.1128/JB.00347-10
6. Marshall KT, Morris RM. 2015. Genome sequence of "*Candidatus* Thioglobus singularis" strain PS1, a mixotroph from the SUP05 clade of marine Gammaproteobacteria. Genome Announc 3:e01155-15. https://doi.org/10.1128/genomeA.01155-15
7. Shah V, Chang BX, Morris RM. 2017. Cultivation of a chemoautotroph from the SUP05 clade of marine bacteria that produces nitrite and consumes ammonium. ISME J 11:263–271. https://doi.org/10.1038/ismej.2016.87
8. Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV. 2012. Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. Science 335:587–590. https://doi.org/10.1126/science.1212665
9. Walsh DA, Zaikova E, Howes CG, Song YC, Wright JJ, Tringe SG, Tortell PD, Hallam SJ. 2009. Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. Science 326:578–582. https://doi.org/10.1126/science.1175309
10. Tully BJ, Graham ED, Heidelberg JF. 2018. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. Sci Data 5:170203. https://doi.org/10.1038/sdata.2017.203
11. Stepanauskas R, Sieracki ME. 2007. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. Proc Natl Acad Sci USA 104:9052–9057. https://doi.org/10.1073/pnas.0700496104
12. Stepanauskas R, Fergusson EA, Brown J, Poulton NJ, Tupper B, Labonté JM, Becraft ED, Brown JM, Pachiadaki MG, Povilaitis T, Thompson BP, Mascena CJ, Bellows WK, Lubys A. 2017. Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. Nat Commun 8:84. https://doi.org/10.1038/s41467-017-00128-z
13. Landry Z, Swan BK, Herndl GJ, Stepanauskas R, Giovannoni SJ. 2017. SAR202 genomes from the dark ocean predict pathways for the oxidation of recalcitrant dissolved organic matter. mBio 8:e00413-17. https://doi.org/10.1128/mBio.00413-17
14. Chang T, Gavelis GS, Brown JM, Stepanauskas R. 2024. Genomic representativeness and chimerism in large collections of single-amplified and metagenome-assembled genomes of marine prokaryoplankton. Microbiome 12:126. https://doi.org/10.1186/s40168-024-01848-3
15. Lim Y, Seo JH, Giovannoni SJ, Kang I, Cho JC. 2023. Cultivation of marine bacteria of the SAR202 clade. Nat Commun 14:5098. https://doi.org/10.1038/s41467-023-40726-8
16. Loman NJ, Quick J, Simpson JT. 2015. A complete bacterial genome assembled *de novo* using only nanopore sequencing data. Nat Methods 12:733–735. https://doi.org/10.1038/nmeth.3444
17. Wick RR, Judd LM, Holt KE. 2023. Assembling the perfect bacterial genome using Oxford Nanopore and illumina sequencing. Plos Comput Biol 19:e1010905. https://doi.org/10.1371/journal.pcbi.1010905
18. Serreze MC, Barrett AP, Slater AG, Woodgate RA, Aagaard K, Lammers RB, Steele M, Moritz R, Meredith M, Lee CM. 2006. The large-scale freshwater cycle of the Arctic. J Geophys Res Oceans 111:C11010. https://doi.org/10.1029/2005JC003424
19. Kwok R. 2018. Arctic sea ice thickness, volume, and multiyear ice coverage: losses and coupled variability (1958–2018). Environ Res Lett 13:105005. https://doi.org/10.1088/1748-9326/aae3ec
20. Timmermans M, Marshall J. 2020. Understanding Arctic Ocean circulation: a review of ocean dynamics in a changing climate. JGR Oceans 125:e2018JC014378. https://doi.org/10.1029/2018JC014378
21. Raymond-Bouchard I, Goordial J, Zolotarov Y, Ronholm J, Stromvik M, Bakermans C, Whyte LG. 2018. Conserved genomic and amino acid traits of cold adaptation in subzero-growing Arctic permafrost bacteria. FEMS Microbiol Ecol 94:fiy023. https://doi.org/10.1093/femsec/fiy023
22. Torstensson A, Young JN, Carlson LT, Ingalls AE, Deming JW. 2019. Use of exogenous glycine betaine and its precursor choline as osmoprotectants in Antarctic sea-ice diatoms. J Phycol 55:663–675. https://doi.org/10.1111/jpy.12839
23. Ewert M, Deming JW. 2014. Bacterial responses to fluctuations and extremes in temperature and brine salinity at the surface of Arctic winter

