UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI BIOLOGIA

Corso di Laurea magistrale in Molecular Biology

TESI DI LAUREA

Marine sediments: interaction partners of cable bacteria revealed through metagenomics

Relatore: Prof. Stefano Campanaro
Dipartimento di Biologia

Laureando: Artur Sannikov

ANNO ACCADEMICO 2021/2022

Abstract

Cable bacteria from the family *Desulfobulbaceae* are multicellular filamentous electroactive bacteria found worldwide in aquatic sediments. They are capable of coupling sulfide oxidation and oxygen or nitrate reduction by transferring electrons over centimeter-long distances. It was hypothesized that other microorganisms might exploit this feature to access otherwise unreachable environments via direct interspecies electron transfer. In this thesis, a screening for extracellular electron transfer proteins was performed on high-quality genomic bins obtained through the integration of four binning tools. This screening led to a shortlist of ten taxonomic groups which have the potential to conduct electron exchange with cable bacteria, thrive in their presence, or compete with them. The *Sulfurimonas* genus was the most interesting group. There is evidence that this group may use cable bacteria to indirectly access oxygen in shortage in marine sediments. Simultaneously, *Sulfurimonas* contain proteins for iron reduction and oxidation, suggesting a potential for electron exchange with cable bacteria. To test this hypothesis, fluorescence in situ hybridization experiment was set up, where two signals were found co-localized with filaments of cable bacteria. Along with previous studies and the presence of electron transfer proteins, this indicates the involvement of *Sulfurimonas* in electron exchange with cable bacteria.

Acknowledgments

First, I would like to thank Ian Marshall for his broad support and close supervision, which inspired me throughout the project and helped me overcome my failures. In addition, I express my thanks to my supervisor Stefano Campanaro for proofreading my thesis and my examiner Alessandro Vezzi who gave me a lot of advice to improve the project. I would also like to thank lab technician Jesper Wulff, who helped me carry out wet lab experiments and patiently addressed all my doubts. I also appreciate the help of my officemate Jamie Lustermans who recommended so many papers without which I could not have imagined this project. I am also grateful to all other members of Center for Electromicrobiology and Microbiology Unit, with whom we had so many exciting moments of talks, games, and social events. I also wish to express my gratitude to the leadership and administrative staff of CEM, who organized professional events which extended my understanding of modern research. Finally, I am grateful to my friends and family for their limitless support.

Table of contents

1 Introduction

1.1 Metagenomics

Microbiology is a quickly developing biological discipline that has profoundly transformed our understanding of diseases and geochemical cycles. Until recently, it was only possible to study microorganisms through microscopy, culturing, and small-scale genetic experiments. The discovery of the structure of DNA and the subsequent development of DNA sequencing have opened a path to a deep study of microbial genetics. Furthermore, it was now possible to discover and investigate thousands of new species from environmental and clinical samples, which was unfeasible before due to their uncultivability. Historically, the sequences of 16S (for bacteria) and 18S (for eukaryotes) of ribosomal RNA (or internal transcribed spaces for fungi) were used to taxonomically profile samples as well as discern the abundance of organisms. The 16S gene sequencing is widely used because it provides a reliable, cost-effective, and computationally simple way to quickly assess the composition of even the most complex samples. The choice of these genes relies on two concepts: they are universally present and evolutionary stable (i.e., not affected by horizontal gene transfer). However, this method also presents some important disadvantages, such as the generation of chimeras during the amplification, which may create an inaccurate figure of sample composition and lead to wrong conclusions about the impact of an environment.

The development of high-throughput sequencing in the last 15 years has helped circumvent many of the issues and biases of single-gene sequencing. The most popular method of next-generation sequencing is called shotgun sequencing. It consists in cutting DNA into random fragments (indeed, shotgunning it), and sequencing these fragments into so-called reads of a known length. As this method produces relatively short reads (i.e., most commonly 50–300 base pairs), it is called short-read sequencing. However, in the past few years, long-read sequencing methods have appeared on the scene. These methods generate reads that may exceed 10,000 base pairs in length and are very useful to close gaps produced by short-read sequencing and resolve repeats by spanning over them. Long-read sequencing is still relatively new; thus, it is more expensive and error-prone than short-read sequencing. However, short and long reads can be used in combination to generate the best sequencing outputs.

A typical metagenomic workflow consists of sample collection, sequencing, quality control, metagenome assembly, genome binning, and taxonomic and functional profiling. The next few sections will briefly describe each of these steps, focusing mainly on bioinformatics. Additionally, many other analyses can be performed, for example, statistical comparisons within and between samples (i.e., α- and β-diversity).

1.1.1 Sample collection, sequencing, and quality control

Sample collection protocols may significantly affect the composition of an extracted sample, introducing biases and giving a wrong picture about the investigated environment. It is, therefore, important to use standardized and

identical protocols to sample the same environments (for instance, soil) to minimize the effect of differences between methods and used reagents and make the samples as comparable as possible. Furthermore, the enrichment can be used in low-abundant environments, although this method can introduce biases by overrepresenting microbes which benefit from the chosen enrichment approach.

Next, the difference between the abundances of organisms introduces the challenge of insufficient sequencing depth. As deep sequencing is costly, researchers cannot increase the depth as much as they want. Moreover, usually exceeding the sequencing depth of 50–100X does not lead to better results for a particular organism. However, if an organism is low-abundant in a sample, it may not be sequenced enough to reconstruct its genome. Thus, the choice of sequencing depth is challenging and should consider the preliminary knowledge of the studied sample. Another common challenge is repeated sequences which create issues in the assembly step as the tools cannot unambiguously connect different parts of a genome if those are contained within a repeat. However, long-read sequencing can span entire repeats and solve this issue.

After sequencing, it is essential to perform the quality control of the generated reads using tools, such as FastQC, and finally, trim the sequencing adapters and read segments of bad quality (for example, at the $3'$ end).

1.1.2 Metagenome assembly

The assembly of a metagenome is similar to that of a single-organism genome. An assembler should find overlaps between reads and connect them into longer sequences, called contigs. The most modern tools rely on the computational fragmentation of reads into shorter overlapping fragments of a certain length called k-mers. These k-mers are nodes, and overlaps between them are edges in de Bruijn graphs. The task of assembly tools is to find a path through the graph to reconstruct the metagenome. However, metagenomics presents a few challenges compared to the reconstruction of a single genome.

First, in metagenomic assembly, it is inaccurate to assume uniform coverage along the whole sequence because organisms have different abundances. Therefore different genomes within a metagenome are represented in different quantities. It is possible to recover low-abundance genomes by using shorter k-mers, but this solution comes at the cost of higher computational resource usage, a higher number of repetitive k-mer sequences, and less accurate and shorter contigs.

Next, in a metagenome, different strains of the same organism which differ by entire operons may create branches in a graph and cause an assembler to stop at these divergent points, leading to incomplete genome reconstructions.

Specifically, to overcome the first issue, modern assemblers use a range of k-mers of different lengths. For example, metaSPAdes iteratively refines a de Bruijn graph by increasing the length of k-mers at each iteration. A similar approach is implemented in MEGAHIT, which was best among the assembles used to reconstruct genomes in the community competition Critical Assessment of Metagenome Interpretation (CAMI) to assess the performance of different metagenomic tools. Finally, MEGAHIT implements memory-efficient

de Bruijn graphs and can run in parallel on Compute Unified Device Architecture (CUDA), making it a good choice for fast and reliable metagenome reconstructions.

