Supplementary information for

Ecophysiology and global dispersal of the freshwater SAR11-IIIb clade

Clafy Fernandes[1, 2], Markus Haber[1], Paul Layoun[1, 2], Maria-Cecilia Chiriac[1], Paul-Adrian Bulzu[1], Rohit Ghai[1], Vojtech Kasalicky[1], Tanja Shabarova[1], Hans-Peter Grossart[3,4], Jason Woodhouse[3], Kasia Piwosz[5], Cecilia Alonso[6], Juan Zanetti[6], David P. Hamilton[7], Maxon Ngochera[8], Shin-ichi Nakano[9], Yusuke Okazaki[10], Michaela M. Salcher[1*]

Supplementary text

Detailed metabolism of freshwater SAR11-IIIb

Detailed description of 11 new *Allofontibacter* species, one SAR11-I species and one SAR11-II genus according to SeqCode

Supplementary Figures

Fig. S1. Map displaying the origin of *Allofontibacter* genomes.

Fig. S2: Average nucleotide identities (ANI, %) within each species of *Allofontibacter* and max. ANI to the closest relative.

Fig. S3: Conserved synteny despite low nucleotide identity between different *Allofontibacter* species.

Fig. S4: Non-metric multidimensional scaling (NMDS) ordination showing distances of SAR11-IIIb.1-16 species based on coverage per Gb values in metagenomes.

Fig. S5: Hierarchical clustering of SAR11-IIIb species based on metagenomic fragment recruitment.

Fig. S6: Stress plot displaying the relationship between observed and ordinated distances for model quality assessment (stress value: 0.083).

Fig. S7: Latitude variation across 9 SAR11-IIIb distribution clusters.

Fig. S8: Redundancy analyses (RDA) of seasonal distribution of SAR11-IIIb species and environmental factors in six lakes.

Fig. S9: Maximum abundances and growth rates of *A. medardicus* ME-17 and *A. abundans* MiE-29 under different experimental conditions (different media and concentrations of sulfur and carbon sources).

Fig. S10: Growth curves of *A. medardicus* ME-17 and *A. abundans* MiE-29 under different experimental conditions (different media and concentrations of sulfur and carbon sources).

Fig. S11: Growth rates of *A. medardicus* ME-17 and *A. abundans* MiE-29 under different experimental conditions.

Fig. S12: Growth curves of *A. medardicus* ME-17 and *A. abundans* MiE-29 under different experimental conditions (Temperature, salinity, sulfur sources, selective removal of amino acids from the media).

Fig. S13: Phylogenetic tree of isocitrate lyase (AceA).

Fig. S14: Genome synteny and genomic islands in *Allofontibacter* spp.

Fig. S15: Phylogenetic trees of proteins involved in sulfate reduction (CysD, CysNC) and cysteine biosynthesis (CysE) present in HVR2 of *A. medardicus* ME-20.

Supplementary Tables

Additional Supplementary information (Tables S1-S15) can be found as combined excel file (Fernandes_SupplTables.xlsx).

Supplementary text

Detailed description of metabolic pathways of freshwater SAR11-IIIb

The two newly cultivated species, *A. medardicus* (IIIb.8, five genome-sequenced strains) and *A. abundans* (IIIb.9, two genome-sequenced strains) display notable similarities in their metabolic profiles, while also showing some variations in specific traits. Below is a detailed overview of the metabolic characteristics observed in these newly cultured isolates (Fig. 5, Table S14, S15).

Central carbon metabolism

All SAR11-IIIb isolates encode the Embden-Meyerhof pathway (EMP), as previously described for freshwater SAR11 genomes[1] (Fig. 5). Subsequent analysis of the genomes revealed that three of the isolates, namely *A. medardicus* ME-17, *A. abundans* MiE-29, and *A. abundans* MKE-138, lacked the enzyme 6-phosphofructokinase (PfkA), which is a crucial component of the EMP pathway. However, all isolates possessed a non-oxidative branch of the pentose phosphate pathway so they still can transform fructose-6-phosphate to glyceraldehyde-3-phosphate.

All strains also encoded a complete tricarboxylic acid (TCA) cycle. Interestingly, the glyoxylate shunt pathway, characterized by the presence of the isocitrate lyase enzyme AceA, was not identified in any of our culture genomes but was present in some of our new MAGs and also observed in the previously reported isolate[1]. The distribution of AceA has been reported to be patchy in SAR11 especially in the brackish/freshwater subclades[1,2]; with its loss having already occurred in the common ancestor of the freshwater IIIb group and its closest marine relatives[3]. The single gene tree of AceA protein sequences from bacteria (Fig. S13) reaffirms the glyoxylate shunt pathway not conserved within the freshwater SAR11 lineage. The irregular distribution of aceA across the SAR11 lineages can be due to different evolutionary processes: gene loss due to niche adaptation, where reduced selective pressure led to its disappearance in certain environments; or horizontal gene transfer with selective retention, where the gene was sporadically acquired or maintained in subpopulations; convergent evolution or parallel loss, where different lineages independently retained it based on ecological pressures like the subclade Ia or due to genomic streamlining, where selective pressures favoured its deletion where it was redundant or energetically costly.

All genomes contained glycolate oxidase genes (*glcDEF*), which are consistently conserved across the SAR11 clade, and can help assimilate low molecular weight carbon substrates into biomass[3,4]. Genomic analyses further indicated the presence of the oxidative phosphorylation pathway in all genomes. Notably, one genome (*A. medardicus* ME-18) contained an additional cytochrome *bd* complex alongside a cytochrome *c* oxidase, which was shown to enhance survival at oxygen limiting conditions and to degrade harmful reactive oxygen and nitrogen species (ROS, RNS)[5]. In general, all the new SAR11-IIIb isolates seemed to have lost most of the pathways for C1 and methylated compounds which are present in marine SAR11 clades[1,6]. The glycine cleavage pathway was present in all isolates except *A. abundans* MKE-138, which had the tetrahydrofolate-ligase (FtfL) enzyme for formate oxidation. The absence of glycine cleavage in MKE-138 suggests the use of formate as alternate carbon source, possibly due to selection by environmental factors.

Nitrogen Assimilation

In the SAR11-IIIb group, ammonia can serve as the primary nitrogen source, as evidenced by the presence of genes encoding ammonia transporters (*amtB*) and the glutamine

synthetase/glutamate synthase (GS/GOGAT cycle) in all analyzed genomes (Fig. 5). Notably, a glutamate dehydrogenase (*gdhA*) was also present in one species (*A. medardicus*, IIIb.8) of the newly cultured strains, which enables the incorporation of ammonia through oxoglutarate to produce glutamate. Although the glutamate dehydrogenase has a low affinity for NH_4^+, this alternative pathway is energetically favored when ammonium availability is high[7]. This suggests an adaptive strategy in this specific SAR11-IIIb species group allowing them to efficiently utilize ammonia as a nitrogen source under conditions when it is abundant.

Biosynthesis of amino acids, cofactors and vitamins

We identified biosynthesis pathways for sixteen amino acids in the genomes of all cultures, and partial synthesis of cysteine present in ME-20 (Fig. 5, Tables S14-S15). Interestingly, one strain (*A. medardicus* ME-20) contained the serine O-acetyltransferase *cysE* in the large hypervariable region between 16S rRNA-23S rRNA and 5S rRNA (HVR2, Fig. S14). The strains exhibited uniformity in the presence or absence of most biosynthetic pathways for vitamins and cofactors. All seven strains are auxotrophic for cobalamin (vitamin B12), missing nearly all enzymes and no transporters for cobalamin were annotated explicitly, thus they may rely on external sources for vitamin precursors. The SAR11-IIIb strains were also auxotrophic for biotin (vitamin B7); however, the biotin transport system substrate-specific component protein BioY, which can function as a solitary transporter, was present. Gene *mobA,* a molybdenum cofactor required for the activity of most bacterial molybdoenzymes such as formate dehydrogenase, was present only in *A. abundans* MKE-138, which was also the only genome encoding a formate dehydrogenase. All strains are auxotrophic for thiamine, pyridoxal, panthothenate, and prototrophic for other vitamin B compounds, such as riboflavin, coenzyme A and NAD, and for heme.

Sulfur Assimilation

All new isolates were deficient in the assimilatory sulfur reduction pathway except for *A. medardicus* ME-20, that had genes *cysNC* and *cysD* to convert sulfate to PAPS (3-phosphoadenosine-5-phosphosulfate) in HVR2 (Fig. S14, S15). Further, the genomes encoded partial pathways for the incorporation of sulfide into cysteine and methionine, but most lacked predicted serine O-acetyltransferase (*cysE*) and homoserine O-succinyltransferase genes (Fig. 5, Tables S14, S15). There was a conspicuous absence of sulfate transporters, which leads to the prevailing conclusion of SAR11-IIIb group using reduced sulfur compounds like cysteine and methionine to fulfill their metabolic needs. However, all genomes encode genes for a putative thiosulfate/3-mercaptopyruvate sulfurtransferase enzyme (*sseA*) and also an inner membrane protein (*yeeE*), which was recently demonstrated to mediate thiosulfate uptake[8]. This might be one source of inorganic sulfur uptake and synthesis of sulfur based amino acids. Indeed, growth assays revealed that the addition of sodium thiosulfate restored growth of strains in the absence of sulfur based amino acids in a species-specific fashion (Fig. 4).

Membrane transporters and environmental sensing

All SAR11-IIIb genomes harbor the *pstSCAB* operon facilitating high-affinity uptake of inorganic phosphate, regulated by the PhoU[9] and a putative ATPase (PhoH[10]), which was previously described to be part of the phosphate regulon (Fig. 5). Analogous to their marine counterparts, the isolates retained essential sensory systems encompassing acidity sensing, nitrogen response, and redox reaction components (ChvG-ChvI, NtrY-NtrX, and RegB-RegA). All isolates contained the versatile L-amino acid transporter (AapJQMP) capable of conveying diverse acidic, basic, and aliphatic amino acids with broad solute specificity. Intriguingly, strain variability was evident in the presence of additional amino acid

transporters with all but one strain possessing branched chain amino acid transporters, except *A. abundans* MiE-29 which had a polar amino acid transporter. All genomes also encoded two additional putative ABC transporters with unknown functionality.

Apart from ATP dependent transporters, we also identified genes for C4-dicarboxylate (TRAP, tripartite ATP independent periplasmic), magnesium (*mgtE*), and lipopolysaccharide transporters. Rhodopsin genes were present in all culture genomes and were predicted to be green-light absorbing with leucine residues at the retinal binding site in the third transmembrane domain. All genomes contained the *crtEBIY* genes, responsible for biosynthesis of β-carotene, and the gene *blh*, which encodes a dioxygenase that catalyzes the final step in the retinal biosynthesis pathway.

Detailed species description of 11 novel SAR11-IIIb, one SAR11-I species and one SAR11-II genus, names were registered at SeqCode[11]

***Allofontibacter abundans*, sp. nov.**

Etymology: a.bun'dans. L. I. adj. abundans, abundant, referring to high global abundances.

Description: Type strain is *Allofontibacter abundans* MiE-29 (GCA_965235095), isolated from 5 m depth from Lake Milada, Czechia (date: 2019-10-15), *via* high-throughput dilution to extinction cultivation. MiE-29 has a genome size of 1.1 Mbp with a genomic GC content of 29.4%, contains 3 rRNA genes and 31 tRNAs. The genome is complete, consisting of a circular chromosome. The genome contains genes encoding rhodopsins and the biosynthetic pathway for retinal biosynthesis. No genes for flagella or pilus assembly and chemotaxis were annotated. Pathways for glycolate oxidation and the biosynthesis of 16 amino acids were predicted. Further, pathways for riboflavin, NAD, coenzyme A, and heme biosynthesis were identified. The closest cultivated relatives are '*Candidatus* Fonsibacter ubiquis' LSUCC0530 (GCF_002688585.1; later reclassified to '*Ca*. Allofontibacter communis'), with an average amino acid identity of 86.5% and average nucleotide identity of 85.2% and another newly proposed species, *Allofontibacter medardicus* ME-17 (GCA_965235075), with an AAI of 92.7% and an ANI of 91.1%. Current GTDB classification (R220): d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Pelagibacterales; f__Pelagibacteraceae; g__Fonsibacter; s__.

