A Collection of 13 Archaeal and 46 Bacterial Genomes Reconstructed from Marine Metagenomes Derived from the North Sea

Bernd Wemheuer

Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, University of Göttingen, D-37077 Göttingen, Germany; bwemheu@gwdg.de

Abstract: Marine bacteria are key drivers of ocean biogeochemistry. Despite the increasing number of studies, the complex interaction of marine bacterioplankton communities with their environment is still not fully understood. Additionally, our knowledge about prominent marine lineages is mostly based on genomic information retrieved from single isolates, which do not necessarily represent these groups. Consequently, deciphering the ecological contributions of single bacterioplankton community members is one major challenge in marine microbiology. In the present study, we reconstructed 13 archaeal and 46 bacterial metagenome-assembled genomes (MAGs) from four metagenomic data sets derived from the North Sea. Archaeal MAGs were affiliated to Marine Group II within the Euryarchaeota. Bacterial MAGs mainly belonged to marine groups within the Bacteroidetes as well as alpha- and gammaproteobacteria. In addition, two bacterial MAGs were classified as members of the Actinobacteria and Verrucomicrobiota, respectively. The reconstructed genomes contribute to our understanding of important marine lineages and may serve as a basis for further research on functional traits of these groups.

Dataset: The metagenome-assembled genomes have been deposited in NCBI GenBank under the accessions QXXR00000000-QXZX00000000 (submission id SUB4359442). For further details see Supplementary Table S1.

Dataset License: CC-BY.

---

1. Summary

The present data set comprises of 59 archaeal and bacterial metagenome-assembled genomes (MAGs). These MAGs were reconstructed from four metagenomic data sets derived from the North Sea. Four seawater samples were taken at three different sites in the North Sea at 3 and 350 m, respectively. Free-living planktonic communities were harvested from these samples by serial filtration. Environmental DNA was extracted from harvested microbial communities and subjected to next-generation sequencing. The obtained sequencing data were quality filtered and scanned for contaminations prior to metagenome assembly. MAGs were reconstructed from the assembled metagenomic datasets using two independent binning approaches and a subsequent refinement. A total of thirteen archaeal MAGs affiliated to Marine Group II (MGII) and 46 bacterial MAGs mainly assigned to marine groups within the Bacteroidetes as well as alpha- and gammaproteobacteria were extracted. All MAGs have been deposited in Genbank. They provide a basis for further studies aiming at understanding the complex interaction of important marine lineage with their environment.

2. Data Description

Here, we report 59 MAGs extracted from 4 marine water samples taken in the North Sea (Figure 1). The North Sea is a typical coastal shelf sea. Coastal shelf seas of the temperate zone are highly productive because of the continuous nutrient supply by rivers. The North Sea is connected to the Atlantic Ocean via the English Channel in the South and the Norwegian Sea in the North. The southern part has a water depth of less than 50 m and is subjected to strong tidal currents. Nutrient suspension from the sediment and loss of water stratification are results of these currents. The northern part of the North Sea is deeper (up to 725 m) and strong tidal currents do not occur. The southern region has especially undergone high nutrient loading and warming during the last 40 years [1,2].
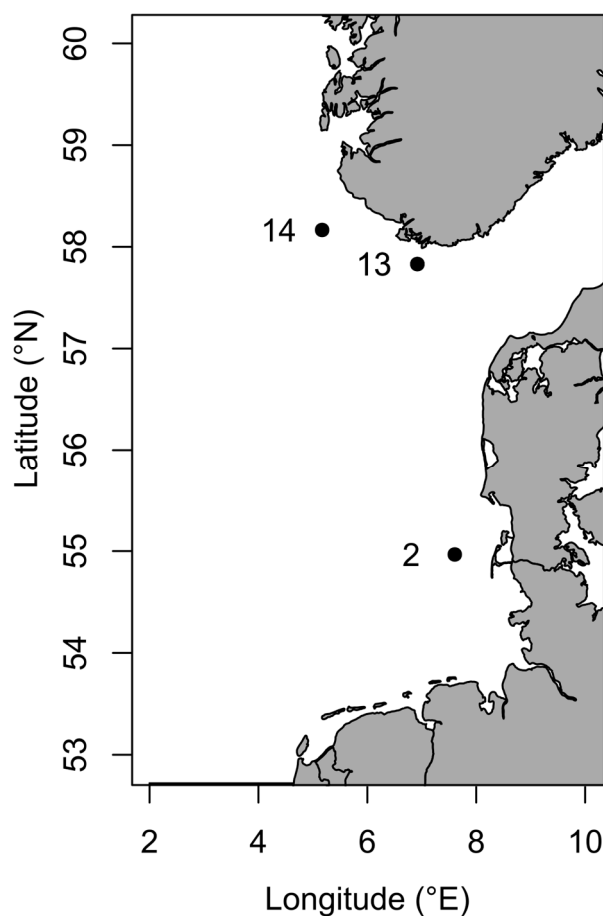

