ISME

ARTICLE

Niche differentiation among annually recurrent coastal Marine Group II Euryarchaeota

Luis H. Orellana[1] · T. Ben Francis[1] · Karen Krüger [1] · Hanno Teeling[1] · Marie-Caroline Müller[1] ·
Bernhard M. Fuchs [1] · Konstantinos T. Konstantinidis[2] · Rudolf I. Amann[1]

Abstract

Since the discovery of archaeoplankton in 1992, the euryarchaeotal Marine Group II (MGII) remains uncultured and less understood than other planktonic archaea. We characterized the seasonal dynamics of MGII populations in the southern North Sea on a genomic and microscopic level over the course of four years. We recovered 34 metagenome-assembled genomes (MAGs) of MGIIa and MGIIb that corroborated proteorhodopsin-based photoheterotrophic lifestyles. However, MGIIa and MGIIb MAG genome sizes differed considerably (~1.9 vs. ~1.4 Mbp), as did their transporter, peptidase, flagella and sulfate assimilation gene repertoires. MGIIb populations were characteristic of winter samples, whereas MGIIa accounted for up to 23% of the community at the beginning of summer. Both clades consisted of annually recurring, sequence-discrete populations with low intra-population sequence diversity. Oligotyping of filtered cell-size fractions and microscopy consistently suggested that MGII cells were predominantly free-living. Cells were coccoid and ~0.7 μm in diameter, likely resulting in grazing avoidance. Based on multiple lines of evidence, we propose distinct niche adaptations of MGIIa and MGIIb *Euryarchaeota* populations that are characteristic of summer and winter conditions in the coastal North Sea.

Introduction

Since the discovery of marine planktonic archaea using culture-independent approaches [1, 2], four major archaeal lineages have been characterized in marine ecosystems. The thaumarchaeotal Marine Group I (MGI) and the euryarchaeotal Marine Group II (MGII) have been found in deep and surface ocean waters, respectively [3–5]. In addition, archaea belonging to the euryarchaeotal Marine Group III (MGIII) [6, 7] and IV (MGIV) [8] have been

mostly associated with deep sea environments. So far, only members of the MGI group have been isolated [9]. Thus, the characterization of other marine archaea has been limited to culture-independent approaches. MGI have been mostly associated with ammonia oxidation [9], whereas less is known about MGII. Nonetheless, sequences recovered from metagenomes belonging to the MGIIa subgroup have suggested a potential role in protein and fatty acid degradation [10–12]. It has been proposed that MGII could thrive under oligotrophic conditions, where particle-associated MGII can outnumber free-living cells of the same group [13]. In addition, a partially recovered genome of the MGIIb group (originally proposed as class 'Thalassoarchaea'), obtained from the Mediterranean deep chlorophyll maximum, showed similar heterotrophic features as those found in MGIIa genomes, but have a lower GC percentage and a smaller genome size [14]. Hence the high diversity recently reported [11, 12] within MGII populations underscores the knowledge gaps for these marine archaeal clades.

Our understanding of the dynamics of MGII populations has until recently [11, 12] been restricted to the use of molecular markers such as 16S rRNA gene sequences [15–17]. Studies in the Mediterranean have suggested niche

separation between MGI and MGII, but also within MGII populations. For instance, while MGIIb (and MGI) dominated at the time of winter sampling, MGIIa were abundant during summer. These seasonal differences were proposed to result from strong changes in biotic (e.g., nutrient and phytoplankton stocks) and abiotic factors (e.g., temperature and water column stratification) over the winter-summer transition [15]. Previous reports have indicated short blooms of MGII populations during summer [3, 15, 18] following phytoplankton bloom events and also coinciding with decreased levels of chlorophyll *a* [17, 19]. Phytoplankton blooms in coastal marine environments often cause successions of bacterioplankton communities participating in the remineralization of the released organic matter [20], which are dominated by members of the classes *Flavobacteriia*, *Alphaproteobacteria* and *Gammaproteobacteria* [21–23]. Still, the diversity of MGII populations occurring before and after these blooms have received much less attention due to their usually lower abundance compared to heterotrophic bacterioplankton [17, 24]. A recent large phylogenomic comparison of MGII flagellum-based adhesion, transport and degradative potential suggests that multiple MGII clades diversified from a surface water-dwelling photoheterotrophic ancestor [11]. However, little is known about how the genomic potential of distinct MGII clades correlates with observed temporal abundance patterns. In addition, not much is known about the level of genetic heterogeneity within recurrent MGII populations, which is important for delimiting MGII ecophysiological niches including ecotypes and their impact on coastal environments.

We combined fluorescence in situ hybridization (FISH) and metagenome sequencing approaches to explore the temporal patterns and metabolic potentials of discrete euryarchaeotal populations in the southern North Sea (German Bight) in the years 2009 to 2012. The recovery of 34 high-quality metagenome-assembled genomes (MAGs) related to MGII allowed us to compare the genomes and ecology of MGIIa and MGIIb. Our results reveal low intra-population diversity of annually recurrent MGII populations and provides novel information on the potential niche differentiation of summer and winter clades of photoheterotrophic Marine Group II *Euryarchaeota*.

Materials and methods

Sampling and sequencing

Surface seawater samples were collected from the 'Kabeltonne' ecological research site off the North Sea island of Helgoland (54° 11.3′ N, 7° 54.0′ E) as described previously [23, 25, 26]. Briefly, samples for DNA sequencing were passed through 10 and 3 μm pore-size filters before cells were collected on 0.2 μm pore-size polycarbonate filters. This approach removed phytoplankton and most of the particle-associated microorganisms. Seawater for cell counting was also collected from the surface, but it was not fractionated [23, 26]. The sequencing of 44 surface water metagenomes from 'Kabeltonne' was performed at the DOE Joint Genome Institute (DOE-JGI) as described previously [25, 26]. Metagenomes from 2009 [23] were sequenced on the GS FLX Ti platform (454 Life Sciences, Branfort, CT, USA), all others on the Illumina HiSeq platform (Illumina, San Diego, CA, USA) with paired-end sequencing [25]. Trimming and processing of raw reads was performed as previously published [26].

Catalyzed reporter deposition fluorescence in situ hybridization

For the enumeration of *Euryarchaeota*, 100 ml of water were fixed with 1% formaldehyde at room temperature for 1 h, filtered onto 47 mm diameter polycarbonate filters (pore size 0.2 μm) with a vacuum of at most 100 mbar. Catalyzed Reporter Deposition Fluorescence in situ hybridization (CARD-FISH) was performed as outlined before [27]. A detailed protocol is available in the supplementary text.

MAG recovery

Archaeal MAGs were obtained from previously assembled 2010–2012 Helgoland metagenomes [26]. Briefly, de novo assembly of short-read metagenomes was performed using SPADES 3.10 [28] and assembled contigs longer than 2.5 kbp were binned using CONCOCT [29] as part of the standard Anvi'o v3 metagenomic workflow [30]. Differential coverage information was generated by mapping reads from four additional randomly selected dates from the same year. Reads were mapped using BBMap v35.14 (http://bbtools.jgi.doe.gov; 'fast' mode, minid = 0.99, and idfilter = 0.97). Manual inspection for congruent genetic composition and coverage of recovered MAGs was performed in Anvi'o [30]. Selected MAGs (%[completion] −4 × % [contamination] ≥ 50 (determined by checkM [31])) were de-replicated using dRep v2.2.3 [32], which combines MASH and average nucleotide identity (ANI > 95%) measures in order to select the best representative from each group of highly similar MAGs (Table S1). Genes were predicted using Prodigal v2.6.3 [33] (metagenome option). Reported taxonomic classification, completion and contamination for archaeal MAGs was determined using the Microbial Genome Atlas webserver (MiGA) [34]. MAGs were uploaded to MiGA using the "popgenome" dataset option. MiGA determines completeness and contamination using previously reported single-copy marker genes [35].

Taxonomic classifications are based on genome-aggregate average amino acid and average nucleotide identity (AAI/ANI) compared to available reference genomes. Additional details about genomic comparisons and gene annotation are available in the supplementary text.

