

1
2
3
4 **Supplementary Information for**
5 **New SAR11 isolate genomes and global marine metagenomes resolve**
6 **ecologically relevant units with the *Pelagibacterales***

7 Kelle C. Freel *et al.*

8
9 *Corresponding author email: rappe@hawaii.edu (Michael S. Rappé)

10
11
12
13
14
15 **This PDF file includes:**

16 Supplementary Methods
17 Supplementary Notes
18 Supplementary Fig. 1 to Supplementary Fig. 10
19 Supplementary References
20
21
22
23
24
25
26
27
28

Supplementary Methods

Genome sequencing of HIMB109 and HIMB123 at the Joint Genome Institute (JGI)

A Nextera XT library prep kit was used to process samples before sequencing on a NextSeq platform (Illumina). Raw data was processed with BBDuk (version 37.36, BBMap – Bushnell B. – sourceforge.net/projects/bbmap/) to remove contamination, trim reads with adapter sequences, and trim the ends of reads with quality scores below 6. This step also removed reads with 2 or more ‘N’s’, less than or equal to 49 base pairs in length, and with an average quality score across the read less than 10. Reads mapped with BBMap (BBMap – Bushnell B. – sourceforge.net/projects/bbmap/) to masked human, cat, dog, and mouse references at 93% identity were separated into a separate file. Reads that aligned to common microbial contaminants [JGI SOP 1077] were also split into a separate file.

16S rRNA gene amplicon sequence analysis

The 16S rRNA gene sequence analysis was previously described¹. Briefly, three Illumina MiSeq runs were imported into QIIME2 v2019.4.0 and demultiplexed, and paired ends were analyzed for sequence quality and merged². The DADA2 software package³ was used to denoise sequences, which included removing chimeras. Due to the low quality at the end of the sequences, 10 bases were truncated from the 3’ end of the reverse reads. Sequence reads from the three runs were then merged post-denoising. Amplicon sequence variant (ASV) identities were defined by DADA2 for all reads that varied by at least 1 bp. Taxonomy was assigned to each ASV using a Naïve Bayes classifier trained on the Silva rRNA v132 database⁴, clustered at 99% similarity, and subsequently modified manually based on phylogenetic analyses and the results of previous work.

The ASV identities were further evaluated in QIIME2, where any ASVs represented by less than 20 reads were removed from the dataset. The proportion of each ASV in an individual culture was determined by dividing the read count for that ASV by the total number of reads from the culture, post-curation. Cultures were identified as (1) monocultures (dominant member $\geq 90\%$ of reads), (2) mixed cultures (dominant member $\leq 90\%$ of reads and an additional member $\geq 5\%$ of reads), or (3) cultures with no dominant member (no ASV $\geq 50\%$ of reads). Cultures in the third category were not included in additional analyses.

Reference genomes

A literature search was used to identify publicly available isolate genomes and high-quality single amplified genomes (SAGs) (Supplementary Data 3). Although the majority of the included SAGs were derived from a single study⁵, the samples used for sequencing were globally distributed. The site from which each genome originated is included in Supplementary Data 3 for all genomes except for HTCC9022, for which no specific latitude or longitude was available. However, it was confirmed that it originated from ocean surface waters, offshore, over the Juan de Fuca Ridge in the Northeast Pacific Ocean. To include a diversity of SAGs and ensure genomes were of reliable quality, we used an 85% completion cutoff for inclusion in this study, except for cases where an established SAR11 lineage had limited or no genomes available that met these criteria. Additional SAGs were included from the Mediterranean Sea⁶, those representing the deep SAR11 Ic clade⁷, genomes from oxygen minimum zones⁸, and the Red Sea⁹. To ensure no high-quality SAGs were excluded from the analyses, we cross-referenced the ‘hq_mimag_genomes’ file for the release of the Genome Taxonomy Database (GTDB) from

April 23, 2024. This file is the complete list of isolate, MAG, and SAG genomes that meet the MIMAG high-quality genome criteria¹⁰ which includes completeness values greater than 90% and contamination less than 5% in addition to the presence of 5S, 16S, and 23S rRNA genes with the minimum lengths of 80 bp, 1200 bp, and 1900 bp, respectively for bacterial genomes. Additionally, these genomes all encode a minimum of 18 tRNAs. For all genomes classified in GTDB as “o__Pelagibacterales”, there were a total of 130 high-quality genomes, 11 of which were MAGs (including 6 from freshwater lake metagenomes), and 91 mislabeled as MAGs but instead are SAGs derived from the study of⁵.

The type material for ‘*Mesopelagibacter carboxydoxydans*’ and *Anoxypelagibacter denitrificans* are MAGs ETNP_2013_S10_SV74_125m_MAG_02 (GCA_017640275.1) and ENTP2013_S02_SV82_300m_MAG_01 (GCA_017640265.1), respectively¹¹. To maintain dataset consistency and avoid the population-level heterogeneity and contamination that accompanies SAR11 MAGs¹², we used high-quality single amplified genomes as representatives of these sublineages in our analyses. To verify the relationship between the designated type material MAGs and substitute SAGs, we constructed phylogenies with the SAR11_16S gene set from the Ic and II subclades, respectively. In both cases, 10 isolate genomes from the *Rhodobacteraceae* were included as an outgroup. To generate a concatenated alignment, the program ‘anvi-get-sequences-for-hmm-hits’ with the custom HMM source was implemented to extract and align genes of interest. For the alignments from each gene set, the program trimAL 1.3¹³ was then used to remove all positions missing in more than 50% of the genomes. Phylogenies were generated with IQ-Tree v2.1.2¹⁴ with the LG+F+R10 model and 1,000 ultrafast bootstraps.

Maps

The maps included in Figure 3d and Supplementary Figure 1a were made in R with the following packages: dplyr v1.1.4, stringr v1.5.2, ggplot2 v4.0.0, maps v.3.4.3, sf v1.0-22, marmap v1.0.12, mapdata v2.3.1, and ggspatial v1.1.10. Images were then saved and imported to Affinity Publisher v1.10.8 (Figure 3d) or Affinity Design v1.10.5 (Supplementary Figure 1a) for finalization. The map included in Supplementary Figure 1b was generated in ArcGIS and finalized in Affinity Design v1.10.5.

Supplementary Notes

Cultured isolates represent dominant and persistent SAR11 populations in Kāneʻohe Bay

We investigated the ecological relevance of the recently isolated strains by recruiting metagenomic reads from a three-year time series of coastal and offshore tropical marine plankton collected as a part of the Kāneʻohe Bay Time-series (Supplementary Fig.1; Supplementary Fig. 8), as well as public sequence data from microbial communities at open-ocean Station ALOHA in the adjacent North Pacific Subtropical Gyre (NPSG). Within the tropical Pacific, fifteen genomes from cultivated HIMB strains are among the limited number of *Pelagibacterales* genomes persistently detected nearshore (Supplementary Fig. 8). Six of these genomes were sequenced from cultures first grown in the laboratory in 2005¹⁵, and their detection in metagenomes from samples collected more than 10 years later (2017 through 2019) underscores the temporal persistence of cultivated populations of SAR11 in Kāneʻohe Bay. The dominant and

119 persistent environmental SAR11 populations are well represented by isolate genomes, and the
120 nearshore environment within Kāneʻohe Bay retains limited SAR11 diversity compared to the
121 adjacent open ocean.

122 **Relative evolutionary distances do not define meaningful units of SAR11 diversity**

123 We examined how the taxonomic classification system proposed here compares to the
124 taxonomic groupings of GTDB and those resulting from the common methodology of Relative
125 Evolutionary Distance (RED) (Supplementary Fig. 9, Supplementary Data 16). The three
126 classification systems (GTDB, RED values, and the one proposed here) reported non-congruent
127 groupings of SAR11 diversity. The GTDB classification system results in polyphyletic clusters
128 at the levels of family, genus, and species when overlaid on our phylogenomic tree. To assess the
129 RED approach, we compared the results of categorizing the genomes using RED value
130 thresholds that in other microbial lineages roughly correspond to family (0.77), genus (0.88), and
131 species levels (0.92) of classification. Using RED values, we found instances of both under- and
132 over-splitting in the family- and genus-level designations compared to our proposed taxonomic
133 framework. For example, the RED value of 0.77 leads to a clade encompassing all SAR11
134 subclade Ic, Ib, and Ia genomes into a single family, whereas we delineate two families through
135 our ecological and evolutionarily-guided analyses. Furthermore, implementing a RED value of
136 0.92 placed all Ia.3 and Ia.4 sublineages together even though they exhibit distinct
137 spatiotemporal patterns of distribution, contain well-supported branching patterns concomitant
138 with distinct genome properties such as genome size and GC content, experience variable
139 selective pressure¹⁶, and harbor variable gene content associated to distinct metabolic
140 characteristics¹⁶. While RED values can be crucial tools to reflect meaningful classification
141 schemes¹⁷, including for the enigmatic SAR86 clade¹⁸, this metric fails to distinguish meaningful
142 units of diversity within SAR11.

143 **Relationship between ‘*Mesopelagibacter carboxydoxydans*’ and *Anoxypelagibacter* 144 *denitrificans* type material and SAGs of the curated dataset**

145 To compare the SAGs included in our analyses with the MAGs designated as type
146 material for ‘*Mesopelagibacter carboxydoxydans*’ and *Anoxypelagibacter denitrificans*, we
147 generated phylogenies that included the type genomes, the related SAGs included in our dataset,
148 and the *Rhodobacteraceae* outgroup. The type material for ‘*Mesopelagibacter carboxydoxydans*’
149 (GCA_017640275.1) and *Anoxypelagibacter denitrificans* (GCA_017640265.1) are most closely
150 related to GCA_902582665.1 (88.6% gANI) and GCA_902578935.1 (79.7% gANI),
151 respectively. In both cases, the type material genomes originate from within the monophyletic
152 clades that include the SAG genomes used in our study (Supplementary Fig. 10).

153 **Other taxonomic considerations of note**

154 In order to validly publish our taxonomic framework for SAR11 through the SeqCode¹⁹,
155 some amendments to existing taxa were needed. First, the initially proposed species name for the
156 type material of the order, ‘*Pelagibacter ubique*’²⁰, has long been known to be invalid. While
157 other options have been proposed [e.g. ‘*communis*’²¹], we sought a species name that retained a
158 linguistic connection to the original and so amended the nomenclatural type from ‘*Pelagibacter*
159 *ubique*’ to *Pelagibacter ubiqueversans*.

160 The type material for ‘*Mesopelagibacter carboxydoxydans*’, MAG
161 ETNP_2013_S10_SV74_125m_MAG_02 (GCA_017640275.1), is of insufficient quality to be

used as type material within the SeqCode. We sought to retain the genus name as proposed by the authors of the effective publication¹¹ and so designated a new species, *Mesopelagibacter profundus*, that includes a high quality SAG (GCA_000504625.1) as type material. This allowed us to then propagate this genus name to the family-level lineage *Mesopelagibacteraceae* to encompass all of SAR11 subclade Ic.

Protologues

Description of *Mesopelagibacteraceae* fam. nov.

