219 metagenome-assembled genomes of microorganisms from Icelandic marine waters

Clara Jégousse [1, 2] , **Pauline Vannier** [2] , **René Groben** [2] , **Frank Oliver Glöckner** [3, 4] , **Viggó Marteinsson** Corresp. [1, 2]

[1] School of Health Sciences, University of Iceland, Reykjavik, Iceland

[2] Microbiology Group, Matís ohf., Reykjavík, Iceland

[3] Data at the Computing Center, Alfred Wegener Institute, Bremenhaven, Germany

[4] MARUM - Center for Marine Environmental Sciences, Universität Bremen, Bremen, Germany

Corresponding Author: Viggó Marteinsson
Email address: viggo@matis.is

Marine microorganisms contribute to the health of the global ocean by supporting the marine food web and regulating biogeochemical cycles. Assessing marine microbial diversity is a crucial step towards understanding the global ocean. The waters surrounding Iceland are a complex environment where relatively warm salty waters from the Atlantic cool down and sink down to the deep. Microbial studies in this area have focused on photosynthetic micro- and nanoplankton mainly using microscopy and chlorophyll measurements. However, the diversity and function of the bacterial and archaeal picoplankton remains unknown. Here, we used a co-assembly approach supported by a marine mock community to reconstruct metagenome-assembled genomes (MAGs) from 31 metagenomes from the sea surface and seafloor of four oceanographic sampling stations sampled between 2015 and 2018. The resulting 219 MAGs include 191 bacterial, 26 archaeal and two eukaryotic MAGs to bridge the gap in our current knowledge of the global marine microbiome.

219 Metagenome-assembled genomes of microorganisms from Icelandic marine waters

Clara Jégousse[1,2], Pauline Vannier[2], René Groben[2], Frank Oliver Glöckner[3,4], and Viggó Marteinsson[1,2]

[1]School of Health Sciences, University of Iceland, Reykjavík, Iceland
[2]Microbiology Group, Matís ohf., Reykjavík, Iceland
[3]Data at the Computing Center, Alfred Wegener Institute, Bremenhaven, Germany
[4]MARUM - Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany

Corresponding author:
Viggó Marteinsson[1]

Email address: viggo@matis.is

ABSTRACT

Marine microorganisms contribute to the health of the global ocean by supporting the marine food web and regulating biogeochemical cycles. Assessing marine microbial diversity is a crucial step towards understanding the global ocean. The waters surrounding Iceland are a complex environment where relatively warm salty waters from the Atlantic cool down and sink down to the deep. Microbial studies in this area have focused on photosynthetic micro- and nanoplankton mainly using microscopy and chlorophyll measurements. However, the diversity and function of the bacterial and archaeal picoplankton remains unknown. Here, we used a co-assembly approach supported by a marine mock community to reconstruct metagenome-assembled genomes (MAGs) from 31 metagenomes from the sea surface and seafloor of four oceanographic sampling stations sampled between 2015 and 2018. The resulting 219 MAGs include 191 bacterial, 26 archaeal and two eukaryotic MAGs to bridge the gap in our current knowledge of the global marine microbiome.

INTRODUCTION

Marine microorganisms are crucial to the global ecosystem as they regulate the carbon cycle (Azam, 1998; Falkowski et al., 2008) and support the marine food web (Pomeroy, 1974; Azam et al., 1983). The study of microorganisms within complex environments, such as the ocean, was accelerated by the emergence of sequencing technologies. In particular, metagenomics – the study of the total genetic material recovered from an environmental sample – have provided previously unavailable information on the functional diversity and ecology of the microbial communities within their environments (Hugenholtz and Tyson, 2008; Quince et al., 2017).

Large-scale metagenomics projects, such as the Global Ocean Sampling (Venter et al., 2004; Rusch et al., 2007), Ocean Sampling Day (Kopf et al., 2015) and Tara Oceans (Sunagawa et al., 2015, 2020), have provided fascinating new insights, but also revealed the gaps in our knowledge of marine microbial species, their geographical distribution, and their organisation in complex and dynamic communities. These and other large-scale initiatives have so far also not covered the oceanic regions around Iceland, a complex marine environment that is characterized by distinct water masses and powerful currents: the cold Polar Water of the East Greenland Current and the Arctic Water of the East Icelandic Current from the north and the warm North Atlantic Water of the Irminger Current from the south (Malmberg et al., 1995; Valdimarsson and Malmberg, 1999). Most microbial studies in Icelandic waters have so far been conducted with traditional methods, like chlorophyll measurements or microscopy, and were therefore mainly focused on larger heterotrophs and photosynthetic microorganisms (Thórdardóttir, 1986; Gudmundsson, 1998; Astthorsson et al., 2007). To establish the baseline knowledge of microbial ecology in Icelandic marine waters, we assembled metagenomic sequence data into draft microbial genomes often called metagenome-assembled genomes (MAGs).

The recovery of MAGs opens the route to further analysis such as comparative genomics to understand the roles of these microorganisms within their community and ecosystem (Sangwan et al., 2016). MAGs are particularly valuable for yet uncultured marine lineages as they reveal the metabolic potential and environmental adaptation of these microorganisms and give clues about trophic interactions and ecology within the environment. Several marine metagenomic studies recovered MAGs from marine environments with - among others - 136 MAGs from the Red Sea (Haroon et al., 2016), 290 from the Mediterranean Sea (Tully et al., 2017), and 2,631 from the global oceans with data harvested by Tara Oceans (Tully et al., 2018).

Here, we report 219 MAGs from 31 samples collected in the Arctic Ocean north of Iceland and in the warmer Atlantic waters south of Iceland. The samples were collected between 2015 and 2018 at four established oceanographic sampling stations visited during six research cruises with two depths sampled at each station. A set of metadata is available for these samples following the best practices recommended by Ten Hoopen et al. (2017), offering an opportunity to further understand the environmental conditions that shape the microbial communities in the waters off the Icelandic coasts.

MATERIALS & METHODS

Sampling

Seawater samples were collected between May 2015 and May 2018 from four stations, two in the North Atlantic Ocean, Selvogsbanki 2 and 5 (SB2 and SB5), and two in the Arctic Ocean, Siglunes 3 and 8 (SI3 and SI8) (Figure 1A and Table 1). Sampling was conducted on board of the oceanographic research vessel Bjarni Sæmundsson RE 30 operated by the Icelandic Marine Research Institute (MRI) by collecting 5 L of seawater from the surface and the seafloor of the ocean, using Niskin bottles on a CTD rosette sampler. Seawater samples were directly filtered onto 0.22 μ m Sterivex filter units (Merck Millipore) and immediately flash frozen in liquid nitrogen before stored at -80 C until further processing (full workflow in Figure 1B).

Mock community

A marine mock community was included in the analysis for quality control, consisting of 20 bacterial and two archaeal species. Strains were cultivated according to Table 2. After 12 to 24 hours of growth (to obtain 10e6 to 10e8 cell/ml), cells were counted on a Thoma cell BRAND (ref. 718020; 0.100 mm depth) to achieve a final concentration of 1.2910e9 cell/L by dilutions. Synthetic seawater was prepared by adding 150 g of sea salts (Sigma-Aldrich, S9883 and 17.25 g of PIPES (Sigma-Aldrich, P1851) to 5 L of autoclaved MilliQ water. The mock community was immediately treated in the same manner as the other seawater samples and filtered onto Sterivex filters for DNA extraction.