1.1.3 Genome binning

Genome binning is a process of placing individual genomic contigs of a metagenome into individual "bins", which ideally should represent individual genomes.

Recent advances in sequencing techniques and a significant drop in cost allowed quickly generating millions of reads which then can be assembled using high-power server clusters and then relatively quickly binned into metagenome-assembled genomes (MAGs). Thus, the main objective of genome binning is to reconstruct known and unknown microbial genomes to power the description of a microbial community.

The most modern tools use hybrid approaches which combine the differential abundance (i.e., contigs with similar coverage should fall in one bin) and sequence composition (i.e., similar proportions of GC and nucleotide frequencies). These tools can be tested during community competition CAMI which uses high-, medium-, and low-complexity realistic metagenomic datasets generated from real data. Moreover, a tool for the assessment of genome binning, Assessment of Metagenome BinnERs (AMBER), has been developed. It uses four evaluations metrics:

1. *Completeness* or sensitivity
2. *Purity* or precision
3. *Contamination*
4. *F1-score* which is a combination of *completeness* and *contamination*

This tool was used to assess twelve non-integrative genome binners and three integrative genome binners using six other custom metrics [1]. The integrative approaches accept bins generated by multiple tools and use different methods to increase the quality of the output bins. Overall, the binners using integrative approaches performed much better both in real-world and CAMI datasets. Among them, DAS Tool recovered the greatest number of high-quality bins (contamination $< 5\%$, completeness $> 90\%$). Moreover, it performed well in terms of the *F1-score*, even though the boxplot ranges were wider and had a lot of outliers compared to the bin refinement module included in MetaWRAP.

In this project, an integrative approach was used by combining the results of four traditional binning tools such as MetaBAT 2, VAMB, MaxBin, and MetaDecoder. All the tools differ in their approaches, although, like many other tools, they share common metrics to split contigs among genomic bins. For example, MetaBAT 2 uses tetranucleotide frequency and abundance score and compared to its previous iteration, it further normalizes these scores using a graph-based clustering algorithm. Similarly, MaxBin uses tetranucleotide

frequency and genome coverage for the binning and employs an expectation-maximization algorithm for binning. On the other hand, VAMB uses deep learning to bin contigs according to their abundance and k-mer distribution. Finally, MetaDecoder employs a two-layer technique that puts contigs through two different models.

For this project, four tools (MetaBAT 2, MaxBin, VAMB, and MetaDecoder) were selected to generate the input bins for DAS Tool, which employs a scoring function based on the number of single-copy genes to score the bins and then uses an iterative procedure to extract a non-redundant set of bins. This tool accepts a flexible number of bins generated by traditional binning tools and is able to output near-complete, less contaminated genomic bins.

After the binning step, it is essential to evaluate the quality of the resulted bins. The most commonly used tool, CheckM, uses lineage-specific sets of single-copy marker genes to evaluate the bin completeness and contamination. In this context, a complete MAG should contain 100% of all marker genes specific to the lineage of this genome. The contamination is determined by counting the number of multicopy marker genes and dividing this number by the number of collocated marker genes within the lineage of the evaluated genome. As such, the contamination may exceed 100% if the number of multicopy genes exceeds the number of collocated genes.

Another way to evaluate the bin quality is to measure the number of contigs (an ideal MAG should have only one circular contig), total contig length, N50, L50, etc. These metrics are usually estimated with QUAST.

1.1.4 Taxonomic classification and profiling

Strictly speaking, taxonomic classification and taxonomic profiling are two distinct procedures. The former is the assignment of taxonomic information to genomic bins using one of the taxonomic databases, while the latter is the estimation of relative abundances of different species within a sample. These two steps in a metagenomic workflow complement each other as we are often interested in both pieces of information.

One of the most used databases for taxonomic classification, Genome Taxonomy Database (GTDB), contains more than 250 thousand genomes classified into almost 48 thousand species clusters (as of version r202). The genomes are pulled from RefSeq and GenBank, and the database is updated regularly. Importantly, this database also incorporates the genomes of uncultured microorganisms mainly discovered with metagenomics and thus, provides a more comprehensive picture of the microbial world. To aid with the classification, the GTDB-Tk tool was developed, which places query genomes into the GTDB reference tree, integrates this information with other genome characteristics, and finally provides the most plausible taxonomic classification.

The taxonomic profiling can be performed through the calculation of the average sequencing depths of all contigs belonging to a specific genomic bin. Note that this will provide only the relative taxa abundances, and it is impossible to obtain absolute abundances due to the nature of sampling.

1.1.5 Functional profiling

Functional profiling is required to establish the main functions performed by a microbial community. This step can be performed by aligning genome sequences to reference genes, computing gene abundances, and grouping them by gene or protein families using databases such as UniProt or KEGG. This process can be easily conducted in a flexible functional profiling pipeline like HUMAnN.

Another way to perform functional profiling is to search for specific genes of interest, for example, those involved in iron or sulfur processing. By linking these genes to pathways they are required for, it is possible to determine a very specific and zoomed functional potential of a limited number of investigated species.

The latter approach may be helpful when more accurate information is needed for the investigation of a specific characteristic of a microbial community, such as electron transfer or antibiotic resistance.

1.2 Marine sediments

Marine sediments cover most of the seabed all over the globe. This environment hosts one of the world's most complex and extensive ecosystems. They host organisms from all major domains of life (Eucarya, Archaea, and Bacteria) with dozens of phyla from each of them. The examples include species from Animalia, Fungi, Protozoa, Chromista, and even Plantae. However, the vast majority of biomass is represented by bacterial and archaeal communities, which are less studied than other taxonomic groups. Marine sediments are important for sulfur, carbon, and nitrogen cycles, influencing life cycles outside marine environments. Microbial communities play a crucial role in these cycles. For example, it is estimated that around 1/3 of organic matter is remineralized by sulfate-reducing microorganisms. Most bacterial and archaeal cells living in marine sediments use sulfur compounds at different stages of transformation as electron donors and acceptors for energy conservation. Some of the most common pathways include dissimilatory reduction of sulfate and sulfur intermediates, sulfur oxidation, and anaerobic oxidation of methane coupled to sulfate reduction. Furthermore, many microbes contribute to mitigating global warming through methane oxidation (which arises from marine sediments) by methanotrophic archaea, which perform this process in concert with sulfate-reducing bacteria.

However, metabolic processes in marine sediments are not defined by only sulfur-related organisms. Bacterial cells have had almost 4 billion years to evolve and exploit every possible energy source. As such, they can also use iron, nitrogen, and carbon compounds as electron acceptors and donors. In addition, their ability to live in the most extreme conditions on Earth (i.e., deep-ocean hydrothermal vents, saline lakes, upper layers of the atmosphere) makes them one of the most important components of today's ecological cycles and therefore, research on them is essential in order to mitigate the most pressing issues, such as global warming, ocean acidification, and air pollution.

Marine sediments can differ substantially in their structure as well as the

conditions above them, for example, pH, porosity, temperature, pressure, and oxygen availability. In this regard, microorganisms of marine sediments can be divided into two main groups: either living in an oxic or anoxic environment. Regarding the bacterial community (the main focus of the present thesis), species of Alphaproteobacteria, Gammaproteobacteria, and Firmicutes prefer oxic environments, while those of Chloroflexota tend to stay in anoxic conditions. In such regard, the main protagonist of this work, cable bacteria, represents an exciting example of microorganisms as they are able to live at the interface of oxic and anoxic environments.