Mesopelagibacteraceae ([Me.so.pe.la.gi.bac.te.ra'ce.ae] N.L. masc. n. *Mesopelagibacter*, the type genus of the family; L. fem. pl. suff. *-aceae*, ending to denote a family; N.L. fem. pl. n. *Mesopelagibacteraceae*, the *Mesopelagibacter* family).

The family *Mesopelagibacteraceae* encompasses what has previously been characterized as the monophyletic sublineage Ic of SAR11 marine bacteria^{7,11,22}. The members of this family are free-living cells that predominantly reside in the mesopelagic region of the open ocean. The type genus is *Mesopelagibacter*.

Registry URL: <https://seqco.de/i:51441>

Description of *Cosmipelagibacteraceae* fam. nov.

Cosmipelagibacteraceae ([Cos.mi.pe.la.gi.bac.te.ra'ce.ae] N.L. masc. n. *Cosmipelagibacter*, the type genus of the family; L. fem. pl. suff. *-aceae*, ending to denote a family; N.L. fem. pl. n. *Cosmipelagibacteraceae*, the *Cosmipelagibacter* family).

The family *Cosmipelagibacteraceae* encompasses what has previously been described as SAR11 subclade II²³. The type genus *Cosmipelagibacter* contains the only strain currently isolated from this family, HIMB58^{15,24}. One other previously described genus (*Anoxypelagibacter*) is also included within the family *Cosmipelagibacteraceae*¹¹.

Registry URL: <https://seqco.de/i:51442>

Description of *Fontibacteriaceae* fam. nov.

Fontibacteriaceae ([Fon.ti.bac.te.ri.a'ce.ae] N.L. neut. n. *Fontibacterium*, the type genus of the family; L. fem. pl. suff. *-aceae*, ending to denote a family; N.L. fem. pl. n. *Fontibacteriaceae*, the *Fontibacterium* family).

The family *Fontibacteriaceae* encompasses previously described SAR11 subclades IIIa and IIIb, which have also been referred to as LD12 or subgroup IV²⁵. To note, subgroup IV of Morris et al., (2005)²⁵ does not correspond to subclade IV. The family currently circumscribes two genera that contain strains isolated from either freshwater environments (type genus *Fontibacterium*²⁶, or surface seawater (*Elongatipelagibacter*, newly described here).

Registry URL: <https://seqco.de/i:51443>

Emendation to the description of the family *Pelagibacteraceae*

We amend the description of the family *Pelagibacteraceae* to encompass historical subclades Ia and Ib of the SAR11 lineage. Previously, the family *Pelagibacteraceae* was described to encompass the entire SAR11 lineage²⁷. The type genus remains *Pelagibacter*.

Emendation to the description of the genus *Pelagibacter*

We amend the description of the genus *Pelagibacter* to include the type species, *P. ubiqueversans* str. HTCC1062^{T 20}, and its closely related monophyletic relatives within the sublineage previously recognized as SAR11 Ia.1.I. Previously, the genus *Pelagibacter* encompassed the entire SAR11 Ia sublineage²⁴.

Emendation to the description of the genus *Mesopelagibacter*

We amend the description of the genus *Mesopelagibacter* to include the type species, *Mesopelagibacter profundus*. The genome of the previously proposed type species, ‘*Mesopelagibacter carboxydoxydans*’, could not be validated because the type genome was of insufficient quality (MAG ETNP_2013_S10_SV74_125m_MAG_02; GCA_017640275.1)¹¹. The type genome of *Mesopelagibacter profundus* is the high quality SAG GCA_000504625.1.

Registry URL: <https://seqco.de/i:51491>

Description of *Xanthinipelagibacter* gen. nov.

Xanthinipelagibacter ([Xan.thi.ni.pe.la.gi.bac'ter] N.L. fem. n. *xanthina*, xanthine, a purine base; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n.

Xanthinipelagibacter, a rod shaped bacterium from the open sea that can metabolize xanthine).

Genomes of the genus *Xanthinipelagibacter* occupy a monophyletic node within the family *Pelagibacteraceae* that is currently known as SAR11 subclade Ia.3.V, and share a minimum gANI of ~85% to the type genome. Cells of this genus contain the genetic machinery for metabolizing xanthine¹⁶, and have been isolated from the tropical Pacific Ocean and Tannan Bay, China. The type species is *Xanthinipelagibacter aumoanae*.

Registry URL: <https://seqco.de/i:51392>

Description of *Undatipelagibacter* gen. nov.

Undatipelagibacter ([Un.da.ti.pe.la.gi.bac'ter] L. masc. perf. part. *undatus*, surged, risen in waves, flooded, welled up; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Undatipelagibacter*, a surging oceanic bacterium, referring both to the oceanic habitat and to the comparatively large genome size found in this lineage).

Genomes of the genus *Undatipelagibacter* occupy a monophyletic node within the family *Pelagibacteraceae* that is currently known as SAR11 subclade Ia.3.VI, and share a minimum gANI of ~88% to the type genome. This group harbors the largest mean genome size compared to other genera of the *Pelagibacteraceae* genera (~1.43 Mbp). Cells of this genus have been isolated from the tropical Pacific Ocean. The type species is *Undatipelagibacter huakaii*.

Registry URL: <https://seqco.de/i:51394>

Description of *Angustipelagibacter* gen. nov.

Angustipelagibacter ([An.gus.ti.pe.la.gi.bac'ter] L. neut. n. *angustum*, a narrow place; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Angustipelagibacter*, a rod from a narrow place in the sea, referring to the limited geographic locations from which most genomes originate).

Genomes of the genus *Angustipelagibacter* occupy a monophyletic node within the family *Pelagibacteraceae* that is currently known as SAR11 subclade Ia.3.VII, and share a minimum gANI of ~86% to the type genome. There are no known cultivated isolates reported for this genus; the majority of publicly available single amplified genomes (SAGs) have been

recovered from the Mediterranean Sea⁶ and Atlantic Ocean⁵. The type species is *Angustipelagibacter atlanticus*.

Registry URL: <https://seqco.de/i:51396>

Description of *Atlantikopelagibacter* gen. nov.

Atlantikopelagibacter ([At.lan.ti.ko.pe.la.gi.bac'ter] Gr. masc. adj. *atlantikos*, Atlantic; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Atlantikopelagibacter*, bacteria of the open sea from the Atlantic ocean).

Genomes of the genus *Atlantikopelagibacter* occupy a monophyletic node within the family *Pelagibacteraceae* that is currently known as SAR11 subclade Ia.3.I, and share a minimum gANI of ~92% to the type genome. Strains of this genus have been isolated from the site of the Bermuda Atlantic Time-series Study in the Sargasso Sea of the Atlantic Ocean. The type species is *Atlantikopelagibacter sargassoseensis*.

Registry URL: <https://seqco.de/i:49683>

Description of *Accedentipelagibacter* gen. nov.

Accedentipelagibacter ([Ac.ce.den.ti.pe.la.gi.bac'ter] L. masc. pres. part. *accedentus*, approaching; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Accedentipelagibacter*, the approaching oceanic bacterium, referring to the identification of this bacterium close to the coast in Newport, Oregon).

Genomes of the genus *Accedentipelagibacter* occupy a monophyletic node within the family *Pelagibacteraceae* that is currently known as SAR11 subclade Ia.3.IV, and share a minimum gANI of ~86% to the type genome. Cells of this genus have been isolated from the temperate northeast Pacific Ocean. The type species is *Accedentipelagibacter yaqonii*.

Registry URL: <https://seqco.de/i:51398>

Description of *Colligatipelagibacter* gen. nov.

Colligatipelagibacter ([Co.lli.ga.ti.pe.la.gi.bac'ter] L. masc. perf. part. *colligatus*, bound together; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Colligatipelagibacter*, a rod bound together across the open sea, referring to the cohesive distribution of this genus).

Genomes of the genus *Colligatipelagibacter* occupy a previously uncharacterized monophyletic node within the Ia.4 cluster of the family *Pelagibacteraceae*, and share a minimum gANI of ~91% to the type genome. There are no known cultivated isolates yet reported for this genus. The type species is *Colligatipelagibacter wilhelmii*.

Registry URL: <https://seqco.de/i:51400>

Description of *Tantillipelagibacter* gen. nov.

Tantillipelagibacter ([Tan.ti.li.pe.la.gi.bac'ter] L. masc. adj. *tantillus*, very small and little; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Tantillipelagibacter*, a very small oceanic bacterium, referring to the small genome size of members of this group compared to close relatives).

Genomes of the genus *Tantillipelagibacter* occupy a previously uncharacterized monophyletic node within the Ia.4 cluster of the family *Pelagibacteraceae*, and share a minimum

gANI of ~87% to the type genome. Cells of this genus have been isolated from the central Red Sea and tropical Pacific Ocean. The type species is *Tantillipelagibacter marisrubii*.
Registry URL: <https://seqco.de/i:51401>

Description of *Proventipelagibacter* gen. nov.

Proventipelagibacter ([Pro.ven.ti.pe.la.gi.bac'ter] L. masc. n. *proventus*, increase, yield, crop; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Proventipelagibacter*, a marine bacterium with increased yield, referring to its increase in detection in select samples).

Genomes of the genus *Proventipelagibacter* occupy a previously uncharacterized monophyletic node within the Ia.4 cluster of the family *Pelagibacteraceae*, and share a minimum gANI of ~88% to the type genome. There are no known cultivated isolates yet reported for this genus. The type species is *Proventipelagibacter oceanicus*.

Registry URL: <https://seqco.de/i:51403>

Description of *Punctipelagibacter* gen. nov.

Punctipelagibacter ([Punc.ti.pe.la.gi.bac'ter] L. neut. n. *punctum*, a point, a spot; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Punctipelagibacter*, a bacterium with spotty distribution globally).

Genomes of the genus *Punctipelagibacter* occupy a previously uncharacterized monophyletic node within the Ia.4 cluster of the family *Pelagibacteraceae*, and share a minimum gANI of ~93% to the type genome. There are no known cultivated isolates yet reported for this genus. The type species is *Punctipelagibacter variabilis*.

Registry URL: <https://seqco.de/i:51405>

Description of *Pantolocipelagibacter* gen. nov.

Pantolocipelagibacter ([Pan.to.lo.ci.pe.la.gi.bac'ter] Gr. masc. adj. *pas*, whole; L. masc. n. *locus*, a place; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Pantolocipelagibacter*, a bacterium found across multiple latitudes and locations).

Genomes of the genus *Pantolocipelagibacter* occupy a previously uncharacterized monophyletic node within the Ia.4 cluster of the family *Pelagibacteraceae*, and share a minimum gANI of ~94% to the type genome. There are no known cultivated isolates yet reported for this genus. The type species is *Pantolocipelagibacter undulae*.

Registry URL: <https://seqco.de/i:51407>

Description of *Alipelagibacter* gen. nov.

Alipelagibacter ([A.li.pe.la.gi.bac'ter] L. masc. pron. *alius*, other; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Alipelagibacter*, the other bacterium from the open sea, referring to the presence of unique (or "other") molybdenum enzymes not found in closely related species).

