

ARTICLE OPEN



Closed genomes uncover a saltwater species of *Candidatus* *Electronema* and shed new light on the boundary between marine and freshwater cable bacteria

Mantas Sereika^{1,3}, Francesca Petriglieri^{1,3}, Thomas Bygh Nymann Jensen¹, Artur Sannikov², Morten Hoppe², Per Halkjær Nielsen¹, Ian P. G. Marshall², Andreas Schramm^{1,2} and Mads Albertsen¹✉

© The Author(s) 2023

Cable bacteria of the *Desulfobulbaceae* family are centimeter-long filamentous bacteria, which are capable of conducting long-distance electron transfer. Currently, all cable bacteria are classified into two candidate genera: *Candidatus* *Electronema*, typically found in freshwater environments, and *Candidatus* *Electrothrix*, typically found in saltwater environments. This taxonomic framework is based on both 16S rRNA gene sequences and metagenome-assembled genome (MAG) phylogenies. However, most of the currently available MAGs are highly fragmented, incomplete, and thus likely miss key genes essential for deciphering the physiology of cable bacteria. Also, a closed, circular genome of cable bacteria has not been published yet. To address this, we performed Nanopore long-read and Illumina short-read shotgun sequencing of selected environmental samples and a single-strain enrichment of *Ca. Electronema aureum*. We recovered multiple cable bacteria MAGs, including two circular and one single-contig. Phylogenomic analysis, also confirmed by 16S rRNA gene-based phylogeny, classified one circular MAG and the single-contig MAG as novel species of cable bacteria, which we propose to name *Ca. Electronema halotolerans* and *Ca. Electrothrix laxa*, respectively. The *Ca. Electronema halotolerans*, despite belonging to the previously recognized freshwater genus of cable bacteria, was retrieved from brackish-water sediment. Metabolic predictions showed several adaptations to a high salinity environment, similar to the “saltwater” *Ca. Electrothrix* species, indicating how *Ca. Electronema halotolerans* may be the evolutionary link between marine and freshwater cable bacteria lineages.

The ISME Journal (2023) 17:561–569; <https://doi.org/10.1038/s41396-023-01372-6>

INTRODUCTION

Cable bacteria of the *Desulfobulbaceae* family (*Desulfobacterota*) are centimeter-long, multicellular filamentous bacteria capable of long-distance electron transfer [1–3]. According to the current taxonomic framework, they belong to genus *Candidatus* *Electronema* (freshwater-based) or *Candidatus* *Electrothrix* (saltwater-based) [4]. Cable bacteria can be found globally in freshwater and saltwater sediments [5–7] as well as around oxygen-releasing roots of aquatic plants [8, 9].

The ecological borderline between saltwater and freshwater habitats continues to be an unresolved challenge in microbiology [10]. Even though these aquatic environments share some ecological features, the radical shift in salinity and ionic concentration suggests that the transition from high to low salinity habitats must be accompanied by substantial changes in the metabolic repertoire and cellular complexes, in response to physicochemical shifts and substrate availability [10, 11]. These mechanisms have not yet been clarified in cable bacteria, even though the Na⁺/H⁺ antiporter NhaA has been previously suggested as a potential discriminant between marine and freshwater cable bacteria [12].

The current metabolic model proposes that cells belonging to one cable filament can exhibit two different kinds of physiological features: at the anoxic zone within the deeper layers of the sediment, cells oxidize sulfide and the resulting electrons are transferred along the conductive structure to the oxic zone, where the cells utilize them to reduce oxygen [12, 13]. This long-distance electron transport (LDET) has been demonstrated using Raman microscopy [3] and the conductive structure involved has recently been proposed to consist of carbohydrates and proteins containing a sulfur-ligated nickel group, which would be an unprecedented form of biological electron transport [14].

Despite enrichment efforts [12, 15], cable bacteria species have not been isolated in pure culture. Therefore, the currently available approach for acquiring cable bacteria genomic sequences is through shotgun sequencing of microbial communities and recovery of metagenome-assembled genomes (MAGs). Although MAGs of cable bacteria have been published by previous studies [12, 13, 15], the publically available cable bacteria genomes are presently highly fragmented, with multiple draft genomes missing 16S rRNA genes and exhibiting reduced genome completeness.

¹Center for Microbial Communities, Aalborg University, Aalborg, Denmark. ²Center for Electromicrobiology, Aarhus University, Aarhus, Denmark. ³These authors contributed equally: Mantas Sereika, Francesca Petriglieri. ✉email: ma@bio.aau.dk

Received: 7 November 2022 Revised: 11 January 2023 Accepted: 13 January 2023
Published online: 25 January 2023

The overall significant fragmentation and lower genome quality of existing cable bacteria MAGs can be explained by the reliance on short-read sequencing for genome-centric metagenomics, which often results in highly fragmented assemblies [16, 17], usually due to the reads not being able to span genomic repeats, resulting in contig breaks during assembly [18]. Binning of short contigs can also lead to incomplete genomes [19–21] or genome bins contaminated with contigs from other organisms [22–24]. Long-read sequencing (e.g., Oxford Nanopore or Pacific Biosciences) has been demonstrated by multiple studies to improve metagenome assembly contiguity, resulting in less fragmented genome bins, as well as enabling the recovery of complete, circular bacterial genomes from complex samples [25–33].

Here, we used deep Nanopore long-read metagenomic sequencing to recover closed and high-quality genome drafts of cable bacteria. Through metabolic annotation of the recovered MAGs and visualization of new cable bacteria species, we provide novel insights into the functional potential, morphology, and ecological niches of these enigmatic bacteria.

RESULTS AND DISCUSSION

Recovery and phylogenetic analysis of cable bacteria MAGs

One *Ca. Electronema aureum* GS enrichment (ENR) and two environmental (brackish-water sediment—BRK; marine sediment—MAR) samples, containing complex microbial communities with known cable bacteria populations [12, 34], were sequenced at varying depths (Table S1). In total, 162.4 Gbp and 148.3 Gbp of Nanopore and Illumina read data, respectively, were generated and, following the Minimum information about metagenome-assembled genome standards (see Methods), 103 high-quality (HQ) and 195 medium-quality (MQ) MAGs were recovered with automated binning. Overall, HQ and MQ MAGs accounted for ~40% of the cumulative relative abundance in long and short-read datasets (per sample), with the MAR sample featuring the lowest-yielding automated binning metrics, while the enrichment sample exhibited the most contiguous metagenome assembly statistics, yielding 66 closed bacterial and archaeal MAGs. After manual inspection and reassembly, a total of five cable bacteria MAGs were recovered (Table S2), spanning both candidate genera of cable bacteria (Fig. 1).

The ENR sample yielded a first-ever circular MAG (ENR-cMAG) of an enriched *Ca. Electronema aureum* GS strain, described in a previous study [12]. The circular genome featured <0.1% SNP rate, suggesting minimal presence of strain microdiversity [35]. Compared to the previously published short-read MAG of the same strain, the ENR-cMAG featured 100% average nucleotide identity (ANI, Figs. S1, S2) and identical 16S rRNA gene sequences (Fig. S3), even though the closed MAG for *Ca. Electronema aureum* GS was found to contain 292 more genes as well as 125 more insertion sequence elements than the short-read-based equivalent MAG (Table S3). Furthermore, the previously published and fragmented *Ca. Electronema aureum* MAG had four contigs (encoding 47 genes) that did not align with the closed ENR-cMAG, which indicates minor contamination in the short-read MAG.