DNA extractions

DNA was extracted from all samples using the QIAGEN AllPrep kit according to the manufacturer's instructions with modifications. Sterivex filters were aseptically removed from their plastic casing as described by Cruaud et al. (2017). Filters were transferred to tubes containing 600 μl RTL buffer from the kit and 0.2 g of 0.1 mm zirconia/silica beads (BioSpec, cat. 11079101z) for mechanical disruption of the cells (bead-beating) using a Disrupt MixerMill MM400 by Retsch with the program P9 (300 Hz) three times for 10 seconds each, cooling down tubes in icy water in between each bead-beating step. DNA quality was assessed with a NanoDrop 1000 Spectrophotometer (ThermoFisher) and DNA was quantified with a Qubit fluorometer (Qubit DNA BR assay, Invitrogen).

Library preparation and sequencing

High-throughput sequencing of the samples was performed by Genome Quebec using the HiSeq system (Illumina). Libraries were prepared using NEBNext UltraTM II DNA Library Prep Kit for Illumina (New England Biolabs) followed by sequencing on two lanes of an Illumina HiSeq 4000 PE150 system (Illumina) allocating 1/20 and 1/25 of a lane for each sample. Demultiplexing and conversion to FASTQ files were performed using bcl2fastq Conversion Software v1.8.4 (Illumina) resulting in 32 metagenomic datasets.

Co-assembly and binning

The quality of the raw sequencing reads was assessed using FastQC v0.11.8 (Andrews et al., 2012) (Supplemental Fig. S1). Quality control of the raw reads was performed with Sunbeam v2.0.2 (Clarke et al., 2019) which includes trimming with Trimmomatic v0.36 (Bolger et al., 2014), adapter removal with Cutadapt v2.6 (Martin, 2011) (parameters PE -phred33 ILLUMINACLIP: NexteraPE-PE.fa:2:30:10:8:true LEADING: 3 TRAILING: 3 SLIDINGWINDOW: 4:15 MINLEN: 36), removal of low complexity sequences using Sunbeam Komplexity (default parameter) and removal of contaminating human sequences using the Genome Reference Consortium Human Build 38 patch release 13 GRCh38.p13 (Lander et al., 2001; Schneider et al., 2017). Resulting quality-filtered metagenomic data were divided into surface and seafloor datasets as the surface of the ocean can be considered a different environment compared to the seafloor (Supplemetal Fig. S2). Both datasets also included the mock community. After quality filtering, MEGAHIT v1.2.9 (Li et al., 2015, 2016) (parameters: –min-contig-len 1000 -m 0.85) co-assembled both datasets of samples with a minimum contig length of 1000 bp, resulting in two FASTA files of community contigs. Quality-filtered short reads from each sample were mapped back to the contigs of both co-assemblies respectively using Bowtie v2 (default parameters and –no-unal flag) with default parameters (Langmead and Salzberg, 2012). The resulting SAM files were indexed and converted to BAM files with SAMTOOLS v0.3.3 (parameters: view -F 4 -bS) (Li et al., 2009). For both co-assemblies, the FASTA files containing the contigs were formatted with the script reformat-fasta from Anvi'o v6.2 (Eren et al., 2015). The two contigs databases (the surface and the seafloor databases) were generated with Anvi'o, BAM files were profiled and merged to the respective databases. Automated binning was performed using Anvi'o script anvi-cluster-contigs with default parameters with three binning algorithms: CONCOCT v1.1.0 (Alneberg et al., 2013), MaxBin2 v2.2.6 (Wu et al., 2016), and MetaBAT 2 v2:2.15 (Kang et al., 2019). For all binning results, completeness and redundancy of the bins were estimated with Anvio's script anvi-estimate-genome-completeness which relies on CheckM v1.1.3 (Parks et al., 2015). Based on the comparison of the three binning algorithms, we selected the ”good quality bins” from MetaBAT 2 with an estimated completion above 50% and an estimated redundancy below 10% according to standards suggested by Bowers et al. (2017). The relative proportions of good quality bins in the total number of bins was assessed by chi^2 test.

Functional assignment, taxonomy and phylogenomic trees

We used PRODIGAL v2.6.3 (Hyatt et al., 2010) to identify Open Reading Frames (ORFs) within the contigs. The resulting ORFs were processed with Kaiju v1.7.3 (Menzel et al., 2016) and NCBI nr+euk database (nr_euk 2019-06-25, 46GB, available for download at http://kaiju.binf.ku.dk/server) for taxonomic assignment. Beside the contig-based taxonomic assignment, we used GTDB-Tk v1.3.0 (Genome Taxonomy Database Toolkit) (Chaumeil et al., 2019) to construct two bacterial and two archaeal phylogenomic trees containing good quality MAGs (completeness $\geq$50%; contamination $\leq$10%) and Genome Taxonomy Data Bank (GTDB) R95 (released in July 2020) reference genomes to confirm taxonomic assignments of the MAGs (Parks et al., 2018). The trees were reconstructed using ARB (Ludwig et al., 2004) for comprehensive visualisation.

Data availability

The raw Illumina sequencing paired-end reads are available in the ENA under project accession number PRJEB41565 (ERP125360). MAGs are available under accession numbers ERS5621908 to ERS5622126. Code is available at `https://github.com/clarajegousse/`.

RESULTS

Co-assemblies

The co-assembly of the 16 samples of the surface of the ocean yielded 445,328 contigs, with a minimal length of 1,000 bp, representing a total length of 1.06 Gb (1,060,942,783 nucleotides) with N50 of 2,627 bp and 1,271,859 gene calls (Table 3).

The co-assembly of the 17 samples of the seafloor of the ocean yielded 554,104 contigs, with a minimal length of 1,000 bp, representing a total of length of 1.23 Gb (1,233,390,295 nucleotides) with N50 of 2,327 bp and 1,532,800 gene calls (Table 3).

Binning

A comparison of the three binning algorithms - CONCOCT, MaxBin2 and MetaBAT 2 - was conducted on the surface and seafloor co-assemblies based on the number of good quality bins (Figure 2). Good quality bins have an estimated completion above 50% and an estimated redundancy (also called estimated contamination) below 10% (Bowers et al., 2017). The relative proportions of good quality bins is significantly different for the three binning methods (χ^2 = 135.23, df = 2, p-value $<$ 2.2e-16). The results of the binning showed that MetaBAT 2 resulted in a lower number of bins compared to CONCOCT and MaxBin2. Yet the number of good quality bins was much higher with MetaBAT 2 compared with CONCOCT and MaxBin2.

MetaBAT 2 gave the best results which were used for further analysis and shown in more detail in Figure 3. Out of the 279 bins identified by MetaBAT 2 for the surface samples, 42.4% (118) of them are good quality bins that can be considered draft MAGs according to Bowers et al. (2017). Within the 118 good quality MAGs (Figure 3B), 16 represent genomes of organisms from the mock community and 102 are assembled from the surface seawater. In the same manner, out of the 299 bins identified by MetaBAT 2 for the seafloor samples, 45.81% (134) of can be considered good draft MAGs. Within the 134 good quality MAGs (Figure 3D), 17 represent genomes of organisms from the mock community and 117 are assembled from the seawater at the seafloor. The relative proportions of MAGs out of the total number of bins is the same out of the two co-assemblies datasets (χ^2 = 0.27784, df = 1, p-value = 0.5981) which means that the environments do not seem to impact significantly the number of MAGs. In the same manner, the relative proportions of MAGs associated to the mock community out of the total number of MAGs is the same in the two co-assemblies datasets (χ^2 = 0.0003, df = 1, p-value = 0.9858).

Taxonomy

When excluding members of the mock community based on taxonomic assignment and differential coverage, we identified 102 MAGs reconstructed from the surface co-assembly and 117 MAGs from the seafloor co-assembly. The surface MAGs include two eukaryotes (*Bathycoccus* and *Micromonas*), 92 bacteria, and eight archaea while the seafloor MAGs include 99 bacteria, 18 archaea and no eukaryotes.