Genomes of the genus *Alipelagibacter* occupy a previously uncharacterized monophyletic node within the Ia.4 cluster of the family *Pelagibacteraceae*. Its members encode the genetic pathway for synthesizing the molybdenum cofactor, while close evolutionary neighbors do not¹⁶. This genus is currently represented by a single isolate genome from the tropical Pacific Ocean. The type species is *Alipelagibacter loemeaikii*.

Registry URL: <https://seqco.de/i:51409>

Description of *Lacunipelagibacter* gen. nov.

Lacunipelagibacter ([La.cu.ni.pe.la.gi.bac'ter] L. fem. n. *lacuna*, a hole, an empty space, deficiency, loss; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Lacunipelagibacter*, a bacterium missing enzymes that closely related groups harbor).

Genomes of the genus *Lacunipelagibacter* occupy a monophyletic node within the family *Pelagibacteraceae* that is currently known as SAR11 subclade Ia.4.II, and share a minimum gANI of ~90% with the type genome. Its members lack the genetic pathway required to synthesize the molybdenum cofactor, while some close evolutionary neighbors contain the required genes¹⁶. Cells of this genus have been isolated from the tropical Pacific Ocean. The type species is *Lacunipelagibacter aumakalihii*.

Registry URL: <https://seqco.de/i:51411>

Description of *Coralipelagibacter* gen. nov.

Coralipelagibacter ([Co.ra.li.pe.la.gi.bac'ter] L. neut. n. *coralium*, coral; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Coralipelagibacter*, a bacterium that is coastally restricted where coral reefs are found).

Genomes of the genus *Coralipelagibacter* occupy a monophyletic node within the family *Pelagibacteraceae* that is currently known as SAR11 subclade Ia.3.II, and share a minimum gANI of ~84% to the type genome. Cells of this genus have been isolated from the tropical Pacific Ocean and this group is restricted globally to nearshore waters. The type species is *Coralipelagibacter heeiaensis*.

Registry URL: <https://seqco.de/i:51413>

Description of *Litoralipelagibacter* gen. nov.

Litoralipelagibacter ([Li.to.ra.li.pe.la.gi.bac'ter] L. masc. adj. *litoralis*, of the seashore; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Litoralipelagibacter*, a coastal bacterium).

Genomes of the genus *Litoralipelagibacter* occupy a monophyletic node within the family *Pelagibacteraceae* that is currently known as SAR11 subclade Ia.3.III, and share a minimum gANI of ~95% to the type genome. Cells of this genus have been isolated from the tropical Pacific Ocean. The type species is *Litoralipelagibacter koolauensis*.

Registry URL: <https://seqco.de/i:51416>

Description of *Stellipelagibacter* gen. nov.

Stellipelagibacter ([Ste.lli.pe.la.gi.bac'ter] L. fem. n. *stella*, a star; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Stellipelagibacter*, a bacterium of the starry sea, referring to the broad distribution of these bacteria as the vast starry sky).

Genomes of the genus *Stellipelagibacter* occupy a monophyletic node within the family *Pelagibacteraceae* that is currently known as SAR11 subclade Ib.2.I, and share a minimum gANI of ~85% to the type genome. There are no known cultivated isolates yet reported for this genus. The type species is *Stellipelagibacter aequoris*.

Registry URL: <https://seqco.de/i:51419>

Description of *Parvipelagibacter* gen. nov.

Parvipelagibacter ([Par.vi.pe.la.gi.bac'ter] L. masc. adj. *parvus*, small; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Parvipelagibacter*, a bacterium with a tight and constrained distribution, which is smaller than that of its close neighbor.)

Genomes of the genus *Parvipelagibacter* occupy a previously uncharacterized monophyletic node within the family *Pelagibacteraceae*, and share a minimum gANI of ~90% to the type genome. There are no known cultivated isolates yet reported for this genus. The type species is *Parvipelagibacter dignotus*.

Registry URL: <https://seqco.de/i:51421>

Description of *Superipelagibacter* gen. nov.

Superipelagibacter ([Su.pe.ri.pe.la.gi.bac'ter] L. masc. adj. *superus*, above; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Superipelagibacter*, a bacterium that is often correlated with higher temperatures).

Genomes of the genus *Superipelagibacter* occupy a monophyletic node within the family *Pelagibacteraceae* that is currently known as SAR11 subclade Ib.1.I, and share a minimum gANI of ~84% to the type genome. Cells of this genus have been isolated from the Red Sea and the tropical Pacific Ocean. The type species is *Superipelagibacter microcenae*.

Registry URL: <https://seqco.de/i:51423>

Description of *Copipelagibacter* gen. nov.

Copipelagibacter ([Co.pi.pe.la.gi.bac'ter] L. fem. n. *copia*, abundance, supply; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Copipelagibacter*, an abundant rod from the open sea).

Genomes of the genus *Copipelagibacter* occupy a monophyletic node within the family *Pelagibacteraceae* that is currently known as SAR11 subclade Ib.1.III, and share a minimum gANI of ~90% to the type genome. There are no known cultivated isolates yet reported for this genus. The type species is *Copipelagibacter splendidus*.

Registry URL: <https://seqco.de/i:51425>

Description of *Fusipelagibacter* gen. nov.

Fusipelagibacter ([Fu.si.pe.la.gi.bac'ter] L. masc. adj. *fuscus*, spread out; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Fusipelagibacter*, a bacterium spread across the global oceans).

Genomes of the genus *Fusipelagibacter* occupy a previously uncharacterized monophyletic node within the Ib.4 genome cluster of the family *Pelagibacteraceae*, and share a minimum gANI of ~93% to the type genome. There are no known cultivated isolates yet reported for this genus. The type species is *Fusipelagibacter tranquillus*.

Registry URL: <https://seqco.de/i:51427>

Description of *Extensipelagibacter* gen. nov.

Extensipelagibacter ([Ex.ten.si.pe.la.gi.bac'ter] L. masc. perf. part. *extensus*, spread out; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Extensipelagibacter*, a bacterium widely distributed across the ocean).

Genomes of the genus *Extensipelagibacter* occupy a monophyletic node within the family *Pelagibacteraceae* that is currently known as SAR11 subclade Ib.4.I, and share a

minimum gANI of ~96% to the type genome. There are no known cultivated isolates yet reported for this genus. The type species is *Extensipelagibacter unicus*.

Registry URL: <https://seqco.de/i:51429>

Description of *Medipelagibacter* gen. nov.

Medipelagibacter ([Me.di.pe.la.gi.bac'ter] L. masc. adj. *medius*, middle; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Medipelagibacter*, an open-ocean bacterium generally found at mid and low latitudes).

Genomes of the genus *Medipelagibacter* occupy a previously uncharacterized monophyletic node within the Ib.4 genome cluster of the family *Pelagibacteraceae*, and share a minimum gANI of ~84% to the type genome. There are no known cultivated isolates yet reported for this genus. The type species is *Medipelagibacter bermudensis*.

Registry URL: <https://seqco.de/i:51431>

Description of *Amplipelagibacter* gen. nov.

Amplipelagibacter ([Am.pli.pe.la.gi.bac'ter] L. masc. adj. *amplus*, great in number; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Amplipelagibacter*, an abundant rod from the sea).

Genomes of the genus *Amplipelagibacter* occupy a previously uncharacterized monophyletic node within the Ib.4 genome cluster of the family *Pelagibacteraceae*, and share a minimum gANI of ~90% to the type genome. There are no known cultivated isolates yet reported for this genus. The type species is *Amplipelagibacter globalis*.

Registry URL: <https://seqco.de/i:51433>

Description of *Rubripelagibacter* gen. nov.

Rubripelagibacter ([Ru.bri.pe.la.gi.bac'ter] L. masc. adj. *ruber*, red, ruddy; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Rubripelagibacter*, a bacterium of the red sea, referring to the Red Sea in which this genus is often found).

Genomes of the genus *Rubripelagibacter* occupy a previously uncharacterized monophyletic node within the family *Pelagibacteraceae*, and share a minimum gANI of ~85% to the type genome. There are no known cultivated isolates yet reported for this genus. The type species is *Rubripelagibacter caeruleus*.

Registry URL: <https://seqco.de/i:51435>

Description of *Cosmipelagibacter* gen. nov.

Cosmipelagibacter ([Cos.mi.pe.la.gi.bac'ter] L. masc. n. *cosmos*, universe; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Cosmipelagibacter*, a bacterium found throughout the global oceans).

Genomes of the genus *Cosmipelagibacter* occupy a previously uncharacterized monophyletic node within the IIa.B genome cluster of the family *Cosmipelagibacteraceae*. A single strain from the tropical Pacific Ocean has been isolated to date. The type species is *Cosmipelagibacter regressus*.

Registry URL: <https://seqco.de/i:51437>

Description of *Elongatipelagibacter* gen. nov.

Elongatipelagibacter ([E.lon.ga.ti.pe.la.gi.bac'ter] L. masc. perf. part. *elongatus*, elongated; L. neut. n. *pelagus*, the open sea; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Elongatipelagibacter*, an elongated rod from the open sea, referring to the elongated morphology of this genus compared to other isolates from the *Pelagibacterales*).

Genomes of the genus *Elongatipelagibacter* occupy a previously uncharacterized monophyletic node within the IIIa.I genome cluster of the family *Fontibacteriaceae*. Multiple strains from this genus have been isolated. The type species is *Elongatipelagibacter longiformis*.

Registry URL: <https://seqco.de/i:51439>

Description of *Xanthipelagibacter aumoanae* sp. nov.

Xanthipelagibacter aumoanae ([a.u.mo.a'nae] N.L. gen. n. *aumoanae*, pertaining to one who swims the ocean, derived from the 'Ōlelo Hawai'i "au moana" meaning "swims the ocean").

The reference strain, HIMB83^T, was isolated from seawater taken from a depth of 2 m depth within Kāne'ōhe Bay, O'ahu, HI, USA at the Hawai'i Institute of Marine Biology located on the island of Moku o Lo'e, in the tropical Pacific Ocean. The DNA G+C content of the reference strain is 29.2 mol% (determined from the genome sequence), and the genome is 1.4 Mb in length. The GenBank accession number for the genome sequence of strain HIMB83^T is GCA_000504225.1.

Registry URL: <https://seqco.de/i:51393>

Description of *Undatipelagibacter huakaii* sp. nov.

Undatipelagibacter huakaii ([hu.a.ka'i.i] N.L. gen. n. *huakaii*, of a microbe from (water) traveling around coasts, from the 'Ōlelo Hawai'i "huaka'i" meaning "traveling, around coasts").

The reference strain, HIMB1556^T, was isolated from seawater taken from a depth of 2 m within Kāne'ōhe Bay, O'ahu, HI, USA at the Hawai'i Institute of Marine Biology located on the island of Moku o Lo'e, in the tropical Pacific Ocean. The DNA G+C content of the reference strain is 29.3 mol% (determined from the genome sequence), and the genome is 1.5 Mb in length. The GenBank accession number for the genome sequence of strain HIMB1556^T is JBNAGE000000000.

Registry URL: <https://seqco.de/i:51395>

Description of *Angustipelagibacter atlanticus* sp. nov.