The MAR sample was found to contain genomic sequences of multiple cable bacteria and three MAGs were recovered of varying contiguity and quality: a single-contig, HQ MAG (MAR-scMAG), a HQ MAG of 37 contigs (MAR-hqMAG) as well as a MQ MAG of nine contigs (MAR-mqMAG). All cable bacteria MAGs from the MAR sample featured above 90% genome completeness and higher than >0.5% SNP rate, indicating the presence of some strain heterogeneity. Using ANI of 95% for species-level demarcation [36], the MAR-mqMAG and MAR-hqMAG were found to be members of *Ca. Electrothrix aarhusensis* and *Ca. Electrothrix gigas* [34], respectively (Fig. S1). Species-level assignments for MAR-mqMAG as well as MAR-hqMAG were further confirmed with 16S rRNA phylogeny (Fig. S3) and the MAR-hqMAG was also observed

to supplement the *Ca. Electrothrix gigas* species pangenome with more genes (Table S4).

Moreover, the MAR-scMAG was found to be a member of *Ca. Electrothrix* (Fig. 1), although the MAG featured less than 90% ANI to all existing HQ and MQ cable bacteria MAGs (Fig. S1). The 16S rRNA tree further confirmed MAR-scMAG as being phylogenetically distinct from other *Ca. Electrothrix* 16S rRNA gene sequences. Thus, we suggest that MAR-scMAG belongs to a novel species in *Ca. Electrothrix*, which we have proposed to name *Ca. Electrothrix laxa* due to the large filament size of the species (see later).

The third sample sequenced in this study, BRK, yielded another circular cable bacterium MAG (BRK-cMAG) with <0.1% SNP rate (Table S2). Surprisingly, the MAG was phylogenetically classified as *Ca. Electronema* using the genome tree (Fig. 1) as well as 16S rRNA gene phylogeny (Fig. S3), even though the sample was collected from brackish-water sediment, where *Ca. Electrothrix* was expected to be present [37]. Using 75–77% ANI and 50% percentage of conserved proteins (POCP) for genus boundary [36, 38, 39], the BRK-cMAG belongs to the *Ca. Electronema* genus (Figs. S1, 2). While just inside the established ANI and POCP genus boundaries, the BRK-cMAG had less than 78% ANI to all cable bacteria MAGs (Fig. S1). Clustering all cable bacteria MAG genes at different identity cutoffs revealed that BRK-cMAG featured the highest fraction of unique genes, indicating that the BRK-cMAG exhibited the most novel gene content of the cable bacteria MAGs recovered in this study (Fig. S4). Based on its ability to survive in a saline environment, we suggest to name the BRK-cMAG as *Ca. Electronema halotolerans*. Furthermore, undefined *Ca. Electronema* species have been previously detected by 16S rRNA gene amplicon sequences in root and associated bulk samples of the seagrass species *Halophila ovalis* and *Zostera marina* [9]. Therefore, we compared these undefined *Ca. Electronema* ASVs with the 16S rRNA gene sequence of *Ca. Electronema halotolerans*. The phylogenetic analysis (Fig. S3) confirmed the association of such ASVs with the new *Ca. Electronema* species and further indicated its potential tolerance of higher salinity.

Adaptation to marine environment

Typically, halophilic microorganisms can use two different strategies to maintain osmotic balance inside their cells: the first, known as salt-in strategy, involves the intracellular accumulation of high concentrations of salts, in particular, potassium and chloride; the second, also known as the “compatible solute” strategy, involves the intracellular accumulation of organic compounds like polyols, betaines, or ectoines, which can be synthesized by the cells themselves or taken up from the environment [40]. Therefore, we screened the cable bacteria genomes for potential adaptations to high salinity habitats, including mechanisms to maintain a high osmotic pressure inside the cells, tolerance to heavy metal ions or accumulation of osmotic solutes [41, 42]. Ion pumps are usually involved in achieving ion gradients across the cell membrane, in particular to exclude Na^+ and accumulate K^+ and Cl^- [42]. One of the most common is the Na^+/H^+ antiporter *NhaA*, which is detected in the genetic potential of all *Ca. Electrothrix* genomes and in *Ca. Electronema halotolerans*, and has been previously suggested as a potential discriminant between marine and freshwater cable bacteria [12] (Figs. 2, 3, Dataset S1). The medium-quality MAGs KV [43] and SY1 [44], which are closely related to *Ca. Electronema halotolerans*, encoded also the *NhaA* gene (Fig. 2), confirming its importance for the survival of bacteria in saltwater environment [45].

We found three distinct clades of the gene encoding the *NhaA* Na^+/H^+ antiporter in *Ca. Electronema* and *Ca. Electrothrix* (Fig. S5a), which appear to have been present in the common ancestor to the two genera. *NhaA_1* persisted in all *Ca. Electrothrix*, *NhaA_2* in *Ca. Electrothrix aarhusensis* and *Ca. Electrothrix laxa*, and *NhaA_3* in *Ca. Electronema* (Fig. S5b). The high identity between the *NhaA_3*