1.3 Inferring interactions between microorganisms

Previously, Morten Hoppe investigated possible interactive partners of cable bacteria from three locations (Baltic Sea, Darwin Harbour, Berre Lagoon) using a graphical model-based tool FlashWeave [2]. This tool is able to construct high-fidelity co-occurrence networks to infer microbial interactions.

The identification of interactions between microorganisms that build ecological networks is not an easy task. Historically, Pearson's or Spearman's correlation coefficients have been used, although recently, co-occurrence networks claimed to be more efficient and accurate in determining microbial associations.

Although co-occurrence methods are widely used as they seem more accurate in predicting interaction patterns, they were also criticized. For example, Hirano and Takemoto compared six correlation-based and three graphical model-based co-occurrence network methods on generated relative abundance data using a dynamical model and tested five different types of interaction: random, mutualistic, competitive, parasitic, and combination of mutualistic and competitive [3]. The results showed that co-occurrence-based methods performed not much better than traditional correlation-based methods like Pearson's or Spearman's correlation. Moreover, the parasitic interaction type decreased the performance of co-occurrence networks. Thus, the usefulness of this method is smaller than previously thought. Manual construction of ecological networks is required to achieve higher accuracy.

1.4 Cable bacteria

Cable bacteria are filamentous multicellular members of the *Desulfobulbaceae* family of the Desulfobacterota phylum, which includes the species capable of anaerobic sulfate reduction. Cable bacteria were discovered in 2012 in an attempt to explain a long-distance transfer of electrons across the marine sediment in Aarhus Bay, Denmark. These bacteria are able to couple the oxidation of hydrogen sulfide, H_2S, in the anoxic zone deep in the marine sediment to the reduction of oxygen, O_2, near the surface [4] over centimeter-long distances. Later, cable bacteria were also confirmed to be able to reduce nitrate as an alternative to oxygen.

These bacteria were also found in freshwater and are now classified into two genera: *Candidatus* Electrothrix and *Candidatus* Electronema. The former is found mainly in the marine environment, whereas the latter is primarily a

freshwater species. Until now, this division was thought to be sharp.

Many questions remain open regarding cable bacteria, such as their conductive mechanisms, metabolic pathways, and ecology. The present work will focus on the ecological aspect, particularly the interactions between bacterial cells of marine sediments and cable bacteria, as not much work has been done in this field.

Cable bacteria belong to electroactive microorganisms, thus it is reasonable to hypothesize that they may exchange electrons with other microbes. One of the most common and proven methods to do so is through direct close contact, also known as direct interspecies electron transfer (DIET).

1.5 Direct interspecies electron transfer

DIET is a process of electron exchange between two or more species without an intermediate electron carrier like it appears in mediated interspecies electron transfer. It is a form of syntrophy, a mutualistic interaction between two species that confers trophic benefits to both interactive partners. For example, in 1967, Barker et al. demonstrated that a "pure" culture of *Methanobacillus omelianskii* was actually a mixture of two interdependent species. One of them, "S organism", produced hydrogen and acetate from ethanol, while the second organism used hydrogen to grow and generate methane. Interestingly, the fermentation of ethanol was inhibited by the high partial pressure of hydrogen; thus, its continuous transformation was required for the survival of both species [5]. Syntrophy may also be beneficial, particularly in environments with a limited number of electron donors and/or electron acceptors. An example of DIET was described between *Methanosarcina barkeri* and *Geobacter metallireducens* [6]. Pure cultures of these two species were unable to metabolize ethanol, but when grown in coculture, they started to transform ethanol to methane and carbon dioxide with a transient formation of acetate. Moreover, methane was also produced from the resulting carbon dioxide. Therefore, the authors suggested that *M. barkeri* used outer surface cytochromes to accept electrons from *G. metallireducens* via DIET.

For example, one organism is required to donate electrons to another organism (which does not have direct access to other electron donors) to ultimately transfer them to a terminal electron acceptor. Thus, a whole microbial community can thrive by using other microbes for their metabolic needs.

Several mechanisms for DIET have been proposed: nanowires, abiotic conduits, or outer membrane porin-cytochrome networks.

1.5.1 Nanowires

Nanowires are biotic or abiotic nanostructures with diameters of a few nanometers. In bacteria, they are mainly made of OmcS and OmcZ *c*-type cytochromes and are used to transfer electrons to extracellular electron acceptors. This type of DIET is mainly studied in *G. sulfurreducens* which uses the multiheme *c*-type cytochrome OmcS to form nanowires. OmcZ is another cytochrome that may also constitute nanowires in *G. sulfurreducens*. The ability

to transmit electrons along nanowires derives from a tight packaging of aromatic amino acids known as π-π stacking (the term is misleading, as direct stacking of aromatic amino acids causes a repulsive interaction, and so-called staggered stacking or π-teeing conformations are much more common).

1.5.2 Abiotic conductive conduits

Another option for DIET is the usage of abiotic conductive conduits such as magnetite or carbon derivatives. For example, *G. sulfurreducens* is capable of using magnetite as a substitute for the OmcS cytochrome and may even use this mineral as a preferred way to transfer electrons. However, it does not seem that magnetite can completely substitute longer conductive structures as it was not proven that it can form long magnetite chains, thus, the close proximity of the membranes of two species is still required, whereas nanowires may extend to longer distances. Nonetheless, carbon materials such as biochar may provide an option for microbes to be distant from each other as these materials appear to facilitate electron exchange between bacteria attached to them [7].

1.5.3 Outer membrane proteins

Finally, porin-cytochrome networks localized in the outer membranes of interacting species may participate in DIET. This possibility was confirmed in the *Prosthecochloris aestuarii* and *G. sulfurreducens* cocultures which were unable to grow when the genes necessary for this network were deleted from *G. sulfurreducens* [7]. Furthermore, it is suspected that these two species are able to fuse their membranes which can be another strategy for direct electron exchange.

1.5.4 Searching for more species performing DIET

Up to now, only a few species have been identified to perform DIET, such as different species of the *Geobacter* and *Methanothrix* genera; thus, new methods to enhance the search strategies should be developed to enable faster discovery of DIETers or at least pave the way for faster identification of microorganisms showing this potential (for example, using metagenomics).

One option to improve this process is through the enrichment of electrodes or minerals. As described earlier, some organisms are able to use magnetite or carbon as electron exchange facilitators; thus, growing organisms suspected in DIET may directly demonstrate their ability to exchange electrons. Electrodes can also be used for the same purpose as some bacteria were shown to grow on them.

However, in order not to waste resources on the organisms which are definitely not able to perform DIET, it is a viable idea to screen for genes involved in extracellular electron exchange (required for DIET) in metagenomic datasets. To some extent, it is correct to say that if a near-complete genome does not contain any of such genes, attempts to identify its DIET mechanisms can be a waste of resources and time. Thus, this strategy allows a much

faster and cheaper way to rule out irrelevant species even in the most complex environments.

Finally, if organisms have distinct and clearly identifiable morphologies, electron microscopy may be used to directly observe physical attachment between organisms (as required by DIET by definition).