sea ice. FEMS Microbiol Ecol 89:476–489. https://doi.org/10.1111/1574-6941.12363

24. Boysen AK, Durham BP, Kumler W, Key RS, Heal KR, Carlson LT, Groussman RD, Armbrust EV, Ingalls AE. 2022. Glycine betaine uptake and metabolism in marine microbial communities. Environ Microbiol 24:2380–2403. https://doi.org/10.1111/1462-2920.16020

25. Deming JW, Somers LK, Straube WL, Swartz DG, Macdonell MT. 1988. Isolation of an obligately barophilic bacterium and description of a new genus, *Colwellia* gen. nov. Syst Appl Microbiol 10:152–160. https://doi.org/10.1016/S0723-2020(88)80030-4

26. Gosink JJ, Woese CR, Staley JT. 1998. *Polaribacter* gen. nov., with three new species, *P. irgensii* sp. nov., *P. franzmannii* sp. nov. and *P. filamentus* sp. nov., gas vacuolate polar marine bacteria of the Cytophaga-Flavobacterium-Bacteroides group and reclassification of "*Flectobacillus glomeratus*" as *Polaribacter glomeratus* comb. nov. Int J Syst Bacteriol 48:223–235. https://doi.org/10.1099/00207713-48-1-223

27. Brinkmeyer R, Knittel K, Jürgens J, Weyland H, Amann R, Helmke E. 2003. Diversity and structure of bacterial communities in Arctic versus Antarctic pack ice. Appl Environ Microbiol 69:6610–6619. https://doi.org/10.1128/AEM.69.11.6610-6619.2003

28. Zeng Y, Zou Y, Grebmeier JM, He J, Zheng T. 2012. Culture-independent and -dependent methods to investigate the diversity of planktonic bacteria in the northern Bering Sea. Polar Biol 35:117–129. https://doi.org/10.1007/s00300-011-1044-8

29. Rantanen M, Karpechko AY, Lipponen A, Nordling K, Hyvärinen O, Ruosteenoja K, Vihma T, Laaksonen A. 2022. The Arctic has warmed nearly four times faster than the globe since 1979. Commun Earth Environ 3:168. https://doi.org/10.1038/s43247-022-00498-3

30. Stroeve J, Notz D. 2018. Changing state of Arctic sea ice across all seasons. Environ Res Lett 13:103001. https://doi.org/10.1088/1748-9326/aade56

31. Polyakov IV, Pnyushkov AV, Alkire MB, Ashik IM, Baumann TM, Carmack EC, Goszczko I, Guthrie J, Ivanov VV, Kanzow T, Krishfield R, Kwok R, Sundfjord A, Morison J, Rember R, Yulin A. 2017. Greater role for Atlantic inflows on sea-ice loss in the Eurasian Basin of the Arctic Ocean. Science 356:285–291. https://doi.org/10.1126/science.aai8204

32. Jackson VLN, Grevesse T, Kilias ES, Onda DFL, Young KF, Allen MJ, Walsh DA, Lovejoy C, Monier A. 2024. Vulnerability of Arctic Ocean microbial eukaryotes to sea ice loss. Sci Rep 14:28896. https://doi.org/10.1038/s41598-024-77821-9

33. Comeau AM, Li WKW, Tremblay JÉ, Carmack EC, Lovejoy C. 2011. Arctic Ocean microbial community structure before and after the 2007 record sea ice minimum. PLOS One 6:e27492. https://doi.org/10.1371/journal.pone.0027492

34. Giovannoni SJ. 2017. SAR11 Bacteria: the most abundant plankton in the oceans. Annu Rev Mar Sci 9:231–255. https://doi.org/10.1146/annurev-marine-010814-015934

