SUPPLEMENTARY INFORMATION

Supplementary Figure S1. Schematic representation of the assembly and binning procedure.

Supplementary Figure S2. Binning of the MG-III genomic fragments. a. Principal component analysis of tetranucleotide frequencies of the MG-III DNA fragments. Reference sequences are shown as larger circles: *Nitrosopumilus maritimus SCM1*, *Aciduliprofundum boneii* T469, MGII-GG3, MGII-Thalassoarchaea, the single amplified genome (SAG) SCGC-AAA288-E19 and the DNA fragments from Cayman92, Cayman93, Guaymas31 and Guaymas32. **b.** Heat-map of the number of reads per kilobase per gigabase of metagenome collection (rpkg) of the MG-III sequences recruiting in 33 different metagenomes (only those in which any of the MG-III sequences recruited over rpkg>3 were considered for the binning analysis).

Supplementary Figure S3. Distribution of the assembled MG-III sequences among the metagenomes of the Mediterranean series and TARA collections ordered by depth. Size fraction is indicated above each column.

Supplementary Figure S4. Average nucleotide identity (ANI) among the MG-III genome bins (in bold the sequences published in this work). Dendrogram showing the similarity among the bins is shown in the *y* axis.

Supplementary Figure S5. 16S and 23S-rRNA phylogeny. a. Maximum-likelihood 16S-rRNA gene tree showing the relationship of the MG-III with other archaea. Circles at nodes in major branches indicate bootstrap support (see legend). Scale bar represents the estimated number of substitutions per site. The different MG-III clusters are indicated by different colours and named following Galand *et al.* 2008 and this work. Sampling location is indicated in each of the sequences. Those sequences from samples from more than 500m deep are underlined. **b.** Maximum likelihood 23S-rRNA gene tree showing the relationship of the MG-III with other archaea. (Picture features as explained before).

Supplementary Figures S6-S12. Maximum likelihood phylogenetic trees for the housekeeping genes RecA, RpoB, SecY, geranylgeranylglyceryl phosphate synthase, DnaK, GyrA and GyrB. Protein sequences from this study are indicated in bold and coloured accordingly with the sequence bin (see legend in Supplementary Figure S4). Genomic bins of MG-II and MG-IIII groups are indicated. Bootstrap values over 50 are indicated.

Supplementary Figure S13. Maximum likelihood phylogenetic tree of the photolyases and cryptochromes found among the MG-III bins. Protein sequences from this study are indicated in bold and coloured accordingly with their kingdom affiliation (see legend). Bootstrap values over 50 are indicated.

Supplementary Figure S14. Maximum likelihood phylogenetic tree showing the relationship of the MG-III rhodopsins with other bacterial and archaeal rhodopsins. Protein sequences from this study are indicated in bold and coloured accordingly with the sequence bin (see legend in Supplementary Figure S4). Following the nomenclature of Iverson et al. (2013), Clade A and Clade B of rhodopsins is shown. In blue are marked Pop, Pop-1, Pop-2, Pop-3 and Pop-4 euryarchaeal rhodopsins previously described. Numbers at nodes in major branches indicate bootstrap support (shown as percentages and only those >50%). Scale bar represents the estimated number of substitutions per site.

Supplementary Figure S15. Distribution of arCOG functional classes. Percentage of arCOGs predicted in the MGIII bins described in this work and MG-II marine euryarchaeal genomes MG2-GG3 and Thalasoarchaea. All genes (**a**) and genes found only in one of the MGIII bins (**b**) are indicated. Asterisks indicate categories where a significant variation was found comparing the epipelagic and pelagic MGIII.

Supplementary Figure S16. Alignment of the MG-III rhodopsins with other cloned rhodopsins sequences. Identical residues are indicated in red. Residues in blue are conserved in more than 70% of the sequences. Key amino acids for rhodopsins functionality (listed herein with *G. pallidula* numbering) are marked by colours: Lys336 (K) binds retinal, and Asp164 (D) and Glu175 (E) function as Schiff base proton acceptor and donor, respectively. Glutamine (Q) in position 172 (*) in the MG-III rhodopsins sequences indicates an absorption maxima at the blue spectrum range. Letters (G) and (B) in the name of the sequences indicate the range of the spectrum. (The GenBank accession numbers of the sequences used for the alignment are as follows: Pop-2 HF10_3D09, 82548293; Pop-3 HF70_19B12, 82548286; Pop-4 HF70_59C08, 77024964; eBAC49C08, AAY82659; HF130_81H07, 119713419; HF10_49E08, 119713779; eBAC20E09, AAS73014; HOT75m4, AAK30179; eBAC31A08 (SAR86), AAG10475; SAR86E, WP_008490645; *C.* Pelagibacter ubique HTCC1062 (SAR11), YP_266049; *Pelagibacter* sp. HTCC7211, WP_008544914; gammaproteobacteria HTCC2207 (SAR92), EAS48197; *G. pallidula*, WP_006008821; *Dokdonia donghaensis* MED134, ZP_01049273; MedDCM-JUL2012-C3793, KP211865; MedDCM-OCT2007-C1678, KP211832; MedDCM-OCT-S08-C16, ADD93192; *Exiguobacterium sibiricum* 255-15, ACB60885; *Exiguobacterium* sp. AT1b, WP_012726785; *Haloarcula marismortui* ATCC 43049, YP_136594.

Supplementary Figure S17. Classification of the DCM-metagenomes reads using the RDP (16S-rRNA) database. Only those genera which represented more than 1% were represented.

Supplementary Figure S18. Number of reads per kilobase per gigabase of metagenome collection (rpkg) for the composite genome CG-Epi1 MG-III (ordered from minus to major) compared with those obtained for the MG-II (MG2-GG3 (Iverson *et al.* 2013) and the Thalassoarchaea (Martin-Cuadrado *et al.* 2014)). Other CG MG-III were also included in the graph.

Supplementary Figure S19. Interactions (showed as percentages) calculated by Mendez-Lima *et al.,* (2015) for the MG-III with other organisms. Data extracted from Supplementary table W7 in Mendez-Lima *et al.,* (2015).

SUPPLEMENTARY TABLES

Supplementary Table S1. List of metagenomes used for recruitments in Figure 4a and Supplementary Figure 2b, sorted by temperature and depth.

Supplementary Table S2. List of sequences of each of the MG-III sequence bins.

Supplementary Table S3. List of sequences of the MG-III composite genomes (CG-MGIII).

Supplementary Table S4. Housekeeping genes found in the MG-III bins and the CG-MG-III bins (as Narasingarao *et al.*, (2012).

Supplementary Table S5. CG-MGIII-protein categories based on the arCOG database.

Supplementary Table S6. Classification of the CG-MGIII unique CDSs based on the arCOG classification.

Supplementary Table S7. List of genes involved in MG-III metabolic pathways.