The surface co-assembly yielded a total of 92 bacterial MAGs (Figure 4). These MAGs are members of seven phyla (number of MAGs in brackets): Proteobacteria (52), Bacteroidota (31), Actinobacteriota (2), Verrumicrobiota (2), Planctomycetota (2), SAR324 (1) and Cyanobacteria (1). The MAG within the Cyanobacteria phylum belongs to the genus Synechococcus. Within the phylum Actinobacteriota, we retrieved two MAGs: one from a member of the genus *Aquiluna* and one of the genus *Pontimonas*. We reconstructed two MAGs within the phylum Planctomycetota. The two MAGs within the Verrumicrobiota belong to the family Akkermansiaceae. The Bacteroidota phylum includes 31 MAGs reconstructed from the sea surface co-assembly. Most of these Bacteroidota MAGs belong to the Flavobacteriaceae family (18), including one representant of the genus *Polaribacter*. Many MAGs within the Flavobacteriaceae family are related to MAGs revealed by Tara Ocean Consortium such as Cryomorphaceae bacterium and Flavobacteriales bacterium (CFB group bacteria). We also reconstructed 52 MAGs belonging to the phylum of Proteobacteria, including nine Rhodobacteraceae, ten SAR86 and ten Porticoccaceae. Within the three MAGs of the Burkholderiales order, one is within the Burkholderia genus, and the two others belong to the Methylophilaceae family according to GTDB.

The seafloor co-assembly yielded a total of 99 bacterial MAGs spanning across 12 phyla: Proteobacteria (46), Verrumicrobiota (9), Bacteroidota (9), Marinisomatota (8), Actinobacteria (5), Planctomycetota (5), Gemmatimonadota (4), Nitrospinota (3), Chloroflexota (2), SAR324 (2), Myxococcota (1), Lactescibacterota (1). Six of these phyla include exclusively MAGs from the seafloor (Nitrospinota, Myxococcota, Gemmatimonadota, Marinisomatota, Chloroflexa, Lactescibacterota). Within the Proteobacteria, most of the MAGs belong to the Gammaproteobacteria class with 32 MAGSs while the remaining 14 are part of the Alphaproteobacteria. Five orders within the Proteobacteria exclusively include MAGs reconstructed from the seafloor co-assembly (Rhizobiales, Rhodospirillales, TMED109, UBA10353, UBA4486) and none from the surface co-assembly.

Out of the 21 bacterial species of the mock community, 12 of them were re-assembled and given the correct taxonomic assignment down to species level (if available for the strain used) for *Alteromonas sp.*, *Geobacillus marinus*, *Colwellia sp.*, *Escherichia coli*, *Marinobacter sp.*, *Photobacterium sp.*, *Pseudoalteromonas sp.*, *Reinekea marinisedimentorum*, *Sulfitobacter donghicola*, *Sulfitobacter guttiformis*, *Sulfitobacter pontiacus* and *Thermus thermophilus*. However, some distinct species of the mock community that belong to the same genus do not match any specific MAGs but seem to have been reassembled as one single MAG within the genus in question, such as *Reinekea aestuarii* and *Reinekea sp. 84* as well as *Sulfitobacter undariae* and *Sulfitobacter sp. 87*. The genomes of *Bacillus thermoleovorans*, *Dietzia sp.*, *Halomonas sp.* and *Vibrio cyclitrophicus* were not reassembled.

The surface co-assembly yielded only eight archaeal MAGs (Figure 5), all within the Thermoplasmota phylum, including three MAGs within the genus MGIIb-O2 of the Thalassarchaeaceae family and five within the Poseidoniaceae family. The seafloor co-assembly resulted in 18 archaeal MAGs including one representant of the Thermoproteota phylum: this MAGs belongs to the UBA57 phylum within the order of the Nitrososphaerales. The 17 other archaeal MAGs are all comprised in the Thermoplasmatota phylum, within the class Poseidoniia, including representatives of the Poseidoniaceae and Thalassarchaeaceae families. The two archaeal members within the mock community (*Pyrococcus abyssi* and *Thermococcus barophilus*) were successfully reconstructed in both co-assemblies.

DISCUSSION

Mock communities are used to quantify and characterise biases introduced in the sample processing pipeline (Brooks et al., 2015) and are indispensable to benchmark sequencing methods and downstream analysis (Singer et al., 2016; Sevim et al., 2019). Mock communities can also be used as a positive control for metagenomic studies. Our mock community confirmed that MetaBAT 2 was able to resolve genomes of species within the same genus, thus making it the most suitable binning algorithms out of the three tested in this study: CONCOCT, MaxBin2 and MetaBAT 2. This result is consistent with previous studies (Yue et al., 2020).

The ocean is a vast continuum and the samples were taken within a relatively small section/fraction of the North Atlantic Ocean at several sampling depths: the surface and the seafloor (90 m, 470 m, 1,006 m, and 1,060 m depending on the station). The differences in the sampling depth implies differences in lighting, pressure and temperature compared to the surface of the ocean. While the surface of the ocean is subjected to seasonal variations in day light and temperature, the seafloor remains darker and colder than the surface, and such parameters are driving microbial community structure and function. Therefore, we considered the surface and the seafloor of the ocean as two different types of environments which justifies our approach of two co-assemblies rather than assembling all of the 32 samples together. The fact that a number of MAGs were exclusively found in only one of the two environments, confirmed this.

CONCLUSIONS

The goal of this study was to reconstruct MAGs from 31 samples from Icelandic sea waters. The 219 MAGs span across 13 bacterial and two archaeal phyla and contribute to a more define picture of the global marine microbiome. Moreover, this study confirms, thanks to the inclusion of a mock community in the analysis, that the combination of co-assembly and binning with MetaBAT 2 allows, despite a relatively shallow sequencing depth, the recovery of quality MAGs that are a precious resource for further ecological and environmental studies.

ACKNOWLEDGMENTS

The authors would like to thank Kristinn Gudmundsson and Bjarni Saemundsson's crew from the Marine Research Institute, and Pauline Bergsten and Mia Cerfonteyn from the University of Iceland & Matís for sampling, Antonio Fernandez Guerra from the Max Plank Institute for Marine Microbiology and Arnar Pálsson from the University of Iceland for advice and Elvar Örn Jónsson from the University of Iceland for technical support. The analyses presented in the study were performed using the resources provided by the Icelandic High Performance Computing Centre at the University of Iceland.

REFERENCES

Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Loman, N. J., Andersson, A. F., and Quince, C. (2013). Concoct: clustering contigs on coverage and composition. *arXiv preprint arXiv:1312.4038*.

Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C., and Wingett, S. (2012). FastQC. Babraham Institute.

Astthorsson, O. S., Gislason, A., and Jonsson, S. (2007). Climate variability and the icelandic marine ecosystem. *Deep Sea Research Part II: Topical Studies in Oceanography*, 54(23-26):2456–2477.

Azam, F. (1998). Microbial control of oceanic carbon flux: the plot thickens. *Science*, 280(5364):694–696.

Azam, F., Fenchel, T., Field, J. G., Gray, J., Meyer-Reil, L., and Thingstad, F. (1983). The ecological role of water-column microbes in the sea. *Marine ecology progress series*, pages 257–263.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, 30(15):2114–2120.

Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., Schulz, F., Jarett, J., Rivers, A. R., Eloe-Fadrosh, E. A., Tringe, S. G., Ivanova, N. N., Copeland, A., Clum, A., Becraft, E. D., Malmstrom, R. R., Birren, B., Podar, M., Bork, P., Weinstock, G. M., Garrity, G. M., Dodsworth, J. A., Yooseph, S., Sutton, G., Glöckner, F. O., Gilbert, J. A., Nelson, W. C., Hallam, S. J., Jungbluth, S. P., Ettema, T. J. G., Tighe, S., Konstantinidis, K. T., Liu, W.-T., Baker, B. J., Rattei, T., Eisen, J. A., Hedlund, B., McMahon, K. D., Fierer, N., Knight, R., Finn, R., Cochrane, G., Karsch-Mizrachi, I., Tyson, G. W., Rinke, C., Schriml, L., Hugenholtz, P., Yilmaz, P., Meyer, F., Lapidus, A., Parks, D. H., Murat Eren, A., Banfield, J. F., Woyke, T., and Consortium, T. G. S. (2017). Minimum information about a single amplified genome (misag) and a metagenome-assembled genome (mimag) of bacteria and archaea. *Nature Biotechnology*, 35(8):725–731.