***Allofontibacter medardicus*, sp. nov.**

Etymology: L. masc. adj. medardicus, pertaining to Lake Medard (Czechia), the isolation source of the type strain.

Description: Type strain is *Allofontibacter medardicus* ME-17 (GCA_965235075), isolated from 5 m depth from Lake Medard, Czechia (date: 2019-10-22), *via* high-throughput dilution to extinction cultivation. ME-17 has a genome size of 1.1 Mbp with a genomic GC content of 29.6%, contains 3 rRNA genes and 31 tRNAs. The genome is complete, consisting of a circular chromosome. The genome contains genes encoding rhodopsins and the biosynthetic pathway for retinal biosynthesis. No genes for flagella or pilus assembly and chemotaxis were annotated. Pathways for glycolate oxidation and the biosynthesis of 16 amino acids were predicted. Further, pathways for riboflavin, NAD, coenzyme A, and heme biosynthesis were identified. The closest cultivated relatives are '*Candidatus* Fonsibacter ubiquis' LSUCC0530 (GCF_002688585.1; later reclassified to '*Ca*. Allofontibacter communis'), with an average amino acid identity of 87.2% and average nucleotide identity of 85.2% and another newly proposed species, *Allofontibacter abundans* MiE-29 (GCA_965235095), with an AAI of 92.7% and an ANI of 91.1%. Current GTDB classification (R220): d__Bacteria;

p__Pseudomonadota; c__Alphaproteobacteria; o__Pelagibacterales; f__Pelagibacteraceae; g__Fonsibacter; s__Fonsibacter sp018882565.

***Allofontibacter africanus*, sp. nov.**

Etymology: a.fri.ca'nus, L. masc. adj. africanus, of Africa, pertaining to the isolation source of the MAG (Lake Malawi) and a prevalence in the African Great Lakes.

Description: Type genome is *Allofontibacter africanus* N-Mw6-13nov23-081 (GCA_965235885), a metagenome-assembled genome (MAG) assembled from 50 m depth from Lake Malawi, Malawi (date: 2023-11-13). N-Mw6-13nov23-081 has a genome size of 1.1 Mbp with a genomic GC content of 30%, contains 1 rRNA gene and 28 tRNAs. The genome is of high quality, consisting of 7 contigs, with a completeness of 95.2%, contamination of 0% and strain heterogeneity of 0% as assessed with checkM. The metagenome was assembled with FLYE from combined long- and short-read sequencing (Oxford Nanopore and Illumina NovaSeq). Metagenomic fragment recruitment of >600 samples from five continents indicate that the species is present only in the African Great Lakes Malawi, Tanganyika, and Kivu. The closest cultivated relatives are '*Candidatus* Fonsibacter ubiquis' LSUCC0530 (GCF_002688585.1; later reclassified to '*Ca*. Allofontibacter communis'), with an average amino acid identity of 68.64 % and average nucleotide identity of 71.78 % and another newly proposed species, *Allofontibacter abundans* MiE-29 (GCA_965235095), with an AAI of 68.51 % and an ANI of 71.6 %. Current GTDB classification (R220): d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Pelagibacterales; f__Pelagibacteraceae; g__Fonsibacter; s__Fonsibacter sp016882225.

***Allofontibacter scandinavicus*, sp. nov.**

Etymology: scan.di.na'vi.cus, N.L. masc. adj., scandinavicus, pertaining to the Scandinavian region, the isolation source of the MAG and a main occurrence in Scandinavian lakes.

Description: Type genome is *Allofontibacter scandinavicus* AM1_bin-0028 (GCA_903874225.1), a metagenome-assembled genome (MAG) assembled from 0.5 m depth from Lake Alinen Mustajärvi, Finland (date: 2015-08-8). AM1_bin-0028 has a genome size of 1.1 Mbp with a genomic GC content of 28.9%, contains 3 rRNA genes and 31 tRNAs. The genome is of high quality, consisting of 55 contigs, with a completeness of 100%, contamination of 0% and strain heterogeneity of 0% as assessed with checkM. The metagenome was assembled with Megahit (version 1.1.13) from short-read sequencing (Illumina MiSeq). Metagenomic fragment recruitment of >600 samples from five continents indicate that the species is mainly present in boreal lakes in Scandinavia. The closest cultivated relatives are '*Candidatus* Fonsibacter ubiquis' LSUCC0530 (GCF_002688585.1; later reclassified to '*Ca*. Allofontibacter communis'), with an average amino acid identity of 72.71 % and average nucleotide identity of 75.18 % and another newly proposed species, *Allofontibacter abundans* MiE-29 (GCA_965235095), with an AAI of 73.3 % and an ANI of 75.56 %. Current GTDB classification (R220): d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Pelagibacterales; f__Pelagibacteraceae; g__Fonsibacter; s__Fonsibacter sp903869725.

***Allofontibacter borealis*, sp. nov.**

Etymology: bo.re.a'lis, L. masc. adj., pertaining to the boreal region of the Northern hemisphere.

Description: Type genome is *Allofontibacter borealis* Umea-bin-09620 (GCA_903909545.1), a metagenome-assembled genome (MAG) assembled from 1-5 m depth from Lake Bjarntjarnan, Sweden (date: 2018-01-01/07). Umea-bin-09620 has a genome size of 1 Mbp with a genomic GC content of 30.2% and contains 20 tRNAs. The genome is of high quality, consisting of 243 contigs, with a completeness of 96.4%, contamination of 0.1% and strain

heterogeneity of 0% as assessed with checkM. The metagenome was assembled with Megahit (version 1.1.13) from short-read sequencing (Illumina MiSeq). Metagenomic fragment recruitment of >600 samples from five continents indicate that the species is present mainly in boreal lakes in Europe and North America. The closest cultivated relatives are '*Candidatus* Fonsibacter ubiquis' LSUCC0530 (GCF_002688585.1; later reclassified to '*Ca*. Allofontibacter communis'), with an average amino acid identity of 82.49% and average nucleotide identity of 81.54 % and another newly proposed species, *Allofontibacter abundans* MiE-29 (GCA_965235095), with an AAI of 84.48% and an ANI of 83.54 %. Current GTDB classification (R220): d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Pelagibacterales; f__Pelagibacteraceae; g__Fonsibacter; s__.

***Allofontibacter oligotrophicus*, sp. nov.**

Etymology: o.li.go.tro'phi.cus, Gr. masc. adj. oligo, little; Gr. masc. adj. trophikos, nursing, tending; N.L. masc. adj. oligotrophicus, oligotrophic, referring to the low nutrient content of the isolation sites and the high abundance in oligotrophic lakes.

Description: Type genome is *Allofontibacter oligotrophicus* N-Balt2-05jul22-047 (GCA_965235975), a metagenome-assembled genome (MAG) assembled from 0.5 m depth from the slightly brackish part of Vistula Lagoon, Poland (date: 2022-07-05). N-Balt2-05jul22-047 has a genome size of 0.92 Mbp with a genomic GC content of 29.4% and contains 30 tRNAs. The genome is of high quality, consisting of 3 contigs, with a completeness of 96.4%, contamination of 0% and strain heterogeneity of 0% as assessed with checkM. The metagenome was assembled with FLYE from combined long- and short-read sequencing (Oxford Nanopore and Illumina NovaSeq). Metagenomic fragment recruitment of >600 samples from five continents indicate that the species is highly abundant in oligotrophic lakes in temperate and subtropical regions. The closest cultivated relatives are '*Candidatus* Fonsibacter ubiquis' LSUCC0530 (GCF_002688585.1; later reclassified to '*Ca*. Allofontibacter communis'), with an average amino acid identity of 92.57% and average nucleotide identity of 90.07% and another newly proposed species, *Allofontibacter medardicus* ME-17 (GCA_965235075), with an AAI of 85.45% and an ANI of 88.16%. Current GTDB classification (R220): d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Pelagibacterales; f__Pelagibacteraceae; g__Fonsibacter; s__Fonsibacter sp947497305.

***Allofontibacter subtropicus*, sp. nov.**

Etymology: sub.tro'pi.cus, L. masc. adj., subtropicus, pertaining to subtropical zone, the isolation source of the MAG.

Description: Type genome is *Allofontibacter subtropicus* N-SamH-20apr23-026 (GCA_965235415), a metagenome-assembled genome (MAG) assembled from 15m depth from the Lake Samsonvale, Australia (date:2023-04-20). N-SamH-20apr23-026 has a genome size of 1.15 Mbp with a genomic GC content of 29.3%, contains 3 rRNA genes and 34 tRNAs. The genome is of high quality, consisting of 100 contigs, with a completeness of 100%, contamination of 1.2% and strain heterogeneity of 100% as assessed with checkM. The metagenome was assembled with FLYE from combined long-and short-read sequencing (Oxford Nanopore and Illumina NovaSeq). Metagenomic fragment recruitment of >600 samples from five continents indicate that the species is present mainly in subtropical lakes. The closest cultivated relatives are '*Candidatus* Fonsibacter ubiquis' LSUCC0530 (GCF_002688585.1; later reclassified to '*Ca*. Allofontibacter communis'), with an average amino acid identity of 88.13% and average nucleotide identity of 85.63% and another newly proposed species, *Allofontibacter medardicus* ME-17 (GCA_965235075), with an AAI of 90.92% and an ANI of 88.79%. Current GTDB classification (R220): d__Bacteria;

p__Pseudomonadota; c__Alphaproteobacteria; o__Pelagibacterales; f__Pelagibacteraceae; g__Fonsibacter; s__Fonsibacter sp023257975.

***Allofontibacter universalis*, sp. nov.**

Etymology: u.ni.ver.sa'lis, L. masc. adj., universalis, universal, pertaining to the widespread distribution of the species.

Description: Type genome is *Allofontibacter universalis* N-InaE-25sep22-010 (GCA_965236175), a metagenome-assembled genome (MAG) assembled from 5 m depth from Lake Inawashiro, Japan (date: 2022-09-25). N-InaE-25sep22-010 has a genome size of 1.0 Mbp with a genomic GC content of 29.3% and contains 29 tRNAs. The genome is of high quality, consisting of 4 contigs, with a completeness of 98.8%, contamination of 0% and strain heterogeneity of 0% as assessed with checkM. The metagenome was assembled with FLYE from combined long-and short-read sequencing (Oxford Nanopore and Illumina NovaSeq). Metagenomic fragment recruitment of >600 samples from five continents indicate that the species is present in many lakes around the world. The closest cultivated relatives are '*Candidatus* Fonsibacter ubiquis' LSUCC0530 (GCF_002688585.1; later reclassified to '*Ca*. Allofontibacter communis'), with an average amino acid identity of 87.08% and average nucleotide identity of 85.18% and another newly proposed species, *Allofontibacter abundans* MiE-29 (GCA_965235095), with an AAI of 92.44% and an ANI of 90.87%. Current GTDB classification (R220): d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Pelagibacterales; f__Pelagibacteraceae; g__Fonsibacter; s__Fonsibacter sp000371845.

***Allofontibacter meridianamericanus*, sp. nov.**

Etymology: me.ri.di.a.na.me.ri.ca'nus, L. masc. adj. meridianus, Southerly, to the south; N.L. masc. adj. americanus, American; L. masc. adj., meridianamericanus, South American, referring to continent from where the MAG was isolated.