MGII MAGs previously reported [11, 12] were also included in the phylogenetic reconstructions. Identical and redundant MAGs based on ANI were de-replicated using FastANI [36] and quality (%[completion]−4 × %[contamination] ≥ 50). Briefly, using an ANI value ≥99% reduced the MGII MAG collection [11, 12] from 643 to 294 de-replicated MAGs. Representative MAGs for each 99% ANI group were selected based on highest completion and lowest contamination (Table S2 and S3).

MAG abundances and intra-population sequence diversity

Bowtie2 [37] was used to map short-reads to MAG contigs, and SamTools [38] to convert the resulting SAM files to BAM files. Genomecov (-bga option) from the Bedtools package [39] was used to determine sequencing depth. In order to exclude biases introduced by highly conserved regions and contig edges, the 80% central truncated average of the sequencing depth (TAD) of all bases was determined using the 'BedGraph.tad.rb' script (option range 80) from the enveomics collection [35]. The option 'range 80' removes the top 10% and bottom 10% positions of per base sequencing coverage in each MAG. The *rpoB* sequencing depth was used as a proxy for the sequencing depth for all microbial genomes in each metagenome assuming that both whole genome and *rpoB* sequencing depths were equal. The sequencing depths of *rpoB* genes was determined for each metagenome by first identifying short-reads related to *rpoB* genes using ROCker [40]. Briefly, a custom database previously constructed for RpoB sequences [41] was used in BLASTx searches [42] for each metagenome. BLASTx outputs were then filtered for high-quality matches using a ROCker model specific to the RpoB database. Finally, MAG abundance values were calculated as the quotient between the determined TAD for each MAG and the sequencing depth of its *rpoB* gene.

Intra-population sequence diversity was determined as previously reported [43], by first mapping the short-reads from different metagenomes to representative MAGs using BLASTn (-outfmt 6 option). Samples from different time points where individual MAGs represented >0.01%—corresponding to ~2× coverage or more of the total population—were selected. Read recruitment plots were generated using the BLASTn outputs and processed with the "BlastTab.catsbj.pl" and "enve.recplot2" scripts from the enveomics collection for each MAG and sample [35]. The nucleotide identity value denoting the sequence

discontinuity for mapped reads was selected upon visual inspection as previously described [44], identifiable as a drop in coverage by 3–4 orders of magnitude (see example in Fig. S5). For all *Euryarchaeota* MAGs, 98% identity represented the cutoff for sequence-discrete populations. Reads mapping with identities above this threshold based on BLASTn searches were thus used to calculate read-based average nucleotide identity (ANIr), as a proxy for intra-populations diversity using the "enve.recplot2.ANIr" R script from the enveomics collection.

MAG and gene phylogenies

MAG phylogenetic analyses were performed using 15 syntenic ribosomal proteins (L2, L3, L4, L5, L6, L14, L15, L18, L22, L24 and S3, S8, S10, S17, S19) [45–47]. Predicted protein sequences were detected in each MAG using target hidden Markov models (HMMs) for each protein obtained from TIGRFAM [48], AMPHORA2 [49], and Pfam [50] using HMMER v3.2.1 [51]. Multiple alignments for each protein were generated using ClustalΩ [52] and a concatenated alignment was generated using the 'Aln.cat.rb' script [35]. Maximum-likelihood phylogenetic estimations were determined in RAxML v8.0.19 [53] (PROT-GAMMAAUTO, -N 500) and visualized in the interactive Tree of Life (iTol) [54].

16S rRNA gene oligotyping

Sample collection at Helgoland and processing for amplicon analysis has been described previously [19]. Briefly, the two fractions corresponding to size ranges of 0.2 to 3 μm, and 3 to 10 μm were separated via filtration of the surface water samples (~1 m depth). Amplicons were generated for both fractions via PCR amplification of the V4 region of the 16S rRNA gene and sequenced using Illumina MiSeq 2 × 250 bp chemistry at the DOE-JGI. Differentially abundant MGII oligotypes were determined with the DESeq2 package [55]. Additional details and complete methods can be found in the supplementary material.

Data availability

Metagenomes were previously deposited in NCBI (see Table S1 for BioProject accession numbers). Helgoland MAGs used in this study were previously deposited at the European Nucleotide Archive (Study PRJEB28156) using the data brokerage service of the German Federation for Biological Data (GFBio) [56]. Other MGII MAGs published were obtained using the accession numbers provided [11, 12]. Raw 16S rRNA gene amplicon data are stored by JGI in the GOLD database under the project IDs

Gp0056779 ('free living'; from 0.2 to 3 μm), and Gp0072732 and Gp0072733 ('attached'; from 3 to 10 μm), as part of the community sequencing project COGITO.

Results

Metagenome-assembled genomes from Helgoland

During the spring phytoplankton blooms of 2009–2012, we collected water samples at the 'Kabeltonne' station, Helgoland Island. We have previously reported an extensive time-series of the characteristics of this shallow (water depth ~10 m) coastal marine site and the sampling strategies [25]. Given that all metagenomes from the time-series were assembled and binned separately, many highly similar MAGs were obtained (see ANI values between MGII MAGs in Fig. S1). MAGs classified as archaeal were reduced from a total of 49 down to 11 representative MAGs using an ANI threshold of 95% and quality over 50 (see Methods) (Table 1, S1). The reconstruction of the MAG phylogeny using a syntenic block of conserved ribosomal genes [45–47] provided an overview of the archaeal diversity at Helgoland (Fig. 1a). Eight of the representative MAGs belonged to the *Euryarchaeota* phylum, two to the TACK superphylum, and one to the DPANN group (Fig. 1a). The topology of a phylogenetic tree using recovered 16S rRNA gene sequences from some of the MAGs (5/11; see Table S1) was congruent with the tree using conserved ribosomal genes (Fig. S2).

The two thaumarchaeotal MAGs (TACK superphylum), Thau1 and Thau2, were placed within the MGI.1a clade and shared 88.7% AAI (SD 9.8%, 1207 protein sequences) and probably represent distinct species of the same genus. Their genome sizes (~1.1 Mbp) and numbers of predicted genes (~1400) were similar. Thau1 and Thau2 shared 88.9 and 95.7% ANI with *Ca.* Nitrosomarinus catalina, respectively. This MGI archaeon, possibly representing the same species as Thau2, was isolated from the San Pedro Ocean time-series (temperate Pacific Ocean waters) off the coast of California [57]. The DPANN MAG was placed within the *Woesearchaeota*, but shared only ~40% AAI with previously reported *Woesearchaeota* MAGs.

The eight *Euryarchaeota* MAGs were divergent from the available reference archaeal genomes and only shared AAI values between ~47 to 80% to the genome of an uncultured MGII euryarchaeote [10] when compared to 11,568 classified reference genomes and isolates [34] (Table S1). Nonetheless, higher AAI values for Helgoland MAGs were obtained (~71.5 and ~79.3% for MGIIa and MGIIb, respectively) when compared to previously recovered MGII MAGs of the highlighted clades on Fig. 1b. The completion for all *Euryarchaeota* MAGs was high, with values up to

96.2% and low levels of contamination (Table 1). Interestingly, the *Euryarchaeota* genomes were separated into two distinct clades, previously described as MGIIa and MGIIb [11, 12, 14, 16, 58]. Similar phylogenetic relationships were obtained using a conserved syntenic block of ribosomal proteins and 120 single-copy markers previously described for MGII genomes [11, 59] (Fig. S3a, b). AAI among MGIIa MAGs ranged between 64–73 and 63–85% for MGIIb MAGs (Fig. S4). Larger average genome sizes (1.9 vs. 1.4 Mbp), and higher %GC values (50.5 vs. 37.5%) were characteristic of MAGs from MGIIa compared to MGIIb, respectively, in agreement with previous findings [11, 12].

A recent study using rank-normalized phylogeny [11] proposed the order level *Ca.* Poseidoniales for the MGII lineage and the subgroups MGIIa and MGIIb as *Ca.* Poseidoniaceae and *Ca.* Thalassarchaeaceae families, respectively, in addition to 21 new genera within these families. According to this classification, the Helgoland MGIIa MAGs belong to the *Ca.* Poseidoniaceae family and the genera L1 (MGIIa_c5 and MGIIa_c10) and M (MGIIa_c4 and MGIIa_c6). MGIIb MAGs are members of the *Ca.* Thalassarchaeaceae family in the genera O3 (MGIIb_c11) and O2 (MGIIb_c8, MGIIb_cB, and MGIIb_c7). MGIIa_c6 corresponds to the proposed type species *Ca.* Poseidonia alphae as both MAGs share 99.98% ANI and were recovered from the same metagenome. Thus, these two MAGs likely represent the same population yet both MAGs are not identical since they were recovered using different assembly and binning strategies.