Brooks, J. P., Edwards, D. J., Harwich, M. D., Rivera, M. C., Fettweis, J. M., Serrano, M. G., Reris, R. A., Sheth, N. U., Huang, B., Girerd, P., Strauss, J. F., Jefferson, K. K., Buck, G. A., and (additional members), V. M. C. (2015). The truth about metagenomics: quantifying and counteracting bias in 16s rrna studies. *BMC Microbiology*, 15(1):66.

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2019). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, 36(6):1925–1927.

Clarke, E. L., Taylor, L. J., Zhao, C., Connell, A., Lee, J.-J., Fett, B., Bushman, F. D., and Bittinger, K. (2019). Sunbeam: an extensible pipeline for analyzing metagenomic sequencing experiments. *Microbiome*, 7(1):46.

Cruaud, P., Vigneron, A., Fradette, M.-S., Charette, S. J., Rodriguez, M. J., Dorea, C. C., and Culley, A. I. (2017). Open the sterivex casing: An easy and effective way to improve dna extraction yields. *Limnology and Oceanography: Methods*, 15(12):1015–1020.

Erauso, G., Reysenbach, A.-L., Godfroy, A., Meunier, J.-R., Crump, B., Partensky, F., Baross, J. A., Marteinsson, V., Barbier, G., Pace, N. R., and Prieur, D. (1993). Pyrococcus abyssi sp. nov., a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *Archives of Microbiology*, 160(5):338–349.

Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., and Delmont, T. O. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ*, 3:e1319.

Falkowski, P. G., Fenchel, T., and Delong, E. F. (2008). The microbial engines that drive earth's biogeochemical cycles. *science*, 320(5879):1034–1039.

Gudmundsson, K. (1998). Long-term variation in phytoplankton productivity during spring in icelandic waters. *ICES Journal of Marine Science*, 55(4):635–643.

Haroon, M. F., Thompson, L. R., Parks, D. H., Hugenholtz, P., and Stingl, U. (2016). A catalogue of 136 microbial draft genomes from red sea metagenomes. *Scientific data*, 3(1):1–6.

Hugenholtz, P. and Tyson, G. W. (2008). Metagenomics. *Nature*, 455(7212):481–483.

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics*, 11(1):119.

Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019). Metabat 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, 7:e7359.

Kopf, A., Bicak, M., Kottmann, R., Schnetzer, J., Kostadinov, I., Lehmann, K., Fernandez-Guerra, A., Jeanthon, C., Rahav, E., Ullrich, M., Wichels, A., Gerdts, G., Polymenakou, P., Kotoulas, G., Siam, R., Abdallah, R. Z., Sonnenschein, E. C., Cariou, T., O'Gara, F., Jackson, S., Orlic, S., Steinke, M., Busch, J., Duarte, B., Caçador, I., Canning-Clode, J., Bobrova, O., Marteinsson, V., Reynisson, E., Loureiro, C. M., Luna, G. M., Quero, G. M., Löscher, C. R., Kremp, A., DeLorenzo, M. E., Øvreås, L., Tolman, J., LaRoche, J., Penna, A., Frischer, M., Davis, T., Katherine, B., Meyer, C. P., Ramos, S., Magalhães, C., Jude-Lemeilleur, F., Aguirre-Macedo, M. L., Wang, S., Poulton, N., Jones, S., Collin, R., Fuhrman, J. A., Conan, P., Alonso, C., Stambler, N., Goodwin, K., Yakimov, M. M., Baltar, F., Bodrossy, L., Van De Kamp, J., Frampton, D. M., Ostrowski, M., Van Ruth, P., Malthouse, P., Claus, S., Deneudt, K., Mortelmans, J., Pitois, S., Wallom, D., Salter, I., Costa, R., Schroeder, D. C., Kandil, M. M., Amaral, V., Biancalana, F., Santana, R., Pedrotti, M. L., Yoshida, T., Ogata, H., Ingleton, T., Munnik, K., Rodriguez-Ezpeleta, N., Berteaux-Lecellier, V., Wecker, P., Cancio, I., Vaulot, D., Bienhold, C., Ghazal, H., Chaouni, B., Essayeh, S., Ettamimi, S., Zaid, E. H., Boukhatem, N., Bouali, A., Chahboune, R., Barrijal, S., Timinouni, M., El Otmani, F., Bennani, M., Mea, M., Todorova, N., Karamfilov, V., ten Hoopen, P., Cochrane, G., L'Haridon, S., Bizsel, K. C., Vezzi, A., Lauro, F. M., Martin, P., Jensen, R. M., Hinks, J., Gebbels, S., Rosselli, R., De Pascale, F., Schiavon, R., dos Santos, A., Villar, E., Pesant, S., Cataletto, B., Malfatti, F., Edirisinghe, R., Silveira, J. A. H., Barbier, M., Turk, V., Tinta, T., Fuller, W. J., Salihoglu, I., Serakinci, N., Ergoren, M. C., Bresnan, E., Iriberri, J., Nyhus, P. A. F., Bente, E., Karlsen, H. E., Golyshin, P. N., Gasol, J. M., Moncheva, S., Dzhembekova, N., Johnson, Z., Sinigalliano, C. D., Gidley, M. L., Zingone, A., Danovaro, R., Tsiamis, G., Clark, M. S., Costa, A. C., El Bour, M., Martins, A. M., Collins, R. E., Ducluzeau, A.-L., Martinez, J., Costello, M. J., Amaral-Zettler, L. A., Gilbert, J. A., Davies, N., Field, D., and Glöckner, F. O. (2015). The ocean sampling day consortium. *GigaScience*, 4(1):27.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.-F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H.-C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G. R., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F. A., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S.-P., Yeh, R.-F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Patrinos, A., Morgan, M. J., Consortium, I. H. G. S., Whitehead Institute for Biomedical Research, C. f. G. R., Centre:, T. S., Center, W. U. G. S., Institute:, U. D. J. G., of Medicine Human Genome Sequencing Center:, B. C., Center:, R. G. S., Genoscope, UMR-8030:, C., Department of Genome Analysis, I. o. M. B., Center:, G. S., Center:, B. G. I. G., Multimegabase Sequencing Center, T. I. f. S. B., Center:, S. G. T., of Oklahoma's Advanced Center for Genome Technology:, U., for Molecular Genetics:, M. P. I., Cold Spring Harbor Laboratory, L. A. H. G. C., for Biotechnology:, G.-G. R. C., *Genome Analysis Group (listed in alphabetical order,

a. i. i. l. u. o. h., Scientific management: National Human Genome Research Institute, U. N. I. o. H., Center:, S. H. G., of Washington Genome Center:, U., Department of Molecular Biology, K. U. S. o. M., of Texas Southwestern Medical Center at Dallas:, U., Office of Science, U. D. o. E., and Trust:, T. W. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nature methods*, 9(4):357.

Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). Megahit: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de bruijn graph. *Bioinformatics*, 31(10):1674–1676.

Li, D., Luo, R., Liu, C.-M., Leung, C.-M., Ting, H.-F., Sadakane, K., Yamashita, H., and Lam, T.-W. (2016). Megahit v1. 0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods*, 102:3–11.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–2079.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Förster, W., Brettske, I., Gerber, S., Ginhart, A. W., Gross, O., Grumann, S., Hermann, S., Jost, R., König, A., Liss, T., Lüßmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A., and Schleifer, K. (2004). ARB: a software environment for sequence data. *Nucleic Acids Research*, 32(4):1363–1371.