Description: Type genome is *Allofontibacter meridianamericanus* N-IMU-22jan24-050 (GCA_965235635), a metagenome-assembled genome (MAG) assembled from 0.5 m depth from a freshwater dam, Represa de India Muerta, Uruguay (date: 2024-01-22). N-IMU-22jan24-050 has a genome size of 1.0 Mbp with a genomic GC content of 29.3% and contains 30 tRNAs. The genome is of high quality, consisting of 4 contigs, with a completeness of 95.8%, contamination of 1.2% and strain heterogeneity of 100% as assessed with checkM. The metagenome was assembled with FLYE from combined long-and short-read sequencing (Oxford Nanopore and Illumina NovaSeq). Metagenomic fragment recruitment of >600 samples from five continents indicate that the species is present mainly in subtropical and tropical lakes in South America. The closest cultivated relatives are '*Candidatus* Fonsibacter ubiquis' LSUCC0530 (GCF_002688585.1; later reclassified to '*Ca*. Allofontibacter communis'), with an average amino acid identity of 86.94% and average nucleotide identity of 84.39% and another newly proposed species, *Allofontibacter abundans* MiE-29 (GCA_965235095), with an AAI of 90.5% and an ANI of 88.11%. Current GTDB classification
(R220):d__Bacteria;p__Pseudomonadota;c__Alphaproteobacteria;o__Pelagibacterales;f__Pelagibacteraceae;g__Fonsibacter;s__

***Allofontibacter temperatus*, sp. nov.**

Etymology: tem.pe.ra'tus, L. masc. adj., temperatus, referring to temperate climate, the species mainly occurs in lakes of the temperate region of the Northern Hemisphere.

Description: Type genome is *Allofontibacter temperatus* ZE-03apr19-LR-3 (GCA_964203055.1), a metagenome-assembled genome (MAG) assembled from 5 m depth from Lake Zurich, Switzerland (date: 2019-04-03). ZE-03apr19-LR-3 has a genome size of

0.9 Mbp with a genomic GC content of 29.4% and contains 28 tRNAs. The genome is of high quality, consisting of 3 contigs, with a completeness of 94%, contamination of 0% and strain heterogeneity of 0% as assessed with checkM. The metagenome was assembled with FLYE from combined long-and short-read sequencing (Oxford Nanopore and Illumina NovaSeq). Metagenomic fragment recruitment of >600 samples from five continents indicate that the species is highly abundant in temperate lakes of the Northern Hemisphere. The closest cultivated relatives are '*Candidatus* Fonsibacter ubiquis' LSUCC0530 (GCF_002688585.1; later reclassified to '*Ca*. Allofontibacter communis'), with an average amino acid identity of 87.44% and average nucleotide identity of 85.23% and another newly proposed species, *Allofontibacter abundans* MiE-29 (GCA_965235095), with an AAI of 93.82% and an ANI of 92.5%. Current GTDB classification (R220): d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Pelagibacterales; f__Pelagibacteraceae; g__Fonsibacter; s__Fonsibacter sp000510845.

***Allofontibacter baikalensis*, sp. nov.**

Etymology: ba.i.ka.len'sis L. masc. adj., baikalensis, pertaining to Lake Baikal, the origin of the metagenome-assembled genome.

Description: Type genome is *Allofontibacter baikalensis* Baikal-deep-G36 (GCA_009693745.1), a metagenome-assembled genome (MAG) co-assembled from 1250 m and 1350 m depth from Lake Baikal, Russia (date: 2018-03-29). Baikal-deep-G36 has a genome size of 1 Mbp with a genomic GC content of 29.4%, contains 2 rRNA genes and 25 tRNAs. The genome is of high quality, consisting of 66 contigs, with a completeness of 95.2%, contamination of 3.9% and strain heterogeneity of 100% as assessed with checkM. The metagenome was assembled with IDBA-UD assembler from short-read sequencing (Illumina HiSeq 3000/4000). Metagenomic fragment recruitment of >600 samples from five continents indicate that the species is present mainly in Lake Baikal. The closest cultivated relatives are '*Candidatus* Fonsibacter ubiquis' LSUCC0530 (GCF_002688585.1; later reclassified to '*Ca*. Allofontibacter communis'), with an average amino acid identity of 85.75% and average nucleotide identity of 84.65% and another newly proposed species, *Allofontibacter medardicus* ME-17 (GCA_965235075), with an AAI of 88.73% and an ANI of 87.93%. Current GTDB classification (R220): d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Pelagibacterales; f__Pelagibacteraceae; g__Fonsibacter; s__Fonsibacter sp009693745.

***Pelagibacter malawensis*, sp. nov.**

Etymology: N.L. masc. adj., malawensis, pertaining to Lake Malawi, the origin of the metagenome-assembled genome.

Description: Type genome is *Pelagibacter malawensis* N-Mw13-23nov23-053 (GCA_965235955), a metagenome-assembled genome (MAG) assembled from 5 m depth from Lake Malawi, Malawi (date: 2023-11-23). N-Mw13-23nov23-053 has a genome size of 1 Mbp with a genomic GC content of 29.4%, contains 3 rRNA genes and 28 tRNAs. The genome is of high quality, consisting of 7 contigs, with a completeness of 99.5%, contamination of 0.5% and strain heterogeneity of 0% as assessed with checkM. The metagenome was assembled with FLYE from combined long-and short-read sequencing (Oxford Nanopore and Illumina NovaSeq). The closest cultivated relatives are '*Candidatus* Pelagibacter ubique SAR11 HTCC9022 (GCF_000472565.1), with an average amino acid identity of 75% and average nucleotide identity of 76.4% and '*Candidatus* Pelagibacter ubique SAR11 HTCC7211 (GCF_000155895.1), with an AAI of 74.2% and an ANI of 76.4%. Current GTDB classification (R220): d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Pelagibacterales; f__Pelagibacteraceae; g__Pelagibacter; s__Pelagibacter sp016870175.

***Pelagilacustribacter*, gen. nov.**

Etymology: Pe.la.gi.la.cus.tri.bac'ter, L. neut. n. pelagus, of the sea, N.L. masc. adj. lacustris, belonging to a lake, N.L. masc. n. bacter, a rod; Pelagilacustribacter, a freshwater genus of otherwise marine Pelagibacterales.

Description: Type species is *Pelagilacustribacter hypolimneticus.* A genus within marine SAR11-II (Pelagibacterales) that was obtained from deep freshwater lakes.

***Pelagilacustribacter hypolimneticus*, sp. nov.**

Etymology: hy.po.lim.ne'ti.cus G. hypo, below, under; N.L. neut. adj. limneticus, of a lake; N.L. masc. adj. hypolimneticus, referring to the deep zone of lakes, the hypolimnion

Description: Type genome is *Pelagilacustribacter hypolimneticus* TrH-25oct19-165 (GCA_965235125), a metagenome-assembled genome (MAG) assembled from 150 m depth from Lake Traunsee, Austria (date: 2019-10-25). TrH-25oct19-165 has a genome size of 1 Mbp with a genomic GC content of 29.2% and contains 19 tRNAs. The genome is of high quality, consisting of 87 contigs, with a completeness of 93.4%, contamination of 4.8% and strain heterogeneity of 20% as assessed with checkM. The metagenome was assembled with megahit from short-read sequencing (Illumina NovaSeq). The closest cultivated relative is *Cosmipelagibacter malulaniensis* HIMB058 (GCA_000419545.1), with an average amino acid identity of 58.8% and average nucleotide identity of 68.1%. Current GTDB classification (R220): d__Bacteria; p__Pseudomonadota; c__Alphaproteobacteria; o__Pelagibacterales; f__Pelagibacteraceae; g__SYDM01; s__SYDM01 sp005801485.

References:

1 Henson, M. W., Lanclos, V. C., Faircloth, B. C. & Thrash, J. C. Cultivation and genomics of the first freshwater SAR11 (LD12) isolate. *The ISME journal* **12**, 1846-1860 (2018).

2 Lanclos, V. C. *et al.* Ecophysiology and genomics of the brackish water adapted SAR11 subclade IIIa. *The ISME Journal* **17**, 620-629 (2023).

3 Eiler, A. *et al.* Tuning fresh: radiation through rewiring of central metabolism in streamlined bacteria. *The ISME journal* **10**, 1902-1914 (2016).

4 Carini, P., Steindler, L., Beszteri, S. & Giovannoni, S. J. Nutrient requirements for growth of the extreme oligotroph 'Candidatus Pelagibacter ubique'HTCC1062 on a defined medium. *The ISME journal* **7**, 592-602 (2013).

5 Borisov, V. B. *et al.* Bacterial oxidases of the cytochrome bd family: Redox enzymes of unique structure, function, and utility as drug targets. *Antioxidants & redox signaling* **34**, 1280-1318 (2021).

6 Sun, J. *et al.* One carbon metabolism in SAR11 pelagic marine bacteria. *PloS one* **6**, e23973 (2011).

7 Yan, D. Protection of the glutamate pool concentration in enteric bacteria. *Proceedings of the National Academy of Sciences* **104**, 9475-9480 (2007).

8 Tanaka, Y. *et al.* Crystal structure of a YeeE/YedE family protein engaged in thiosulfate uptake. *Science advances* **6**, eaba7637 (2020).

9 diCenzo, G. C., Sharthiya, H., Nanda, A., Zamani, M. & Finan, T. M. PhoU allows rapid adaptation to high phosphate concentrations by modulating PstSCAB transport rate in Sinorhizobium meliloti. *Journal of Bacteriology* **199**, 10.1128/jb. 00143-00117 (2017).

10 Kim, S., Makino, K., Amemura, M., Shinagawa, H. & Nakata, A. Molecular analysis of the phoH gene, belonging to the phosphate regulon in Escherichia coli. *Journal of bacteriology* **175**, 1316-1324 (1993).

11 Hedlund, B. P. *et al.* SeqCode: a nomenclatural code for prokaryotes described from sequence data. *Nat. Microbiol.* **7**, 1702-1708 (2022). https://doi.org:10.1038/s41564-022-01214-9