35. Bano N, Hollibaugh JT. 2002. Phylogenetic composition of bacterio-plankton assemblages from the Arctic Ocean. Appl Environ Microbiol 68:505–518. https://doi.org/10.1128/AEM.68.2.505-518.2002

36. de Sousa AGG, Tomasino MP, Duarte P, Fernández-Méndez M, Assmy P, Ribeiro H, Surkont J, Leite RB, Pereira-Leal JB, Torgo L, Magalhães C. 2019. Diversity and composition of pelagic prokaryotic and protist communi-ties in a thin Arctic sea-ice regime. Microb Ecol 78:388–408. https://doi.org/10.1007/s00248-018-01314-2

37. Collins RE, Rocap G, Deming JW. 2010. Persistence of bacterial and archaeal communities in sea ice through an Arctic winter. Environ Microbiol 12:1828–1841. https://doi.org/10.1111/j.1462-2920.2010.02179.x

38. Carlson CA, Morris R, Parsons R, Treusch AH, Giovannoni SJ, Vergin K. 2009. Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. ISME J 3:283–295. https://doi.org/10.1038/ismej.2008.117

39. Grote J, Thrash JC, Huggett MJ, Landry ZC, Carini P, Giovannoni SJ, Rappé MS. 2012. Streamlining and core genome conservation among highly divergent members of the SAR11 clade. mBio 3:e00252-12. https://doi.org/10.1128/mBio.00252-12

40. Haro-Moreno JM, Rodriguez-Valera F, Rosselli R, Martinez-Hernandez F, Roda-Garcia JJ, Gomez ML, Fornas O, Martinez-Garcia M, López-Pérez M. 2020. Ecogenomics of the SAR11 clade. Environ Microbiol 22:1748–1763. https://doi.org/10.1111/1462-2920.14896

41. Thrash JC, Temperton B, Swan BK, Landry ZC, Woyke T, DeLong EF, Stepanauskas R, Giovannoni SJ. 2014. Single-cell enabled comparative

genomics of a deep ocean SAR11 bathytype. ISME J 8:1440–1451. https://doi.org/10.1038/ismej.2013.243

42. Vergin KL, Beszteri B, Monier A, Thrash JC, Temperton B, Treusch AH, Kilpert F, Worden AZ, Giovannoni SJ. 2013. High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time-series study site by phylogenetic placement of pyrosequences. ISME J 7:1322–1332. https://doi.org/10.1038/ismej.2013.32

43. Kraemer S, Ramachandran A, Colatriano D, Lovejoy C, Walsh DA. 2020. Diversity and biogeography of SAR11 bacteria from the Arctic Ocean. ISME J 14:79–90. https://doi.org/10.1038/s41396-019-0499-4

44. Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, Bibbs L, Eads J, Richardson TH, Noordewier M, Rappé MS, Short JM, Carrington JC, Mathur EJ. 2005. Genome streamlining in a cosmopolitan oceanic bacterium. Science 309:1242–1245. https://doi.org/10.1126/science.1114057

45. Wilhelm LJ, Tripp HJ, Givan SA, Smith DP, Giovannoni SJ. 2007. Natural variation in SAR11 marine bacterioplankton genomes inferred from metagenomic data. Biol Direct 2:27. https://doi.org/10.1186/1745-6150-2-27

46. López-Pérez M, Haro-Moreno JM, Coutinho FH, Martinez-Garcia M, Rodriguez-Valera F. 2020. The evolutionary success of the marine bacterium SAR11 analyzed through a metagenomic perspective. mSystems 5:e00605-20. https://doi.org/10.1128/mSystems.00605-20

47. López-Pérez M, Martin-Cuadrado A-B, Rodriguez-Valera F. 2014. Homologous recombination is involved in the diversity of replacement flexible genomic islands in aquatic prokaryotes. Front Genet 5:147. https://doi.org/10.3389/fgene.2014.00147

48. Roda-Garcia JJ, Haro-Moreno JM, Huschet LA, Rodriguez-Valera F, López-Pérez M. 2021. Phylogenomics of SAR116 clade reveals two subclades with different evolutionary trajectories and an important role in the ocean sulfur cycle. mSystems 6:e0094421. https://doi.org/10.1128/mSystems.00944-21