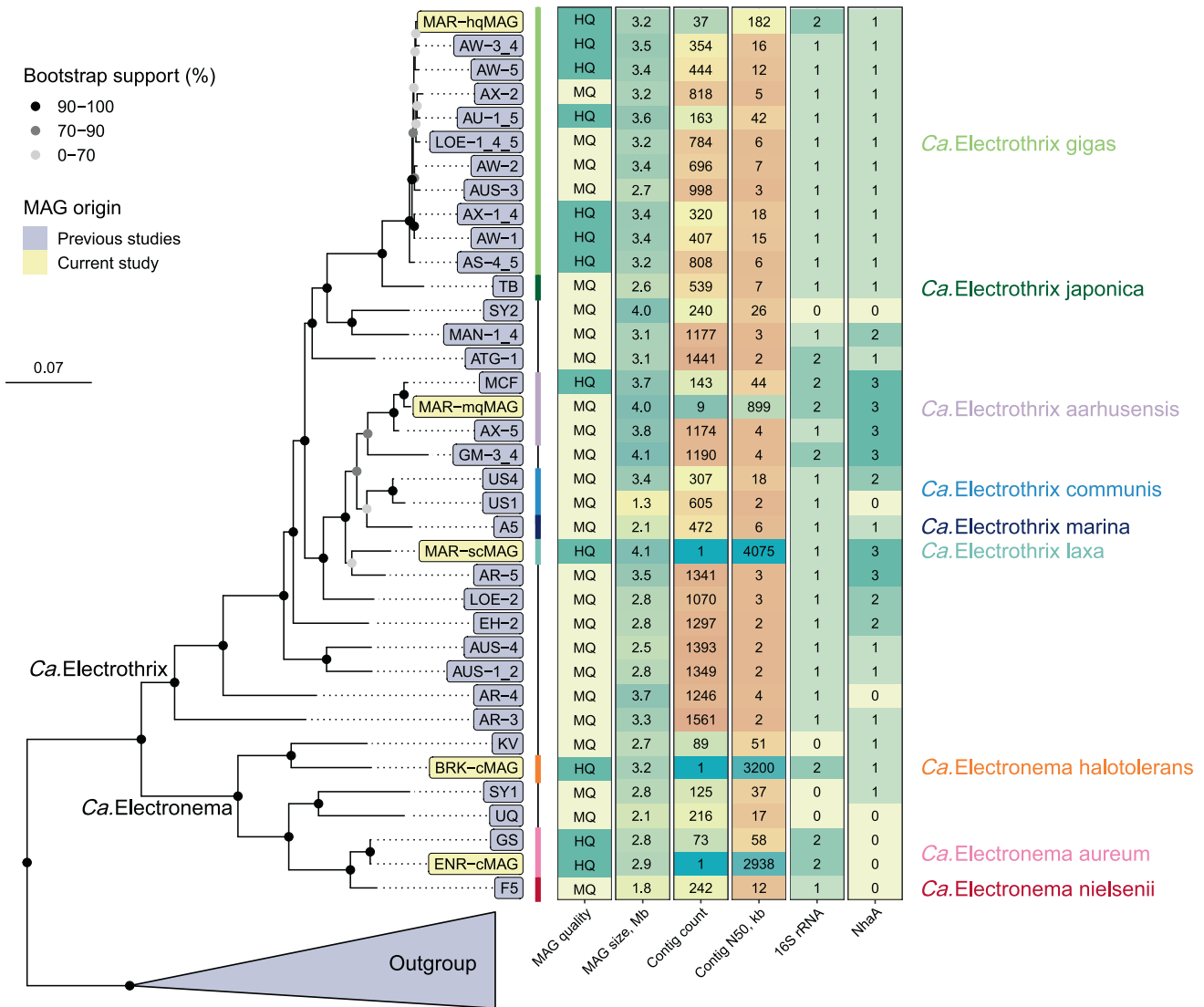


Fig. 1 Cable bacteria genome tree and selected MAG features. The maximum-likelihood tree was built from 120 universal bacterial marker genes, 100 bootstraps, and multiple outgroups were included (Table S5). Only HQ and MQ MAGs were used for building the tree. Nodes suspected of delineating relevant phylogenetic groups are highlighted. MAG quality—quality ranking according to MIMAG standards. MAG size, Mb—total MAG size in megabases. Contig count—number of contigs per MAG. Contig N50, kb—MAG N50 values in kilobases. 16S rRNA—count of 16S rRNA gene sequences detected in MAGs. *NhaA*—count of *NhaA* gene sequences predicted in MAGs. Additional cable bacteria MAG statistics are provided in Dataset S2.

sequence in *Ca. Electrothrix aarhusensis* and *Ca. Electrothrix laxa* suggests that there has been a horizontal gene transfer event between *Ca. Electrothrix laxa* and *Ca. Electrothrix aarhusensis*. Furthermore, *Ca. Electronema aureum* lost all copies of *NhaA*, which likely contribute to the restriction of its habitat to freshwater. *NhaA* has been shown to confer saltwater tolerance when overexpressed in a freshwater cyanobacterium in the past [46] and it appears that the presence of this gene is critical for the survival of *Ca. Electronema* in saltwater. It is unclear what the functional difference is between the three different types of *NhaA*, but it is worth noting that all saltwater *Ca. Electronema* genomes found so far have been found in brackish water (salinity <30), suggesting that *NhaA_3* may be adapted to lower Na^+ concentrations.

Moreover, all *Ca. Electrothrix laxa*, *Ca. Electrothrix gigas* and *Ca. Electronema halotolerans* MAGs also encoded homologs (56% amino acid identity) of the two subunits of the $\text{Na}^+/\text{Li}^+/\text{K}^+:\text{H}^+$ antiporter UmpAB, a transporter firstly characterized in *Halomonas zhaodongensis* [47] (Figs. 2, 3, Dataset S1). Homologs of a putative

$\text{Ca}^{2+}/\text{Na}^+:\text{H}^+$ antiporter (41% amino acid identity with a transporter first identified in *Alkalimonas amylolytica* [48]) and a putative phosphate: Na^+ symporter were present in *Ca. Electrothrix* and *Ca. Electronema halotolerans* MAGs, but absent in all the other *Ca. Electronema* genomes (Figs. 2, 3, Dataset S1).

The detection of several genes involved in ion translocation through the membrane may indicate the preference of the marine/brackish cable bacteria for a “salt-in” strategy [42]. This seems further supported by the fact that none of the cable bacteria MAGs encode the biosynthetic pathways for osmotic solutes and only one subunit of a putative glycine betaine/carnitine/choline transporter is detected in some of the MAGs, including the freshwater *Ca. Electronema* (Dataset S1).

Closed genomes clarify key features of the cable bacteria metabolism

Metabolic prediction included a total of ten HQ MAGs, while the remaining low and middle-quality genomes were used as

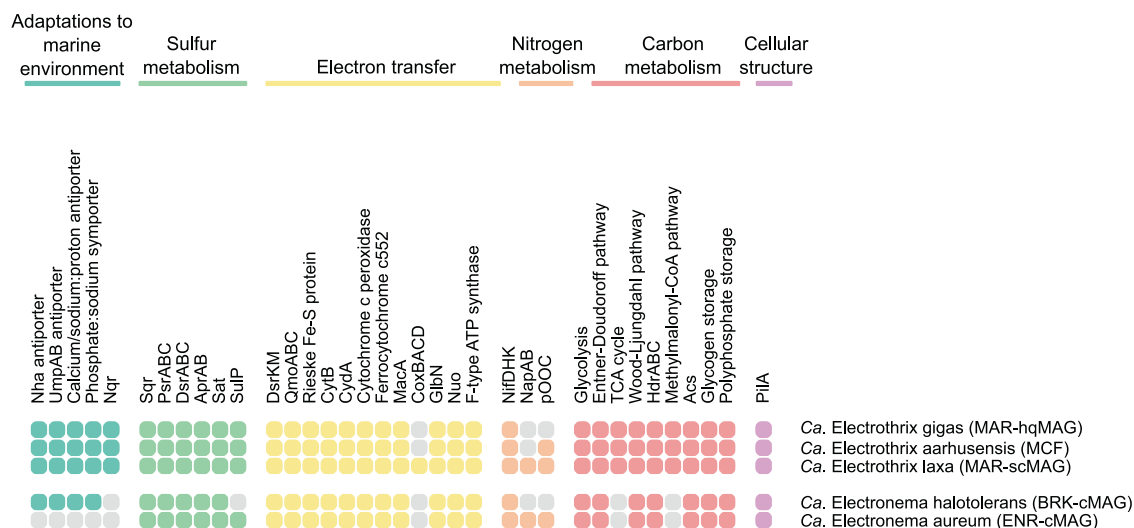


Fig. 2 General overview of the functional potential of HQ cable bacteria MAGs. Pathways are considered to be present if more than 80% of the genes are predicted. For the full list of gene names and associated KO numbers, see Dataset S1. The MAGs and genomes are ordered as in the genome tree in Fig. 1.

additional information to support the proposed metabolic model. All the MAGs retrieved in this study presented a metabolic potential consistent with the previously proposed model for *Ca. Electronema* and *Ca. Electrothrix* [12] (Figs. 2, 3, Dataset S1). Furthermore, the availability of closed genomes allowed us to shed light on metabolic features of cable bacteria that were previously uncertain. One of the most controversial features of the LDET is the lack of cable bacteria genomes of a terminal oxidase, indicating the reduction of terminal acceptors without energy conservation [12]. This is supported by our findings, with no known complete terminal cytochrome *bd-II* oxidase detected in the closed genomes and other MAGs. Instead, the MAGs encoded only homologs of the cytochrome *bd-II* oxidase subunit, *CydA*, as previously reported [12] (Figs. 2, 3, Dataset S1). By contrast, the genome of *Ca. Electrothrix laxa* encodes all four subunits of the membrane-bound cytochrome *c* oxidase, with ~60% amino acid sequence identity to homologs in other *Desulfobulbaceae* species and most likely resulting from horizontal gene transfer, as previously observed for its close relative *Ca. Electrothrix communis* [12] (Figs. 2, 3, Dataset S1). We can therefore hypothesize that these species may be able to couple oxygen reduction to proton translocation with energy conservation.