Supplementary Figures

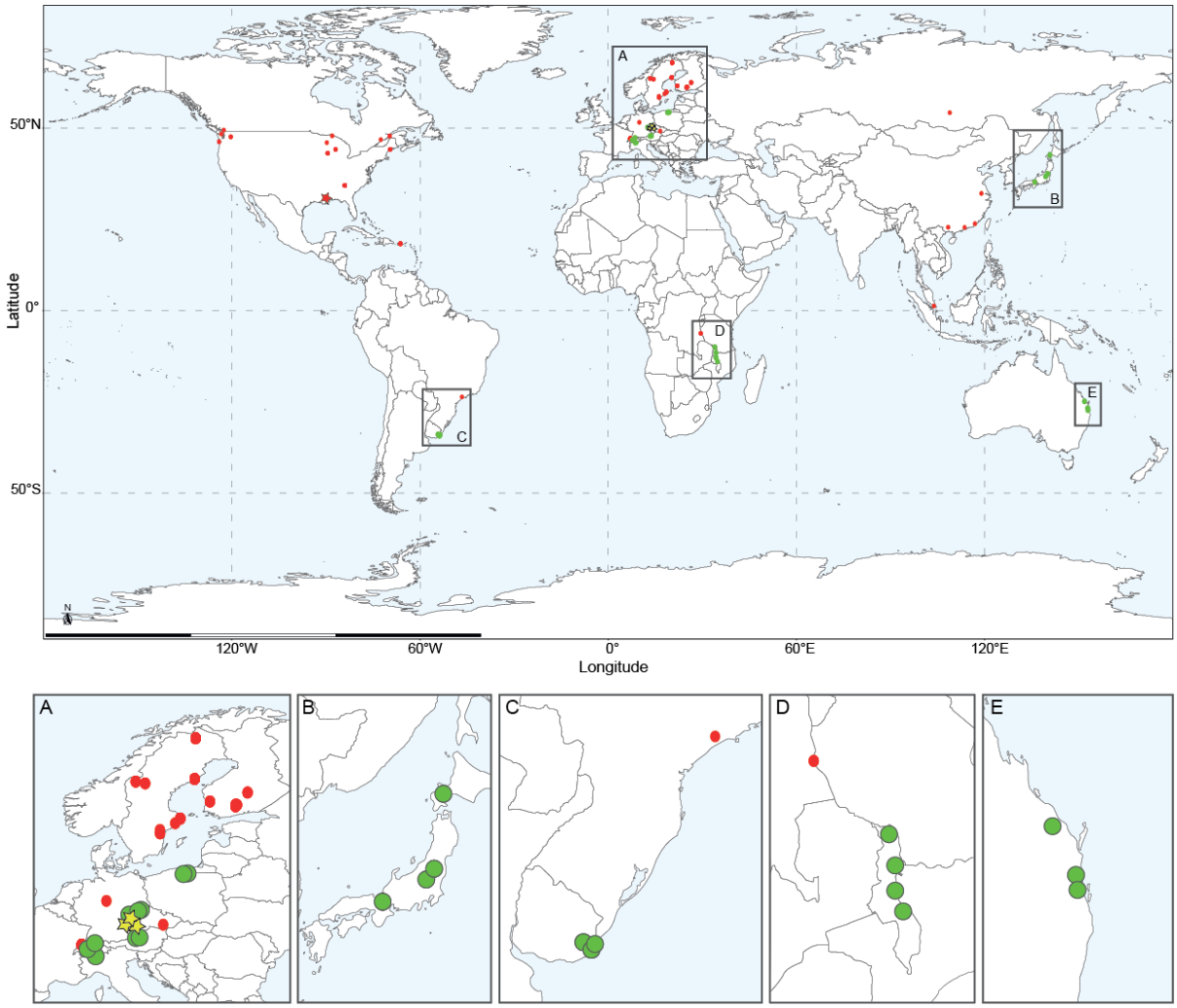

Fig. S1. Map displaying the origin of *Allofontibacter* genomes. A-E show enlarged parts of the map with novel genomes gained in this study. Yellow stars indicate novel cultures; green circles indicate novel high-quality MAGs from A. Europe, B. Japan, C. Uruguay, D. Malawi and E. Australia. Red circles indicate genomes from public databases and a red star indicates the culture *A. communis* LSUCC0530.

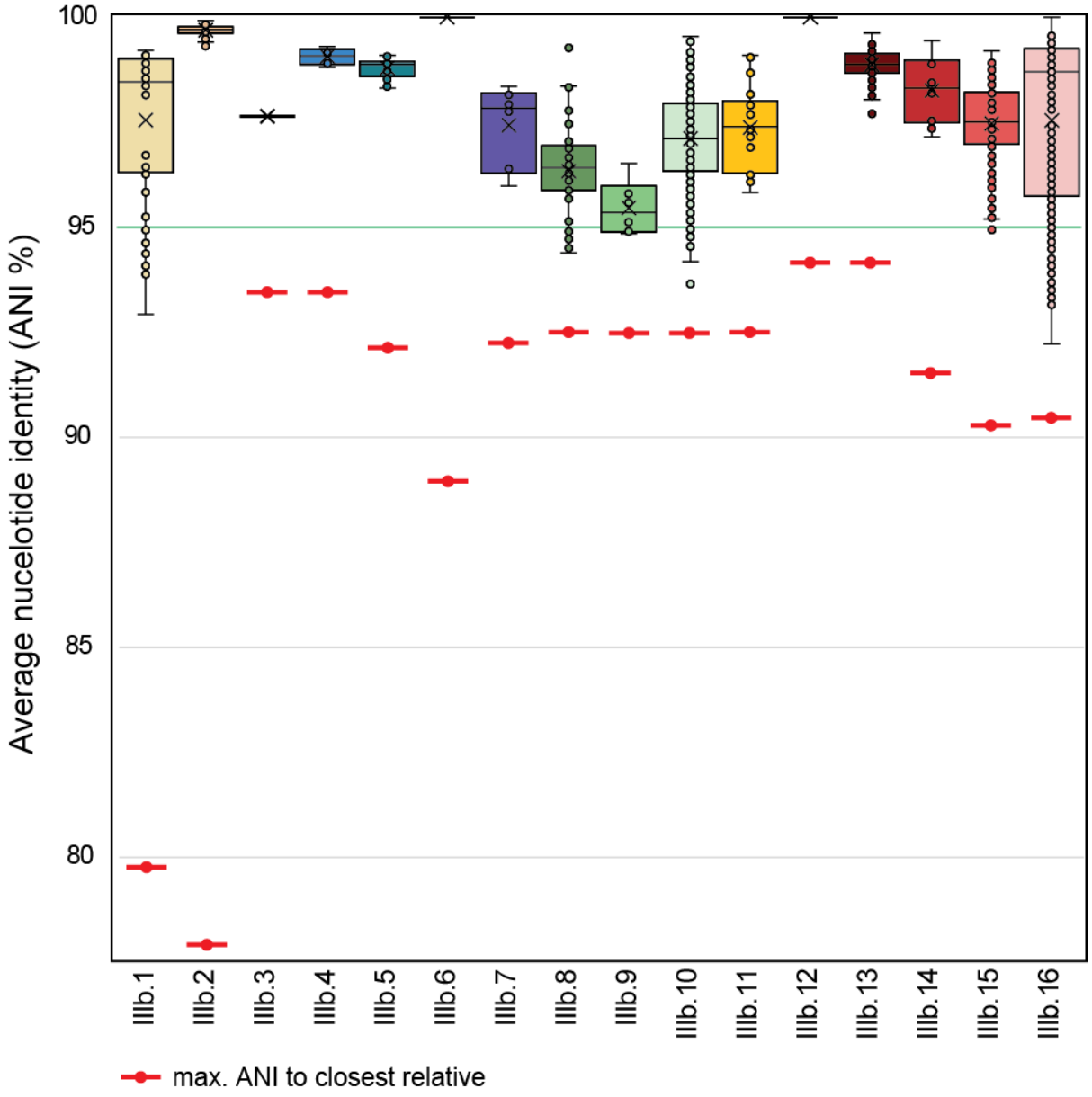

Fig. S2: Average nucleotide identities (ANI, %) within each species of *Allofontibacter* and max. ANI to the closest relative.

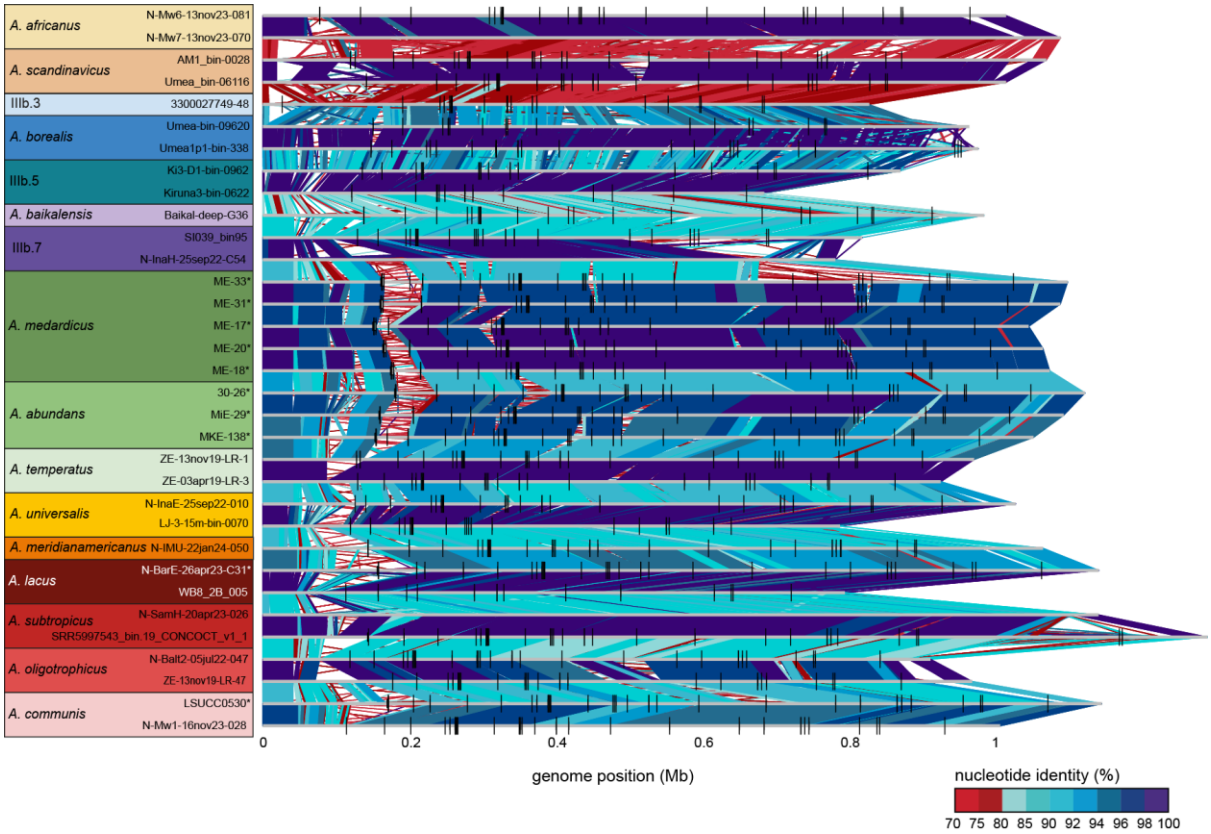

Fig. S3: Conserved synteny despite low nucleotide identity between different *Allofontibacter* species. Whole-genome alignments and BLASTN comparison of one to two representatives per *Allofontibacter* species and all complete genomes analyzed in this study. Genomes are sorted according to taxonomy and color coded as in Fig. 1. Asterisks indicate complete genomes; all others were concatenated in order based on the closest complete relative and turned to start with DnaA for an easier display of synteny and nucleotide identity. Note that nucleotide identities >95% were rarely detected between different species and that HVR2 (located at approx. 0.1-0.2 Mb) was not assembled in multiple MAGs.

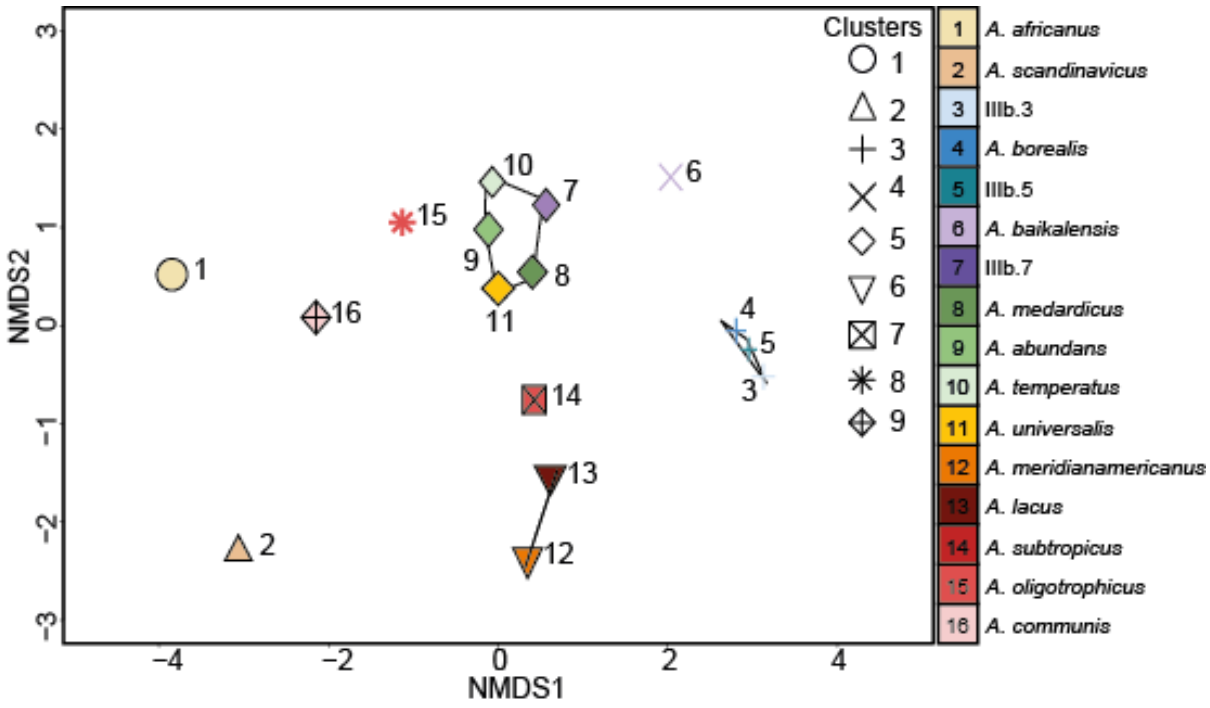

Fig. S4: Non-metric multidimensional scaling (NMDS) ordination showing distances of SAR11-IIIb.1-16 species based on coverage per Gb values in metagenomes (stress value: 0.071). Nine clusters were identified through hierarchical clustering using Ward's method, with the optimum number of clusters determined by the average silhouette width and evaluation of the dendrogram structure (see Fig. S5). Points in the NMDS plot represent species that are coloured and shaped according to their assigned cluster.