Abundance and intra-population sequence diversity dynamics of MGII MAGs

The metagenome time-series allowed us to evaluate the abundance and persistence of the recovered MAGs over the course of the 2009–2012 spring blooms. Previous studies have focused on the rapid succession patterns of the *Flavobacteriia*, *Gammaproteobacteria*, and *Roseobacter* bacterial classes [23, 25]. The abundances of the recovered archaeal MAGs were low for most of the samples (Fig. 2a, b). Nonetheless, MGII MAGs were recurrent and most abundant in late winter and early summer, before and after the phytoplankton blooms. Specifically, MGI and MGIIb MAGs were detected in pre-bloom metagenomes, whereas MGIIa were dominant in post-bloom metagenomes in two of the four years. The MGIIa_c6 MAG represented the most prevalent population as it was detected in 90% of the post-bloom metagenomes (9/10) and 65.7% of all metagenomes (25/38). Metagenomes obtained during 2010 were mostly obtained during the spring phytoplankton bloom, therefore MGII populations were almost undetectable in the samples from this year (Fig. 2a, b). MGIIb

Table 1 Genome statistics for eleven representative archaeal MAGs

| MAG	Phylum	MAG length (Mbp)	Number of contigs	Number of predicted genes	GC content (%)	Completeness (%)	Contamination (%)	Sampling date
MGIIa_c4	Euryarchaeota	1.79	43	1511	51.9	88.5	3.8	08-Mar-2012
MGIIa_c5		1.93	48	1631	51.5	96.2	0	31-May-2012
MGIIa_c6		1.98	38	1652	50.1	96.2	0	26-May-2011
MGIIa_c10		1.96	49	1619	49.3	96.2	0	24-May-2012
MGIIb_c7		1.39	73	1267	38.2	92.3	0	08-Mar-2012
MGIIb_c8		1.41	31	1244	38.6	88.5	0	21-Mar-2011
MGIIb_c11		1.62	25	1448	36.8	96.2	0	28-Apr-2011
MGIIb_cB		1.33	187	1231	37.4	80.8	0	24-Mar-2011
Thau_1	TACK	1.14	86	1483	30.9	88.5	0	21-Mar-2011
Thau_2		1.07	82	1404	31.3	92.3	0	03-Mar-2010
Woes_A	DPANN	0.71	158	869	36.0	50	0	04-May-2010
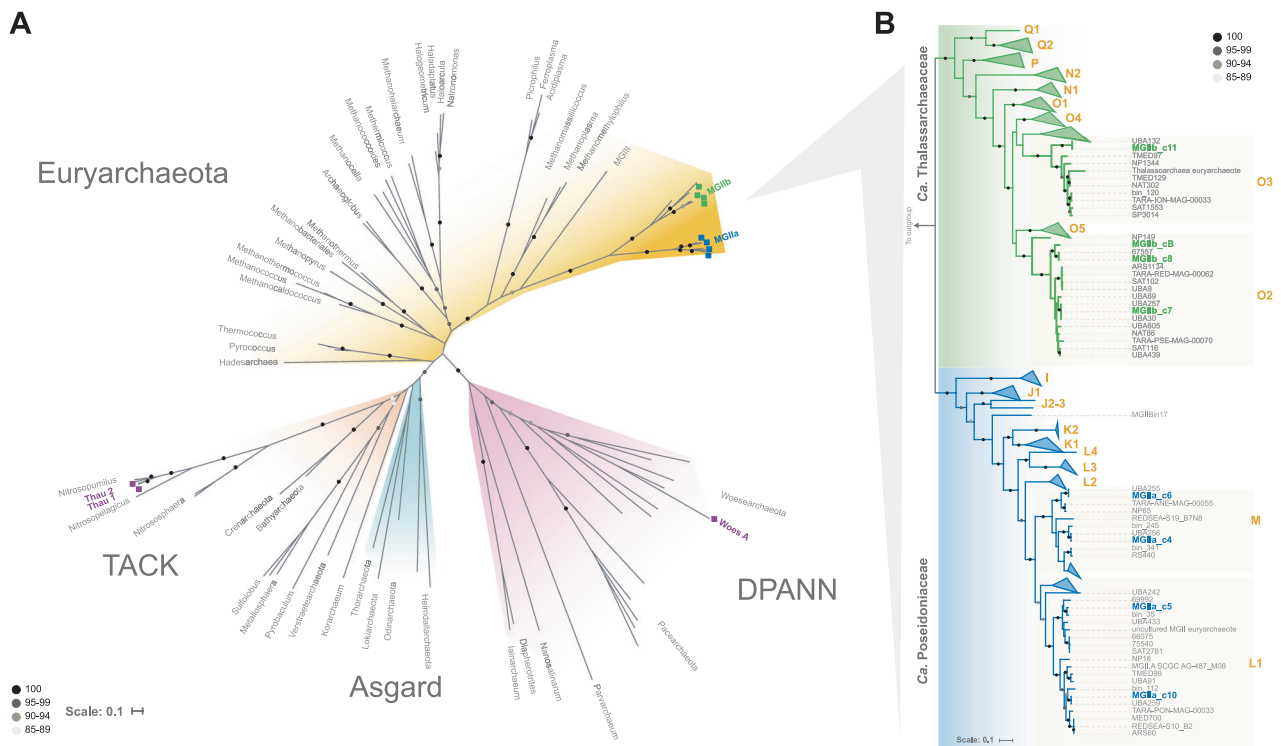
Fig. 1 Phylogenetic reconstruction of archaeal genomes including MAGs from Helgoland. **a** The reconstruction of the maximum likelihood phylogeny used representative genomes of DPANN, TACK, *Euryarchaeota*, and Asgard groups and was based on a syntenic block of 15 conserved ribosomal archaeal genes (see methods). Colored squares represent recovered MAGs in this study. Main proposed genera names are displayed. Branch support values between 85–100% are represented by dots. **b** Phylogenetic reconstruction of MGII MAGs from Helgoland and others recovered from elsewhere around the world [11, 12]. Clades were collapsed according to genera previously reported [11]. Clades L1, M, O2, and O3 were not collapsed to highlight the phylogenetic relationship of Helgoland MAGs to other previously reported MGII MAGs

MAGs were only detected in samples obtained during 2011–2012, with MGIIb_c11 being most prevalent.

In pre-bloom metagenomes from 2011 and 2012, MGIIa and MGIIb MAGs comprised, on average, 0.1 and 0.3% of the total microbial fraction, respectively. Notably, MAGs affiliated to the MGIIa populations comprised up to 23% of the total microbial community in post-bloom metagenomes during 2012. In fact, individually, MGIIa_c6 and
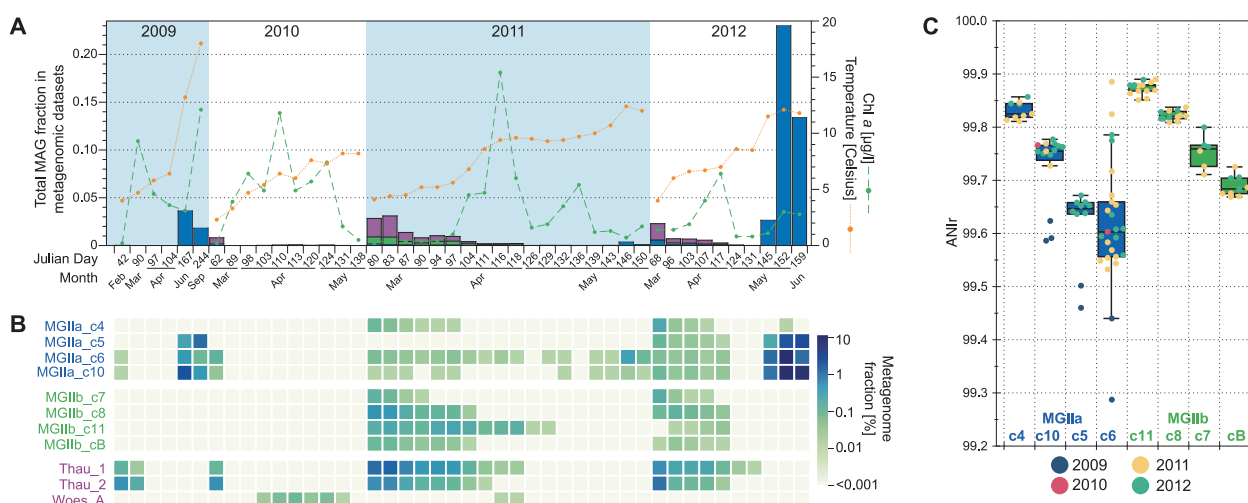
Fig. 2 Abundance of MAGs during coastal spring algae blooms at Helgoland in 2011 and 2012. **a** Fraction of main archaeal groups in each spring bloom metagenome. Chlorophyll *a* and temperature measurements of the same samples are also displayed. **b** Individual abundance of each MAG (measured as fraction of the metagenome).