All cable bacteria show the metabolic potential for nitrogen fixation (Figs. 2, 3, Dataset S1) via the molybdenum-dependent nitrogenase *NifDHK*. The ammonium, which can also be taken up by the *AmtB* ammonium transporter, can then be incorporated into amino acids via glutamine/glutamate synthesis (Dataset S1). Experimental evidence has shown that cable bacteria can couple sulfide oxidation with nitrate or nitrite reduction [13, 49]. Only two species (*Ca. Electronema aureum* and *Ca. Electrothrix laxa*) have the potential for nitrate reduction via the periplasmic nitrate reductase *NapAB*, and nitrate reduction by the *NapAB* system has been recently confirmed by transcriptomics in *Ca. Electronema aureum* [49]. The *nap* operon is lacking the gene encoding for the subunit *NapB* in *Ca. Electrothrix laxa*, but *NapB* has been shown to be non-essential in the nitrate reduction of *Shewanella oneidensis* [50]. No *Nir-* or *Nrf-*type nitrite reductase was detected in *Ca. Electronema aureum* and *Ca. Electrothrix laxa* (Figs. 2, 3, Dataset S1), but a periplasmic multiheme cytochrome adjacent to the *nap* operon, which is also predicted in *Ca. Electrothrix aarhusensis*, has been suggested as the enzyme responsible for nitrite reduction to ammonium [12, 49], suggesting a linked function and preference of cable bacteria for dissimilatory nitrate reduction to ammonium.

All the MAGs encoded nearly complete central carbon processing pathways through glycolysis, Entner–Doudoroff pathway, pentose phosphate pathway, and TCA cycle (Figs. 2, 3, Dataset S1). As previously reported for *Ca. Electronema aureum* and *Ca. Electrothrix aarhusensis*, and now confirmed by the availability of closed genomes, no genes encoding for the enzyme *enolase* were detected in the MAGs, indicating a potential alternative and unknown pathway to complete glycolysis [12]. A key difference between the two cable bacteria genera appears to be in their carbon substrate utilization potential. While *Ca. Electrothrix* MAGs encoded the potential for full TCA cycle, the two closed genomes for *Ca. Electronema* MAGs lacked several genes, including succinyl-CoA-synthase 2-oxoglutarate ferredoxin oxidoreductase (Figs. 2, 3, Dataset S1). Similarly, all *Ca. Electrothrix* MAGs have the potential to assimilate propionate via the methylmalonyl-CoA pathway to succinyl-CoA, absent in *Ca. Electronema* (Figs. 2, 3, Dataset S1).

Visualization of the new species

A set of new FISH probes (Table S6) was designed to distinguish the species in situ (Fig. 4). The diameter of filaments ranged from 1 to 6 μm for *Ca. Electrothrix laxa* and 1–2 μm for *Ca. Electronema halotolerans*. Raman microspectroscopy has recently proved to be an excellent technique in helping to clarify the conductive structure of cable bacteria [14]. Therefore, we applied it in combination with the newly designed FISH probes to confirm the chemical composition of the cable's cell envelope in the new species. Raman spectra of FISH-identified *Ca. Electrothrix laxa* showed peaks typical of biological components, such as phenylalanine (1005 cm^{-1}), CH_2 bond (1462 cm^{-1}) and the amide I peak of protein (1672 cm^{-1}), as well as peaks linked to cytochromes (750 cm^{-1} , 1129 cm^{-1} , 1314 cm^{-1} , and 1586 cm^{-1}). Additionally, two large peaks were present at 371 and 492 cm^{-1} (Fig. S6), which may be linked to the presence of a S-ligated metal group in the conductive fiber [14].

CONCLUSION

With this study, we provide the first-ever closed genomes of cable bacteria. The closed genome for *Ca. Electronema aureum* enabled the identification of novel genes and genomic repeats, which were absent from the previously published fragmented MAGs, as well as confirmed metabolic models proposed for cable bacteria. Furthermore, we also provide the first-ever single-contig HQ MAG