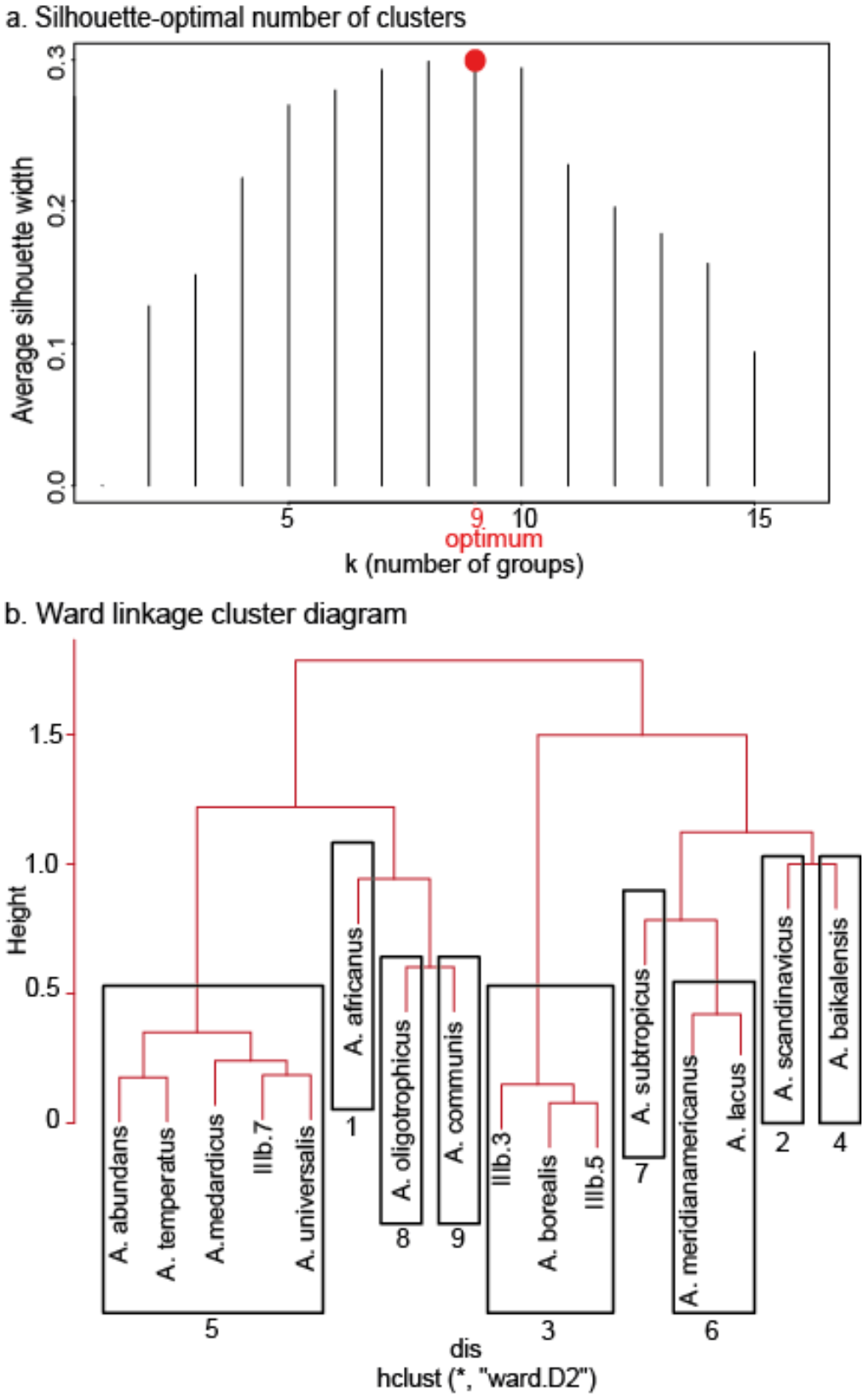

Fig. S5: Hierarchical clustering of SAR11-IIIb species based on metagenomic fragment recruitment. a: Barplot showing the average silhouette widths for k up to 15 groups. The best partition by this criterion is the one with the largest average silhouette width, i.e., in 9 groups. **b:** Hierarchical Clustering Dendrogram; the dendrogram displays clustering of species using Ward's linkage method; rectangles highlight the nine clusters identified.

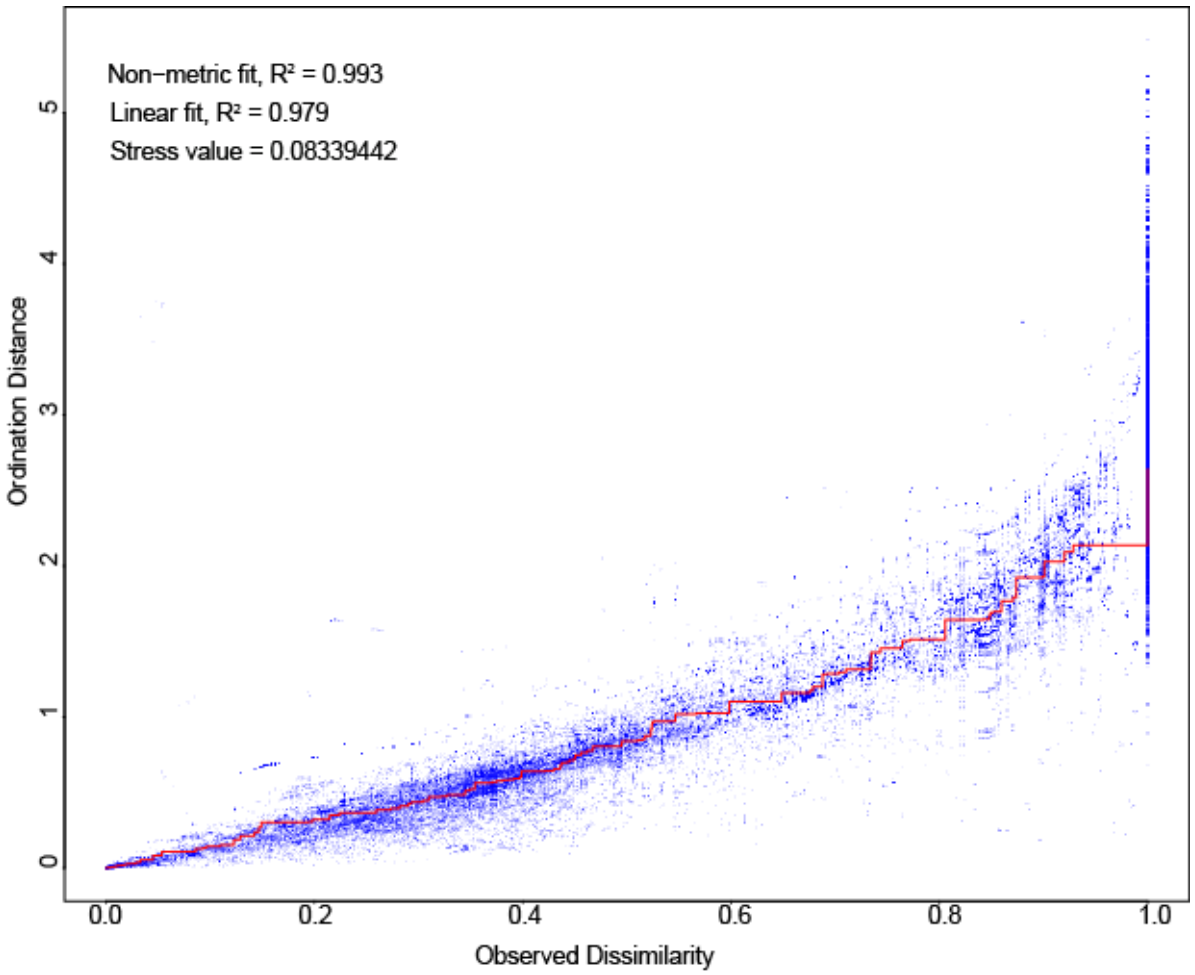

Fig. S6: Stress plot displaying the relationship between observed and ordinated distances for model quality assessment (stress value: 0.083).

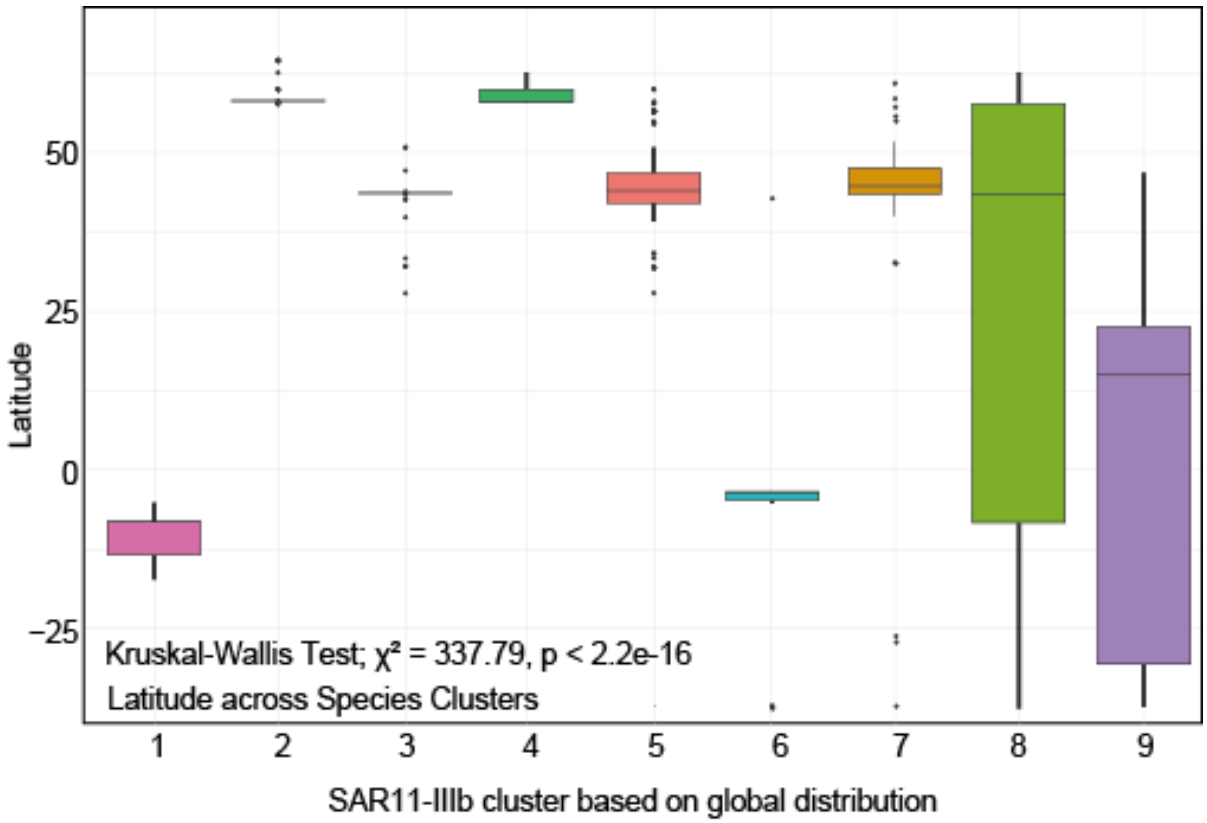

Fig. S7: Latitude variation across SAR11-IIIb distribution clusters. Boxplots illustrate the variation in latitude for each identified cluster (see Fig. S5 for clusters). Clusters are shown on the x-axis, and Latitude values are on the y-axis.