Note that separation between samples does not represent a continuous time scale on the *x*-axis. **c** Average nucleotide identity for reads mapping *Euryarchaeota* MAGs (ANIr) representing >0.01% of the microbial communities (~2×) in 2009–2012 metagenomes

MGIIa_c10 comprised up to ~10% of the total microbial community in post-bloom metagenomes of 2012. The exception to the pattern described above for MGIIa MAGs was the MGIIa_c4 MAG, which had an abundance distribution similar to MGIIb populations (i.e., mostly detected during winter) despite both MGIIa_c4 and MGIIa_c6 MAGs belong to the same genus (Fig. 1b).

We analyzed the intra-population sequence diversity of discrete populations by mapping short-metagenomic reads from metagenomes to the MGII MAGs and determining ANIr values. This approach allows a quantitative evaluation of sequence diversity for sequence-discrete populations and is less prone to systematic issues carried by missassemblies or differences in the degree of completion when comparing recovered MAGs. In addition, the use of MAGs as reference sequences in recruitment plots that were assembled from single metagenomes allowed us to preserve the genetic characteristics of discrete archaeal populations at each sampling point (e.g., recruitment plot for MGIIa_c6 in Fig. S5). The ANIr values ranged from 99.2 to 99.8%, indicating a low intra-population sequence diversity (or highly clonal, Fig. 2c). Even though the variation in ANIr values among metagenomes (when the same reference genome was used in the recruitment of reads from different metagenomes) was small for both MGII populations, the interquartile range of ANIr values for MGIIa was higher compared to MGIIb MAGs (two tailed *t*-test, $P < 0.01$), indicating a higher intra-population sequence variability for MGIIa populations. For instance, the interquartile range of ANIr values for MGIIa_c6 was the highest (IQR = 0.6) among MGII MAGs.

Visualization and quantification of MGII abundance

We visualized and quantified marine *Euryarchaeota* populations using FISH and CARD-FISH. The CARD-FISH approach allowed us to confirm and expand on the temporal distribution patterns of archaeal populations observed in the metagenomic data. In addition to quantifying all MGII *Euryarchaeota* with class-specific oligonucleotide probe EURY806 [60], we chose to follow clade MGIIa_c6 because it was among the abundant populations that also contained a 16S rRNA gene sequence for probe design (Table S1). The vast majority of the cells detected using probe EURY806 and MGIIa_c6 were coccoid, with a diameter of <1 μm, suggesting that MGIIa and MGIIb had similar cell morphologies. The cells were neither accumulated in aggregates nor attached to particles, and appeared to be mostly planktonic (Fig. 3a and S6). The average cell diameter for MGIIa_c6 cell was 0.7 ± 0.2 μm. *Euryarchaeota* represented a relatively small fraction of the microbial community before and during the phytoplankton spring blooms (i.e., from March to May), but dominated by the end of May and the beginning of June of 2011 and 2012, respectively (Fig. 3b and S7).

The relative abundance patterns observed for *Euryarchaeota* and MGIIa_c6 populations using CARD-FISH were consistent with the abundances determined in metagenomes. For instance, on May 31st of 2012 (Julian day 152), *Euryarchaeota* peaked in abundance, accounting for 23 and 14% of the microbial populations in metagenomes and cell counts, respectively (Fig. 3c, top circle). Similarly, MGIIa_c6 cells with respect to
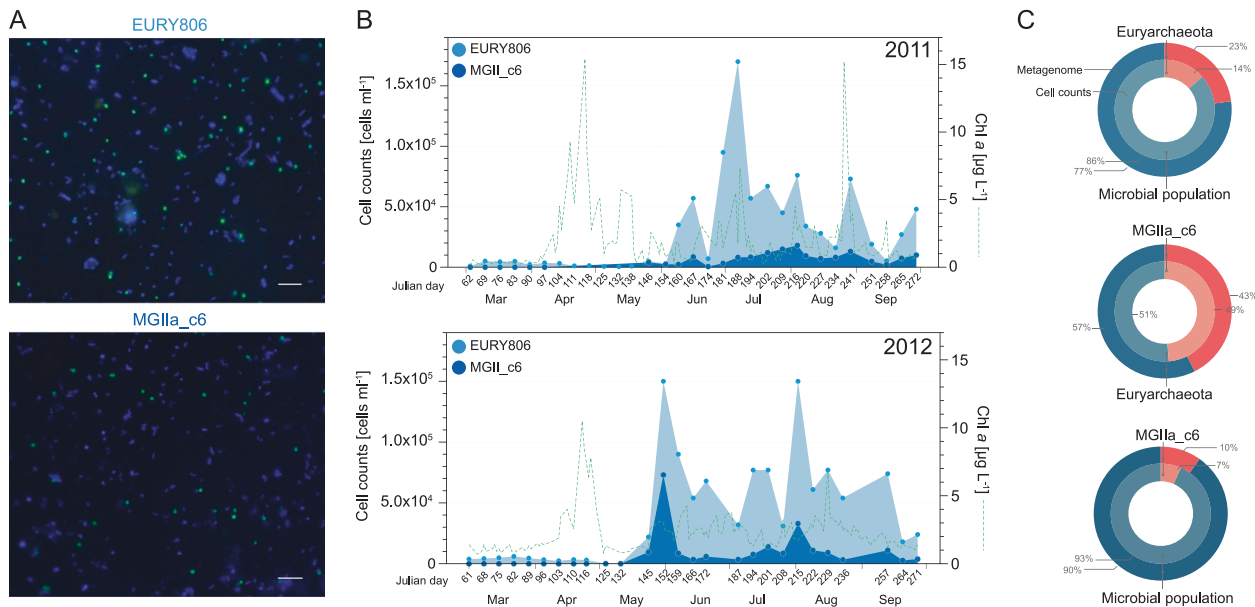
Fig. 3 Cell morphology and in-situ abundances of MGIIa_c6 populations at Helgoland based on microscopy. **a** CARD-FISH results for sample from 31-May-2012. DAPI and MGIIa_c6 probe. The scale represents 5 μm. **b** Counts of total *Euryarchaeota* (light blue) and MGIIa_c6 (blue) cells using the EURY806 and MGIIa_c6 probes, respectively. **c** Comparison of relative abundances obtained from metagenome (outer circle) and cell counts (inner circle) approaches for *Euryarchaeota* from 31-May 2012 with respect to the total microbial population (top), MGIIa_c6 with respect to *Euryarchaeota* (middle) and total microbial population (bottom)

euryarchaeotal cells accounted for 43, 49, and 10% on May 24th, 31st and June 6th of 2012 (145, 152, and 159 Julian days) while at the metagenome level the corresponding values were 52, 43, and 9% for the same dates (Julian day 152 showed in Fig. 3c, middle circle). Interestingly, MGIIa_c6 represented, as a single species-level clade, 7 and 10% of visualized cells and metagenome, respectively (Fig. 3c, bottom circle).