1.5.5 DIET of cable bacteria

Cable bacteria are a great example of electroactive microorganisms because even the "cable" term in their name derives from their main property: electron transfer at very long distances. Their strength lies in easily accessing a very efficient electron donor (sulfide) and electron acceptor (oxygen or nitrate). Thus, many other microorganisms may benefit from this long-distance coupling by attaching themselves to long, thick biotic cables. Meysman et al. have demonstrated that air-dried filaments of cable bacteria can be interfaced with electrodes [8]. This study, therefore, opens the possibility of electron transfer to solid materials in addition to dissolved compounds. Even though the filaments used in this were air-dried, they still were able to conduct electrons; thus, there is no reason they cannot do it in natural conditions. However, this capability is yet to be demonstrated. Consequently, cable bacteria may also donate electrons to other microorganisms. In syntrophic terms, when cable bacteria lack access to their prominent electron donors or acceptors, they may utilize other species as sources or sinks, while the latter can greatly benefit from this process by accessing wider environments and not being restricted by narrower zones of interfaces between electron acceptors and donors.

In a study of the effect exerted by cable bacteria on a microbial community in order to improve the remediation of polluted with polycyclic aromatic hydrocarbons (PAH) environments, many positive interactions were discovered between three bacterial phyla and cable bacteria during the artificially elevated oxygen treatment (which promotes the growth of cable bacteria) [9].

Furthermore, in a study by Vasquez-Cardenas et al., a more substantial inorganic carbon uptake was observed in sediments with cable bacteria. When cables were cut, it was associated with a drastic reduction in carbon uptake, suggesting that the chemoautotrophic community in the suboxic zone relies on access to oxygen in the top layers of the sediment (provided by cable bacteria). The study suggested a possibility that these microorganisms may utilize cable bacteria as an electron acceptor, therefore indirectly accessing oxygen [10].

Finally, Sarah Jannie de Roode obtained scanning electron microscope images of cable bacteria in which attached microorganisms are clearly observed (Fig. 1, unpublished data).

In conclusion, the indications that cable bacteria may transfer electrons to solid materials, have a huge impact on the surrounding microbial communities, and direct observation of attached microorganisms suggest that it is worth investigating which microbes may potentially exchange electrons with cables and, in general, expand the knowledge about the microbial community surrounding cable bacteria. This project will investigate the ecology of cable bacteria mainly by searching for the relevant EET proteins in high-quality draft

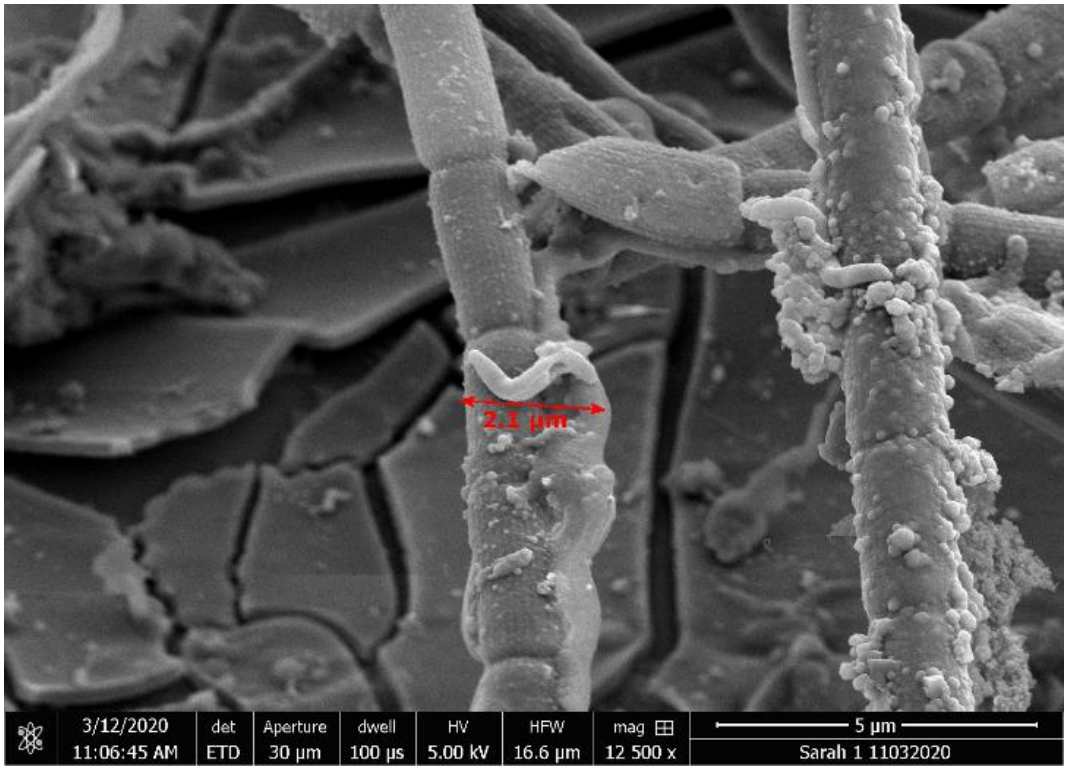


Figure 1: Image of cable bacteria with a microorganism (length 2.1 µm) attached to a filament. Adopted from the report "A method to visualize bacteria attached to cable bacteria with scanning electron microscopy" by Sarah Jannie de Roode.

genomes of species present in three different marine samples in the context of collaboration and competition between bacteria.

2 Materials and methods

2.1 Samples

Two samples were collected in March 2020 from Kalø Vig Boat Harbor (56°14′33.7″N 10°20′18.6″E) and Løgten Strand (56°17′13.9″N 10°22′58.6″E) in Central Jutland, in Denmark. The DNA was subsequently extracted and sequenced on Illumina HiSeq machines as described in the Master's thesis by Morten Hoppe [11].

The third sample (*Candidatus* Electrothrix communis RB) was extracted from a single-strain enrichment of cable bacteria and sequenced on an Illumina MiSeq machine. Compared to the first two samples, these are clumps of filaments extracted from the sediment.

2.2 Trimming, quality control, and assembly

The present work started from raw untrimmed reads and assembled metagenomes from Kalø Vig and Løgten. The trimming was repeated with Trimmomatic

0.39 using the same parameters as in Morten Hoppe's thesis (`CROP:150, HEADCROP:20, SLIDINGWINDO:4:20, MINLEN:100`). The adapter sequences to remove were retrieved from the file `adapters.fa` the BBMap's GitHub page. The quality control was performed with FastQC 0.11.9. The assembly files generated by MEGAHIT 1.2.9 were available beforehand, and the assembly procedure is described in Hoppe's thesis.

The *Ca.* Electrothrix communis RB sample was sequenced by Lea Plum-Jensen on Illumina MiSeq. She used Trimmomatic 0.39 with the following parameters, `CROP:290, HEADCROP:20, SLIDINGWINDOW:4:20, MINLEN:100`, and the same `adapters.fa` file to trim the reads. The quality control was performed with FastQC 0.11.5. The genome assembly was conducted with metaSPAdes v.3.15.4 with the following k-mer lengths for graph construction: `-k 21,33,55,77,99,127`.

2.3 Genome binning and data integration

Four different tools were used for the initial binning: MetaBAT 2.12.1 (as part of the MetaWrap 1.3.2 package), VAMB 3.0.2, MaxBin 2.2.6 (as part of the MetaWrap 1.3.2 package), and MetaDecoder 1.0.7.

MetaBAT 2 (inside MetaWrap) was used with default parameters.

VAMB required creating FASTA catalogs out of contig files using an included script `concatenate.py`. Next, mapping and indexing of the trimmed reads on the resulting catalogs were done with bbmap 38.90 with the parameter `usemodulo=t` (to save memory). In the VAMB documentation, it is stated that MetaBAT is a better tool to estimate the depths of `bam` files, which can be passed to VAMB using the `--jgi` parameters. Thus, the `bam` files were sorted by read position with samtools 1.15 and then their depths were estimated with the MetaBAT's `jgi_summarize_bam_contig_depths` parameter. Finally, the tool was run with the `--minfasta 200000` parameter to remove the bins shorter than 200 kilobase pairs.