Malmberg, S.-A., Valdimarsson, H., and Mortensen, J. (1995). Long time series in icelandic waters, in relation to physical variability in the northern north atlantic. *Ocean Challenge*, 6:48–51.

Marteinsson, V. T., Birrien, J.-L., Reysenbach, A.-L., Vernet, M., Marie, D., Gambacorta, A., Messner, P., Sleytr, U. B., and Prieur, D. (1999). Thermococcus barophilus sp. nov., a new barophilic and hyperthermophilic archaeon isolated under high hydrostatic pressure from a deep-sea hydrothermal vent. *International Journal of Systematic and Evolutionary Microbiology*, 49(2):351–359.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, 17(1):10–12.

Menzel, P., Ng, K. L., and Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with kaiju. *Nature communications*, 7(1):1–9.

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., and Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature biotechnology*, 36(10):996–1004.

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). Checkm: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research*, 25(7):1043–1055.

Pomeroy, L. R. (1974). The ocean's food web, a changing paradigm. *Bioscience*, 24(9):499–504.

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., and Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature biotechnology*, 35(9):833–844.

Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., Wu, D., Eisen, J. A., Hoffman, J. M., Remington, K., Beeson, K., Tran, B., Smith, H., Baden-Tillson, H., Stewart, C., Thorpe, J., Freeman, J., Andrews-Pfannkoch, C., Venter, J. E., Li, K., Kravitz, S., Heidelberg, J. F., Utterback, T., Rogers, Y.-H., Falcón, L. I., Souza, V., Bonilla-Rosso, G., Eguiarte, L. E., Karl, D. M., Sathyendranath, S., Platt, T., Bermingham, E., Gallardo, V., Tamayo-Castillo, G., Ferrari, M. R., Strausberg, R. L., Nealson, K., Friedman, R., Frazier, M., and Venter, J. C. (2007). The sorcerer ii global ocean sampling expedition: Northwest atlantic through eastern tropical pacific. *PLOS Biology*, 5(3):1–34.

Sangwan, N., Xia, F., and Gilbert, J. A. (2016). Recovering complete and draft population genomes from metagenome datasets. *Microbiome*, 4(1):8.

Schneider, V. A., Graves-Lindsay, T., Howe, K., Bouk, N., Chen, H.-C., Kitts, P. A., Murphy, T. D., Pruitt, K. D., Thibaud-Nissen, F., Albracht, D., Fulton, R. S., Kremitzki, M., Magrini, V., Markovic, C., McGrath, S., Steinberg, K. M., Auger, K., Chow, W., Collins, J., Harden, G., Hubbard, T., Pelan, S., Simpson, J. T., Threadgold, G., Torrance, J., Wood, J. M., Clarke, L., Koren, S., Boitano, M., Peluso, P., Li, H., Chin, C.-S., Phillippy, A. M., Durbin, R., Wilson, R. K., Flicek, P., Eichler, E. E., and Church, D. M. (2017). Evaluation of grch38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome research*, 27(5):849–864.

Sevim, V., Lee, J., Egan, R., Clum, A., Hundley, H., Lee, J., Everroad, R. C., Detweiler, A. M., Bebout, B. M., Pett-Ridge, J., Göker, M., Murray, A. E., Lindemann, S. R., Klenk, H.-P., O'Malley, R., Zane, M., Cheng, J.-F., Copeland, A., Daum, C., Singer, E., and Woyke, T. (2019). Shotgun metagenome data of a defined mock community using oxford nanopore, pacbio and illumina technologies. *Scientific Data*, 6(1):285.

Singer, E., Andreopoulos, B., Bowers, R. M., Lee, J., Deshpande, S., Chiniquy, J., Ciobanu, D., Klenk, H.-P., Zane, M., Daum, C., Clum, A., Cheng, J.-F., Copeland, A., and Woyke, T. (2016). Next generation sequencing data of a defined microbial mock community. *Scientific Data*, 3(1):160081.

Sunagawa, S., Acinas, S. G., Bork, P., Bowler, C., Acinas, S. G., Babin, M., Boss, E., Cochrane, G., de Vargas, C., Follows, M., Gorsky, G., Grimsley, N., Guidi, L., Hingamp, P., Iudicone, D., Jaillon, O., Kandels, S., Karp-Boss, L., Karsenti, E., Lescot, M., Not, F., Ogata, H., Pesant, S., Poulton, N., Raes, J., Sardet, C., Sieracki, M., Speich, S., Stemmann, L., Sullivan, M. B., Wincker, P., Eveillard, D., Lombard, F., Pesant, S., Sullivan, M. B., and Tara Oceans Coordinators (2020). Tara oceans: towards global ocean ecosystems biology. *Nature Reviews Microbiology*, 18(8):428–445.

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D. R., Alberti, A., Cornejo-Castillo, F. M., Costea, P. I., Cruaud, C., d'Ovidio, F., Engelen, S., Ferrera, I., Gasol, J. M., Guidi, L., Hildebrand, F., Kokoszka, F., Lepoivre, C., Lima-Mendez, G., Poulain, J., Poulos, B. T., Royo-Llonch, M., Sarmento, H., Vieira-Silva, S., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Bowler, C., de Vargas, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Jaillon, O., Not, F., Ogata, H., Pesant, S., Speich, S., Stemmann, L., Sullivan, M. B., Weissenbach, J., Wincker, P., Karsenti, E., Raes, J., Acinas, S. G., and Bork, P. (2015). Structure and function of the global ocean microbiome. *Science*, 348(6237).

Ten Hoopen, P., Finn, R. D., Bongo, L. A., Corre, E., Fosso, B., Meyer, F., Mitchell, A., Pelletier, E., Pesole, G., Santamaria, M., Willassen, N. P., and Cochrane, G. (2017). The metagenomic data life-cycle: standards and best practices. *Gigascience*, 6(8):1–11.

Thórdardóttir, T. (1986). Timing and duration of spring blooming south and southwest of iceland. In *The role of freshwater outflow in coastal marine ecosystems*, pages 345–360. Springer.

Tully, B. J., Graham, E. D., and Heidelberg, J. F. (2018). The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Scientific data*, 5:170203.

Tully, B. J., Sachdeva, R., Graham, E. D., and Heidelberg, J. F. (2017). 290 metagenome-assembled genomes from the mediterranean sea: a resource for marine microbiology. *PeerJ*, 5:e3558.

Valdimarsson, H. and Malmberg, S.-A. (1999). Near-surface circulation in icelandic waters derived from satellite tracked drifters. *Rit Fiskideild*, 16:23–40.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.-H., and Smith, H. O. (2004). Environmental genome shotgun sequencing of the sargasso sea. *Science*, 304(5667):66–74.

Wu, Y.-W., Simmons, B. A., and Singer, S. W. (2016). Maxbin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*, 32(4):605–607.

Yue, Y., Huang, H., Qi, Z., Dou, H.-M., Liu, X.-Y., Han, T.-F., Chen, Y., Song, X.-J., Zhang, Y.-H., and Tu, J. (2020). Evaluating metagenomics tools for genome binning with real metagenomic datasets and cami datasets. *BMC bioinformatics*, 21(1):1–15.

Table 1. Sampling dates and locations with corresponding seawater temperature and salinity.