Oligotyping of size fractions

The examination of MGII 16S rRNA gene oligotypes obtained from 0.2–3 and 3–10 μm pore-size fractions allowed us to further investigate the association of MGII populations with particles in an expanded time scale. A total of 27 recurrent MGII oligotypes, of which 21 corresponded to MGIIa and 6 to MGIIb, had above 0.1% abundance in at least one sample from 2010, 2011, and 2012 (Fig. S8). MGIIa oligotypes were more abundant during summer/autumn, whereas MGIIb were mostly detected in winter/autumn, in agreement with our metagenomic results. Interestingly, 59 to 89% of the MGII oligotypes were 4-fold higher in the 0.2–3 μm fraction in all seasons (log2-fold > 2; Fig. 4). Although 9 MGII oligotypes were slightly enriched in the 3–10 μm fraction mostly during spring, they were never above 2-fold higher (log2-fold < −1), indicating a higher prevalence of MGII populations in the smaller 0.2–3 μm pore-size fractions.

Metabolic potential of North Sea Euryarchaeota

Core metabolic functions indicating a heterotrophic lifestyle, such as the capacity to perform glycolysis, and the citric acid cycle were detected in all MGII MAGs from Helgoland, and are consistent with previous analyses based on globally recovered MGII MAGs [11, 12, 58]. Nonetheless, differences such as the absence or reduced number of genes for key metabolic functions were noted among MGII MAGs from Helgoland. For instance, while MGIIa MAGs encode all necessary enzymes for non-oxidative pentose formation, MGIIb MAGs from Helgoland of O2 genera (MGIIb_c7, c8, and cB) lack the ribose-phosphate pyrophosphokinase gene (*prs*), necessary for the phosphoribosyl pyrophosphate (PRPP) and de novo generation of purine nucleotides and several amino acids. The lack of the *prs* gene was also consistent in the majority of the representative MGII O2 MAGs used for the phylogenetic reconstruction (3/17) but not in the O3 genus (9/10), where this feature was prevalent. In addition, different loci containing the genes *flaB*, *flaH*, *flaI*, and *flaJ*, which commonly form part of the flagellar structure [61], were present in all MGIIa MAGs. In contrast, only MGIIb_c11 among MGIIb MAGs contained a locus encoding these flagellar genes (Table S4). Helgoland MGIIa MAGs possessed additional components of the flagella machinery including the *flaJ*, the ATPases *flaI* and *tadC* genes. However, no known chemotaxis components were detected in any of the Helgoland MGII MAGs, in agreement with previous findings [11].
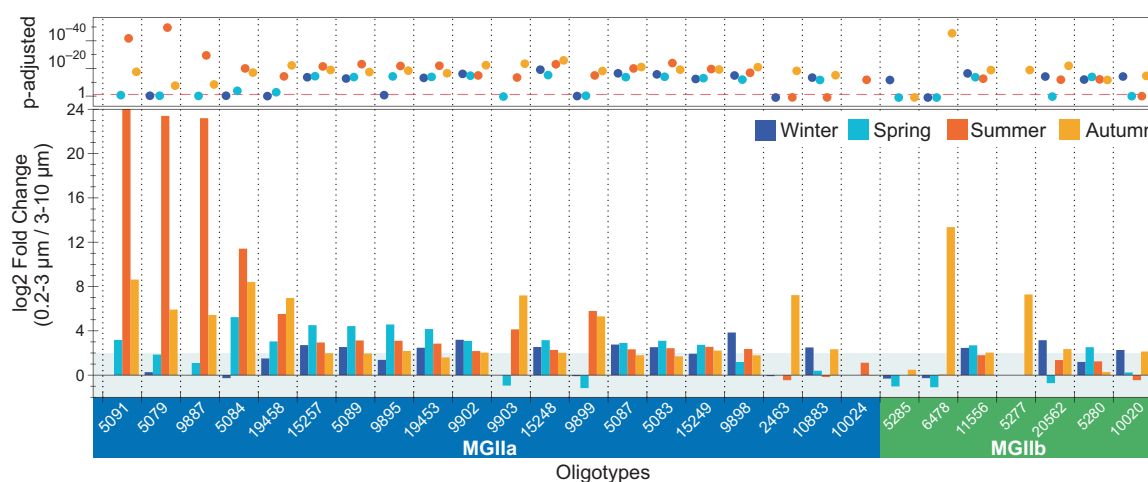
Fig. 4 Distribution of MGII oligotypes in Helgoland surface waters. Differential abundance of MGII oligotypes was performed between 0.2–3 μm and 3–10 μm pore-size fractions within seasons in DESeq2. The lower panel shows the log2-fold values between 0.2–3 μm (values above zero) and 3–10 μm (values below zero) pore-size fractions for each of the 27 oligotypes. Oligotypes were selected based on relative abundance of at least 0.1% in at least one sample. The top panel shows the p-adjusted values determined for the comparison between size fractions and the dashed line demarcates the p-adjusted $= 0.05$ value

Unlike heterotrophic bacteria linked to phytoplankton blooms, a low number of glycoside hydrolases (GH) were detected in MGII MAGs. In fact, only a single glycoside hydrolase belonging to family 1 (GH1) was detected in MGIIa MAGs (MGIIa_c4, MGIIa_c5, and MGIIa_c10) and no previously described GHs were found in MGIIb MAGs. The GH1 belongs to the wide-spread family of beta-glucosidase enzymes (EC 3.2.1.-), involved in the processing of polysaccharides and oligosaccharides. In contrast, both MGII groups encoded several copies of glycosyl-transferases (GT) involved in the formation of glycoside linkages, among other predicted functions (Table S4). A more detailed description of the metabolic potential (Fig. 5) and predicted protein orthologs (Fig. S9) is available in the supplementary text.

Potential for assimilatory sulfate reduction in MGIIb MAGs

Another difference between the two family-level MGII groups in Helgoland was the potential for assimilatory sulfate reduction predicted only for MGIIb MAGs. This family encoded an ATP sulfurylase (*cysD* and *cysN*) that, in conjunction with an APS kinase (*cysC*), could convert sulfate molecules into adenosine 5'-phosphosulfate (APS), followed by the generation of phosphoadenosine phospho-sulfate (PAPS), a sulfate source for sulfotransferases involved in the generation of sulfated compounds (Fig. 5). Interestingly, all MGIIa MAGs lacked the ATP sulfurylase subunits and the APS kinase. This finding was consistent for the MGIIa genera L1 (0/40) and M (0/26), but in sharp contrast to MGIIb genera where these genes were prevalent. Known sulfate transporters (e.g., SulT, SulP, CysP, and CysZ) were not detected in MGIIa Helgoland MAGs [62, 63]. These results suggest there may be alternative pathways for sulfur assimilation or incorporation of reduced sulfur-containing organic compounds.

Peptidases and membrane transport proteins in MGII MAGs

Previous studies proposed a role in organic matter degradation for MGII populations, partly due to the presence of peptidases and several transporters in their genomes [10–12, 58]. Peptidase prediction according to MEROPS classification [64] resulted in 66 sub-families detected in recovered MAGs, comprising, on average, 4.8 and 5.7% of the predicted coding sequences of MGIIa and MGIIb genomes respectively. Families C (cysteine), M (metallo), and S (serine) were among the most abundant. Extracellular peptidases were detected in MGII MAGs using a subcellular localization analysis based on amino acid sequences. For instance, all MGII MAGs encoded a predicted number of one to four copies of the extracellular subtilisin S8, likely involved in degradation of external peptides. Other peptidases are specific to the different MGII clades and might contribute to niche differentiation by targeting specific peptides. Furthermore, M28D and M14A were among the extracellular peptidases most commonly detected in MGIIa and MGIIb Helgoland MAGs (Table S5).