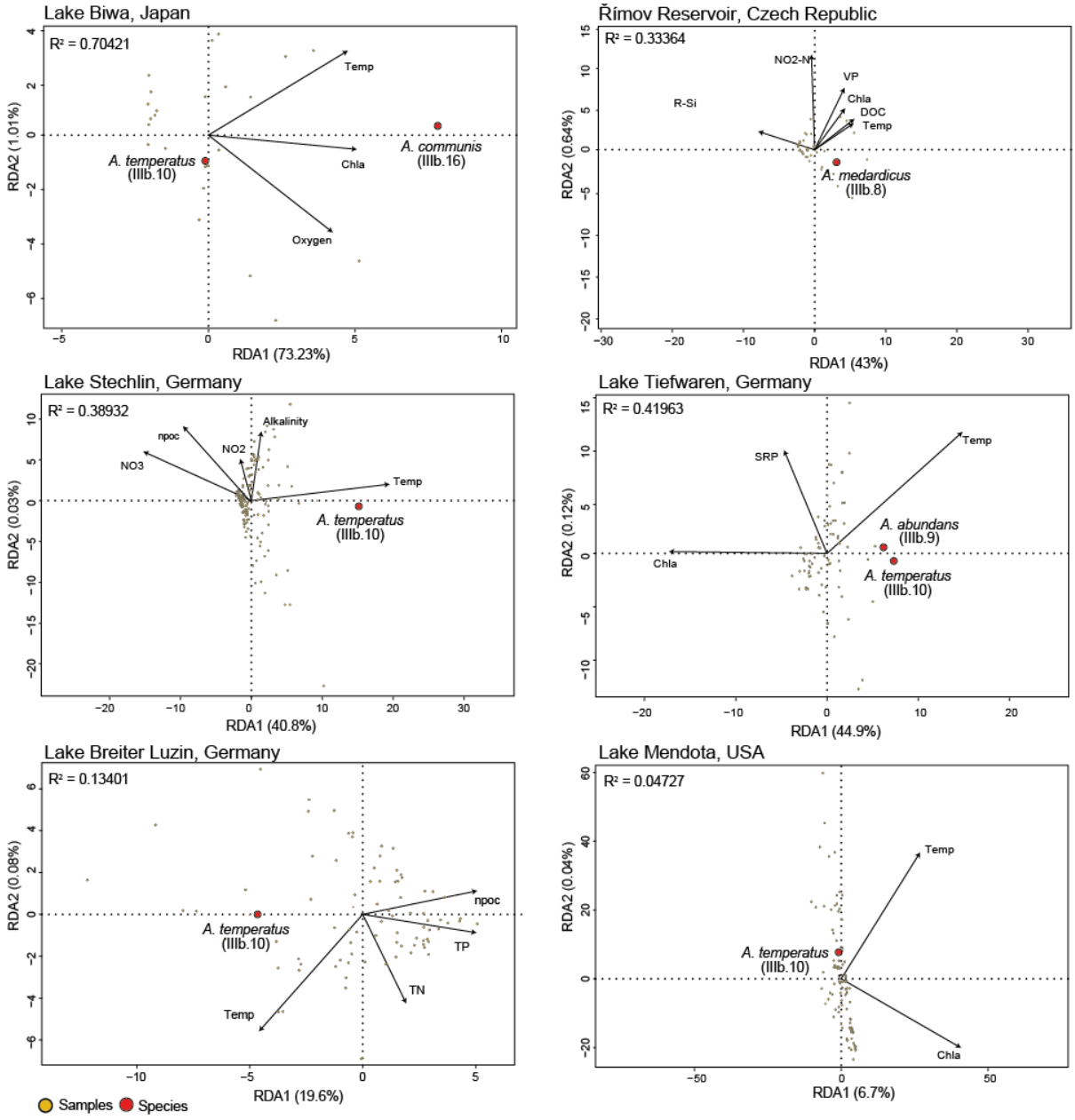

Fig. S8: Redundancy analyses (RDA) of seasonal distribution of SAR11-IIIb species and environmental factors in six lakes. Key environmental factors were identified through forward selection ($p < 0.05$). Sampling points and the dominant SAR11-IIIb species in each of the six lakes are color-coded by yellow and red circles respectively. Temp: water temperature (°C); Chla: chlorophyll *a* concentrations (µg l^{-1}); SRP: soluble reactive phosphorus concentrations (µg l^{-1}); NO3: nitrate concentrations (mg l^{-1}); NO2: nitrite concentrations (mg l^{-1}); npoc: non-purgeable particulate organic carbon (total organic carbon); TP: total phosphorus concentrations (µg l^{-1}); TN: total nitrogen concentrations (mg l^{-1}); VP: volatile phosphorus; DOC: dissolved organic carbon concentrations (mg l^{-1}); R-Si: dissolved reactive silicon concentrations (mg l^{-1}); Oxygen: dissolved oxygen concentrations (mg l^{-1}).

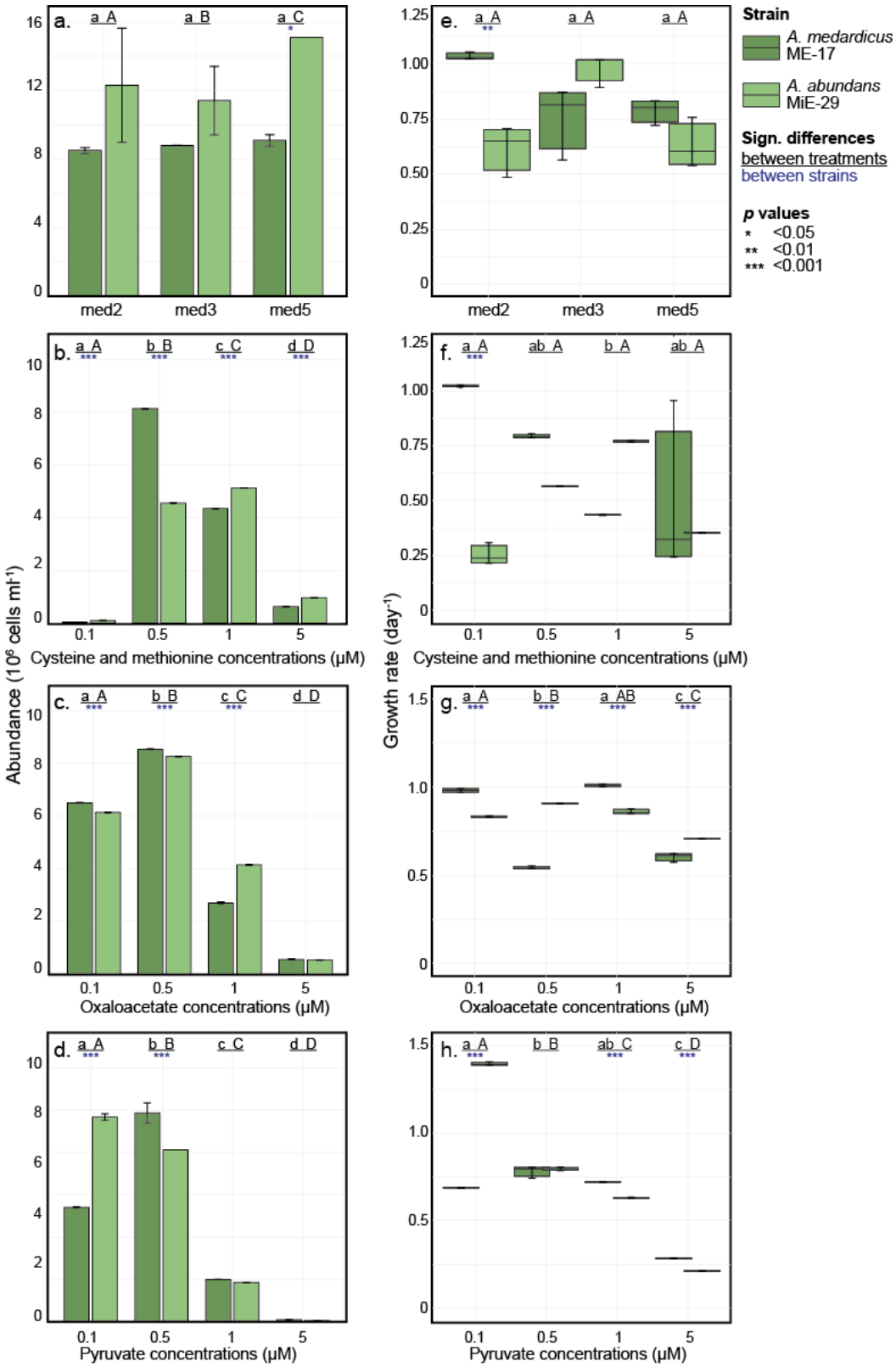

Fig. S9: Maximum abundances (a-d) and growth rates (e-h) of *A. medardicus* ME-17 and *A. abundans* MiE-29 under different experimental conditions. a, e. Test of three different media (med2, med3, med5); **b, f.** cysteine and methionine in different concentrations (0.1-5 µM); **c, g.** pyruvate in different concentrations (0.1-5 µM); **d, h.** oxaloacetate in different concentrations (0.1-5 µM). Asterisks or letters indicate significant differences between treatments.