MaxBin (inside MetaWrap) has had multiple issues while trying to run it. The main issue was detected: it could not find abundance information for contigs that were not present in the mapped `sam` files. Thus, a custom Python command-line program, `extract_fasta.py` (available on the project's GitHub repository) generated FASTA contig files with only the contigs present in the mapping files. Next, MaxBin was used with the default parameters of MetaWrap.

Finally, MetaDecoder was used with default parameters.

The next step was the bin refinement using DAS Tool 1.1.4. Two different sets of bins were used as an input for DAS Tool: one set including MetaDecoder results and the other one excluding them. The parameters used are the following: `--write_bin_evals` and `--write_unbinned`.

The genomic binning of *Ca.* Electrothrix communis RB enrichment was performed only with MetaBAT 2.12.1 using default parameters, and bin refinement described below was not applied.

2.4 Bin refinement with custom Python scripts

To decide in which setup (including or excluding MetaDecoder's bins) the DAS Tool algorithm performed better, a comparison was made between different metrics generated by CheckM 1.1.3 and Quast 5.0.2:

1. Completeness
2. Contamination
3. N50
4. L50
5. Number of contigs
6. Maximum contig length
7. Total contig length

These quality metrics are recommended by the Genomic Standards Consortium. However, neither of the two setups greatly outperformed the other one, and it was difficult to select only one. Thus, using only the completeness and contamination metrics, a procedure was developed to extract the best genomic bins from the two datasets:

1. Use FastANI 1.33 to select the bins which are identical between the two setups (consider identical only if average nucleotide identity (ANI) $>$ 99.5%).
2. Extract the bins which are exclusive to either the dataset from the MetaDecoder or non-MetaDecoder approach.
3. If the completeness and contamination are identical between the two (nearly) identical bins, select only one of them.
4. Select the bins with the highest quality using the following algorithm:
   (a) If one of the two (nearly) identical bins has lower contamination, select it.
   (b) Otherwise, if the contamination percentages are equal, select the bin with the highest completeness.
   (c) Simultaneously, filter out all the bins with contamination $> 10\,\%$ and completeness $\leq 50\,\%$.
5. Combine the exclusive to each setup bins with the selected bins with the highest quality.

The comparison and selection procedures are described in the Jupyter notebooks `statistics.ipynb` and `selecting_best_bins.ipynb` available on the project's GitHub. Python 3.10.4 and pandas 1.4.1 were used for these notebooks. The same versions were used throughout the project in addition to NumPy 1.22.3, Biopython 1.79, Matplotlib 3.5.1, and Seaborn 0.11.2.

2.5 Taxonomic classification and taxa abundance

GTDB-Tk 1.7.0 with reference data r202 was used to perform the taxonomic classification of the data.

Next, using the MetaBAT's contig depth files (generated with the `jgi_summarize_bam_contig_depths` parameter), relative taxa abundances were computed as the average depth value of all contigs contained within a certain bin. The process for the Kalø Vig and Løgten samples is performed in the Jupyter notebook `compute_abundance.ipynb`. The same procedure was used for the *Ca.* Electrothrix communis RB sample and described in the `compute_electrothrix_communis_illumina_abundance.ipynb` notebook.

2.6 Phylogenetic tree

The shortlisted for interactions genomic bins were chosen to build a phylogenetic tree. To put them into the context of phylogeny, the most complete and similar genomes were chosen in GTDB (release 07-RS207, 08 April 2022). The complete list of shortlisted genomic bins and their references in the database are available in the supplementary tables in the directory `csv/taxonomy` on the project's GitHub page. Additionally, a cyanobacterium *Synechocystis sp. PCC 6803* (assembly ASM792v1 in the Assembly database of National Center for Biotechnology Information) was used to root the tree.

To build the tree, GTDB-Tk 1.7.0 with reference data r202 and default parameters was used to identify marker genes (command `gtdbtk identify` with default parameters) and create a multiple alignment (command `gtdbtk align` with default parameters). Next, IQ-TREE 2.2.0.3 was used to build the tree with the following parameters: `-m TESTNEW -bb 1000`. Ian P. G. Marshall created the protocol.

The resulting `tree` file was opened with FigTree 1.1.4 and rerooted with the cyanobacterium mentioned above, and the resulting file was saved in the `svg` format.

2.7 Searching for relevant proteins

Six different proteins involved in EET were selected to be searched for in the genomic bins:

1. OmcS;
2. OmcB;
3. OmcC;
4. OmcZ;
5. MtrC;
6. PioA.

A brief description of their functions can be found in the Functional profiling section of the Results. The search for these proteins was conducted with a standalone version of BLAST.

In addition, proteins related to iron reduction, oxidation, and dissimilatory sulfur reduction were searched for with FeGenie and InterProScan.

2.7.1 BLAST

The gene sequences of all genomic bins were converted into protein sequences with Prodigal 2.6.3 (default parameters, output files in GenBank format). Then, a local BLAST database was built with the `makeblastdb` tool included in BLAST+ 2.12.0 programs. To build the database, concatenated protein sequences should be used as input. However, during the database build-up, an issue arose: some contigs were present in multiple bins simultaneously, causing `makeblastdb` to malfunction and crash. One of the reasons for this issue was that the same contigs were present in bins with identical names but one of them with the `_sub` suffix. These bins result to be virtually identical but the `_sub` version is slightly shorter (for example, the bins `metabat_bin.119` and `metabat_bin.119_sub` in the Løgten sample have the lengths of 5,195,887 and 5,190,063 base pairs, respectively). It is, however, not clear why these different versions are created by DAS Tool, as the only mention of this suffix is to be found in the tool's source code with a code comment `remove contigs of highest scoring bin`.

Thus, the duplicate protein sequences were removed directly from the concatenated file using the command-line script `remove_duplicate_seq.py` (on project's GitHub). This procedure allowed a smooth generation of a local BLAST database.

The result of BLASTing the EET proteins against this database resulted in a list of bin protein sequences that hit the EET protein sequences. However, to understand how much proteins hits were attributed to each protein in each bin, these tables were used to generate new tables where rows correspond to bin names and columns to protein accession numbers, and the rest of the table is filled with numbers corresponding to the number of hits of a protein in a bin. Briefly, the algorithm works as follows: search for each protein sequence inside each bin, and if found, increase the total number of protein hits for that protein in that bin by 1. An example of the resulting table is given below.

	Protein 1	Protein 2	Protein 3	Protein 4	Protein 5	Protein 6
bin 1	1	0	2	0	0	0
bin 2	1	2	3	1	0	2

For example, protein 1 had one hit in bin 1 and bin 2, protein 2 had zero hits in bin 1 and two hits in bin 2, etc.

It is also worth noting that sometimes a query protein sequence matched the same contig (converted into a protein sequence) two or more times. This is probably due to the fact that a query could match a sequence in the first few amino acids and report it but then continue and find another match in

the same sequence and also report it. In this case, only the sequences with the lowest e-value were used to generate the above tables to avoid unwanted and artificial increases in the number of protein hits. The algorithm to generate the tables with protein hits are described in the notebooks `protein_hits.ipynb` for the Kalø Vig and Løgten samples and `protein_hits_electrothrix_communis.ipynb` for the *Ca.* Electrothrix communis RB sample.