49. Priest T, von Appen WJ, Oldenburg E, Popa O, Torres-Valdés S, Bienhold C, Metfies K, Boulton W, Mock T, Fuchs BM, Amann R, Boetius A, Wietz M. 2023. Atlantic water influx and sea-ice cover drive taxonomic and functional shifts in Arctic marine bacterial communities. ISME J 17:1612–1625. https://doi.org/10.1038/s41396-023-01461-6

50. Priest T, Oldenburg E, Popa O, Dede B, Metfies K, von Appen WJ, Torres-Valdés S, Bienhold C, Fuchs BM, Amann R, Boetius A, Wietz M. 2025. Seasonal recurrence and modular assembly of an Arctic pelagic marine microbiome. Nat Commun 16:1326. https://doi.org/10.1038/s41467-025-56203-3

51. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil PA, Hugenholtz P. 2021. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank-normalized and complete genome-based taxonomy. Nucleic Acids Res 50:D785–D794. https://doi.org/10.1093/nar/gkab776

52. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol 36:996–1004. https://doi.org/10.1038/nbt.4229

53. Fuhrman JA, Steele JA, Hewson I, Schwalbach MS, Brown MV, Green JL, Brown JH. 2008. A latitudinal diversity gradient in planktonic marine bacteria. Proc Natl Acad Sci USA 105:7774–7778. https://doi.org/10.1073/pnas.0803070105

54. Molina-Pardines C, Haro-Moreno JM, López-Pérez M. 2023. Phosphate-related genomic islands as drivers of environmental adaptation in the streamlined marine alphaproteobacterial HIMB59. mSystems 8:e0089823. https://doi.org/10.1128/msystems.00898-23

55. Layoun P, López-Pérez M, Haro-Moreno JM, Haber M, Thrash JC, Henson MW, Kavagutti VS, Ghai R, Salcher MM. 2024. Flexible genomic island conservation across freshwater and marine *Methylophilaceae*. ISME J 18:wrad036. https://doi.org/10.1093/ismejo/wrad036

56. Delmont TO, Kiefl E, Kilinc O, Esen OC, Uysal I, Rappé MS, Giovannoni S, Eren AM. 2019. Single-amino acid variants reveal evolutionary processes that shape the biogeography of a global SAR11 subclade. Elife 8:e46497. https://doi.org/10.7554/eLife.46497

57. Schloss PD. 2021. Amplicon sequence variants artificially split bacterial genomes into separate clusters. mSphere 6:e0019121. https://doi.org/10.1128/mSphere.00191-21

58. Zehr JP, Bench SR, Mondragon EA, McCarren J, DeLong EF. 2007. Low genomic diversity in tropical oceanic N2-fixing cyanobacteria. Proc Natl Acad Sci USA 104:17807–17812. https://doi.org/10.1073/pnas.0701017104

59. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH, Smith HO. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science 304:66–74. https://doi.org/10.1126/science.1093857

60. Monaghan EA, Freel KC, Rappé MS. 2020. Isolation of SAR11 marine bacteria from cryopreserved seawater. mSystems 5:e00954-20. https://doi.org/10.1128/mSystems.00954-20

61. Bramucci AR, Focardi A, Rinke C, Hugenholtz P, Tyson GW, Seymour JR, Raina J-B. 2021. Microvolume DNA extraction methods for microscale amplicon and metagenomic studies. ISME Communications 1:79. https://doi.org/10.1038/s43705-021-00079-z

62. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–6. https://doi.org/10.1093/nar/gks1219

63. Glöckner FO, Yilmaz P, Quast C, Gerken J, Beccati A, Ciuprina A, Bruns G, Yarza P, Peplies J, Westram R, Ludwig W. 2017. Twenty-five years of serving the community with ribosomal RNA gene reference databases and tools. J Biotechnol 261:169–176. https://doi.org/10.1016/j.jbiotec.2017.06.1198

64. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8

65. Abueg LAL, Afgan E, Allart O, Awan AH, Bacon WA, Baker D, Bassetti M, Batut B, Bernt M, Blankenberg D, et al. 2024. The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. Nucleic Acids Res 52:W83–W94. https://doi.org/10.1093/nar/gkae410

66. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569

67. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, Coulouris G, Chitsaz F, Derbyshire MK, Durkin AS, Gonzales NR, Gwadz M, Lanczycki CJ, Song JS, Thanki N, Wang J, Yamashita RA, Yang M, Zheng C, Marchler-Bauer A, Thibaud-Nissen F. 2020. RefSeq: expanding the prokaryotic genome annotation pipeline reach with protein-family model curation. Nucleic Acids Res 49:D1020–D1028. https://doi.org/10.1093/nar/gkaa1105

68. Morris RM, Cain KR, Hvorecny KL, Kollman JM. 2020. Lysogenic host-virus interactions in SAR11 marine bacteria. Nat Microbiol 5:1011–1015. https://doi.org/10.1038/s41564-020-0725-x

69. Morris RM, Mino S. 2024. Complete genome sequences of *Thioglobus autotrophicus* strains EF2 and EF3. Microbiol Resour Announc 13:e01118–23. https://doi.org/10.1128/mra.01118-23

70. Hall M. 2022. Rasusa: randomly subsample sequencing reads to a specified coverage. JOSS 7:3941. https://doi.org/10.21105/joss.03941

71. Vasimuddin M, Misra S, Li H, Aluru S. 2019. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS); Rio de Janeiro, Brazil: , p 314–324IEEE. https://doi.org/10.1109/IPDPS.2019.00041

72. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLOS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963

73. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics 34:i142–i150. https://doi.org/10.1093/bioinformatics/bty266

74. Villar E, Vannier T, Vernette C, Lescot M, Cuenca M, Alexandre A, Bachelerie P, Rosnet T, Pelletier E, Sunagawa S, Hingamp P. 2018. The Ocean Gene Atlas: exploring the biogeography of plankton genes online. Nucleic Acids Res 46:W289–W295. https://doi.org/10.1093/nar/gky376

75. Vernette C, Lecubin J, Sánchez P, Sunagawa S, Delmont TO, Acinas SG, Pelletier E, Hingamp P, Lescot M, Tara Oceans Coordinators. 2022. The Ocean Gene Atlas v2.0: online exploration of the biogeography and phylogeny of plankton genes. Nucleic Acids Res 50:W516–W526. https://doi.org/10.1093/nar/gkac420

76. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, et al. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 17:261–272. https://doi.org/10.1038/s41592-019-0686-2

77. Seabold S, Perktold J. 2010. Statsmodels: econometric and statistical modeling with python. In Proceedings of the 9th Python in Science Conference; Austin, Texas: , p 92–96. https://doi.org/10.25080/Majora-92bf1922-011

78. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2022. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics 38:5315–5316. https://doi.org/10.1093/bioinformatics/btac672

79. Kim D, Park S, Chun J. 2021. Introducing EzAAI: a pipeline for high throughput calculations of prokaryotic average amino acid identity. J Microbiol 59:476–480. https://doi.org/10.1007/s12275-021-1154-0

80. Hölzer M. 2024. POCP-nf: an automatic nextflow pipeline for calculating the percentage of conserved proteins in bacterial taxonomy. Bioinformatics 40:btae175. https://doi.org/10.1093/bioinformatics/btae175

81. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Anal Methods 8:12–24. https://doi.org/10.1039/C5AY02550H

82. Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLOS One 5:e11147. https://doi.org/10.1371/journal.pone.0011147

83. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113. https://doi.org/10.1186/1471-2105-5-113

84. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033

85. Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. Mol Biol Evol 33:1635–1638. https://doi.org/10.1093/molbev/msw046

86. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, et al. 2021. Community-led, integrated, reproducible multi-omics with anvi'o. Nat Microbiol 6:3–6. https://doi.org/10.1038/s41564-020-00834-3

87. Delmont TO, Eren AM. 2018. Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. PeerJ 6:e4320. https://doi.org/10.7717/peerj.4320