Sampling date	Station ID	Latitude (dd.mm)	Longitude (dd.mm)	Depth (m)	Temperature (C)	Salinity (PSU)
23.05.2015	SI8	67.9993	-18.8313	1,045	-0.481	34.913
30.05.2015	SB5	62.9822	-21.4737	0	7.632	35.195
30.05.2015	SB5	62.9822	-21.4737	1,004	4.391	34.998
23.05.2016	SI8	68.0100	-18.8247	0	1.632	34.869
23.05.2016	SI8	68.0100	-18.8247	1,045	-0.431	34.914
31.05.2016	SB5	62.9936	-21.4839	0	8.147	35.113
31.05.2016	SB5	62.9936	-21.4839	1,004	4.722	35.017
21.05.2017	SI8	68.0094	-18.8325	1,045	2.700	34.852
21.05.2017	SI8	68.0094	-18.8325	0	-0.381	34.914
22.05.2017	SI3	66.5342	-18.8378	470	5.517	34.492
22.05.2017	SI3	66.5342	-18.8378	0	0.151	34.906
30.05.2017	SB5	62.9878	-21.4800	1,004	8.477	34.761
30.05.2017	SB5	62.9878	-21.4800	0	4.801	35.009
09.08.2017	SI3	66.5344	-18.8419	0	9.980	34.310
09.08.2017	SI3	66.5344	-18.8419	470	0.190	34.900
09.08.2017	SI8	68.0006	-18.8375	1,045	7.640	34.650
09.08.2017	SI8	68.0006	-18.8375	0	-0.370	34.910
18.08.2017	SB2	63.4933	-20.9569	0	12.000	33.700
18.08.2017	SB2	63.4933	-20.9569	90	8.470	34.940
18.08.2017	SB5	62.9883	-21.4867	0	12.200	34.980
18.08.2017	SB5	62.9883	-21.4867	1,004	4.730	35.010
16.02.2018	SI3	66.5442	-18.8400	470	0.044	34.901
16.02.2018	SI8	68.0000	-18.8386	0	0.533	34.640
16.02.2018	SI8	68.0000	-18.8386	1,045	-0.410	34.914
18.05.2018	SI8	68.0058	-18.8256	0	1.355	34.727
18.05.2018	SI8	68.0058	-18.8256	1,045	-0.428	34.914
20.05.2018	SI3	66.5439	-18.8406	0	5.108	34.894
29.05.2018	SB2	63.4942	-20.9008	0	7.625	34.913
29.05.2018	SB2	63.4942	-20.9008	90	7.298	35.031
29.05.2018	SB5	62.9858	-21.4731	0	7.740	35.042
29.05.2018	SB5	62.9858	-21.4731	1,004	4.488	34.978

Table 2. List of bacterial and archaeal species in the mock community. Strains were obtained from the Icelandic Strain Collection and Records (ISCAR) or the German Collection of Microorganisms and Cell Cultures (DSMZ). Recipes for growth media can be found at https://www.dsmz.de/ if not otherwise indicated.

Domain	Species name	% identity	Collection number	Growth parameters	Successfully reassembled
Bacteria	*Alteromonas naphthalenivorans*	99.66%	ISCAR-05201	Marine Broth, 22 C, pH 6.8, aerobic condition	Yes
Bacteria	*Jeotgalibacillus marinus*	100%	ISCAR-03118	Marine Broth, 22 C, pH 6.8, aerobic condition	No
Bacteria	*Geobacillus thermoleovorans*	100%	ISCAR-00004	162 media, 65 C, pH 7.0, aerobic condition	No
Bacteria	*Colwellia psychrerythraea*	99%	ISCAR-05175	Marine Broth, 22 C, pH 6.8, aerobic condition	Yes
Bacteria	*Dietzia psychralcaliphila*	99.52%	ISCAR-05191	92 media, 22 C, pH 6.8, aerobic condition	No
Bacteria	*Escherichia coli*	100%	ISCAR-02961	LB media, 37 C, pH 7.0, aerobic condition	Yes
Bacteria	*Pseudomonas salina*	99.83%	ISCAR-05249	Marine Broth media, 22 C, pH 6.8, aerobic condition	No
Bacteria	*Marinobacter psychrophilus*	99.84%	ISCAR-05186	Marine Broth media, 22 C, pH 6.8, aerobic condition	Yes
Bacteria	*Photobacterium indicum*	100%	ISCAR-05002	Marine Broth media, 22 C, pH 6.8, aerobic condition	Yes
Bacteria	*Pseudoalteromonas neustonica*	98.58%	ISCAR-05312	172 media, 22 C, pH 6.8, aerobic condition	Yes
Bacteria	*Reinekea aestuarii*	100%	DSM 29881	Marine Broth media, 22 C, pH 6.8, aerobic condition	No
Bacteria	*Reinekea marinisedimentorum*	100%	DSM 15388	Marine Broth media, 30 C, pH 6.8, aerobic condition	Yes
Bacteria	*Rhodococcus kyotonensis*	99.23%	ISCAR-05221	Marine Broth media,22 C, pH 6.8, aerobic condition	No
Bacteria	*Reinekea sp. 84*	97.75% with *Reinekea marina*	ISCAR-05258	Marine Broth media, 22 C, pH 6.8, aerobic condition	No
Bacteria	*Sulfitobacter sp. 87*	97.73% with *Sulfitobacter donghicola*	ISCAR-05261	Marine Broth media, 22 C, pH 6.8, aerobic condition	No
Bacteria	*Sulfitobacter donghicola*	100%	DSM 23563	Marine Broth media, 22 C, pH 6.8, aerobic condition	Yes
Bacteria	*Sulfitobacter guttiformis*	100%	DSM 11544	Marine Broth media, 22 C, pH 6.8, aerobic condition	Yes
Bacteria	*Sulfitobacter pontiacus*	100%	DSM 10014	Marine Broth media, 22 C, pH 6.8, aerobic condition	Yes
Bacteria	*Sulfitobacter undariae*	100%	DSM 102234	Marine Broth media, 22 C, pH 6.8, aerobic condition	No
Bacteria	*Thermus thermophilus*	100%	ISCAR-03915	166 media, 65 C, pH 7.0, aerobic condition	No
Bacteria	*Vibrio cyclitrophicus*	100%	ISCAR-06209	Marine Broth media, 22 C, pH 6.8, aerobic condition	No
Archaea	*Pyrococcus abyssi*	100%	DSM 25543	YPS[1] media, 90 C, pH 7, anaerobic condition, elemental sulfur	Yes
Archaea	*Thermococcus barophilus*	100%	DSM 11836	TRM[2], 85 C, pH 6.5, anaerobic condition, elemental sulfur	Yes

Growth media recipes in: [1] Erauso et al. (1993); [2]Marteinsson et al. (1999)