2.7.2 Iron reduction and oxidation

Additionally, FeGenie v1.2 was employed to screen the bins for proteins involved in the iron reduction and oxidation. The program is based on its own Hidden Markov model libraries containing a collection of protein sequences involved in iron utilization. FeGenie accepts protein or gene sequences and reports findings split into five categories related to iron processing with genes for 13 different functional categories, including iron reduction and oxidation. This program was not only useful to search for additional proteins with possible EET functions but also to confirm the BLAST result as many protein families overlapped with the previously used sequences, such as PioA, MtrC, OmcS, and OmcZ. As previously with BLAST, protein sequences were used as an input for FeGenie, and the parameters used are: `--orfs` and `-inflation 1000`.

Only the iron oxidation and reduction categories were then selected for further analysis. Note that the PioAB/MtoAB proteins are required for iron oxidation, but in FeGenie, they belong to the iron oxidation/reduction category, so the presence of PioA found by BLAST could not be technically confirmed by FeGenie.

2.7.3 InterProScan

Finally, as dissimilatory sulfate reduction results in the generation of the main substrate of cable bacteria (sulfide), InterProScan 5.54-87.0 was applied on the 50 most abundant genomic bins in the Kalø Vig and Løgten samples (in the former though only 49 bins were used as one of them was removed due to high contamination) to search for the DsrA (InterPro accession number IPR011806) and DsrB (InterPro accession number IPR011808) proteins essential for this pathway. This tool uses the InterPro database, which classifies proteins into functional families. The parameters used are `-pa` (to include pathway in the output) and `-dp` (to disable precalculated lookup service). The full procedure to find the bins containing DsrA or DsrB proteins is described in the `interproscan_Dsr.ipynb` Jupyter notebook on the project's GitHub page.

This tool, however, was not used on the *Ca.* Electrothrix communis RB sample, as it was an enrichment of cable bacteria; thus, its main purpose was to search for proteins involved in EET.

2.8 Data integration

Different pieces of data were subsequently merged to obtain a better overview of the results. First, quality metrics were merged with taxonomic informa-

tion using the script `merge_quality_taxa_tables.py` (for the Kalø Vig and Løgten samples). There all the quality metrics are being used, even though subsequently, only the contamination and completeness metrics remained. For the *Ca.* Electrothrix communis RB sample, the `electrothrix_communis_illumina_taxa_quality_table.ipynb` notebook was used to merge the taxonomic and quality information (in this case, only contamination and completeness were used). Next, the taxa abundance information was added to the tables with the `compute_abundance.ipynb` (for the Kalø Vig and Løgten samples) and `compute_electrothrix_communis_illumina_abundance.ipynb` (for the *Ca.* Electrothrix communis RB sample) notebooks. Finally, the FeGenie information (in `fegenie_quality_table.ipynb` notebook) and BLAST protein hits (in the `fegenie_quality_blast.ipynb` notebook) were added . In the final tables, the number of genes found by FeGenie was reported per each category per each bin.

2.9 Fluorescence in situ hybridization

To test the hypothesis of cable bacteria-*Sulfurimonas* interaction, a fluorescence in situ hybridization (FISH) experiment was set up. However, the first results were unsatisfactory because the signals were too weak. However, the results obtained by the lab technician Jesper Lundsgaard Wulff were of excellent quality, so his images were used for this project.

Cable bacteria can be easily visualized with a phase-contrast microscope on so-called trench slides, custom-manufactured glass slabs glued on top of a microscope slide. This setup creates a trench in the middle, filled with sand or sediment and covered by a coverslip. Thus, cable bacteria can grow from the sediment toward the oxic/suboxic zone at the edge of the slide.

A special setup exists to perform FISH on trench slides. A *Ca.* Electrothrix communis RB enrichment culture was grown on a Poly-D-Lysine coated trench slide for five days. Then, the trench slide was horizontally submerged in a 4% paraformaldehyde (PFA) solution for 15 minutes. After that, the slide was consecutively submerged in 25%, 50%, and 96% of ethanol solutions (diluted with 1 x phosphate-buffered saline buffer) for five minutes each. Then, it was dried at 46°C, the coverslip removed, and sediment carefully scraped off with a scalpel and a metal inoculation loop. The trench slide was horizontally submerged in 0.1% sodium pyrophosphate for ten minutes. Next, it was again submerged horizontally in a pre-heated solution of 0.02% agarose and 0.1% of sodium pyrophosphate for ten minutes and let dry at 46°C. To complete the fixation, the sample was dehydrated by immersing the slide and coverslip in ethanol solutions of 50%, 80%, and 96% for three minutes each and left to dry vertically.

After the fixation step, hybridization buffer (360 µL of 5M NaCl, 40 µL of 1M Tris hydrochloride (Tris-HCl) at pH 8, 700 µL of deionized formamide, 858 µL of sterile H_2O, 2 µL of 10% sodium dodecyl sulfate (SDS), 20 µL of MgCl (5 g/L), and 20 µL of $CaCl_2$ (1 g/L)) with three probes (with relation 8:1, hybridization buffer:probe) was applied, and the slide was incubated for 90 minutes at 46°C. The probes used are:

1. Universal bacterial probe EUB338-III_Atto-488, 5′-GCT GCC ACC CGT AGG TGT-3′.

2. DSB706_Atto-550 probe for *Desulfobulbaceae* (including cable bacteria), 5′-ACC GGT ATT CCT CCC GAT-3′.

3. ESPY549 for Epsilonprotebacteria including *Sulfurimonas*, 5′-CAG TGA TTC CGA GTA ACG-3′.

The washing buffer (1000 µL of 1 M Tris-HCl, 500 µL of 0.5 M ethylenediaminetetraacetic acid (EDTA), 700 µL of 5M NaCl, 50 µL of 10% SDS, and enough Milli-Q water to reach 50 mL) was prepared and preheated at 48°C in a water bath. It was used to rinse the slide after the incubation and perform the washing in falcon tubes for 15 minutes in a water bath at 46°C. Finally, the slide was rinsed with sterile Milli-Q water and dried in the dark at 46°C.

For the DNA staining, a solution of 1 µg/mL DAPI (800 µL of Citifluor, 200 µL of VECTASHIELD Antifade Mounting Medium, 1 µL of 1 mg/mL of DAPI, 4′,6-diamidino-2-phenylindole) was applied on the slides and incubated in the dark for 15 minutes before microscopy.

3 Results

3.1 Integrative approach improved quality of bins compared to traditional MetaBAT method

The application of an approach using four (including MetaDecoder) different binning tools and the integration of their results with DAS Tool generated two sets of high-quality bins (one including MetaDecoder bins and the other one excluding MetaDecoder bins). Interestingly, the integrative approach to the bins from all four tools recovered a greater number of bins: 344 against 279 in Kalø Vig, and 251 against 203 bins in Løgten, with and without MetaDecoder, respectively.

As described in the Materials and methods section, using Python scripts, the bins with the highest quality were selected from the two sets and, subsequently, combined to generate genomic bins with even higher quality that were further filtered to obtain. This method yielded 251 and 185 genomic bins with contamination of $< 10\%$ and completeness $\geq 50\%$, from the Kalø Vig and Løgten samples, respectively. Out of these, 75 and 60 bins (in Kalø Vig and Løgten, respectively) had contamination $< 5\%$ and completeness $> 90\%$, which can be regarded as high-quality drafts of metagenome-assembled genomes. The 10% and 50% thresholds belong to medium-quality drafts and were selected as a trade-off between the quality and quantity to filter genomic bins for future analyses.