Membrane transporters in MGII genomes were mostly classified as alpha-type channels and P-P bond hydrolysis, decarboxylation, and oxido-reduction-driven transporters (primary active transporters) according to the TCDB database (Table S6). Most of the detected proteins were potentially involved in the transport of inorganic nutrients
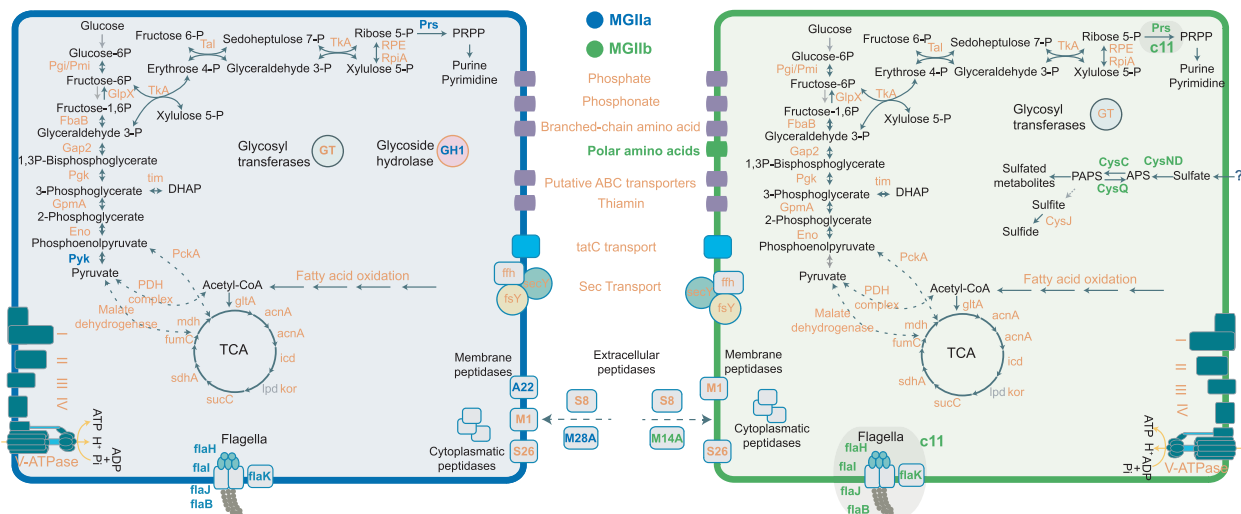
Fig. 5 Graphical representation of the predicted metabolic capabilities of MGII MAGs. Functions exclusively detected in MGIIa and MGIIb MAGs are colored in blue and green, respectively. Predicted functions shared by the two groups are colored in magenta. Substrates and products from depicted pathways are not necessarily balanced. A more detailed gene description is available in Tables S4, S5, and S6

(e.g., phosphate) and organic substrates (e.g., amino acids). Nonetheless, the searches were inconclusive in identifying the exact type of transported molecules, in agreement with previous findings [58]. Interestingly, MGIIb MAGs encoded, on average, ~34% more membrane transporters compared to MGIIa genomes (Fig. 6). Thus, MGIIb MAGs encoded, in general, a higher fraction of peptidases and membrane transporters compared to MGIIa. Exceptions to this trend were seven MAGs for which their peptidase and membrane protein contents resembled MAGs belonging to the opposite group (Fig. 6).

Rhodopsins in MGII MAGs

The presence of single rhodopsin genes in the majority of the MGII MAGs retrieved in this study suggested a photoheterotrophic metabolism (see Figs. S10 and S11 for phylogenetic and sequence analyses). The genetic context surrounding the rhodopsin genes were almost completely syntenic among MGIIa MAGs and also with published MGII genomes (Fig. S12). In MGIIa MAGs, this genomic region was uniformly composed of homologs for the geranylgeranylglyceryl phosphate synthase, a homolog of an ABC-type transporter, genes involved in the metabolism of purines (purQ and purL), and uvrA endonuclease. In contrast, the genetic context of rhodopsin genes in MGIIb MAGs was more variable. For instance, a copy of the rRNA large subunit methyltransferase E (rlmE) preceded the rhodopsin gene in three of the recovered MGIIb MAGs. However, MGIIb_c11 cluster encoded a copy of the pyridoxamine 5'-phosphate oxidase gene, similar to the genetic context observed in Ca. Thalassoarchaea euryarchaeote reported earlier [14].

Discussion

Dynamics and diversity of MGII North Sea populations

Metagenomics, oligotyping of 16S rRNA gene sequences, and microscopy consistently identified MGIIa (Ca. Poseidoniaceae) as summer and MGIIb (Ca. Thalassarchaeaceae) as winter clades. This is consistent with earlier observations in the Mediterranean Sea [15, 16], off the coast of California [17], and MGII populations recovered at a global scale [11]. Also, sharp increases in MGIIa abundance reported here of up to 23% of the microbial community, have been previously observed after phytoplankton blooms [17, 19] or as sporadic high abundances detected during extended sampling periods [3, 15, 18]. In fact, the high relative Euryarchaeota cell counts reported here, reaching up to 1.7×10^5 cells ml^{-1}, agreed with earlier reports describing $1–2 \times 10^5$ cells ml^{-1} in spring and summer seasons of coastal areas of the North Sea [18]. Our microscopy and oligotyping approaches also agreed with a higher year-round frequency of MGIIa cells in the smaller size fraction. Thus, we hypothesize that their small average cell sizes (~0.7 μm diameter) might provide protection against predation by protozoa in the picoplankton, which is often size selective for cells > 1 μm diameter [65] and could drive the short periods of high MGIIa abundance at the start of the summer.

During the 2009–2012 sampling period, ANIr values, a proxy for quantifying intra-population sequence diversity, were remarkably high for all MGII MAGs (but also in coastal metagenomes from the TARA ocean expedition, see Fig. S13). Previous analyses have proposed that populations showing higher intra-population sequence diversity might
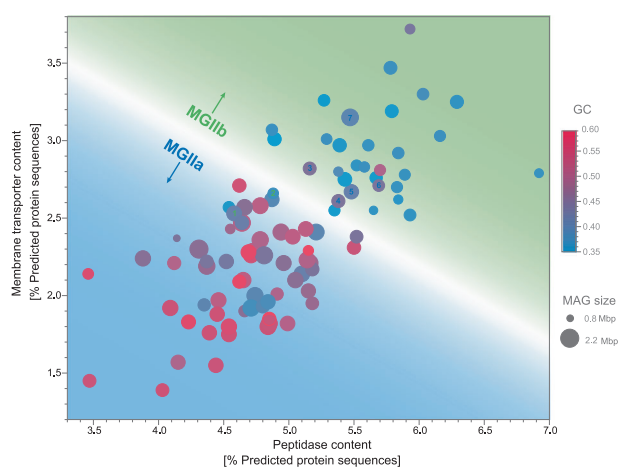
Fig. 6 Membrane transporter and peptidase contents in MGII MAGs. The membrane transporter and peptidase contents were determined in recovered MGII MAGs from Helgoland and others recovered from elsewhere around the world [11, 12]. All displayed MAGs belong to the O2, O3, L1, and M genera (showed in Fig. 1b). Highlighted MAGs that deviate from the differential clustering are TMED129 [1], SAT116 [2], 67673 [3], 69590 [4], 67694 [5], UBA111 [6], and RS406 [7]

represent differentiated/distinct sub-populations characterized by large population sizes and longer evolutionary time since the last intra-population diversity sweep [43]. This previous study has also indicated increased persistence over time for populations showing higher intra-population sequence diversity (i.e., a positive correlation between higher intra-population diversity and increased detection in a series of samples). In contrast, lower intra-population diversity (i.e., more clonal populations) could represent populations that had undergone a recent intra-population diversity sweep event [66]. Our results highlight that almost identical (clonal) MGIIb populations recurred every year in a predictable manner. Thus, we hypothesize that MGIIb populations could be subjected to high selective pressures that promote a particular genetic composition during winter seasons. Their streamlined genomes might also limit the incorporation of new genetic variations (assuming most genetic content has become essential). In contrast, the higher variance in the distribution of ANIr values observed for MGIIa MAGs, in particular to MGIIa_c6 and MGIIa_c5, revealed a higher intra-population diversity for the more persistent MGIIa populations. We hypothesize that MGIIa populations might undergo a higher frequency of selective sweeps compared to MGIIb, potentially due to stronger selection during the summer period when the predation rate is high. Alternatively, stochastic priming events of closely related MGIIa genotypes could emerge during spring and dominate the population throughout the season. The isolation of individual members of the MGII populations and high-resolution population data (e.g., metagenomes and single-cell data) over time would be

necessary in order to distinguish between these two distinct scenarios.