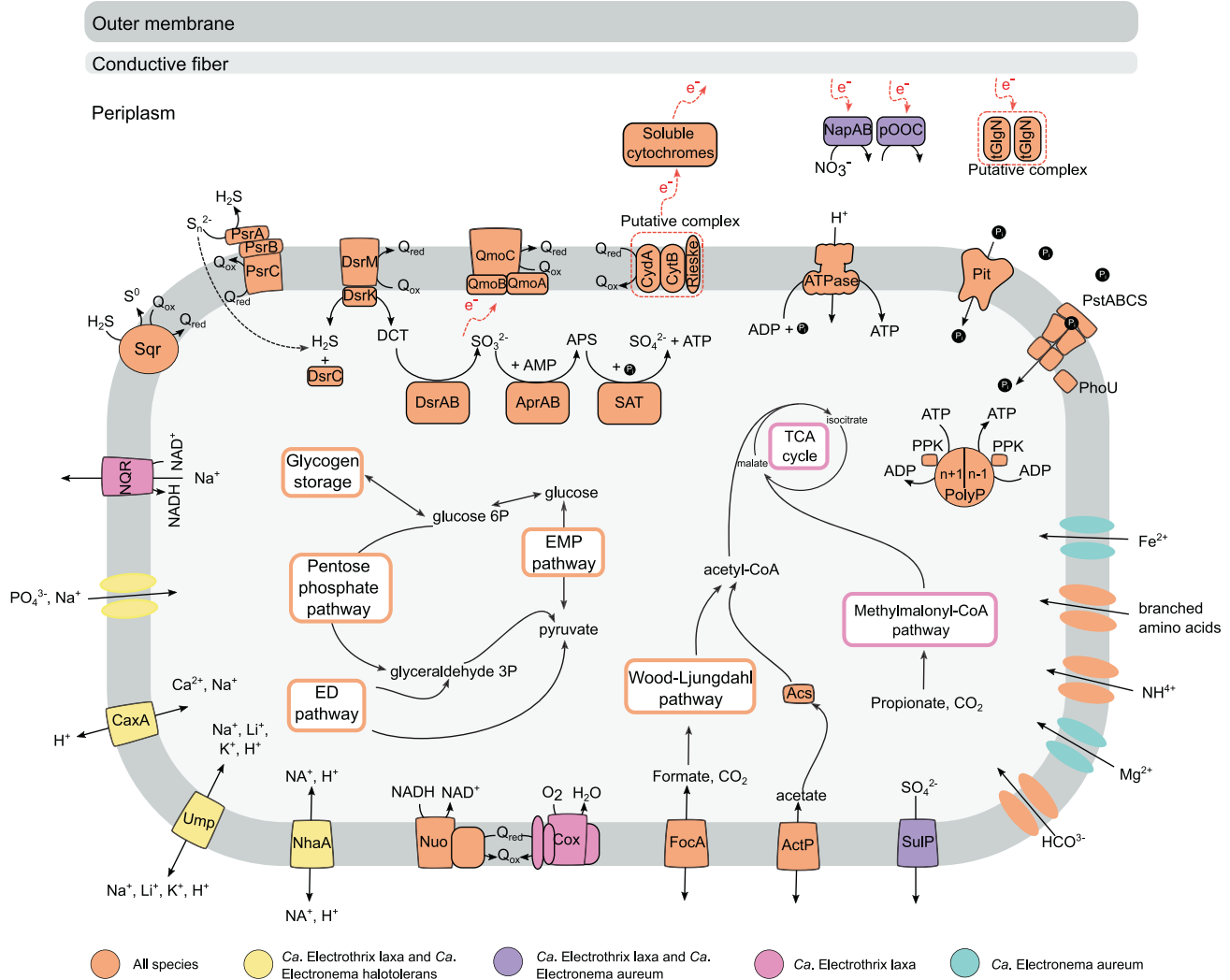


Fig. 3 Metabolic model for *Ca. Electrothrix laxa* and *Ca. Electronema halotolerans*. Gray and red dotted lines indicate putative reactions and complexes. Abbreviations: EMP pathway, Embden–Meyerhof–Parnas pathway (glycolysis); TCA cycle, tricarboxylic acid cycle; EMP pathway, Entner–Doudoroff pathway; Acs, acetyl-CoA synthetase; Dsr, dissimilatory bisulfite reductase; DCT, DsrC–trisulfide; Apr, Adenosine phosphosulfate reductase; Sat, ATP sulfurylase; Psr, polysulfide reductase; SQR, sulfide:quinone oxidoreductase; Qmo, quinone-interacting membrane-bound oxidoreductase complex; Q(ox/red), quinone (oxidized or reduced); CydA, membrane-bound cytochrome bd quinol oxidase–subunit A; CytB, cytochrome bc complex–subunit B; Rieske, Rieske Fe-S domain protein; Nap, periplasmic nitrate reductase; pOOC, periplasmic multiheme cytochrome (nitrite reductase); tGlnN, truncated hemoglobin; PolyP, polyphosphate; PPK, polyphosphate kinase; Pit, inorganic phosphate transporter family; PstABCS, inorganic phosphate ABC transporter, PhoU, phosphate transport system accessory protein; SulP, sulfate permease; ActP, acetate transporter; FocA, putative formate transporter; Cox, cytochrome c oxidase; Nuo, NADH dehydrogenase; NhaA, Sodium:proton antiporter; Ump, putative sodium, lithium, potassium: proton antiporter; Cax, putative calcium, sodium:proton antiporter; NQR, sodium-translocating NADH:ubiquinone oxidoreductase. Further details are provided in the main text and supplementary notes.

for *Ca. Electrothrix* which we propose to name as novel species of *Ca. Electrothrix laxa*, which we suspect is capable of functions atypical for cable bacteria due to horizontal gene transfer. Lastly but most importantly, we provide a closed genome of a novel species we propose to name *Ca. Electronema halotolerans*, which is the first-ever genome draft for a *Ca. Electronema* that was observed to be present and feature genes for survival in a saltwater environment. The existence of a saline-tolerant *Ca. Electronema* species challenges the current freshwater-saltwater-centric taxonomic framework for cable bacteria, necessitating a reassessment for the categorization of cable bacteria candidate genera.

Taxonomic proposal

Description of “*Candidatus Electrothrix laxa*” sp. nov.: “*Candidatus Electrothrix laxa*”, (la’xa, L. fem. adj. *laxa*, large). This taxon is

represented by the MAG MAR-scMAG. The complete protologue can be found in Table S8.

Description of “*Candidatus Electronema halotolerans*” sp. nov.: “*Candidatus Electronema halotolerans*”, (ha.lo.to’le.rans, from Gr. n. *hals*, halos salt; L. part. adj. *tolerans* tolerating; N.L. part. adj. *halotolerans*). This taxon is represented by the MAG BRK-cMAG. The complete protologue can be found in Table S9.

METHODS

Sample collection and manual processing

A marine sediment sample was collected from Hou, Denmark (55°54’36.8’’ N 10°14’46.8’’E) and the brackish-water sediment (typical salinity 18–23) sample was collected from Løgten, Denmark (56°17’17.8’’ N 10°22’54.9’’E). *Ca. Electronema aureum* GS enrichment sample and manual processing of all samples were performed as detailed in a previous study [15].

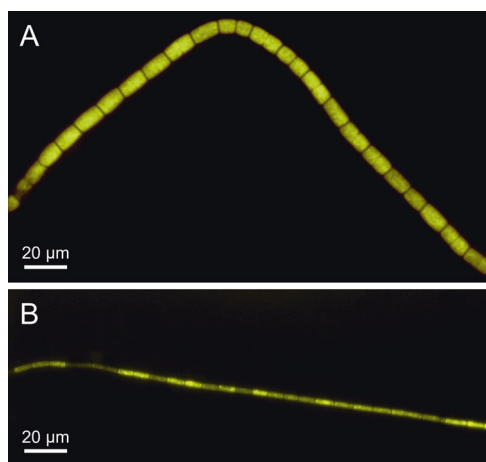


Fig. 4 Composite FISH micrographs of the two novel cable bacteria species. **A** *Ca. Electrothrix laxa* appears in yellow, by the overlap of the species-specific probe Ex-la-189 (FAM-labeled, green) and the probe for the *Desulfobulbaceae* family DSB706 (Cy3-labeled, red); **B** *Ca. Electronema halotolerans* appears in yellow, by the overlap of the species-specific probe EN-ha-80 (FAM-label, green) and the probe for the *Desulfobulbaceae* family DSB706 (Cy3, red).

DNA extraction and QC

DNA from environmental samples was extracted using the DNeasy PowerSoil Pro Kit (QIAGEN, Germany) according to the manufacturer's protocol. The concentration of the extracts was measured using the Qubit dsDNA HS kit (Thermo Fisher Scientific, USA, #Q33231) with a Qubit 3.0 fluorometer (Thermo Fisher Scientific, USA), while DNA purity was assessed with a NanoDrop One Spectrophotometer (Thermo Fisher, USA). Sample DNA fragment sizes were inspected using an Agilent 2200 TapeStation system with Genomic DNA ScreenTapes (Agilent Technologies, USA, #5067-5365). DNA samples were also size-selected (with the exception of BRK sample due to limited DNA amounts) using the Circulomics SRE XS kit (Circulomics, USA), following the manufacturer's instructions to deplete DNA fragments below 10 kb.

Nanopore DNA sequencing

DNA libraries for Nanopore sequencing were prepared using the SQK-LSK110 Ligation Sequencing kit (Oxford Nanopore Technologies, UK) according to the manufacturer's instructions. The libraries were loaded into Nanopore R.9.4.1 chemistry flow cells and sequenced on the MinION Mk1B sequencer (Oxford Nanopore Technologies, UK) using MinKNOW v21.06.13 (<https://community.nanoporetech.com/downloads>) software. Raw Nanopore data was basecalled using the "dna_r9.4.1_450bps_sup.cfg" basecalling model with Guppy (v. 5.0.7, <https://community.nanoporetech.com/downloads>).

Illumina DNA sequencing

Short-read sequencing libraries were prepared using the Illumina DNA Prep kit in combination with IDT® UD Indexes Set A (Illumina, USA, #20018705, #20027213). Between 66.3 ng and 200 ng per sample was used as starting material. Five cycles of PCR were applied in the amplification of the tagmented DNA. All other steps were performed according to the manufacturer's protocol. Final libraries were quantified and insert size evaluated using the Qubit dsDNA HS assay (Thermo Fisher Scientific, USA, #Q33231) and a D1000 screentape (Agilent Technologies, USA, #5067-5582, #5067-5602). Samples were multiplexed using 100 ng of DNA and the final pooled library was evaluated in the same way as for the individual libraries.

The pooled library was sequenced on an S1 flow cell for the NovaSeq 6000 platform (Illumina, USA). 2.0 nM was used as input for the V1.5 300 cycle kit (Illumina, USA, #20012863).