The final numbers of bins are significantly lower than those from the MetaBAT 2 approach, which yielded 615 and 448 bins for the Kalø Vig and Løgten samples, respectively. However, the number of high-quality bins produced by MetaBAT 2 is lower: 53 in Kalø Vig and 38 in Løgten.

Even though the quality metrics of N50, L50, the total number of contigs, largest contig, and the total sequence length were not used to select the bins for subsequent analyses, they are still available on the project's page on GitHub in the `quality` directory (`kalovig_quality_table.csv` and `logten_quality_table.csv`).

Tables 1 and 2 show truncated versions of these tables with only the first three genomic bins (ordered by N50) with their completeness, contamination, number of contigs, and N50.

As an example of the extent to which the quality of genomic bins can be improved using this method, it is worth mentioning the bin *maxbin_bin.577* (genus *Marmoricola* of the phylum Actinobacteriota) from the Kalø Vig sample, which has completeness of 98.28, contamination of 1.72, 23 contigs, and N50 of almost 600,000 (599,905). This result can be expected from the sequencing of a pure culture but not from a complex metagenome of marine sediments. Another good example is the bin *maxbin_bin.515* (genus *Panacagrimonas* of the phylum Proteobacteria) from the Løgten sample, which has the same completeness of 98.28, contamination of zero, 30 contigs, and N50 of 385 thousand (358,576).

Kernel density estimate (KDE) plots (Figs. 2 and 3) demonstrate how the contamination decreased, and the completeness increased in both samples when using the DAS Tool approach (including the results from MetaDecoder) before manual curation of the data.

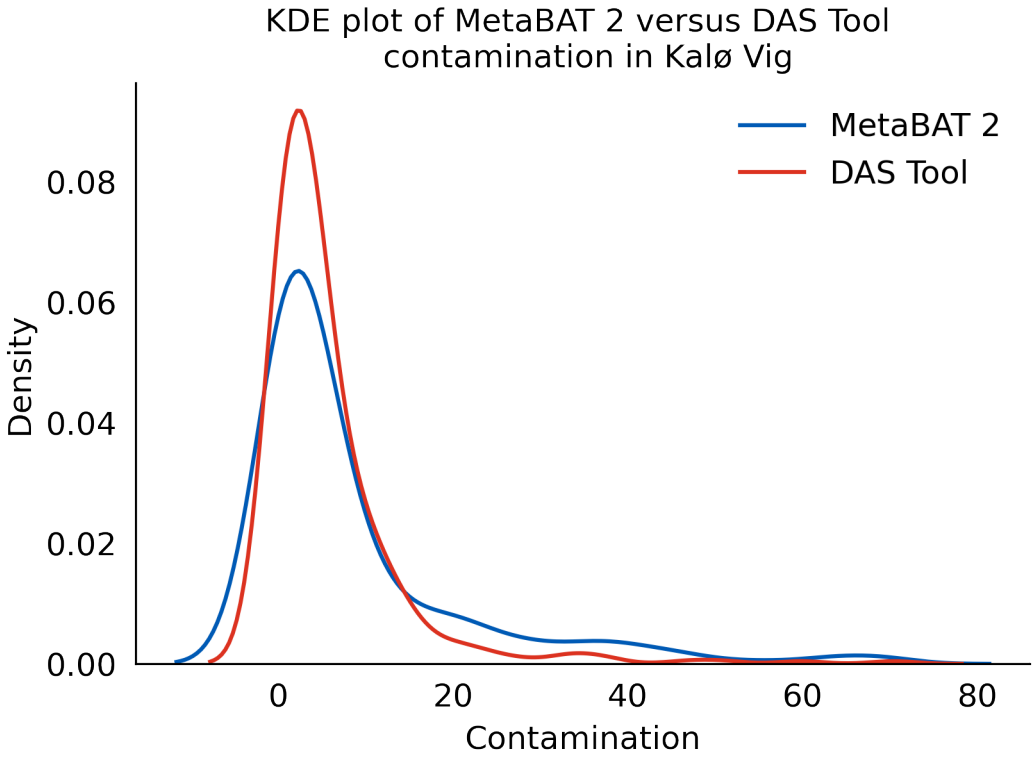
In contrast to the previous approach, the *Ca.* Electrothrix communis RB cable bacteria enrichment was not subjected to the DAS Tool approach and was used as is. The motivation behind this choice stems from the fact that this is an extraction of a filament that provides a more zoomed-in view of the microbial community around cable bacteria. This choice, however, resulted in less complete and more contaminated genomic bins, and only three (out of 48) of them can be considered high-quality genome drafts.

Table 1: Three genomic bins with the highest N50 from Kalø Vig.

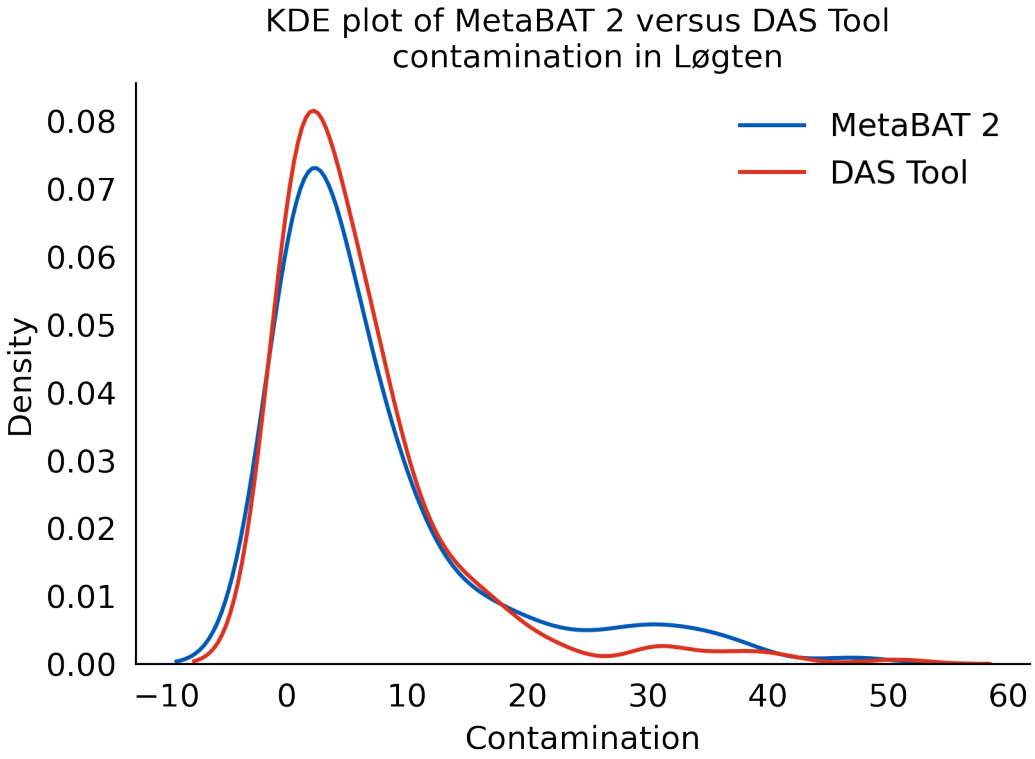
Bin id	Completeness (%)	Contamination (%)	Number of contigs	N50
maxbin_bin.577	98.28	1.72	23	599,905
metabat_bin.448	87.93	0	14	167,221
maxbin_bin.525	86.21	1.72	89	128,483

Table 2: Three genomic bins with the highest N50 from Løgten.