Figure 1. Map of the North Sea showing the three sampling sites. Seawater samples were taken at 3 m (2, 13, 14) and 350 m depth (13). Note that numbering follows ship stations. The map was generated in R using the *maps* and *mapdata* packages [3–5].

2.1. Archaeal Metagenome-Assembled Genomes

A total of 13 archaeal MAGs were reconstructed from the four seawater metagenomes (Supplementary Table S1). Completeness and contamination varied between 54.3% and 84.39% (average 70.84%) and 0% and 8.33% (average 1.39%), respectively. Classification with GTDB-Tk [6] placed all genomes within Marine Group II (MGII; Euryarchaeaota). Members of this group have been frequently observed in various marine ecosystems [7,8]. For instance, Pernthaler et al. observed blooms of MGII archaea during spring and summer in the German Bight with >30% of the total picoplankton abundance [7]. Although frequently observed, our understanding of this abundant marine lineage is still rudimentary [9].

2.2. Bacterial Metagenome-Assembled Genomes

A total of 46 bacterial MAGs were reconstructed (Supplementary Table S1). Completeness and contamination varied between 50.86% and 97.52% (average 77.96%) and 0% and 9.73% (average 2.85%), respectively. Classification with GTDB-Tk [6] affiliated most MAGs to important marine lineages including the Roseobacter clade, the SAR92 clade, as well the OM60/NOR5 clade. Interestingly, four MAGs were assigned to *Planktomarina temperata* within the *Roseobacter* RCA cluster, which is abundant in temperate and polar regions [10–13]. The first published genome belonging to the RCA has been published recently, highlighting the global abundance of this marine pelagic group [14]. Five genomes were affiliated to the marine SAR92 clade, an important gammaproteobacterial marine lineage, which has been frequently observed in the North Sea [15,16]. Five genomes were assigned to the genus Luminiphilus, a genus within the OM60/NOR5 clade, which is abundant in coastal marine ecosystems [11,17].

Deciphering the functional traits of prominent marine lineages is a major challenge in marine microbiology. Information on biogeochemical and functional traits of prominent marine bacterial lineages is often missing and is mostly based on genomic information that is retrieved from single isolates [14,18,19]. These isolates do not necessarily represent abundant marine lineages. This is exemplified by the unexpected discovery of respiratory nitrate reductases in members of the SAR11 clade [20], contributing to an anoxic lifestyle. Here, we present 59 genomes belonging to important marine lineages, such as the marine Roseobacter lineage [14] as well as the SAR 92 clade [18]. These genomes may serve as a basis for further research on functional traits of important marine clades and their contribution to ecosystem services.

3. Methods

3.1. Sampling and Sample Preparation

Seawater samples were collected in the North Sea at three sites on board of the RV Heincke in July 2011 (Figure 1). Three samples were taken in 3 m depth (2, 13, 14) and one sample was taken at 350 m depth (13). Sampling and filtration were performed as described previously [21]. In brief, samples were prefiltered with a glass fiber filter (Whatman GF/D, GE Healthcare, Freiburg, Germany). Bacterioplankton was subsequently harvested from a prefiltered 10 L sample using a filter sandwich consisting of a glass fiber filter (Whatman GF/F, GE Healthcare) and a 0.2-μm polycarbonate filter (Whatman Nuclepore, GE Healthcare). Filter samples were stored at −80 °C or on dry ice during transport from ship to laboratory.

3.2. DNA Extraction and Sequencing

DNA was extracted and purified according to Weinbauer et al. [22]. DNA was subsequently purified, employing the peqGOLD gel extraction kit (Peqlab, Erlangen, Germany). The Göttingen Genomics Laboratory determined the sequences of the extracted DNA using an Illumina Genome Analyzer IIx (San Diego, USA).