Angustipelagibacter atlanticus ([a.tlan'ti.cus] L. masc. adj. *atlanticus*, Atlantic, from the Atlantic ocean).

The type genome, SAG GCA_902580005.1^T, was collected from seawater taken from a depth of 89.9 m within the Mid-Atlantic Ocean. The DNA G+C content of the type genome is 29.8 mol% (determined from the genome sequence), and the genome is 1.18 Mb in length. The GenBank accession number for the genome sequence is GCA_902580005.1.

Registry URL: <https://seqco.de/i:51397>

Description of *Atlantikopelagibacter sargassoseensis* sp. nov.

Atlantikopelagibacter sargassoseensis ([sar.gas.so.se.en'sis] N.L. masc. adj. *sargassoseensis*, of the Sargasso Sea, the site from which multiple strains of this species have been isolated).

The reference strain, HTCC7211^T, was isolated from seawater taken from a depth of 10 m within the western Atlantic Ocean. The DNA G+C content of the reference strain is 29.0 mol% (determined from the genome sequence), and the genome is 1.5 Mb in length. The

GenBank accession number for the genome sequence of strain HTCC7211^T is GCF_000155895.1.

Registry URL: <https://seqco.de/i:49682>

Description of *Accedentipelagibacter yaqonii* sp. nov.

Accedentipelagibacter yaqonii ([ya.qo'ni.i] N.L. gen. n. *yaqonii*, of the Yaqo'n (pronounced Yacona) people that lived along the coast of the Northeastern Pacific Ocean in what is now the central coast of Oregon, USA).

The reference strain, HTCC8051^T, was isolated from seawater taken from a depth of 10 m off of the coast of Newport, OR, USA, in the northeast Pacific Ocean. The DNA G+C content of the reference strain is 29.2 mol% (determined from the genome sequence), and the genome is 1.4 Mb in length. The GenBank accession number for the genome sequence of strain HTCC8051^T is GCA_000472605.1.

Registry URL: <https://seqco.de/i:51399>

Description of *Colligatipelagibacter wilhelmii* sp. nov.

Colligatipelagibacter wilhelmii ([wil.hel'mi.i] N.L. masc. gen. n. *wilhelmii*, of Wilhelm, named after Dr. Larry Wilhelm, instrumental in early SAR11 research).

The type genome, SAG GCA_902524535.1^T, was collected from seawater taken from a depth of 10 m in the western Atlantic Ocean. The DNA G+C content of the type genome is 29.4 mol% (determined from the genome sequence), and the genome is 1.1 Mb in length. The GenBank accession number for the genome sequence is GCA_902524535.1.

Registry URL: <https://seqco.de/i:49684>

Description of *Tantilluspelagibacter marisrubii* sp. nov.

Tantilluspelagibacter marisrubii ([ma.ris.ru'bi.i] L. neut. n. *mare*, the sea; L. masc. adj. *rubius*, red; N.L. gen. n. *marisrubii*, of the Red Sea, where the first strain was isolated from).

The reference strain, HIMB1436^T, was isolated from seawater taken from a depth of 2 m depth from outside of Kāneʻohe Bay, Oʻahu, HI, USA, at the Hawaiʻi Institute of Marine Biology located on the island of Moku o Loʻe, in the tropical Pacific Ocean.. While the first isolated strain (RS39) originated from the central Red Sea at 19 m depth, it was not eligible to be submitted to the SeqCode as a reference strain as the raw DNA genome sequence reads are no longer available. The DNA G+C content of the reference strain is 29.2 mol% (determined from the genome sequence), and the genome is 1.2 Mb in length. The GenBank accession number for the genome sequence of strain HIMB1436^T is JBNGZC000000000.

Registry URL: <https://seqco.de/i:51402>

Description of *Proventipelagibacter oceanicus* sp. nov.

Proventipelagibacter oceanicus ([o.ce.a'ni.cus] N.L. masc. adj. *oceanicus*, of the ocean).

The type genome, SAG GCA_902621345.1^T, was collected from seawater taken from a depth of 10 m within the western Atlantic Ocean. The DNA G+C content of the type genome is 29.5 mol% (determined from the genome sequence), and the genome is 1.2 Mb in length. The GenBank accession number for the genome sequence is GCA_902621345.1.

Registry URL: <https://seqco.de/i:51404>

Description of *Punctumpelagibacter variabilis* sp. nov.

Punctumpelagibacter variabilis ([va.ri.a.bi'lis] L. masc. adj. *variabilis*, variable, referring to a variable global distribution).

The type genome, SAG GCA_902540275.1^T, was collected from seawater taken from a depth of 10 m within the western Atlantic Ocean. The DNA G+C content of the genome is 28.3 mol% (determined from the genome sequence), and the genome is 1.2 Mb in length. The GenBank accession number for the genome sequence is GCA_902540275.1.

Registry URL: <https://seqco.de/i:51406>

Description of *Pantolocipelagibacter undulae* sp. nov.

Pantolocipelagibacter undulae ([un.du'lae] L. gen. n. *undulae*, of small waves, referring to the marine habitat of this bacterium).

The type genome, SAG GCA_902597885.1^T, was collected from seawater taken from a depth of 10 m within the western Atlantic Ocean. The DNA G+C content of the type genome is 28.8 mol% (determined from the genome sequence), and the genome is 1.1 Mb in length. The GenBank accession number for the genome sequence is GCA_902597885.1.

Registry URL: <https://seqco.de/i:51408>

Description of *Alipelagibacter loemeaiki* sp. nov.

Alipelagibacter loemeaiki ([lo.e.me.a.i'ki.i] N.L. gen. n. *loemeaiki*, of a small organism of Lo'e, derived from the 'Ōlelo Hawai'i "Lo'e mea iki" meaning "a small organism from Lo'e", referring to the island of Moku o Lo'e).

The reference strain, HIMB1483^T, was isolated from seawater taken from a depth of 2 m within Kāne'ohe Bay, O'ahu, HI, USA, at the Hawai'i Institute of Marine Biology located on the island of Moku o Lo'e, in the tropical Pacific Ocean. The DNA G+C content of the reference strain is 28.5 mol% (determined from the genome sequence), and the genome is 1.2 Mb in length. The GenBank accession number for the genome sequence of strain HIMB1483^T is JBNAIM000000000.

Registry URL: <https://seqco.de/i:51410>

Description of *Lacunipelagibacter aumakalihii* sp. nov.

Lacunipelagibacter aumakalihii ([a.u.ma.ka.li'hi.i] N.L. gen. n. *aumakalihii*, of a microbe swimming the outside border, from the 'Ōlelo Hawai'i "au ma ka lihi" meaning "swims the outside border").

The reference strain, HIMB1437^T, was isolated from seawater taken from a depth of 2 m within Kāne'ohe Bay, O'ahu, HI, USA at the Hawai'i Institute of Marine Biology located on the island of Moku o Lo'e, in the tropical Pacific Ocean. The DNA G+C content of the reference strain is 29.4 mol% (determined from the genome sequence), and the genome is 1.2 Mb in length. The GenBank accession number for the genome sequence of strain HIMB1437^T is JBNGZD000000000.

Registry URL: <https://seqco.de/i:51412>

Description of *Coralipelagibacter heeiaensis* sp. nov.

Coralipelagibacter heeiaensis ([he.e.i.a.en'sis] N.L. masc. adj. *heeiaensis*, isolated from the ahupua'a (a traditional subdivision of land and sea) of He'eia off the island of O'ahu).

The reference strain, HIMB5^T, was isolated from seawater taken from a depth of 2 m within Kāne'ohe Bay, O'ahu, HI, USA, at the Hawai'i Institute of Marine Biology located on the

island of Moku o Lo‘e, in the tropical Pacific Ocean. The DNA G+C content of the reference strain is 28.6 mol% (determined from the genome sequence), and the genome is 1.3 Mb in length. The GenBank accession number for the genome sequence of strain HIMB5^T is GCF_000299095.1.

Registry URL: <https://seqco.de/i:51414>

Description of *Coralipelagibacter kukahii* sp. nov.

Coralipelagibacter kukahii ([k.uka'hi.i] N.L. gen. n. *kukahii*, referring to a state of uniqueness, derived from the ‘Ōlelo Hawai‘i "kū kahi" meaning "unique, standing alone").

The reference strain, HIMB1412^T, was isolated from seawater taken from a depth of 2 m within Kāne‘ohe Bay, O‘ahu, HI, USA, at the Hawai‘i Institute of Marine Biology located on Moku o Lo‘e, in the tropical Pacific Ocean. The DNA G+C content of the reference strain is 29.3 mol% (determined from the genome sequence), and the genome is 1.3 Mb in length. The GenBank accession number for the genome sequence of strain HIMB1412^T is JBNACKG000000000.

Registry URL: <https://seqco.de/i:51415>

Description of *Litoralipelagibacter koolauensis* sp. nov.

Litoralipelagibacter koolauensis ([ko.o.la.u.en'sis] N.L. masc. adj. *koolauensis*, of Ko‘olau, which is the ‘Ōlelo Hawai‘i term for "windward", referring in this case specifically both to Ko‘olau, the ancient eastern shield volcano on O‘ahu, and to the Ko‘olau Range, the modern-day remnants of the Ko‘olau volcano that overlook the bay from which the first isolate of this species was obtained).

The reference strain, HIMB4^T, was isolated from seawater taken from a depth of 2 m water depth within Kāne‘ohe Bay, O‘ahu, HI, USA at the Hawai‘i Institute of Marine Biology located on the island of Moku o Lo‘e, in the tropical Pacific Ocean. The DNA G+C content of the reference strain is 29.0 mol% (determined from the genome sequence), and the genome is 1.4 Mb in length. The GenBank accession number for the genome sequence of strain HIMB4^T is JBPPFS000000000.

Registry URL: <https://seqco.de/i:51417>

Description of *Pelagibacter yakuwasensis* sp. nov.

Pelagibacter yakuwasensis ([ya.ku.wa.si.en'sis] N.L. masc. adj. *yakuwasensis*, of the vast ocean, derived from "yaku wasi" in the language of the Yaqo'n (pronounced Yacona) people that lived along the coast of the Northeastern Pacific Ocean in what is now the central coast of Oregon, USA).

The reference strain, HTCC9565^T, was isolated from surface seawater of the Pacific Ocean from over the Juan de Fuca Ridge off the coast of Newport, OR, USA. The DNA G+C content of the reference strain is 28.9 mol% (determined from the genome sequence), and the genome is 1.3 Mb in length. The GenBank accession number for the genome sequence of strain HTCC9565^T is GCF_012932795.1.

Registry URL: <https://seqco.de/i:51418>

Description of *Stellipelagibacter aequoris* sp. nov.

Stellipelagibacter aequoris ([æ'quor.is] L. gen. n. *aequoris*, of the sea, referring to the broad distribution of this species).

The type genome, SAG GCA_902583075.1^T, was collected from seawater of the Atlantic Ocean at 10 m water depth. The DNA G+C content of the type genome is 29.4 mol% (determined from the genome sequence), and the genome is 1.2 Mb in length. The GenBank accession number for the genome sequence is GCA_902583075.1. This name refers to the broad distribution of this species.