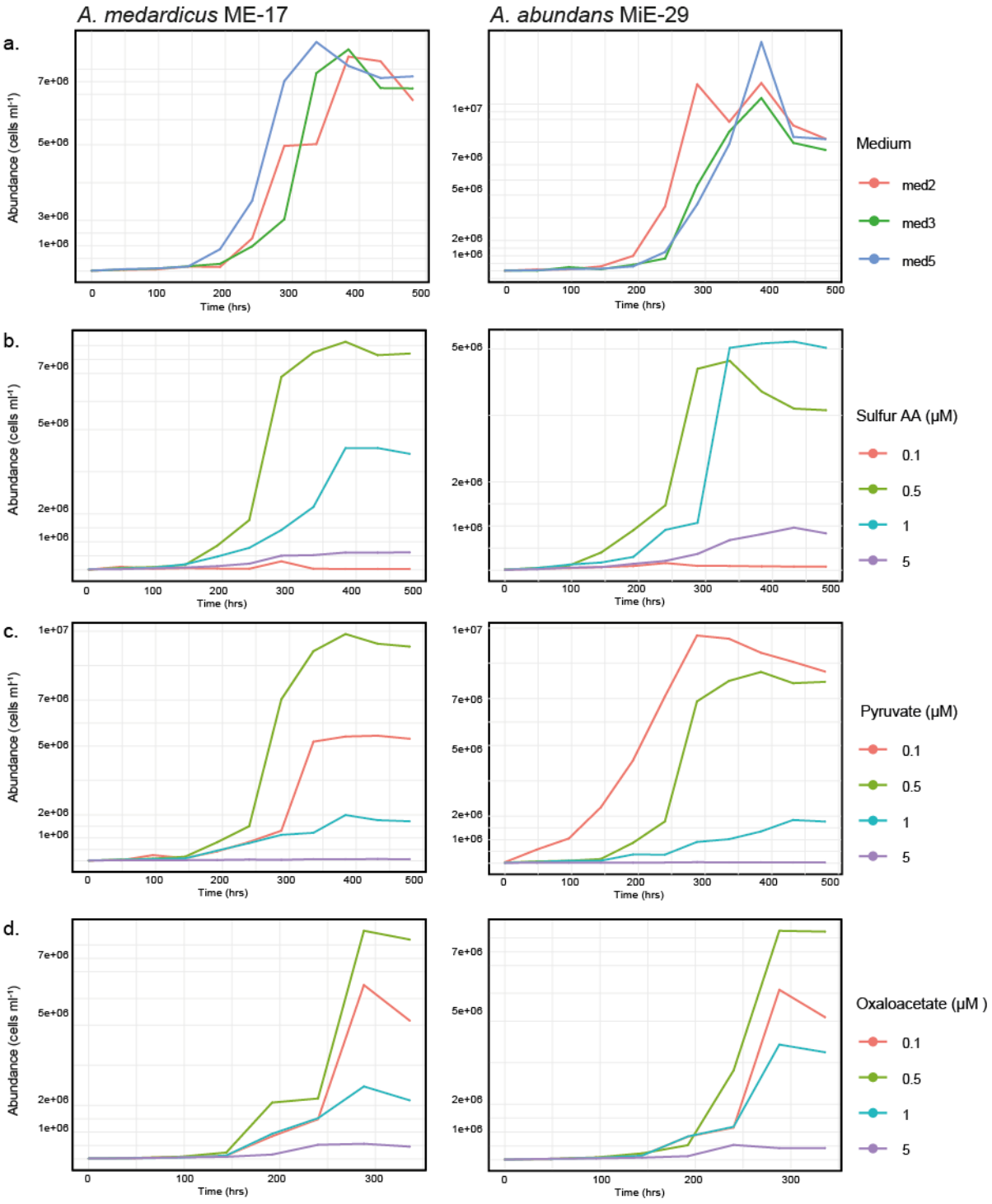

Fig. S10: Growth curves of *A. medardicus* ME-17 and *A. abundans* MiE-29 under different experimental conditions. a. Test of three different media (med2, med3, med5); **b.** cysteine and methionine in different concentrations (0.1-5 µM); **c.** pyruvate in different concentrations (0.1-5 µM); **d.** oxaloacetate in different concentrations (0.1-5 µM). Shown are averages of triplicated growth assays, raw data can be found as Table S11. Note log y-axis.

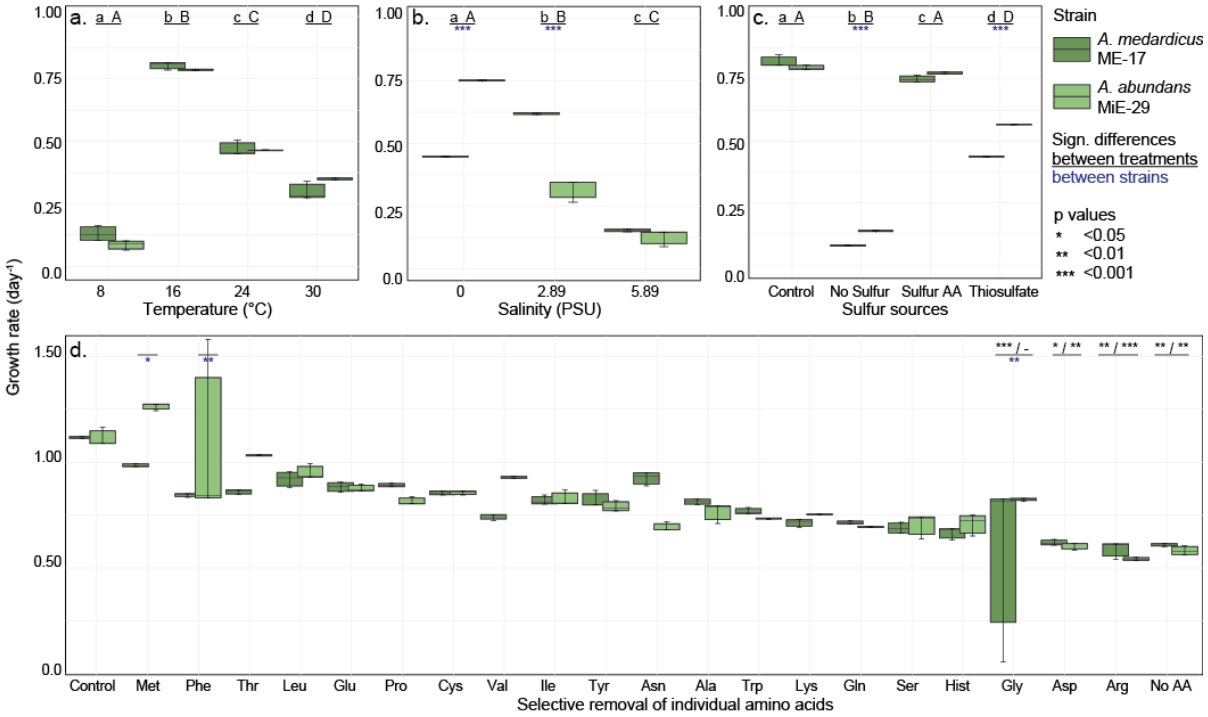

Fig. S11: Growth rates of *A. medardicus* ME-17 and *A. abundans* MiE-29 under different experimental conditions. a: Temperature gradients (8, 16, 24, and 30°C); **b:** Salinity gradients (0, 2.89, and 5.89 PSU); **c:** Different sulfur sources (sulfur-rich amino acids cysteine and methionine and sodium thiosulfate); **d:** Selective removal of individual amino acids. Asterisks indicate significant differences between treatments. Asterisks or letters indicate significant differences between treatments.

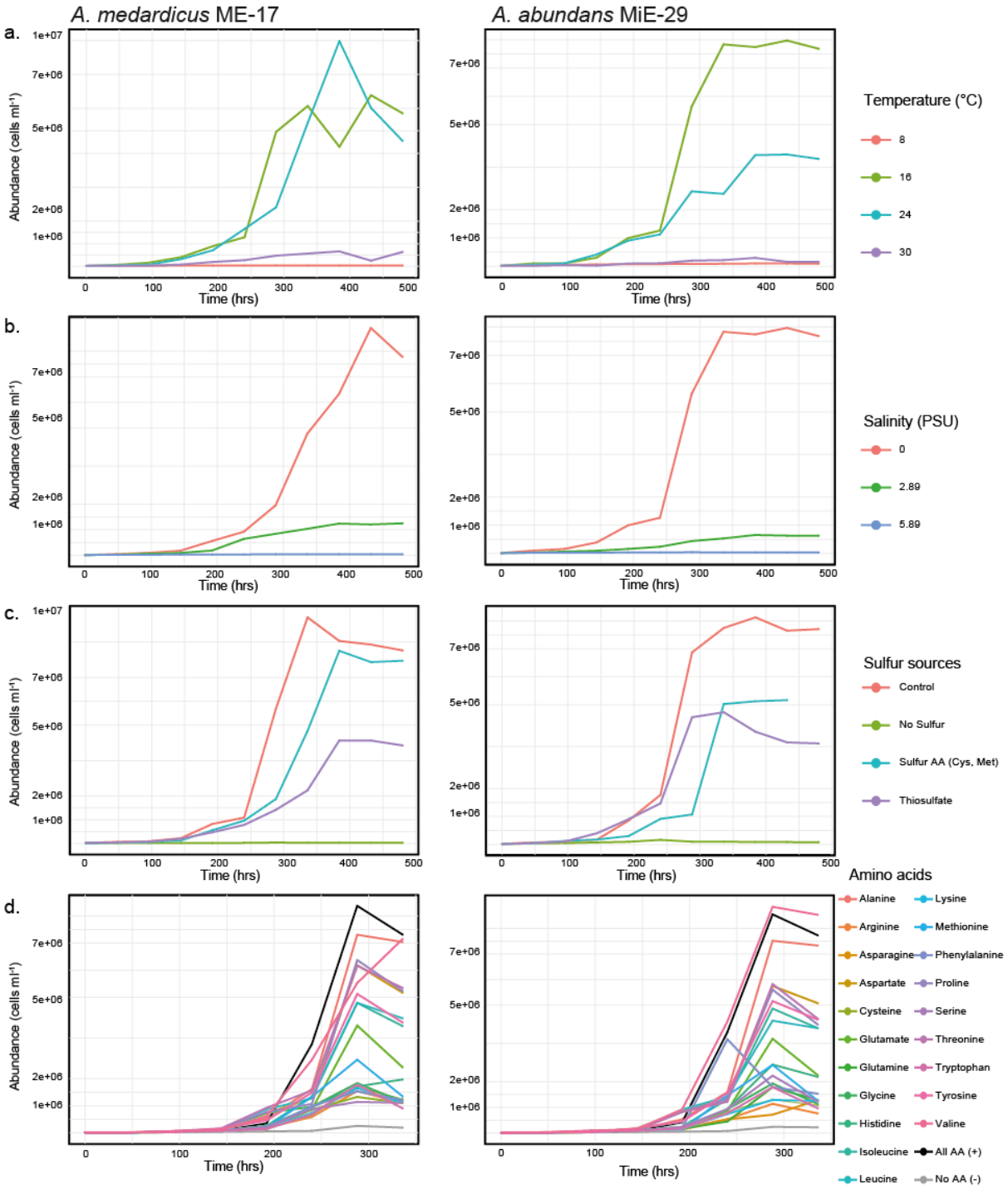

Fig. S12: Growth curves of *A. medardicus* ME-17 and *A. abundans* MiE-29 under different experimental conditions. a. Test of different temperatures (8-30°C); **b.** different salinities (0-5.89 PSU [practical salinity unit]); **c.** sulfur sources (sulfur-rich amino acids cysteine and methionine and sodium thiosulfate); **d.** selective removal of individual amino acids. Shown are averages of triplicated growth assays, raw data can be found as Table S11. Note log y-axis.