Sequencing read processing

Illumina reads were trimmed, quality-filtered and de-duplicated using fastp (v. 0.23.2 [51]) with "-l 100 -dup_calc_accuracy 6" settings. For Nanopore reads, adapter sequences were trimmed off and chimeric reads were split using Porechop (v. 0.2.3 [52]). Reads with a lower Phred score

than 7 or shorter read length than 200 bp were discarded using NanoFilter (v. 2.6.0 [53]). Read length and quality statistics were acquired with NanoPlot (v. 1.24.0 [53]).

Metagenome assembly, polishing, and binning

Trimmed and filtered nanopore reads were assembled into metagenomes via the Flye (v. 2.9-b1768 [54]) assembler using the "--meta", "--nano-hq" and "min_read_cov_cutoff=8" settings. The assembly was then polished first with Nanopore reads using 3 rounds of Racon (v. 1.3.3 [55]) and then two rounds of Medaka (v. 1.4.4, <https://github.com/nanoporetech/medaka>) polishing with "r941_min_sup_g507" model. After polishing the assemblies with Nanopore reads, the metagenome was then polished using Illumina short reads via 1 round of Racon to correct the contigs for frameshift errors. Dependencies for metagenome polishing, read mappings include Minimap2 (v. 2.24 [56]) and SAMtools (v. 1.14 [57]). Contigs with lower length than 1 kb were removed via Bioawk v. 1.0 (<https://github.com/lh3/bioawk>).

The polished contigs were then binned using MetaBAT2 (v. 2.12.1 [58]), with "-s 500000" setting, Vamb (v. 3.0.2 [59]) with "-o C --minfasta 500000" setting, MaxBin2 (v. 2.2.7 [60]) and metaBinner (v. 1.4.2 [61]). Contig coverage profiles from Nanopore and Illumina read data were used as input for binning. The bins were then refined using DAS Tool (v. 1.1.2 [62]) with "--search_engine diamond" setting. For the enrichment sample, all circular contigs longer than 500 kb were selected and used as bins. CoverM (v. 0.6.1, <https://github.com/wood/CoverM>) was used to acquire bin coverage ("m mean" setting) and relative abundance ("m relative_abundance" setting) values.

Bin processing and reassembly

Genome bin completeness and contamination values were acquired with CheckM (v. 1.1.2 [23]) using the lineage-specific workflow and the Deltaproteobacteria marker lineage (UID3217) was used for all cable bacteria MAGs. Bin rRNA genes were predicted with Barrnap (v. 0.9, <https://github.com/tseemann/barrnap>) and tRNA gene sequences were predicted using tRNAscan-SE (v. 2.0.5 [63]). Bin quality classification was performed in accordance with the Genomic Standards Consortium guidelines, where a high-quality genome bin featured greater genome completeness than 90%, lower genome contamination than 5%, a minimum of 18 different tRNA genes, and the 16S, 5S, 23S rRNA genes occurring at least once in the bin [64]. Furthermore, bins with completeness greater than 50% and contamination lower than 10% were assigned as medium quality, whereas the remaining bins with less than 10% contamination were classified as low quality. Bins with contamination estimates greater than 10% were considered contaminated. Bins were de-replicated between the different samples using dRep (v. 2.6.2, <https://github.com/MrOlm/drep>) with "--comp 50 -con 10 -sa 0.95" settings. An overview of the de-replicated medium and high-quality MAGs is provided in Dataset S3.

Bin taxonomy was assigned using GTDB-Tk (v. 2.0.0 [65], R207 database) with the "--full_tree" setting enabled. Contig taxonomic classification was carried out using mmseqs2 (v. 13.45111 [66]) with NCBI nr protein database [67] (download date: 2021-06-26). 16S rRNA genes were classified using usearch (v. 11.0.667 [68]) "usearch_global" command against the SILVA v138 nr99 database [69].

To acquire the circular cable bacteria genome from BRK sample, reads were extracted for contigs present in bins classified as cable bacteria by GTDB-Tk using SAMtools bam2fq command on the read mapping files and the extracted Nanopore reads were assembled with Flye. For the MAR sample, two reassemblies were performed. For the first one, which resulted in the linear cable bacteria genome, reads were extracted for contigs above 5 kb length, classified as *Desulfobulbaceae* by mmseqs2 or containing *Desulfobulbaceae* 16S rRNA genes or from MAGs classified as *Desulfobulbaceae* by GTDB-tk. For the second reassembly, which enabled recovery of two fragmented cable bacteria MAGs from the MAR sample, a subset of reads from the first reassembly was made by extracting reads from contigs with longer length than 10 kb and higher coverage than 100, and after reassembly with Flye (using the "min_read_cov_cutoff=80" setting), the bins were manually refined by cross-referencing contigs with the Flye assembly graph. The acquired cable bacteria MAGs and sequenced reads are publicly available via accession numbers provided in Table S7.

Comparative genomics and MAG annotation

For constructing cable bacteria genome trees, a multiple sequence alignment for 120 bacterial marker genes was generated using GTDB-Tk and provided to IQ-TREE (v. 2.0.6 [70]) to build a tree using the "WAG + G"

model and 100x bootstrapping. Multiple outgroups were included in the tree, summarized in Table S5. ANI and alignment coverage values between different cable bacteria MAGs and outgroups were acquired via pyANI (v. 0.2.11 [71]) with “-m ANIb” setting enabled. AAI and POCP measurements between the genomes were acquired with CompareM (v. 0.1.2, <https://github.com/dparks1134/CompareM>) using the “--evaluate 0.00001 --per_identity 40 --per_aln_len 50 --blastp” options. Quast (v. 4.6.3 [72]) was used to compare MAGs of closely related strains and ISEScan (v. 1.7.2.3 [73]) was applied to identify insertion sequence elements.

MAGs were automatically annotated using Prokka (v. 1.14.0 [74]) with “--metagenome” setting enabled. Protein sequences annotated as *NhaA* were extracted, aligned using MUSCLE (v. 5.0.1428 [75]) with 1000 iterations, followed by protein tree generation using IQ-TREE with “LG + G4” model and 1000 bootstrapping instances. MAG gene clustering and pangenome creation were performed by using Prokka output files as input for Roary (v. 3.13.0 [76]) with “-cd 85” setting and varying identity thresholds, which are described in the text. Additionally, HQ-MAGs were selected based on quality standards proposed by Bowers et al. [64], (completeness and contamination of >90% and <5%, respectively) and uploaded to the “MicroScope Microbial Genome Annotation & Analysis Platform” (MAGE) [77] for manual inspection and metabolic pathways were predicted by KEGG [78] pathway profiling of MAGE annotations.

16S rRNA gene-based phylogenetic analysis, FISH probe design and evaluation, and Raman microspectroscopy

Phylogenetic analysis of 16S rRNA gene sequences and design of FISH probes were performed using the ARB software v.6.0.6 [79]. A phylogenetic tree was calculated based on a comparative analysis of aligned 16S rRNA gene sequences, retrieved from SILVA v138 nr99 database [69] using the maximum likelihood method (PhyML) and a 1000—replicates bootstrap analysis. Coverage and specificity were evaluated and validated in silico with the MathFISH software for hybridization efficiencies of target and potentially weak non-target matches [80]. When needed, unlabeled competitors and helper probes were designed. All probes were purchased from Biomers (Ulm, Germany), labeled with 6-carboxyfluorescein or indocarbocyanine (Cy3) fluorochromes.