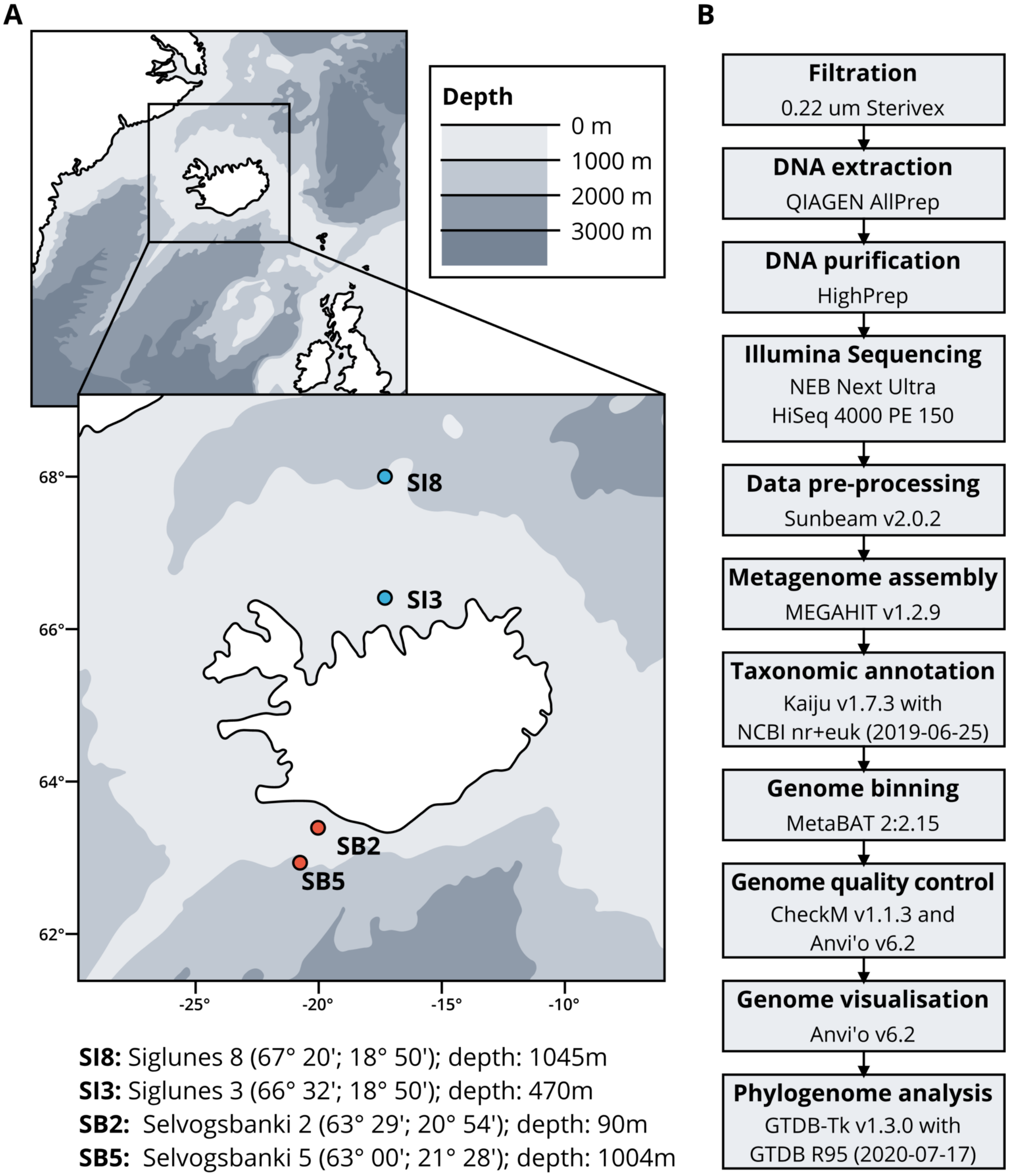

Figure 1. A, Sampling stations location and coordinates. B, Workflow of bio-molecular processes and downstream analysis.

Table 3. Statistics summary of co-assemblies.

	Surface	Seafloor
Total nucleotides	1.06 Gb	1.23 Gb
N50	2,382 bp	2,327 bp
L50	83,272 bp	114,549 bp
Number of contigs	445,328	554,104
Longest contig	864,343 bp	1,302,516 bp
Shortest contig	1,000 bp	1,000 bp
Number of contigs >10 kb	8,521	8,306
Number of genes (Prodigal)	1,271,859	1,532,800

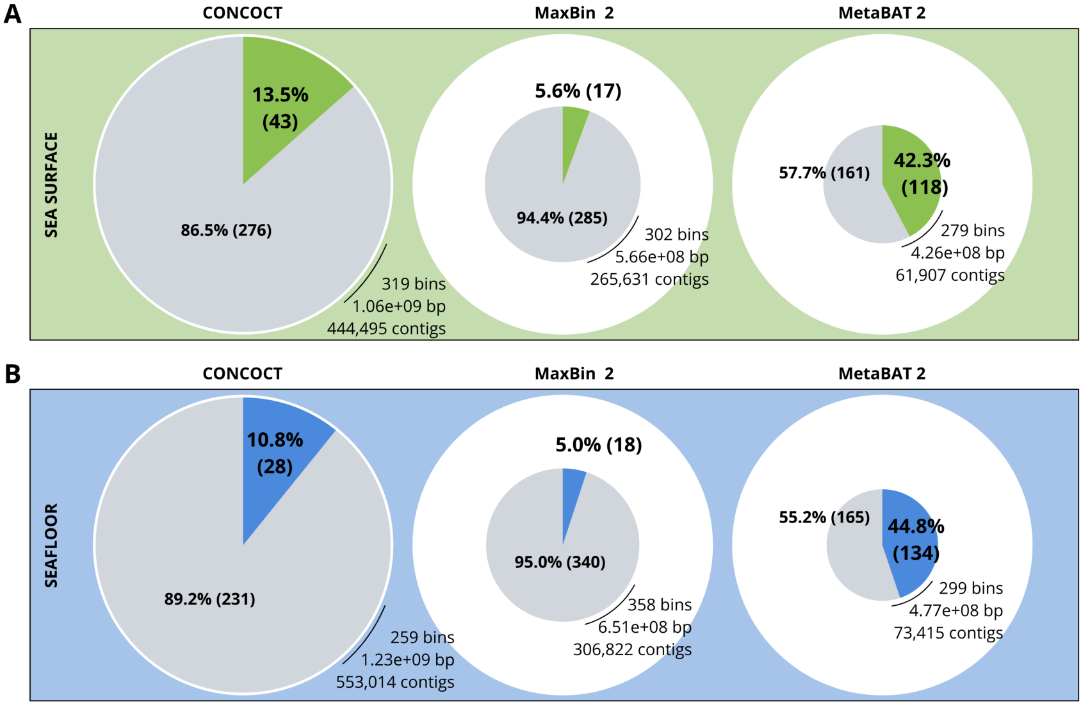

Figure 2. Binning comparison. Numbers of contigs binned and numbers of bad and good quality bins obtained with CONCOCT, MaxBin2 and MetaBAT 2 from the surface co-assembly (A) and the seafloor co-assembly (B). Numbers of contigs binned is represented by the size of the pie plots. Numbers and percentages of bad quality bins and good quality bins are shown within the grey and coloured slices of the chart respectively. Good quality bins have an estimated completion above 50% and an estimated redundancy (also called estimated contamination) below 10% (Bowers et al., 2017).

Table 4. Statistics summary of co-assemblies.

Co-assembly	Binning method	Number of bins	Number of MAGs	Average Completeness (%)	Average Contamination (%)
Surface	CONCOCT	319	43	45.15	49.23
Surface	MaxBin2	302	17	25.77	13.30
Surface	MetaBAT 2	279	118	44.12	3.46
Seafloor	CONCOCT	259	28	51.26	90.39
Seafloor	MaxBin2	358	18	34.59	18.63
Seafloor	MetaBAT 2	299	134	49.90	7.13