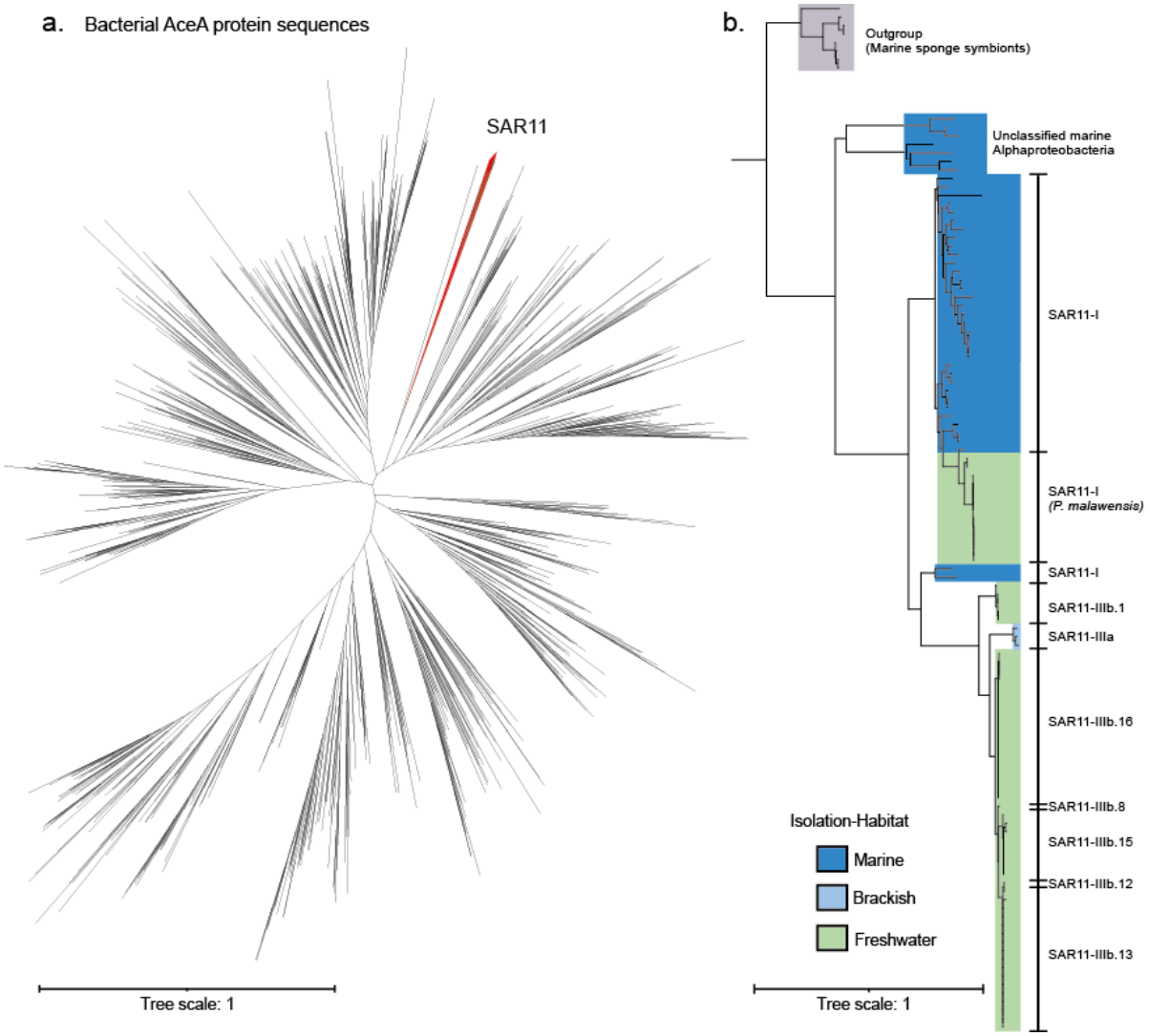

Fig. S13: Phylogenetic tree of isocitrate lyase (AceA). a. Full tree of 1367 AceA protein sequences from all bacterial phyla, the SAR11 clade is highlighted in red. **b.** Subtree of SAR11 AceA protein sequences and closely related taxa. The habitat of origin is indicated by different colours.

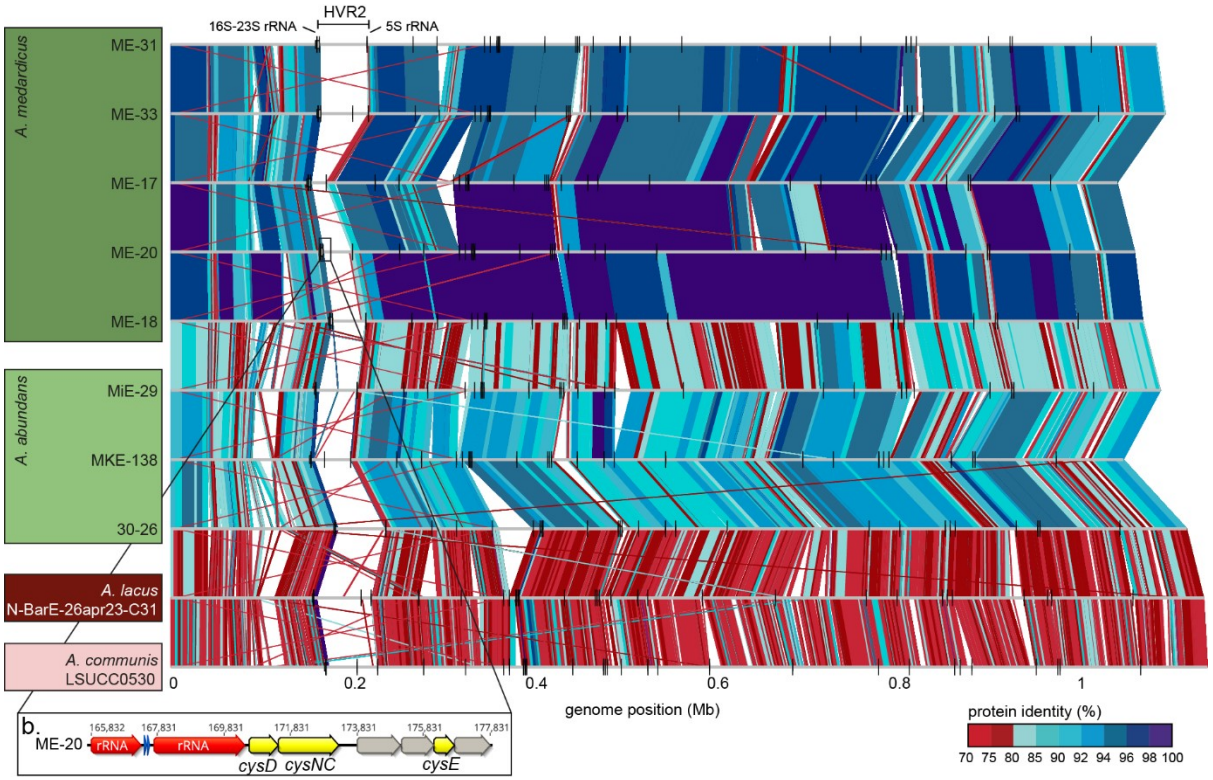

Fig. S14: Genome synteny and genomic islands in *Allofontibacter* spp. a. Whole-genome alignments of all complete *Allofontibacter* genomes analyzed in this study. The hypervariable region 2 (HVR2) in between 16S-23S rRNA and 5S rRNA appears conserved in the genus. Genomes are sorted according to taxonomy as follows: *A. medardicus* (ME-31, ME-33, ME-17, ME-20, ME-18), *A. abundans* (30-26, MiE-29, MKE-138), *A. lacus* (N-BarE-26apr23-C31), *A. communis* (LSUCC0530), tRNA and rRNA genes are indicated by short lines. **b.** Zoom-in of the first few genes in HVR2 of strain ME-20. Genes involved in sulfate reduction (*cysD, cysNC*) and cysteine biosynthesis (*cysE*) are highlighted in yellow.

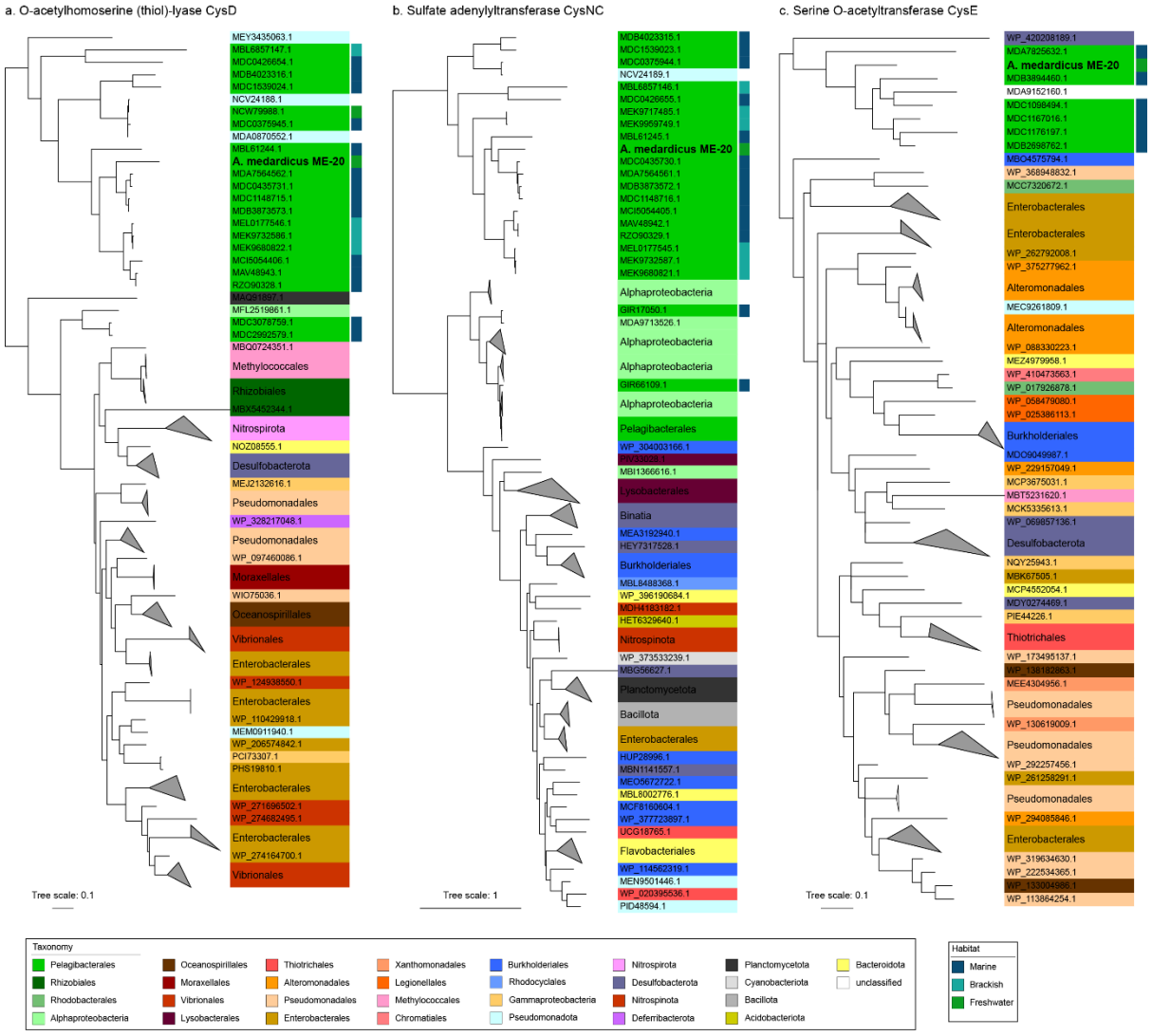

Fig. S15: Phylogenetic trees of proteins involved in sulfate reduction (CysD, CysNC) and cysteine biosynthesis (CysE) present in HVR2 of *A. medardicus* ME-20. a. Phylogenetic tree of O-acetylhomoserine (thiol)-lyase CysD including 100 closely related protein sequences. **b.** Phylogenetic tree of sulfate adenylyltransferase CysNC including 100 closely related protein sequences. c. Phylogenetic tree of serine O-acetyltransferase CysE including 100 closely related protein sequences. Accession numbers are color-coded by taxonomy and collapsed at order level. The habitat of origin of Pelagibacterales is indicated by different colours.

Supplementary Tables:

Tables S1-S15 can be found as combined excel file (Fernandes_SupplTables.xlsx).

Table S1: Sampling details for newly sequenced metagenomes.

Table S2: Details on genomes used in this study.

Table S3: List of TIGRFAMs used for phylogenomic tree reconstruction.

Table S4: Average nucleotide identity (ANI) matrix of all genomes used in this study.

Table S5: Average amino acid identity (AAI) matrix of all genomes used in this study.

Table S6: Metagenomic fragment recruitment results for 16 *Allofontibacter* species.

Table S7: Metagenomic fragment recruitment results for 16 Allofontibacter species in time series metagenomes.

Table S8: Physico-chemical data for time series metagenomes.

Table S9. Statistical tests for seasonal distribution of *Allofontibacter* species.

Table S10: Media components.

Table S11: Raw data of growth assays conducted in this study.

Table S12. ANOVA and post-hoc tests (Turkey-HSD) for growth assays based on maximum abundances.

Table S13. ANOVA and post-hoc tests (Turkey-HSD) for growth assays based on maximum growth rates.

Table S14: Overview of metabolic pathways present in the culture genomes.

Table S15: Presence/absence of individual KEGG IDs (metabolic pathways) detected in different *Allofontibacter* species.