FISH was performed as described by Daims et al. [81]. Optimal formamide concentration for each novel FISH probe was visually determined by microscope observation after performing hybridization at different formamide concentrations in the range 15–70% (with 5% increments). Optimal hybridization conditions are described in Table S6. EUBmix [82, 83] was used to target all bacteria and NON-EUB [84] was used as a negative control for sequence-independent probe binding. Microscopic analysis was performed with Axioskop epifluorescence microscope (Carl Zeiss, Germany) equipped with a LEICA DFC7000 T CCD camera or a white light laser confocal microscope (Leica TCS SP8 X). Raman microspectroscopy was applied in combination with FISH as previously described [85] to identify cellular components and clarify the structure of the conductive fiber.

DATA AVAILABILITY

Sequencing read data as well as the de-replicated high and medium-quality MAGs are available at ENA with bio project ID: PRJEB52550. Code and datasets used to generate the figures as well additional resources are available at <https://github.com/Serka-M/Cable-Bacteria-MAGs>. GTDB-tk database used in the project can be accessed at <https://data.ace.uq.edu.au/public/gtdb/data/releases/release207/>. NCBI nr protein database can be downloaded from <https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/>. The SILVA database can be accessed at <https://www.arb-silva.de/download/arb-files>.

REFERENCES

- Pfeffer C, Larsen S, Song J, Dong M, Besenbacher F, Meyer RL, et al. Filamentous bacteria transport electrons over centimetre distances. *Nature*. 2012;491: 218–21.
- Lovley DR, Holmes DE. Electromicrobiology: the ecophysiology of phylogenetically diverse electroactive microorganisms. *Nat Rev Microbiol*. 2022;20:5–19.
- Bjerg JT, Boschker HTS, Larsen S, Berry D, Schmid M, Millo D, et al. Long-distance electron transport in individual, living cable bacteria. *Proc Natl Acad Sci USA*. 2018;115:5786–91.
- Trojan D, Schreiber L, Bjerg JT, Bøggild A, Yang T, Kjeldsen KU, et al. A taxonomic framework for cable bacteria and proposal of the candidate genera *Electrothrix* and *Electronema*. *Syst Appl Microbiol*. 2016;39:297–306.

- Malkin SY, Rao AM, Seitaj D, Vasquez-Cardenas D, Zetsche EM, Hidalgo-Martinez S, et al. Natural occurrence of microbial sulphur oxidation by long-range electron transport in the seafloor. *ISME J*. 2014;8:1843–54.
- Risgaard-Petersen N, Kristiansen M, Frederiksen RB, Dittmer AL, Bjerg JT, Trojan D, et al. Cable bacteria in freshwater sediments. *Appl Environ Microbiol*. 2015;81:6003–11.
- Burdorf LDW, Tramper A, Seitaj D, Meire L, Hidalgo-Martinez S, Zetsche EM, et al. Long-distance electron transport occurs globally in marine sediments. *Bio-geosciences*. 2017;14:683–701.
- Scholz VV, Müller H, Koren K, Nielsen LP, Meckenstock RU. The rhizosphere of aquatic plants is a habitat for cable bacteria. *FEMS Microbiol Ecol*. 2019;95:fiz062.
- Scholz VV, Martin BC, Meyer R, Schramm A, Fraser MW, Nielsen LP, et al. Cable bacteria at oxygen-releasing roots of aquatic plants: a widespread and diverse plant–microbe association. *N Phytol*. 2021;232:2138–51.
- Paver SF, Muratore D, Newton RJ, Coleman ML. Reevaluating the salty divide: phylogenetic specificity of transitions between marine and freshwater systems. *mSystems*. 2018;3:e00232–18.
- Cabello-Yeves PJ, Rodriguez-Valera F. Marine-freshwater prokaryotic transitions require extensive changes in the predicted proteome. *Microbiome*. 2019;7:117.
- Kjeldsen KU, Schreiber L, Thorup CA, Boesen T, Bjerg JT, Yang T, et al. On the evolution and physiology of cable bacteria. *Proc Natl Acad Sci USA*. 2019;116:19116–25.
- Müller H, Marozava S, Probst AJ, Meckenstock RU. Groundwater cable bacteria conserve energy by sulfur disproportionation. *ISME J*. 2020;14:623–34.
- Boschker HTS, Cook PLM, Polerecky L, Eachambadi RT, Lozano H, Hidalgo-Martinez S, et al. Efficient long-range conduction in cable bacteria through nickel protein wires. *Nat Commun*. 2021;12:3996.
- Thorup C, Petro C, Bøggild A, Ebsen TS, Brokjær S, Nielsen LP, et al. How to grow your cable bacteria: establishment of a stable single-strain culture in sediment and proposal of *Candidatus Electronema aureum* GS. *Syst Appl Microbiol*. 2021;44:126236.
- Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol*. 2017;13:e1005595.
- Stewart RD, Auffret MD, Warr A, Walker AW, Roehe R, Watson M. Compendium of 4941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat Biotechnol*. 2019;37:953–61.
- Koren S, Phillippy AM. One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Curr Opin Microbiol*. 2015;23:110–20.
- Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol*. 2017;2:1533–42.
- Imelfort M, Parks D, Woodcroft BJ, Dennis P, Hugenholtz P, Tyson GW. GroopM: an automated tool for the recovery of population genomes from related metagenomes. *PeerJ*. 2014;2:e603.
- Vasudevan K, Devanga Ragupathi NK, Jacob JJ, Veerarahavan B. Highly accurate-single chromosomal complete genomes using IonTorrent and MinION sequencing of clinical pathogens. *Genomics*. 2020;112:545–51.
- Tennessen K, Andersen E, Clingenpeel S, Rinke C, Lundberg DS, Han J, et al. ProDeGe: a computational protocol for fully automated decontamination of genomes. *ISME J*. 2016;10:269–72.
- 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.
- Nayfach S, Shi ZJ, Seshadri R, Pollard KS, Kyrpides NC. New insights from uncultivated genomes of the global human gut microbiome. *Nature*. 2019;568:505–10.
- Moss EL, Maghini DG, Bhatt AS. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. *Nat Biotechnol*. 2020;38:1–7.
- Bickhart DM, Kolmogorov M, Tseng E, Portik DM, Korobeynikov A, Tolstoganov I, et al. Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities. *Nat Biotechnol*. 2022;40:711–9.
- Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sørensen EA, Wollenberg RD, et al. Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat Methods*. 2022;19:823–6.
- Overholt WA, Hölzer M, Geesink P, Diezel C, Marz M, Küsel K. Inclusion of Oxford Nanopore long reads improves all microbial and viral metagenome-assembled genomes from a complex aquifer system. *Environ Microbiol*. 2020;22:4000–13.
- De Maio N, Shaw LP, Hubbard A, George S, Sanderson ND, Swann J, et al. Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. *Microb Genom*. 2019;5:e000294.
- Hagshenas E, Asghari H, Stoye J, Chauve C, Hach F. HASLR: fast hybrid assembly of long reads. *iScience*. 2020;23:101389.