Registry URL: <https://seqco.de/i:51420>

Description of *Parvipelagibacter dignotus* sp. nov.

Parvipelagibacter dignotus ([dig.no'tus] L. masc. part. adj. *dignotus*, dignified, distinguished, recognized as different from its closest neighbor).

The type genome, SAG GCA_902577075.1^T, was collected from seawater taken from a depth of 90.8 m within the northwest Atlantic Ocean. The DNA G+C content of the type genome is 29.4 mol% (determined from the genome sequence), and the genome is 1.36 Mb in length. The GenBank accession number for the genome sequence is GCA_902577075.1.

Registry URL: <https://seqco.de/i:51422>

Description of *Superipelagibacter microcena* sp. nov.

Superipelagibacter microcena ([mi.cro.ce'nae] Gr. masc. adj. *mikros*, small; L. fem. n. *cena*, meal; N.L. gen. n. *microcena*, of the small meal, referring to the utilization of small compounds including aldehydes and formate)¹⁶.

The reference strain, HIMB2305^T, was isolated from seawater taken from a depth of 2 m outside of Kāneʻohe Bay, Oʻahu, HI, USA at the Hawaiʻi Institute of Marine Biology located on the island of Moku o Loʻe, in the tropical Pacific Ocean. While the first cultured isolate of this genus originated from the central Red Sea at 19 m water depth (RS40), it cannot serve as type material for the SeqCode as the raw reads are no longer available. The DNA G+C content of the reference strain HIMB2305^T is 29.5 mol% (determined from the genome sequence), and the genome is 1.4 Mb in length. The GenBank accession number for the genome sequence of strain HIMB2305^T is JBNKJ000000000.

Registry URL: <https://seqco.de/i:51424>

Description of *Copipelagibacter splendidus* sp. nov.

Copipelagibacter splendidus ([splen.di'dus] L. masc. adj. *splendidus*, splendid, glittering).

The type genome, SAG GCA_902617415.1^T, was collected from seawater taken from a depth of 10 m in the western Atlantic Ocean. The DNA G+C content of the type genome is 29.2 mol% (determined from the genome sequence), and the genome is 1.4 Mb in length. The GenBank accession number for the genome sequence is GCA_902617415.1.

Registry URL: <https://seqco.de/i:51426>

Description of *Fusipelagibacter tranquillus* sp. nov.

Fusipelagibacter tranquillus ([tran.qui'llus] L. masc. adj. *tranquillus*, quiet, calm).

The type genome, SAG GCA_902560075.1^T, was collected from seawater taken from a depth of 10 m within the western Atlantic Ocean. The DNA G+C content of the type genome is 29.1 mol% (determined from the genome sequence), and the genome is 1.2 Mb in length. The GenBank accession number for the genome sequence is GCA_902560075.1.

Registry URL: <https://seqco.de/i:51428>

Description of *Extensipelagibacter unicus* sp. nov.

Extensipelagibacter unicus ([u.ni'cus] L. masc. adj. *unicus*, unique; referring to the high genome similarity found between members of this species, a unique trait compared to closely related lineages).

The type genome, SAG GCA_902576855.1^T, was collected from seawater taken from a depth of 90.8 m within the northwest Atlantic Ocean. The DNA G+C content of the type genome is 28.7 mol% (determined from the genome sequence), and the genome is 1.2 Mb in length. The GenBank accession number for the genome sequence is GCA_902576855.1.

Registry URL: <https://seqco.de/i:51430>

Description of *Medipelagibacter bermudensis* sp. nov.

Medipelagibacter bermudensis ([ber.mu.den'sis] N.L. masc. adj. *bermudensis*, of Bermuda, referring to the sampling location near Bermuda at the Bermuda Atlantic Time-series (BATS) site).

The type genome, SAG GCA_902622545.1^T, was collected from seawater taken from a depth of 10 m within the western Atlantic Ocean. The DNA G+C content of the type genome is 28.9 mol% (determined from the genome sequence), and the genome is 1.2 Mb in length. The GenBank accession number for the genome sequence is GCA_902622545.1.

Registry URL: <https://seqco.de/i:51432>

Description of *Amplipelagibacter globalis* sp. nov.

Amplipelagibacter globalis ([glo.ba'lis] N.L. masc. adj. *globalis*, global, referring to the global distribution of the species).

The type genome, SAG GCA_902580675.1^T, was collected from seawater taken from a depth of 89.9 m within the Atlantic Ocean. The DNA G+C content of the type genome is 29.5 mol% (determined from the genome sequence), and the genome is 1.1 Mb in length. The GenBank accession number for the genome sequence is GCA_902580675.1.

Registry URL: <https://seqco.de/i:51434>

Description of *Rubripelagibacter caeruleus* sp. nov.

Rubripelagibacter caeruleus ([cae.ru.le'us] L. masc. adj. *caeruleus*, blue, cerulean, referring to the ocean environment).

The type genome, SAG GCA_902581445.1^T, was collected from seawater taken from a depth of 10 m within the western Atlantic Ocean. The DNA G+C content of the type genome is 29.2 mol% (determined from the genome sequence), and the genome is 1.1 Mb in length. The GenBank accession number for the genome sequence is GCA_902581445.1.

Registry URL: <https://seqco.de/i:51436>

Description of *Cosmipelagibacter regressus* sp. nov.

Cosmipelagibacter regressus ([re.gres'sus] L. gen. n. *regressus*, of going back, of the return, referring to the *Cosmiplelagibacter* that has returned (to the coast)).

The reference strain, HIMB58^T, was isolated from seawater taken from a depth of 2 m within Kāneʻohe Bay, Oʻahu, HI, USA, at the Hawaiʻi Institute of Marine Biology located on the island of Moku o Loʻe, in the tropical Pacific Ocean. The DNA G+C content of the reference strain is 29.9 mol% (determined from the genome sequence), and the genome is 1.1 Mb in

length. The GenBank accession number for the genome sequence of strain HIMB58^T is GCA_000419545.1.
Registry URL: <https://seqco.de/i:51438>

Description of *Elongatipelagibacter longiformis* sp. nov.

Elongatipelagibacter longiformis ([lon.gi.for'mis] L. masc. adj. *longus*, long; L. masc. adj. suff. -*formis*, in the shape of; N.L. masc. adj. *longiformis*, long-shaped (compared to other *Pelagibacterales*)).

The reference strain, HIMB114^T, was isolated from seawater taken from a depth of 2 m within Kāneʻohe Bay, Oʻahu, HI, USA, at the Hawaiʻi Institute of Marine Biology located on the island of Moku o Loʻe in the tropical Pacific Ocean. The DNA G+C content of the reference strain is 29.6 mol% (determined from the genome sequence), and the genome is 1.2 Mb in length. The GenBank accession number for the genome sequence of strain HIMB114^T is GCA_000163555.2.

Registry URL: <https://seqco.de/i:51440>

Description of *Mesopelagibacter profundus* sp. nov.

Mesopelagibacter profundus ([pro.fun'di] L. gen. n. *profundus*, of the depths, living within the depths of the oceans).

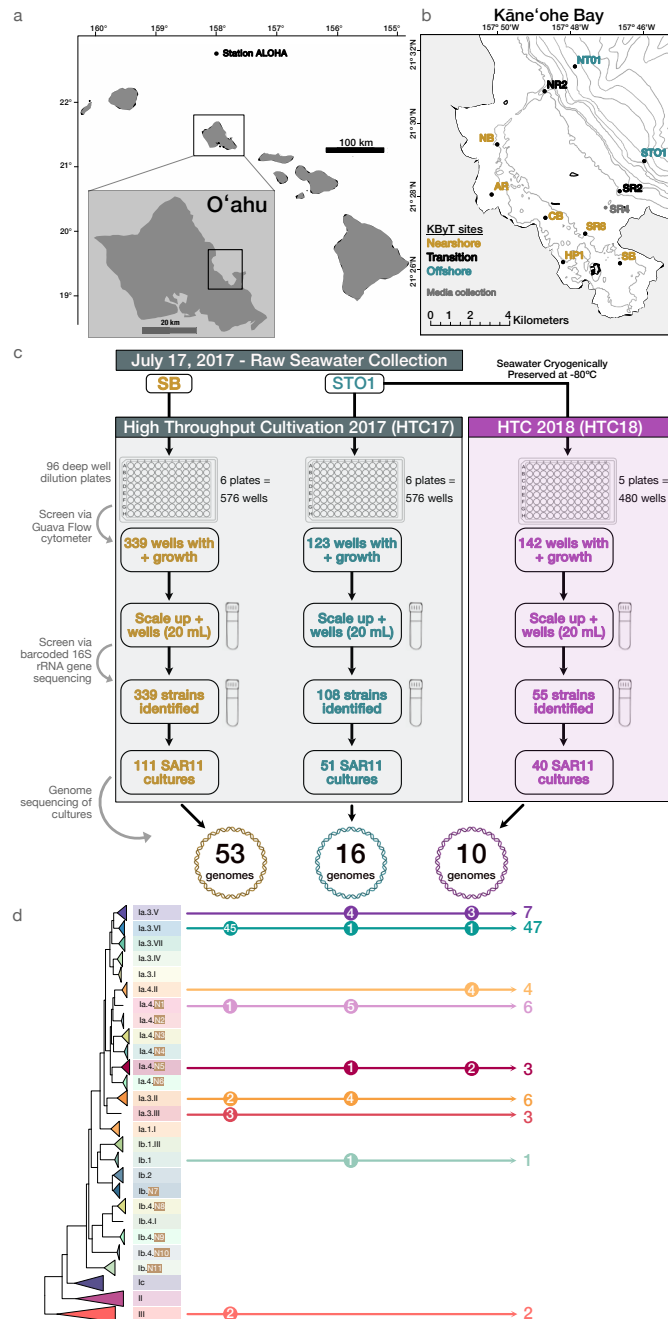
The type genome, SAG GCA_000504625.1^T, was collected from seawater taken from a depth of 770 m within the tropical Pacific Ocean from Station ALOHA (located at 22° 45'N, 158°W). The DNA G+C content of the type genome is 29.0 mol% (determined from the genome sequence), and the genome is 1.4 Mb in length.