Niche differentiation and metabolic potential of MGII populations of the North Sea

The potential participation of MGII archaea in marine biogeochemical processes has recently been recognized as the result of several studies using culture-independent approaches [10–12, 14]. Our analyses corroborate the suggested photoheterotrophic lifestyle involving protein and lipid degradation for both MGII clades, yet we also corroborated significant differences between sympatric clades MGIIa and MGIIb. The most evident of these features was the smaller average genome size of the winter MGIIb clade where MAGs had an average 1.4 Mbp, significantly smaller than those of MGIIa (~1.9 Mbp). A trend that microorganisms dwelling in substrate-poor settings have smaller genomes has also been proposed [67]. The summer-clade MGIIa MAGs had additional genes such as those encoding for the synthesis of phosphoribosyl pyrophosphate (PRPP; nucleotides pathway), a higher number of flagella-related proteins, and specific peptidases were characteristic of MGIIa MAGs (Fig. 5). However, the presence of these features can in itself not explain the success of MGIIa populations since the MGIIb_c11 MAG encoding some of them (e.g., reduced set of flagella proteins) remained below detection limits in spring and summer. MGIIb_c11 belongs to a genus (O3) previously predicted among the few groups within MGIIb encoding flagellar proteins [11, 12]. In agreement with previous findings [11], no known chemotaxis signal transduction systems were detected, yet the ability to swim towards particles could be an advantage in a post-bloom summer situation.

MGII *Euryarchaeota* lack the genetic resources for the oxidation of small carbon molecules or a large set of hydrolases for the degradation of polysaccharides. High numbers of glycoside hydrolases are characteristic for bacterioplankton bloom populations co-occurring with algal blooms [23, 25]. Previous analyses of MAGs recovered at a global scale have also observed a reduced set of GHs in MGIIa clades and an overall lack of these proteins in MGIIb [12]. We found only one GH (GH1) in Helgoland MGIIa MAGs (clades M and L1) but other GHs reported for members of the M genus (e.g., GH13, GH57, and GH77) were not detected in Helgoland MGIIa MAGs. The low number of detected GHs is likely insufficient for effective degradation of complex glycans and we therefore hypothesize that grazing-resistant MGII use organic matter, including peptides, left in the aftermath of spring algal blooms.

Our data suggests that MGIIb populations are well adapted to conditions at the end of the winter, when most

organic substrates have been remineralized [25]. One major advantage they have is a large number of transporters in their genomes [11, 12], which may allow them to better recover what limited organic substrate is available. Although energetically costly, the capacity for sulfate assimilation exclusively found in MGIIb MAGs likely provides a competitive advantage over MGIIa in late winter when reduced sulfur compounds such as DMS, DMSO, and DMSP will likely be at minimum [68]. In contrast, in early summer the massive release of organic matter from phytoplankton during spring blooms in Helgoland [23, 25] will have replenished the pool of readily available reduced sulfur compounds. Therefore, MGIIa populations can forego the sulfate assimilation pathway, similar to other microorganisms dominating the summer season such as SAR11 [69, 70].

The inferred metabolic reconstruction from recovered MAGs highlights the metabolism of major substrates such as fatty acids and proteins. Compared to MGIIa, a higher content of peptidases and membrane transporters were characteristic of MGIIb MAGs from the German Bight and those recovered from elsewhere around the world [11, 12], suggesting that both features significantly contribute to niche differentiation between marine MGII populations (Fig. 6). In addition, extracellular peptidases detected in MGII populations from global surface water analyses [12] and the deep-sea [7] were also detected in MGII MAGs. Nonetheless, only a fraction of them (see Table S5) were predicted as extracellular peptidases using a protein localization prediction algorithm [71]. The extracellular peptidases detected in Helgoland MGII populations (e.g., M28D/ F, M14A, and S08) have been previously detected in metatranscriptomes of MGII populations from the deep-sea [7], thus suggesting a potential role for MGII in the recycling of more recalcitrant organic matter after spring phytoplankton blooms. It is worth mentioning the limitations when attempting to predict metabolic potential and phenotype from recovered MAGs. In our case, only an average of 46% of all genes in the MGII MAGs can be annotated with a function, underscoring our limited knowledge of Archaea, especially in MGII populations which lack cultured isolates. Clearly, additional efforts are needed to better resolve the metabolism of the ecologically successful and recurrent MGII populations.

Previous reports have suggested the existence of MGII populations associated with detritus particulate organic matter (POM) that could outnumber free-living MGII in oligotrophic environments [13]. In fact, flagellar components have also been suggested to mediate cell adhesion [11, 72]. Nonetheless, our microscopy and oligotyping approaches both suggested that coastal MGII populations at Helgoland are mostly free-living, although we also found them associated with particles, but less frequently throughout the year.

Based on multi-year sampling, our study documents the diversity and ecology of marine *Euryarchaeota* characteristic of the winter and summer seasons of the coastal North Sea in unprecedented detail. Based on high-quality MAGs, 16S rRNA oligotyping of size-fractionated cells, and CARD-FISH, we detected a pronounced niche separation of annually recurrent discrete MGII populations. Future studies should expand the study of MGII to more sites, acknowledging that abundances of the different clades change strongly over the course of the year and are markedly seasonal. A deeper understanding of key aspects such as motility and chemotactic behavior will, however, require the cultivation (or enrichment) of a representative of MGII. To do so would provide considerable insight into what is, at least in the North Sea, a significant microbial catalyst of organic matter remineralization.

Compliance with ethical standards

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. DeLong EF. Archaea in coastal marine environments. Proc Natl Acad Sci USA. 1992;89:5685–9.
2. Fuhrman JA, McCallum K, Davis AA. Novel major archaebacterial group from marine plankton. Nature. 1992;356:148–9.

3. Murray A, Blakis A, Massana R, Strawzewiski S, Passow U, Alldredge A, et al. A timeseries assessment of planktonic archaeal variability in the Santa Barbara Channel. Aquat Microb Ecol. 1999;20:129–45.

4. Massana R, DeLong EF, Pedrós-Alió C. A few cosmopolitan phylotypes dominate planktonic archaeal assemblages in widely different oceanic provinces. Appl Environ Microbiol. 2000;66:1777–87.

5. Karner MB, DeLong EF, Karl DM. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. Nature. 2001;409:507–10.

6. Fuhrman JA, Davis AA. Widespread Archaea and novel Bacteria from the deep sea as shown by 16S rRNA gene sequences. Marine Ecol Prog Series. 1997;150:275–85.

7. Li M, Baker BJ, Anantharaman K, Jain S, Breier JA, Dick GJ. Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. Nat Commun. 2015;6:8933.

8. López-García P, Moreira D, López-López A, Rodríguez-Valera F. A novel haloarchaeal-related lineage is widely distributed in deep oceanic regions. Environ Microbiol. 2001;3:72–8.

9. Könneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA. Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature. 2005;437:543–6.

10. Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV. Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. Science. 2012;335:587–90.

11. Rinke C, Rubino F, Messer LF, Youssef N, Parks DH, Chuvochina M, et al. A phylogenomic and ecological analysis of the globally abundant Marine Group IIarchaea (Ca. Poseidoniales ord. nov.). ISME J 2019;13:663–75.

12. Tully BJ. Metabolic diversity within the globally abundant Marine Group II Euryarchaea offers insight into ecological patterns. Nat Commun. 2019;10:271.

13. Orsi WD, Smith JM, Wilcox HM, Swalwell JE, Carini P, Worden AZ, et al. Ecophysiology of uncultivated marine euryarchaea is linked to particulate organic matter. ISME J. 2015;9:1747–63.

14. Martin-Cuadrado AB, Garcia-Heredia I, Moltó AG, López-Úbeda R, Kimes N, López-García P, et al. A new class of marine Euryarchaeota group II from the Mediterranean deep chlorophyll maximum. ISME J. 2015;9:1619–34.

15. Galand PE, Gutierrez-Provecho C, Massana R, Gasol JM, Casamayor EO. Inter-annual recurrence of archaeal assemblages in the coastal NW Mediterranean Sea (Blanes Bay Microbial Observatory). Limnol Oceano. 2010;55:2117–25.

16. Hugoni M, Taib N, Debroas D, Domaizon I, Jouan Dufournel I, Bronner G, et al. Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. Proc Natl Acad Sci USA. 2013;110:6004–9.