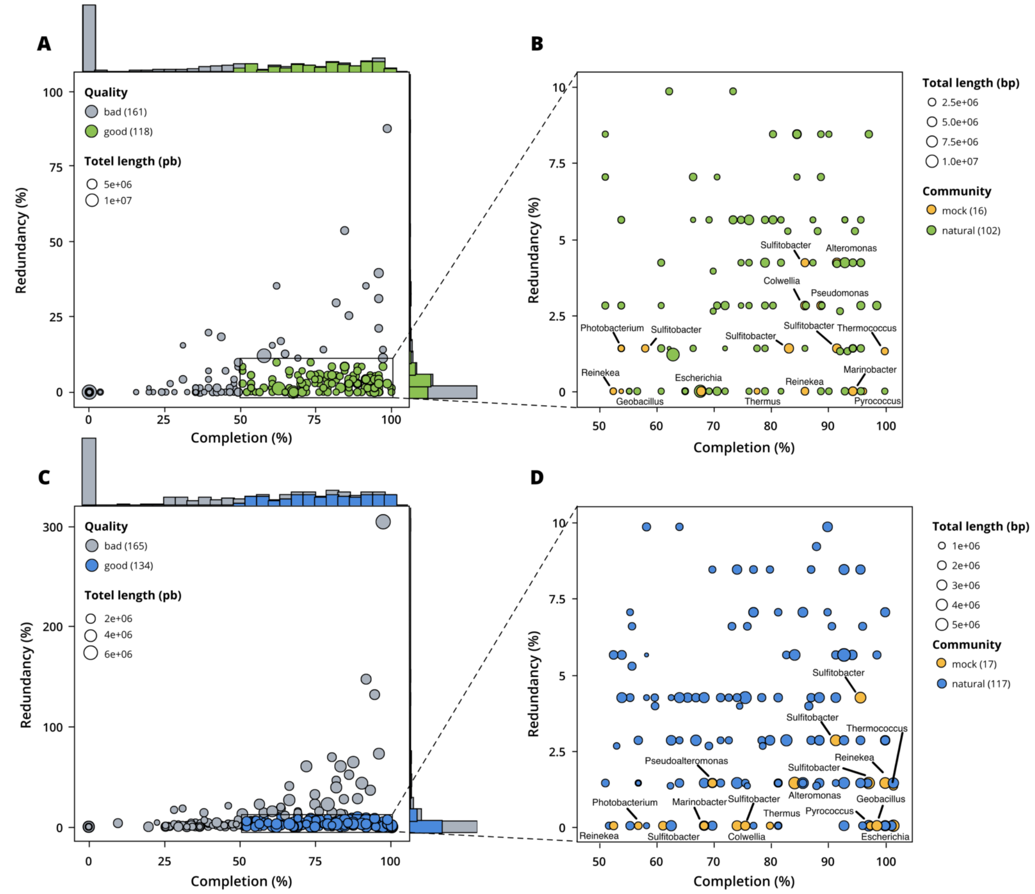

Figure 3. Assessment of bin quality with the estimated completeness as a function of the redundancy. Bad quality bins (completeness below 50% and redundancy above 10%) are shown in grey while good quality bins are in colours (green for surface, blue for seafloor samples). A) 279 bins obtained with MetaBAT 2 from the surface co-assembly with 118 good quality bins. B) Good quality bins from the surface co-assembly with the identification bins corresponding to members of the mock community. C) 299 bins obtained with MetaBAT 2 from the seafloor co-assembly with 134 good quality bins. D) Good quality bins from the seafloor with the identification of the bins corresponding to members of the mock community.

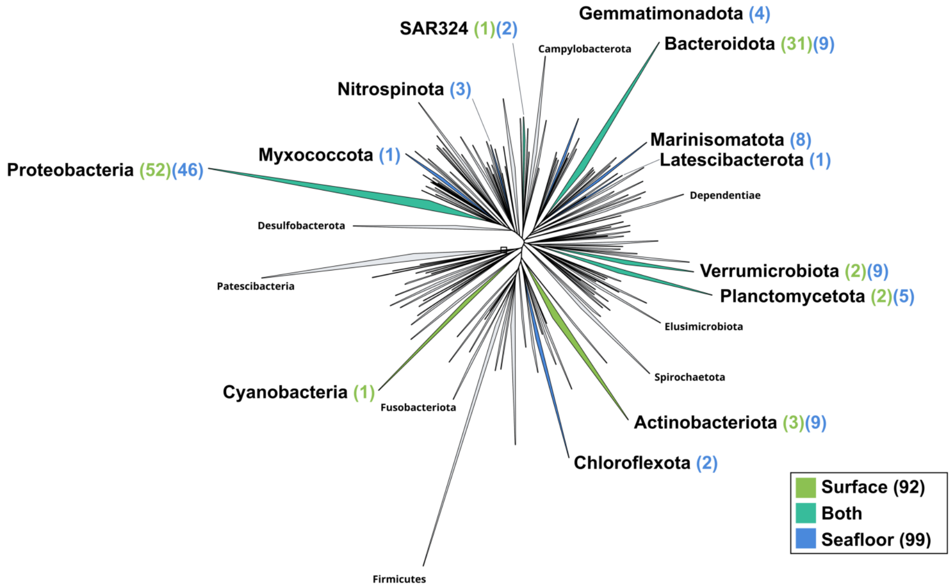

Figure 4. Bacterial phylogenomic tree. Distribution of the Marine Icelandic MAGs across 76 bacterial phyla from GTDB. The maximum likelihood tree was inferred from the concatenation of 120 proteins spanning a dereplicated set of 191,527 bacterial genomes (GTDB 05-RS95 released on the 17th July 2020) and the Marine Icelandic MAGs. Phyla containing MAGs from the surface seawater, seafloor or both are shown in green, blue or teal respectively. Number of Marine Icelandic MAGs from the surface and the seafloor in each phylum are indicated in between parenthesis in green and blue respectively.

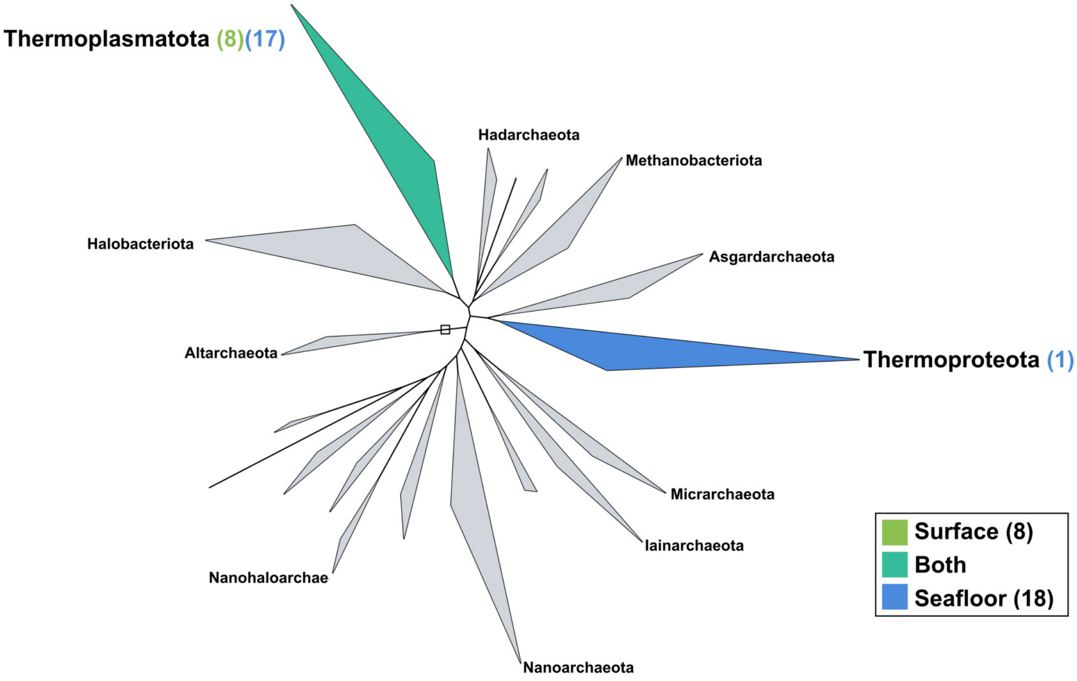

Figure 5. Archaeal phylogenomic tree. Distribution of the Marine Icelandic MAGs across 18 archaeal phyla from GTDB. The maximum likelihood tree was inferred from the concatenation of 122 proteins spanning a dereplicated set of 3,073 archaeal genomes (GTDB 05-RS95 released on the 17th July 2020) and the Marine Icelandic MAGs. Phyla containing MAGs from the surface seawater, seafloor or both are shown in green, blue or teal respectively. Number of Marine Icelandic MAGs from the surface and the seafloor in each phylum are indicated in between parenthesis in green and blue respectively.