31. Singleton CM, Petriglieri F, Kristensen JM, Kirkegaard RH, Michaelsen TY, Andersen MH, et al. Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing. *Nat Commun.* 2021;12:2009.
32. Cuscó A, Pérez D, Viñes J, Fàbregas N, Francino O. Long-read metagenomics retrieves complete single-contig bacterial genomes from canine feces. *BMC Genomics.* 2021;22:330.
33. Liu L, Wang Y, Che Y, Chen Y, Xia Y, Luo R, et al. High-quality bacterial genomes of a partial-nitritation/anammox system by an iterative hybrid assembly method. *Microbiome.* 2020;8:155.
34. Geelhoed JS, Thorup CA, Bjerg JJ, Schreiber L, Ochman H, Nielsen LP, et al. Pangenome analysis of cable bacteria reveals the likely genetic basis of large sized bacteria in *Candidatus Electrothrix gigas* sp. nov [manuscript in preparation]. University of Antwerp and Aarhus University; 2022 [cited 2023 Jan 23].
35. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* 2019;176:649–662.e20.
36. Parks DH, Chuvpochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol.* 2020;38:1079–86.
37. Dam AS, Marshall IPG, Risgaard-Petersen N, Burdorf LDW, Marzocchi U. Effect of salinity on cable bacteria species composition and diversity. *Environ Microbiol.* 2021;23:2605–16.
38. Qin QL, Xie BB, Zhang XY, Chen XL, Zhou BC, Zhou J, et al. A proposed genus boundary for the prokaryotes based on genomic insights. *J Bacteriol.* 2014;196:2210–5.
39. Barco RA, Garrity GM, Scott JJ, Amend JP, Neelson KH, Emerson D. A genus definition for bacteria and archaea based on a standard genome relatedness index. *mBio* 2020;11:e02475–19.
40. Oren A. Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Syst.* 2008;4:2.
41. Stal LJ, Cretoiu MS (eds). *The marine microbiome*. 1st ed. Springer Cham, Yersseke, 2016.
42. Gunde-Cimerman N, Plemenitaš A, Oren A. Strategies of adaptation of microorganisms of the three domains of life to high salt concentrations. *FEMS Microbiol Rev.* 2018;42:353–75.
43. Sannikov A. Marine sediments: interaction partners of cable bacteria revealed through metagenomics [master's thesis on the Internet]. University of Padua; 2021 [cited 2023 Jan 23]. Available from: <http://hdl.handle.net/20.500.12608/34057>.
44. Fang Y, Liu J, Yang J, Wu G, Hua Z, Dong H, et al. Compositional and metabolic responses of autotrophic microbial community to salinity in lacustrine environments. *mSystems.* 2022;7:e0033522.
45. Herz K, Vimont S, Padan E, Berche P. Roles of NhaA, NhaB, and NhaD Na⁺/H⁺ antiporters in survival of *Vibrio cholerae* in a saline environment. *J Bacteriol.* 2003;185:1236–44.
46. Waditee R, Hibino T, Nakamura T, Incharoensakdi A, Takabe T. Overexpression of a Na⁺/H⁺ antiporter confers salt tolerance on a freshwater cyanobacterium, making it capable of growth in sea water. *Proc Natl Acad Sci USA.* 2002;99:4109–14.
47. Meng L, Meng F, Zhang R, Zhang Z, Dong P, Sun K, et al. Characterization of a novel two-component Na⁺(Li⁺, K⁺)/H⁺ antiporter from *Halomonas zhao-dongensis*. *Sci Rep.* 2017;7:4221.
48. Wei Y, Liu J, Ma Y, Krulwich TA. Three putative cation/proton antiporters from the soda lake alkaliphile *Alkalimonas amylolytica* N10 complement an alkali-sensitive *Escherichia coli* mutant. *Microbiology.* 2007;153:2168–79.
49. Marzocchi U, Thorup C, Dam AS, Schramm A, Risgaard-Petersen N. Dissimilatory nitrate reduction by a freshwater cable bacterium. *ISME J.* 2022;16:50–7.
50. Gao H, Yang ZK, Barua S, Reed SB, Romine MF, Neelson KH, et al. Reduction of nitrate in *Shewanella oneidensis* depends on atypical NAP and NRF systems with NapB as a preferred electron transport protein from CymA to NapA. *ISME J.* 2009;3:966–76.
51. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;34:i884–90.
52. Wick RR, Judd LM, Gorrie CL, Holt KEY. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. *Micro Genom.* 2017;3:e000132.
53. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics.* 2018;34:2666–9.
54. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods.* 2020;17:1103–10.
55. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 2017;27:737–46.
56. Li H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics.* 2016;32:2103–10.
57. 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.
58. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ.* 2019;7:e7359.
59. Nissen JN, Johansen J, Allesøe RL, Sønderby CK, Armenteros JJA, Grønbech CH, et al. Improved metagenome binning and assembly using deep variational autoencoders. *Nat Biotechnol.* 2021;39:555–60.
60. Wu YW, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics.* 2016;32:605–7.
61. Wang Z, Huang P, You R, Sun F, Zhu S. MetaBinner: a high-performance and stand-alone ensemble binning method to recover individual genomes from complex microbial communities. *Genome Biol.* 2023;24:1.
62. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol.* 2018;3:836–43.
63. Chan PP, Lowe TM. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol Biol.* 2019;1962:1–14.
64. Bowers RM, Kyrpidides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol.* 2017;35:725–31.
65. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics.* 2020;36:1925–7.
66. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol.* 2017;35:1026–8.
67. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 2007;35:D61–5.
68. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics.* 2010;26:2460–1.
69. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yara P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 2013;41:D590–596.
70. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32:268–74.
71. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods.* 2015;8:12–24.
72. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUASt: quality assessment tool for genome assemblies. *Bioinformatics.* 2013;29:1072–5.
73. Xie Z, Tang H. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics.* 2017;33:3340–7.
74. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014;30:2068–9.
75. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* 2004;5:113.
76. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics.* 2015;31:3691–3.
77. Vallenet D, Calteau A, Dubois M, Amours P, Bazin A, Beuvin M, et al. MicroScope: an integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Res.* 2020;48:D579–89.
78. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28:27–30.
79. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadukumar null, et al. ARB: a software environment for sequence data. *Nucleic Acids Res.* 2004;32:1363–71.
80. Yilmaz LS, Parnerkar S, Noguera DR. mathFISH, a web tool that uses thermodynamics-based mathematical models for in silico evaluation of oligonucleotide probes for fluorescence in situ hybridization. *Appl Environ Microbiol.* 2011;77:1118–22.
81. Daims H, Stoecker K, Wagner M. Fluorescence in situ hybridization for the detection of prokaryotes. In: *Molecular Microbial Ecology*. Taylor & Francis; 2005. p. 213–39.
82. Amann RL, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol.* 1990;56:1919–25.
83. Daims H, Brühl A, Amann R, Schleifer KH, Wagner M. The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development

and evaluation of a more comprehensive probe set. *Syst Appl Microbiol.* 1999;22:434–44.

84. Wallner G, Amann R, Beisker W. Optimizing fluorescent in situ hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. *Cytometry.* 1993;14:136–43.
85. Fernando EY, McLroy SJ, Nierychlo M, Herbst FA, Petriglieri F, Schmid MC, et al. Resolving the individual contribution of key microbial populations to enhanced biological phosphorus removal with Raman-FISH. *ISME J.* 2019;13:1933–46.

ACKNOWLEDGEMENTS

This study was funded by research grants from Poul Due Jensen Foundation (Microflora Danica), and the Danish National Research Foundation (DNRF136). We thank Jesper L. Wulff for help with the sampling.

AUTHOR CONTRIBUTIONS

MA, PHN, IM, and AS designed the study. IM and AS oversaw the manual processing of the samples. MS extracted the DNA and performed Nanopore sequencing, while TBNJ generated the Illumina data. MS performed bioinformatics processing, MAG generation and comparative genomics. FP performed genome annotation, phylogenetic analysis and FISH/RAMAN microscopy. MS and FP wrote the initial manuscript. All authors reviewed the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41396-023-01372-6>.

Correspondence and requests for materials should be addressed to Mads Albertsen.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023