3.3. Assembly and Genome Reconstruction

Generated metagenomic datasets were processed as follows: fastq files derived from Illumina sequencing were processed employing the Trimmomatic tool version 0.36 [23]. Processing included the removal adapter sequences and low-quality regions (settings: ILLUMINACLIP:adaptor.fa:2:30:10:2 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36). Processed reads were assembled using metaSPAdes version 3.12.0 [24]. To determine the coverage for each contig, unassembled reads were mapped on obtained scaffolds using bowtie2 version 2.3.2 [25]. Sam files were converted to sorted bam files using samtools version 1.7 [26]. The depth was calculated with jgi_summarize_depth supplied with MetaBAT [27]. MetaBAT version 0.32.5 [27] and MyCC version 2017 [28] were used to reconstruct archaeal and bacterial genomes with a minimum input sequence length of 2500 bp. In

order to increase the overall accuracy and to remove potential contaminations, obtained genomes were refined using binning_refiner [29]. The completeness and contamination were determined using CheckM version 0.7 [30]. Genomes were taxonomically classified using GTDB-Tk version 1.0.2 and the GTDB release 86 [6,31].

References

1. Wiltshire, K.H.; Kraberg, A.; Bartsch, I.; Boersma, M.; Franke, H.-D.; Freund, J.; Gebühr, C.; Gerdts, G.; Stockmann, K.; Wichels, A. Helgoland Roads, North Sea: 45 Years of Change. *Estuaries Coasts* **2010**, *33*, 295–310. [CrossRef]

2. McQuatters-Gollop, A.; Raitsos, D.E.; Edwards, M.; Pradhan, Y.; Mee, L.D.; Lavender, S.J.; Attrill, M.J. A long-term chlorophyll data set reveals regime shift in North Sea phytoplankton biomass unconnected to nutrient trends. *Limnol. Oceanogr.* **2007**, *52*, 635. [CrossRef]

3. Team, R.C. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018.

4. Brownrigg, R. *Maps*, 3.3.0. 2018. Available online: https://cran.r-project.org/web/packages/maps/ (accessed on 1 January 2020).

5. Brownrigg, R. *MapData*, 2.3.0. CRAN, 2018. Available online: https://cran.r-project.org/web/packages/mapdata/index.html (accessed on 1 January 2020).

6. Chaumeil, P.-A.; Mussig, A.J.; Hugenholtz, P.; Parks, D.H. GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **2019**. [CrossRef] [PubMed]

7. Pernthaler, A.; Preston, C.M.; Pernthaler, J.; DeLong, E.F.; Amann, R. Comparison of Fluorescently Labeled Oligonucleotide and Polynucleotide Probes for the Detection of Pelagic Marine Bacteria and Archaea. *Appl. Environ. Microbiol.* **2002**, *68*, 661. [CrossRef] [PubMed]

8. Wemheuer, B.; Wemheuer, F.; Daniel, R. RNA-Based Assessment of Diversity and Composition of Active Archaeal Communities in the German Bight. *Archaea* **2012**, *2012*, 695826. [CrossRef] [PubMed]

9. Zhang, C.L.; Xie, W.; Martin-Cuadrado, A.-B.; Rodriguez-Valera, F. Marine Group II Archaea, potentially important players in the global ocean carbon cycle. *Front. Microbiol.* **2015**, *6*, 1108. [CrossRef] [PubMed]

10. Giebel, H.A.; Brinkhoff, T.; Zwisler, W.; Selje, N.; Simon, M. Distribution of Roseobacter RCA and SAR11 lineages and distinct bacterial communities from the subtropics to the Southern Ocean. *Environ. Microbiol.* **2009**, *11*, 2164–2178. [CrossRef]

11. Wemheuer, B.; Güllert, S.; Billerbeck, S.; Giebel, H.-A.; Voget, S.; Simon, M.; Daniel, R. Impact of a phytoplankton bloom on the diversity of the active bacterial community in the southern North Sea as revealed by metatranscriptomic approaches. *Fems Microbiol. Ecol.* **2014**, *87*, 378–389. [CrossRef]

12. Giebel, H.A.; Kalhoefer, D.; Lemke, A.; Thole, S.; Gahl-Janssen, R.; Simon, M.; Brinkhoff, T. Distribution of Roseobacter RCA and SAR11 lineages in the North Sea and characteristics of an abundant RCA isolate. *ISME J.* **2011**, *5*, 8–19. [CrossRef]

13. Wemheuer, B.; Wemheuer, F.; Hollensteiner, J.; Meyer, F.-D.; Voget, S.; Daniel, R. The green impact: Bacterioplankton response towards a phytoplankton spring bloom in the southern North Sea assessed by comparative metagenomic and metatranscriptomic approaches. *Front. Microbiol.* **2015**, 6. [CrossRef]