17. Needham DM, Fuhrman JA. Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. Nat Microbiol. 2016;1:16005.

18. Pernthaler A, Preston CM, Pernthaler J, DeLong EF, Amann R. Comparison of fluorescently labeled oligonucleotide and polynucleotide probes for the detection of pelagic marine bacteria and archaea. Appl Environ Microbiol. 2002;68:661–7.

19. Chafee M, Fernàndez-Guerra A, Buttigieg PL, Gerdts G, Eren AM, Teeling H, et al. Recurrent patterns of microdiversity in a temperate coastal marine environment. ISME J. 2018;12:237–52.

20. Buchan A, LeCleir GR, Gulvik CA, González JM. Master recyclers: features and functions of bacteria associated with phytoplankton blooms. Nat Rev Microbiol. 2014;12:686–98.

21. Fandino LB, Riemann L, Steward GF, Long RA, Azam F. Variations in bacterial community structure during a dinoflagellate bloom analyzed by DGGE and 16S rDNA sequencing. Aquat Microb Ecol. 2001;23:119–30.

22. Pinhassi J, Sala MM, Havskum H, Peters F, Guadayol O, Malits A, et al. Changes in bacterioplankton composition under different phytoplankton regimens. Appl Environ Microbiol. 2004;70:6753–66.

23. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science. 2012;336:608–11.

24. Williams TJ, Wilkins D, Long E, Evans F, DeMaere MZ, Raftery MJ, et al. The role of planktonic Flavobacteria in processing algal organic matter in coastal East Antarctica revealed using metagenomics and metaproteomics. Environ Microbiol. 2013;15:1302–17.

25. Teeling H, Fuchs BM, Bennke CM, Krüger K, Chafee M, Kappelmann L, et al. Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. Elife. 2016;5:e11888.

26. Francis TB, Krüger K, Fuchs BM, Teeling H, Amann RI. Candidatus Prosiliicoccus vernus, a spring phytoplankton bloom associated member of the Flavobacteriaceae. Syst Appl Microbiol. 2018;42:41–53.

27. Thiele S, Fuchs B, Amann R. Identification of microorganisms using the ribosomal RNA approach and fluorescence in situ hybridization. In: Wilderer PA, editor. Treatise on Water Science. Elsevier Science; Oxford, United Kingdom; 2011. p. 171–89.

28. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77.

29. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. Nat Methods. 2014;11:1144–6.

30. Eren AM, Esen OC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ. 2015;3:e1319.

31. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043–55.

32. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 2017;11:2864–8.

33. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinforma. 2010;11:119.

34. Rodriguez RL, Gunturu S, Harvey WT, Rosselló-Mora R, Tiedje JM, Cole JR, et al. The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. Nucleic Acids Res. 2018;46:W282–W8.

35. Rodriguez-R LM, Konstantinidis KT. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Preprints. 2016.

36. Jain C, Rodriguez RL, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018;9:5114.

37. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.

38. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.

39. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.

40. Orellana LH, Rodriguez RL, Konstantinidis KT. ROCker: accurate detection and quantification of target genes in short-read

metagenomic data sets by modeling sliding-window bitscores. Nucleic Acids Res. 2017;45:e14.

41. Orellana LH, Chee-Sanford JC, Sanford RA, Löffler FE, Konstantinidis KT. Year-round shotgun metagenomes reveal stable microbial communities in agricultural soils and novel ammonia oxidizers responding to fertilization. Appl Environ Microbiol. 2017;84:1646–17.

42. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinforma. 2009;10:421.

43. Meziti A, Tsementzi D, Rodriguez RL, Hatt JK, Karayanni H, Kormas KA, et al. Quantifying the changes in genetic diversity within sequence-discrete bacterial populations across a spatial and temporal riverine gradient. ISME J. 2018;13:767–79.

44. Caro-Quintero A, Konstantinidis KT. Bacterial species may exist, metagenomics reveal. Environ Microbiol. 2012;14:347–55.

45. Sorek R, Zhu Y, Creevey CJ, Francino MP, Bork P, Rubin EM. Genome-wide experimental determination of barriers to horizontal gene transfer. Science. 2007;318:1449–52.

46. Castelle CJ, Hug LA, Wrighton KC, Thomas BC, Williams KH, Wu D, et al. Extraordinary phylogenetic diversity and metabolic versatility in aquifer sediment. Nat Commun. 2013;4:2120.

47. Castelle CJ, Wrighton KC, Thomas BC, Hug LA, Brown CT, Wilkins MJ, et al. Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. Curr Biol. 2015;25:690–701.

48. Selengut JD, Haft DH, Davidsen T, Ganapathy A, Gwinn-Giglio M, Nelson WC, et al. TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. Nucleic Acids Res. 2007;35:D260–4.

49. Wu M, Scott AJ. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. Bioinformatics. 2012;28:1033–4.

50. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. Nucleic Acids Res. 2019;47:D427–32.

51. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011;7:e1002195.

52. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. 2018;27:135–45.

53. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.

54. Letunic I, Bork P. Interactive tree of life (iTOL)v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44:W242–5.

55. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

56. Diepenbroek M, Glöckner FO, Grobe P, Güntsch A, Huber R, König-Ries B, et al. Towards an integrated biodiversity and ecological research data management and archiving platform: the German federation for the curation of biological data (GFBio). Informatik 2014;2014:1711–24.

57. Ahlgren NA, Chen Y, Needham DM, Parada AE, Sachdeva R, Trinh V, et al. Genome and epigenome of a novel marine Thaumarchaeota strain suggest viral infection, phosphorothioation DNA modification and multiple restriction systems. Environ Microbiol. 2017;19:2434–52.

58. Zhang CL, Xie W, Martin-Cuadrado AB, Rodriguez-Valera F. Marine Group II Archaea, potentially important players in the global ocean carbon cycle. Front Microbiol 2015;6:1108.

59. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol. 2018;36:996–1004.

60. Teira E, Reinthaler T, Pernthaler A, Pernthaler J, Herndl GJ. Combining catalyzed reporter deposition-fluorescence in situ hybridization and microautoradiography to detect substrate utilization by bacteria and Archaea in the deep ocean. Appl Environ Microbiol. 2004;70:4411–4.

61. Ghosh A, Albers SV. Assembly and function of the archaeal flagellum. Biochem Soc Trans. 2011;39:64–9.

62. Marietou A, Roy H, Jorgensen BB, Kjeldsen KU. Sulfate transporters in dissimilatory sulfate reducing microorganisms: a comparative genomics analysis. Front Microbiol. 2018;9:309.

63. Saier MH Jr., Reddy VS, Tsu BV, Ahmed MS, Li C, Moreno-Hagelsieb G. The Transporter Classification Database (TCDB): recent advances. Nucleic Acids Res. 2016;44:D372–9.

64. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res. 2018;46:D624–32.

65. Pernthaler J. Predation on prokaryotes in the water column and its ecological implications. Nat Rev Microbiol. 2005;3:537–46.

66. Garcia SL, Stevens SLR, Crary B, Martinez-Garcia M, Stepanauskas R, Woyke T, et al. Contrasting patterns of genome-level diversity across distinct co-occurring bacterial populations. ISME J. 2018;12:742–55.

67. Giovannoni SJ, Thrash JC, Temperton B. Implications of streamlining theory for microbial ecology. Isme J. 2014;8:1553–65.

68. Archer SD, Gilbert FJ, Allen JI, Blackford J, Nightingale PD. Modelling of the seasonal patterns of dimethylsulphide production and fate during 1989 at a site in the North Sea. Can J Fish Aquat Sci. 2004;61:765–87.

69. Tripp HJ, Kitner JB, Schwalbach MS, Dacey JW, Wilhelm LJ, Giovannoni SJ. SAR11 marine bacteria require exogenous reduced sulphur for growth. Nature. 2008;452:741–4.

70. Giovannoni SJ. SAR11 bacteria: the most abundant plankton in the oceans. Ann Rev Mar Sci. 2017;9:231–55.

71. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics. 2010;26:1608–15.

72. Chaudhury P, Quax TEF, Albers SV. Versatile cell surface structures of archaea. Mol Microbiol. 2018;107:298–311.