Registry URL: <https://seqco.de/i:52920>

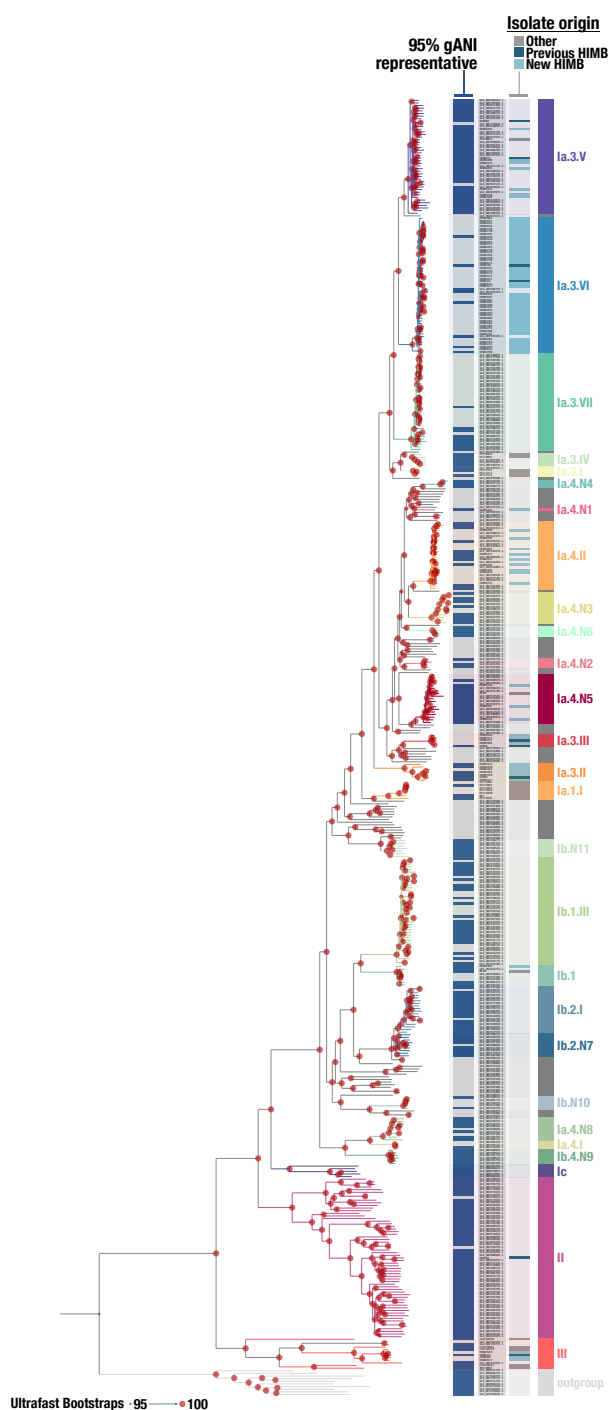
783 **Supplementary Figures**

784 Supplementary Figures are available on FigShare at the following DOI:
785 [10.6084/m9.figshare.30048979](https://doi.org/10.6084/m9.figshare.30048979).

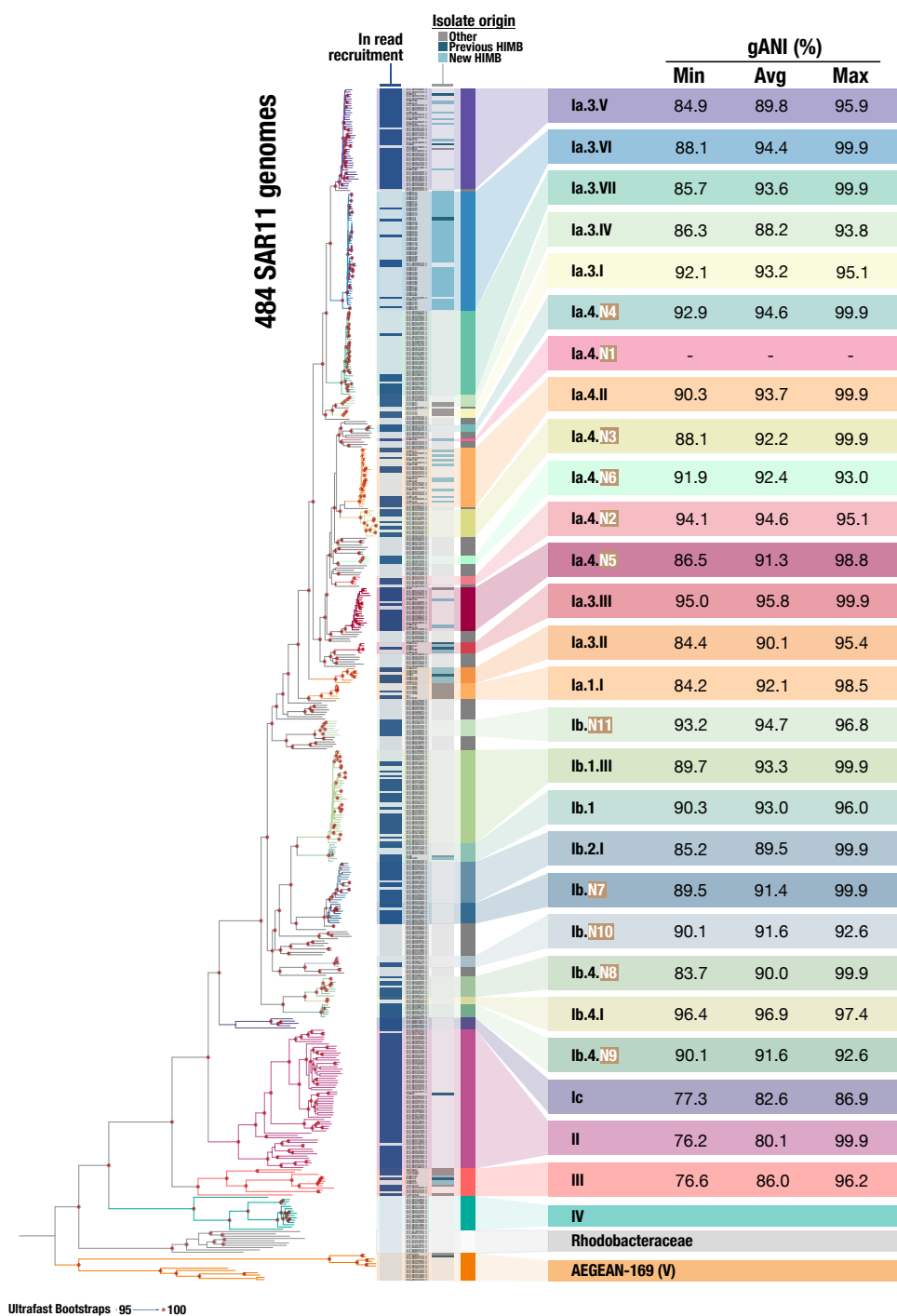
786
787
788
789
790
791
792
793
794



Supplementary Fig. 1. Overview of high-throughput culturing experiments. (A) Location of O'ahu in the Hawaiian archipelago in relation to Station ALOHA (~100 km north). (B) Map of the embayment on the windward side of O'ahu with sites included in the Kāne'ohe Bay Time-series (KByT) that are classified as 'nearshore' (orange), 'transition' (black), or 'offshore' (turquoise). Site SR4 in gray from which seawater media was collected for the cultivation experiments is also indicated. Bathymetry lines are approximate indicating every 10 meters up to 50 meters, then marking 100 meter increments, after Tucker et al.,²⁸. (C) Flowchart outlining the high throughput cultivation (HTC) experiments conducted in 2017 and 2018 leading to the isolation of hundreds of SAR11 cultures and 79 new SAR11 isolate genomes. (D) Schematic SAR11 phylogeny indicating which samples harbored genomes from which subgroups.

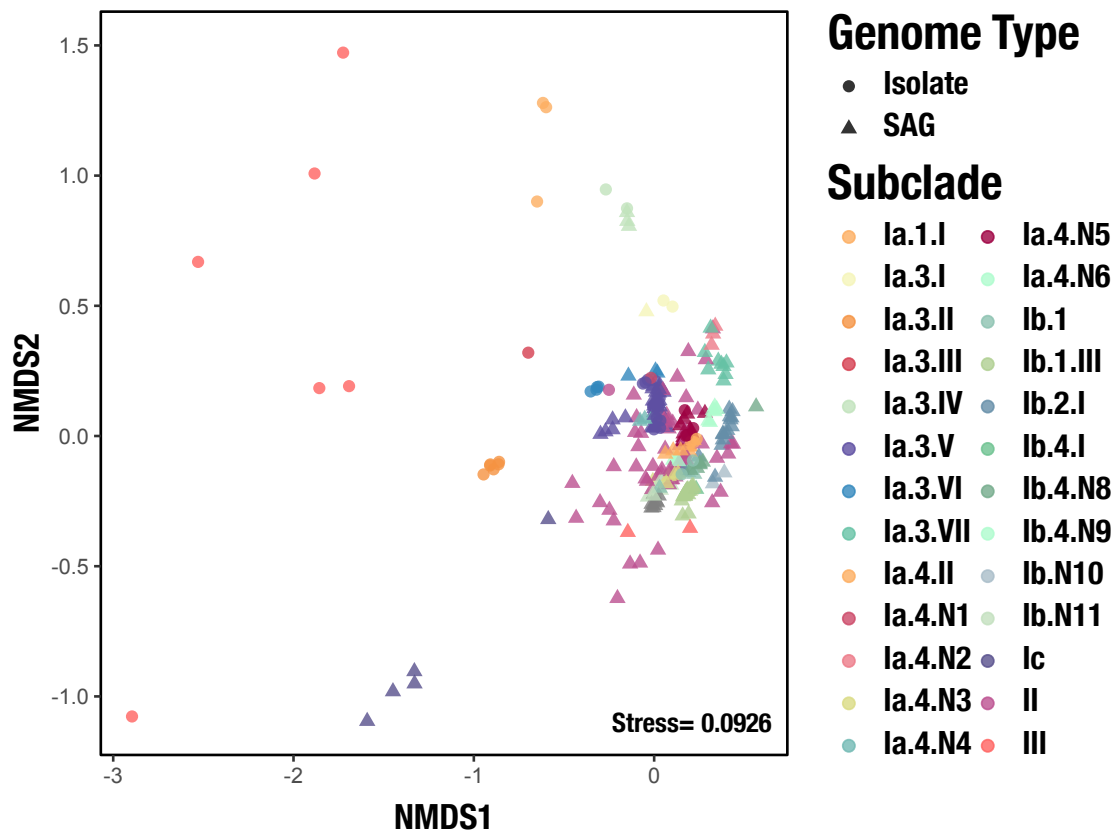


Supplementary Fig. 2. Phylogenomic tree of all 484 *Pelagibacterales* genomes included with accession numbers. Of the 484 SAR11 genomes 109 were isolates and 375 SAGs. The phylogeny is based on protein sequences from a curated SAR11-specific set of 165 genes. Isolate origin indicates if the genome was from this study, a previous isolate from Kāneʻohe Bay, or from another source. Genomes that serve as a 95% gANI representative and were included in the read recruitment are also indicated with a dark blue bar. Circles indicate ultrafast bootstrap support values $\geq 95\%$, from 1000 replicates.