14. Voget, S.; Wemheuer, B.; Brinkhoff, T.; Vollmers, J.; Dietrich, S.; Giebel, H.-A.; Beardsley, C.; Sardemann, C.; Bakenhus, I.; Billerbeck, S.; et al. Adaptation of an abundant Roseobacter RCA organism to pelagic systems revealed by genomic and transcriptomic analyses. *ISME J.* **2015**, *9*, 371–384. [CrossRef] [PubMed]

15. Lucas, J.; Wichels, A.; Teeling, H.; Chafee, M.; Scharfe, M.; Gerdts, G. Annual dynamics of North Sea bacterioplankton: Seasonal variability superimposes short-term variation. *Fems Microbiol. Ecol.* **2015**, *91*. [CrossRef] [PubMed]

16. Wemheuer, B.; Wemheuer, F.; Meier, D.; Billerbeck, S.; Giebel, H.-A.; Simon, M.; Scherber, C.; Daniel, R. Linking Compositional and Functional Predictions to Decipher the Biogeochemical Significance in DFAA Turnover of Abundant Bacterioplankton Lineages in the North Sea. *Microorganisms* **2017**, *5*, 68. [CrossRef] [PubMed]

17. Yan, S.; Fuchs, B.M.; Lenk, S.; Harder, J.; Wulf, J.; Jiao, N.Z.; Amann, R. Biogeography and phylogeny of the NOR5/OM60 clade of Gammaproteobacteria. *Syst. Appl. Microbiol.* **2009**, *32*, 124–139. [CrossRef] [PubMed]

18. Stingl, U.; Desiderio, R.A.; Cho, J.C.; Vergin, K.L.; Giovannoni, S.J. The SAR92 clade: An abundant coastal clade of culturable marine bacteria possessing proteorhodopsin. *Appl. Environ. Microbiol.* **2007**, *73*, 2290–2296. [CrossRef]

19. Billerbeck, S.; Wemheuer, B.; Voget, S.; Poehlein, A.; Giebel, H.-A.; Brinkhoff, T.; Gram, L.; Jeffrey, W.H.; Daniel, R.; Meinhard, S. Biogeography and environmental genomics of the Roseobacter group affiliated pelagic CHAB-I-5 lineage. *Nat. Microbiol.* **2016**, *1*, 16063. [CrossRef]

20. Tsementzi, D.; Wu, J.; Deutsch, S.; Nath, S.; Rodriguez-R, L.M.; Burns, A.S.; Ranjan, P.; Sarode, N.; Malmstrom, R.R.; Padilla, C.C.; et al. SAR11 bacteria linked to ocean anoxia and nitrogen loss. *Nature* **2016**, *536*, 179–183. [CrossRef]

21. Osterholz, H.; Singer, G.; Wemheuer, B.; Daniel, R.; Simon, M.; Niggemann, J.; Dittmar, T. Deciphering associations between dissolved organic molecules and bacterial communities in a pelagic marine system. *ISME J.* **2016**. [CrossRef]

22. Weinbauer, M.G.; Fritz, I.; Wenderoth, D.F.; Höfle, M.G. Simultaneous extraction from bacterioplankton of total RNA and DNA suitable for quantitative structure and function analyses. *Appl. Environ. Microbiol.* **2002**, *68*, 1082–1087. [CrossRef]

23. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]

24. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef] [PubMed]

25. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Meth.* **2012**, *9*, 357–359. [CrossRef] [PubMed]

26. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*. [CrossRef] [PubMed]

27. Kang, D.D.; Froula, J.; Egan, R.; Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *Peer J.* **2015**, *3*, e1165. [CrossRef]

28. Lin, H.-H.; Liao, Y.-C. Accurate binning of metagenomic contigs via automated clustering sequences using information of genomic signatures and marker genes. *Sci. Rep.* **2016**, *6*, 24175. [CrossRef]

29. Song, W.-Z.; Thomas, T. Binning–refiner: Improving genome bins through the combination of different binning programs. *Bioinformatics* **2017**, *33*, 1873–1875. [CrossRef]

30. Parks, D.H.; Imelfort, M.; Skennerton, C.T.; Hugenholtz, P.; Tyson, G.W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **2015**, *25*, 1043–1055. [CrossRef]

31. Parks, D.H.; Chuvochina, M.; Chaumeil, P.-A.; Rinke, C.; Mussig, A.J.; Hugenholtz, P. Selection of representative genomes for 24,706 bacterial and archaeal species clusters provide a complete genome-based taxonomy. *BioRxiv* **2019**, 771964. [CrossRef]