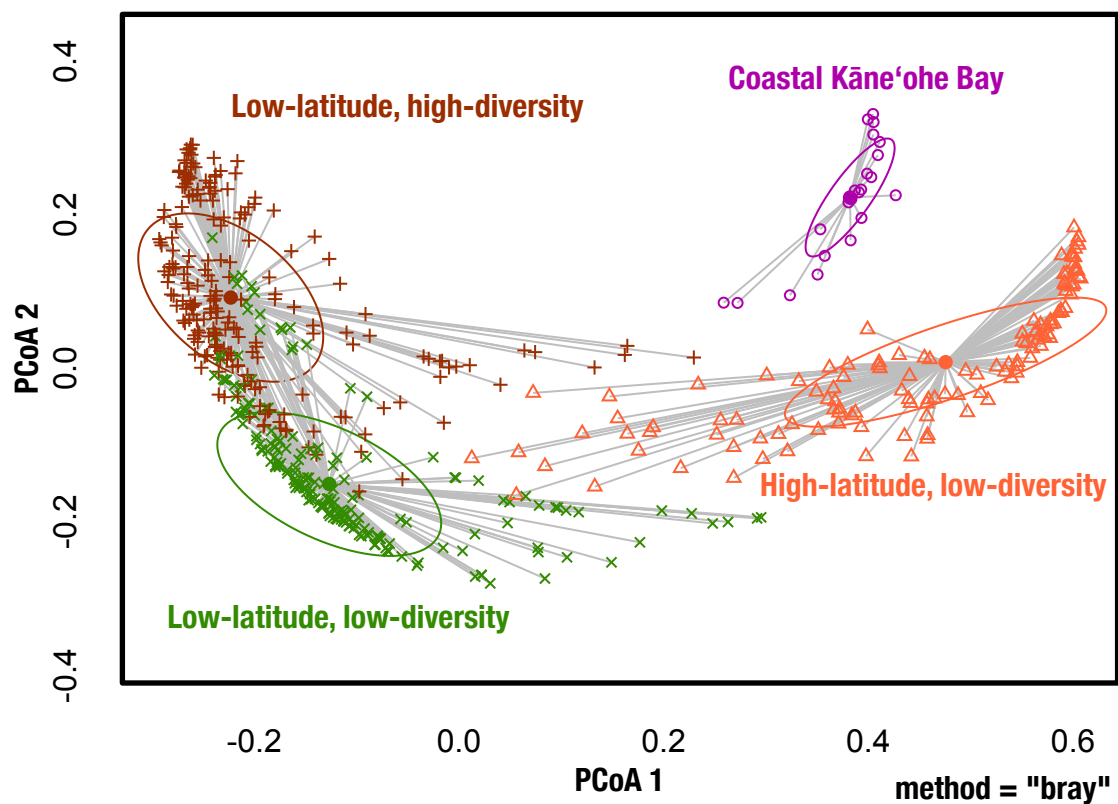
Supplementary Fig. 3. Phylogenomic tree with 484 *Pelagibacterales* genomes and subgroups IV and V genomes. Isolate origin is highlighted on the tree and indicates if the genome was from this study, a previous isolate from Kāneʻohe Bay, or from another source. Genomes that serve as a 95% gANI representative and were included in the read recruitment are also indicated with a dark blue bar.

Genomes by Subclade

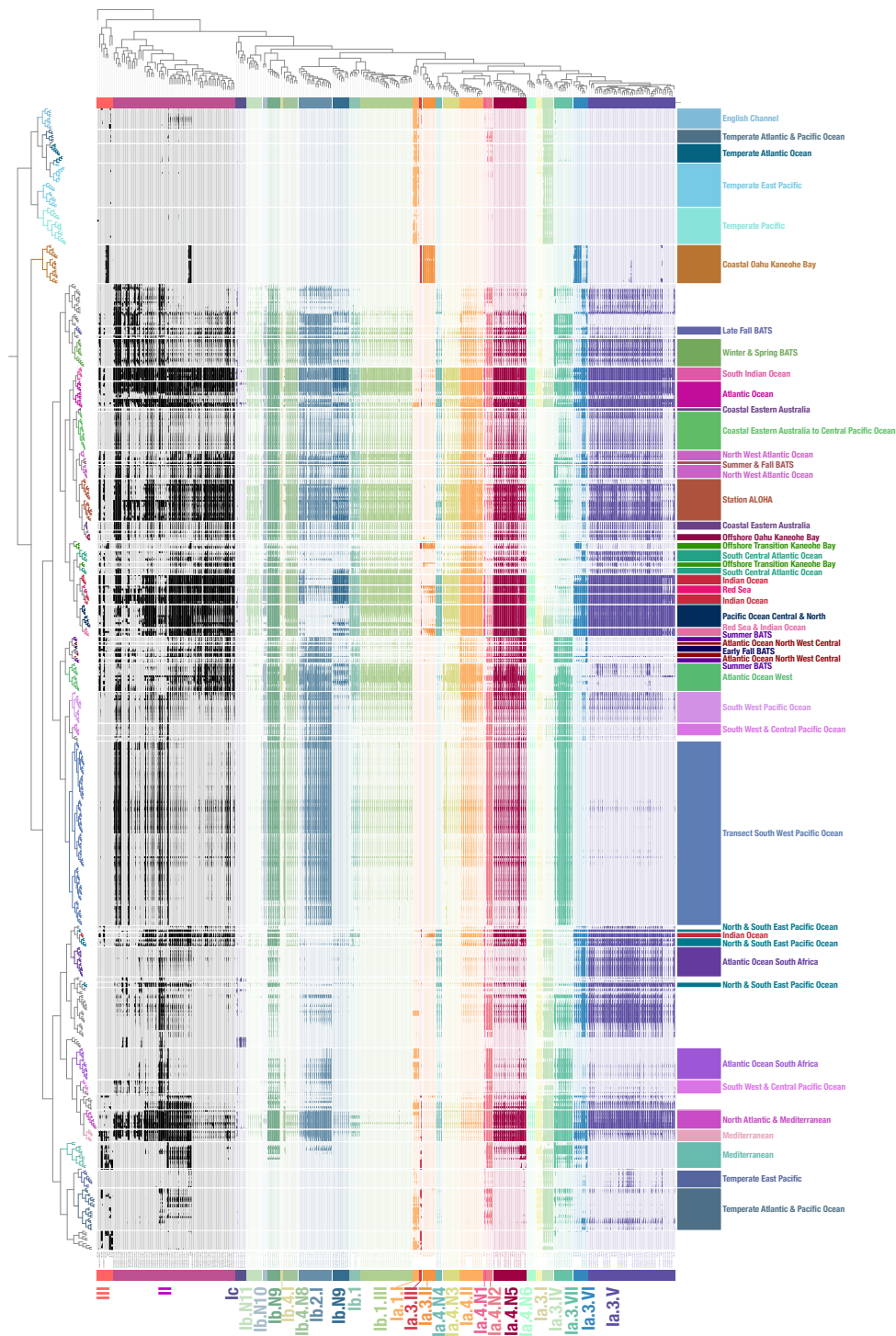


Supplementary Fig. 4. NMDS of genomes with detection data. All of the genomes included in the analysis are included here, with distinction between isolate genome (circle) or SAG (triangle) indicated.

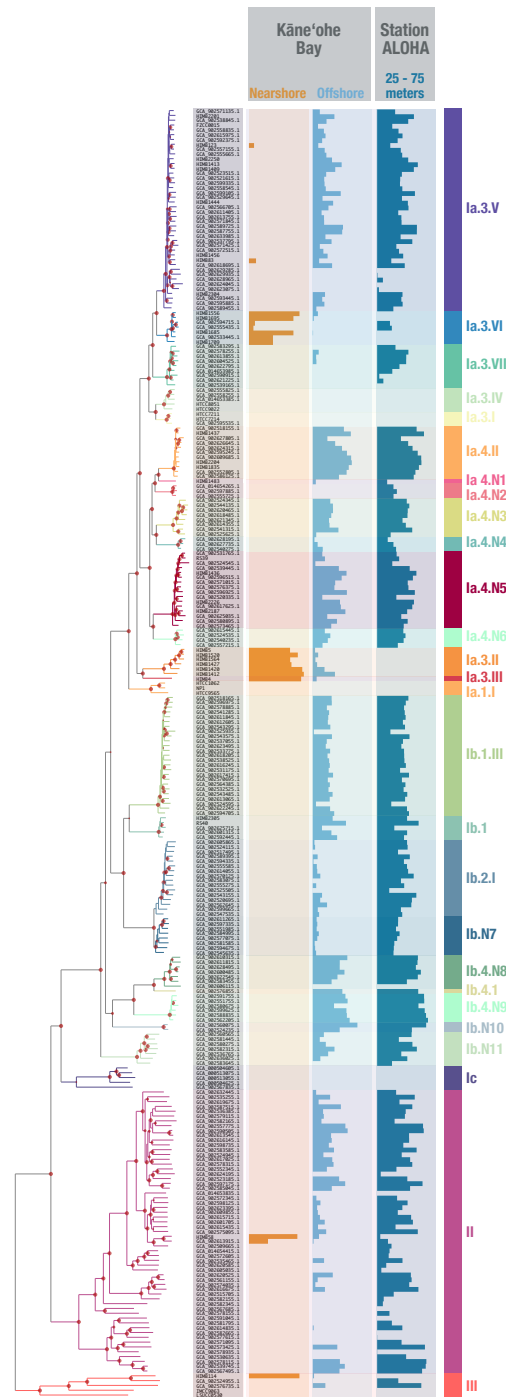
Metagenomes by Fig. 2 Community Group



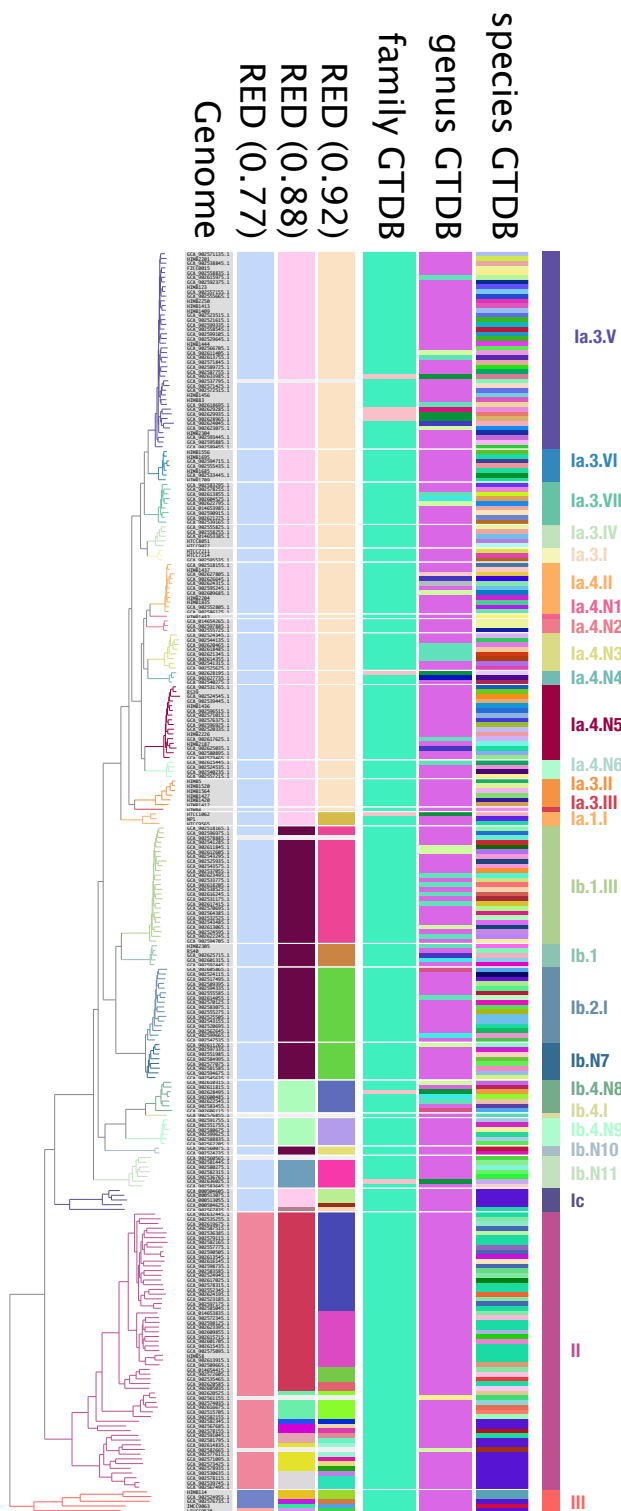
Supplementary Fig. 5. PCoA of metagenomes by community group. All of the genomes included in the read recruitment analysis are included here as well as all metagenomes in Figure 2. This is a visualization of the PERMDISP results.



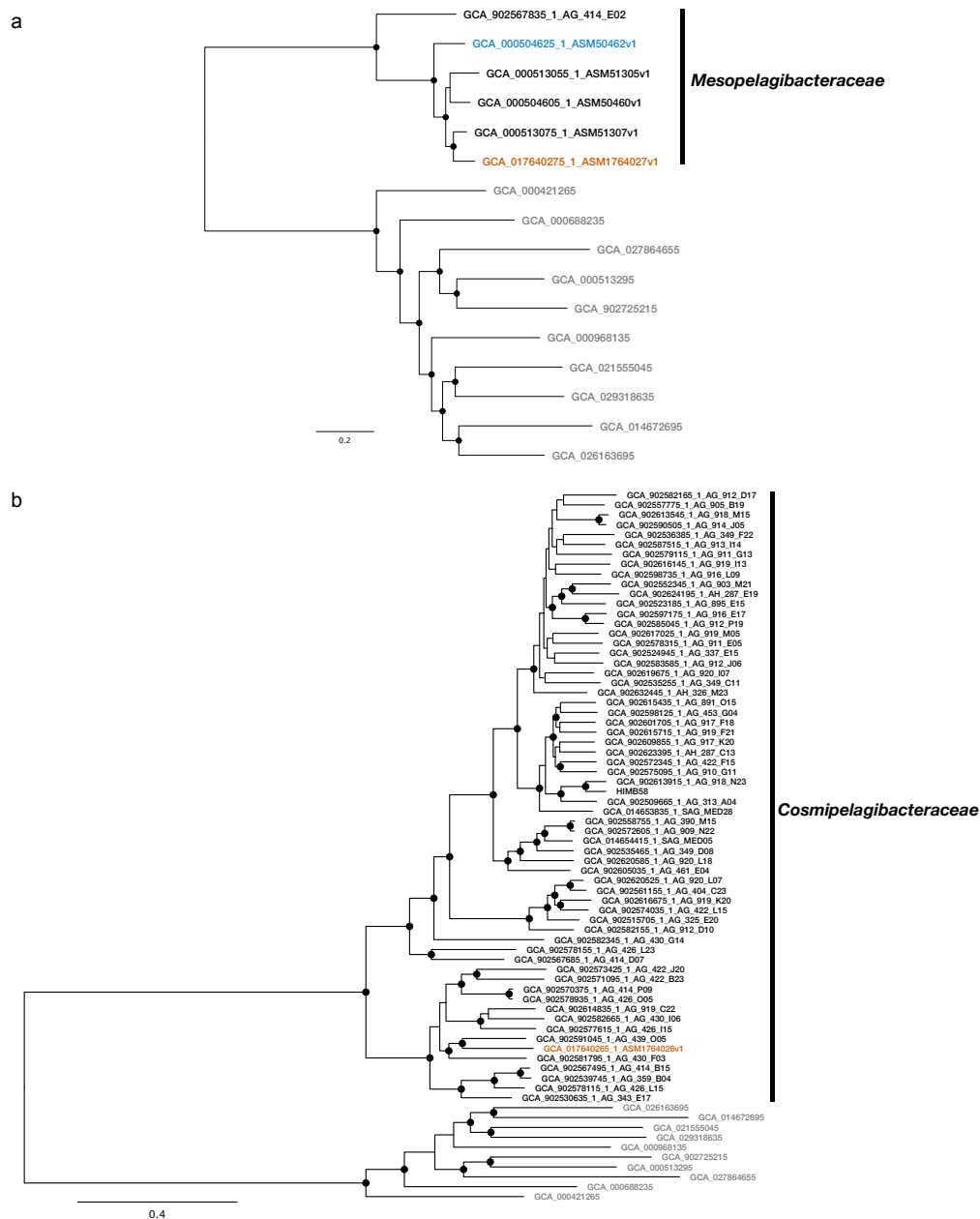
Supplementary Fig. 6 Read recruitment data including all metagenomes included in cluster analysis of metagenomes based on genome detection values using the k-means algorithm. Subgroups are indicated at the bottom of the figure with labels indicating the metagenome groups along the right hand side.



Supplementary Fig. 8. Phylogenomic tree of the *Pelagibacterales* with read recruitment from Kāne'ohe Bay and Station ALOHA. The phylogeny includes 268 genomes (50 isolates and 218 SAGs) included in the pruned phylogeny. Read recruitment data in the form of detection values on a scale of 0.25 to 0.75 are displayed along the right hand side. Circles indicate ultrafast bootstrap support values $\geq 95\%$ or greater from 1000 replicates.



Supplementary Fig. 9. Phylogenomic tree of the *Pelagibacterales* compared to other classification methods. The phylogeny includes 268 genomes (50 isolates and 218 SAGs) included in the pruned phylogeny. Circles indicate ultrafast bootstrap support values $\geq 95\%$ or greater from 1000 replicates.



Supplementary Fig. 10. Phylogenomic placement of previously described type material for ‘*Mesopelagibacter carboxydoxydans*’, *Mesopelagibacter profundus*, and *Anoxypelagibacter denitrificans*. Phylogenies based on the SAR11_165 gene set for the (A) *Mesopelagibacteraceae* and (B) *Cosmipelagibacteraceae*. Both phylogenies include the single amplified genomes used for analyses in this study (black). The type genomes for ‘*Mesopelagibacter carboxydoxydans*’ (MAG GCA_017640275.1) and *Anoxypelagibacter denitrificans* (MAG GCA_017640265.1) are highlighted in orange, while *Mesopelagibacter profundus* (SAG GCA_000504625.1) is highlighted in blue. The outgroup is shown in gray. Ultrafast bootstrap support values from 95 to 100 are indicated by circles at the relevant node.

Supplementary References

1. Monaghan, E. A., Freel, K. C. & Rappé, M. S. Isolation of SAR11 Marine Bacteria from Cryopreserved Seawater. *mSystems* **5**, (2020).
2. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
3. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
4. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–6 (2013).
5. Pachiadaki, M. G. *et al.* Charting the complexity of the marine microbiome through single-cell genomics. *Cell* **179**, 1623–1635.e11 (2019).
6. Haro-Moreno, J. M. *et al.* Ecogenomics of the SAR11 clade. *Environ. Microbiol.* **22**, 1748–1763 (2020).
7. Thrash, J. C. *et al.* Single-cell enabled comparative genomics of a deep ocean SAR11 bathytype. *ISME J.* **8**, 1440–1451 (2014).
8. Tsementzi, D. *et al.* SAR11 bacteria linked to ocean anoxia and nitrogen loss. *Nature* **536**, 179–183 (2016).
9. Thompson, L. R. *et al.* Red Sea SAR11 and Prochlorococcus single-cell genomes reflect globally distributed pangenomes. *Appl. Environ. Microbiol.* **85**, (2019).
10. Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).
11. Ruiz-Perez, C. A. *et al.* Description of *Candidatus Mesopelagibacter carboxydoxydans* and *Candidatus Anoxipelagibacter denitrificans*: Nitrate-reducing SAR11 genera that dominate mesopelagic and anoxic marine zones. *Syst. Appl. Microbiol.* **44**, 126185 (2021).

- 887 12. Chang, T., Gavelis, G. S., Brown, J. M. & Stepanauskas, R. Genomic representativeness and
888 chimerism in large collections of SAGs and MAGs of marine prokaryoplankton. *Microbiome* **12**,
889 126 (2024).
- 890 13. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment
891 trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
- 892 14. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the
893 Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
- 894 15. Brandon, M. L. High-throughput isolation of pelagic marine bacteria from the coastal subtropical
895 Pacific Ocean. (University of Hawai‘i at Mānoa, 2006).
- 896 16. Tucker, S. J., Freel, K. C., Eren, A. M. & Rappé, M. S. Habitat-specificity in SAR11 is associated
897 with a few genes under high selection. *ISME J.* (2025) doi:10.1093/ismejo/wraf216.
- 898 17. Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially
899 revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
- 900 18. Ramfelt, O., Freel, K. C., Tucker, S. J., Nigro, O. D. & Rappé, M. S. Isolate-anchored comparisons
901 reveal evolutionary and functional differentiation across SAR86 marine bacteria. *ISME J.* **18**, (2024).
- 902 19. Hedlund, B. P. *et al.* SeqCode: a nomenclatural code for prokaryotes described from sequence data.
903 *Nat. Microbiol.* **7**, 1702–1708 (2022).
- 904 20. Rappé, M. S., Connon, S. A., Vergin, K. L. & Giovannoni, S. J. Cultivation of the ubiquitous SAR11
905 marine bacterioplankton clade. *Nature* **418**, 630–633 (2002).
- 906 21. Oren, A., Garrity, G. M., Parker, C. T., Chuvochina, M. & Trujillo, M. E. Lists of names of
907 prokaryotic Candidatus taxa. *Int. J. Syst. Evol. Microbiol.* **70**, 3956–4042 (2020).
- 908 22. Vergin, K. L. *et al.* High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time-series
909 Study site by phylogenetic placement of pyrosequences. *ISME J.* **7**, 1322–1332 (2013).
- 910 23. Suzuki, M. T., Béjà, O., Taylor, L. T. & DeLong, E. F. Phylogenetic analysis of ribosomal RNA
911 operons from uncultivated coastal marine bacterioplankton. *Environ. Microbiol.* **3**, 323–331 (2001).

24. Grote, J. *et al.* Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *MBio* **3**, (2012).
25. Morris, R. M. *et al.* Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic Time-series Study site. *Limnol. Oceanogr.* **50**, 1687–1696 (2005).
26. Henson, M. W., Lanclos, V. C., Faircloth, B. C. & Thrash, J. C. Cultivation and genomics of the first freshwater SAR11 (LD12) isolate. *ISME J.* **12**, 1846–1860 (2018).
27. Thrash, J. C. *et al.* Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade. *Sci. Rep.* **1**, 13 (2011).
28. Tucker, S. J. *et al.* Seasonal and spatial transitions in phytoplankton assemblages spanning estuarine to open ocean waters of the tropical Pacific. *Limnol. Oceanogr.* **70**, 1693–1708 (2025).