Bin id	Completeness (%)	Contamination (%)	Number of contigs	N50
maxbin_bin.515	98.28	0	30	358,576
vamb_S1C10321	96.55	0	33	135,972
maxbin_bin.466	100	2.51	59	124,753

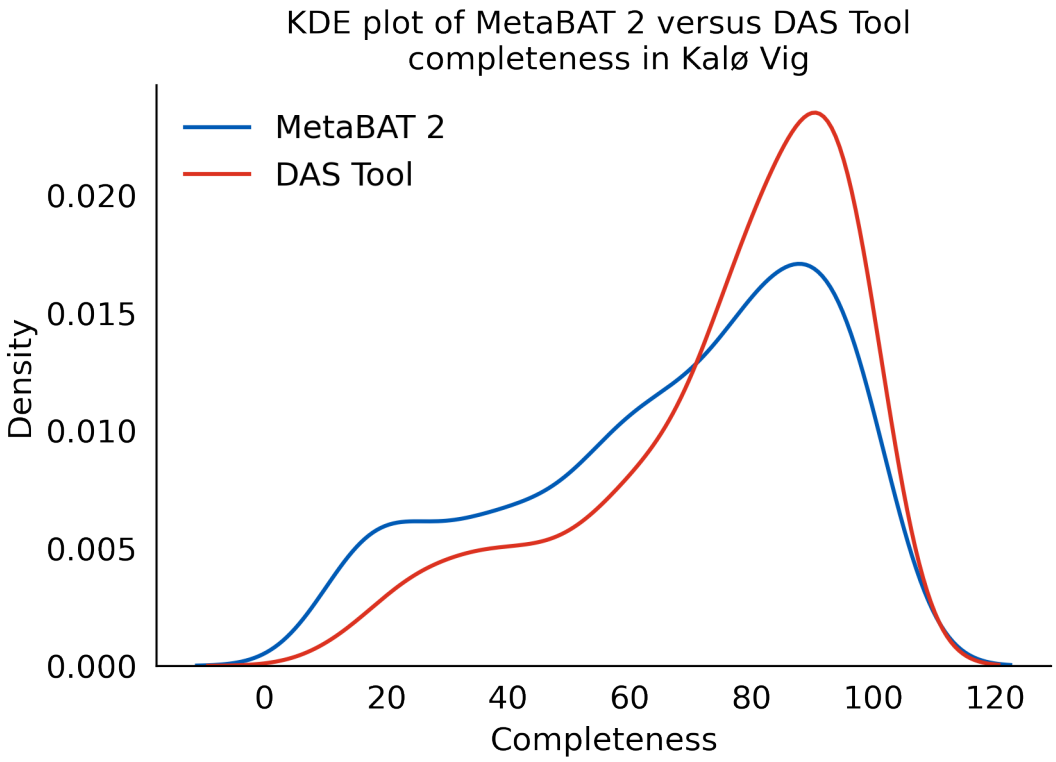


(a) KDE plot of contamination values in Kalø Vig.

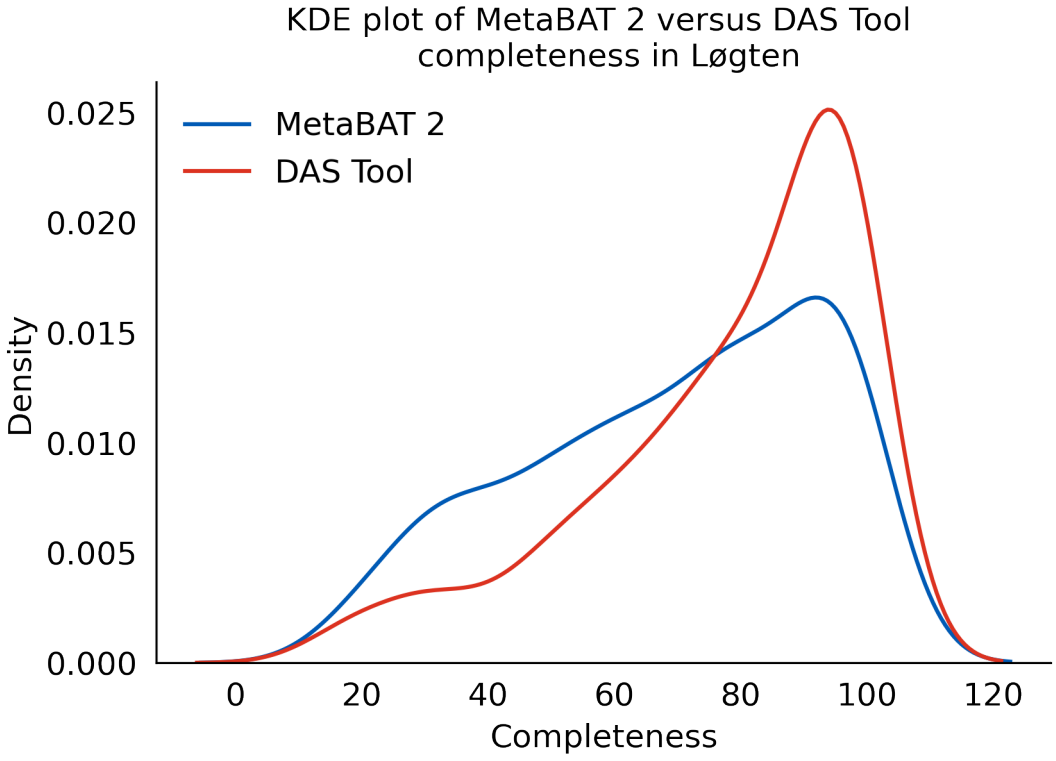


(b) KDE plot of contamination values in Løgten.

Figure 2: KDE plots of contamination values.



(a) KDE plot of completeness values in Kalø Vig.



(b) KDE plot of completeness values in Løgten.

Figure 3: KDE plots of completeness values.

3.2 Taxonomic classification and profiling

The taxonomic assignment of the genomic bins revealed that most bacteria in the three samples are Gram-negative phyla of Bacteroidota and Proteobacteria (also Pseudomonadota from 2021). Bacteroidota constitutes the greatest number in the Kalø Vig sample (68/251), while the Løgten and *Ca.* Electrothrix communis RB samples are dominated by Proteobacteria (59/185 and 22/48, respectively). In the Løgten sample, the Bacteroidota phylum accounts for 39/185 species, and in the *Ca.* Electrothrix communis RB sample for 9/48 species. Finally, in the Kalø Vig sample, Proteobacteria account for 61/251 species. Tables 3, 6 and 9 show classification, completeness, contamination and abundance of selected genomic bins. For full classification results, refer to the tables `kaloevig_genes.csv`, `loegten_genes.csv`, and `marine_gs_illumina_genes.csv` on the project's GitHub page (directory `csv/genes`).

Using the depth files generated by MetaBAT 2, taxa abundances were computed to use them as another criterion in the shortlisting of taxa of interest. The cable bacteria are among the most abundant species in all three samples, which confirms that they are satisfactory for the analysis of cable bacteria-related communities. In the Kalø Vig sample, the genus *Electronema* is the third and the seventh most abundant species, while *Electronema* is the fifth. Similarly, in the Løgten sample, *Electronema* is the fourth species, and in the enrichment *Electronema* is the sixth.

Interestingly, shortlisted for interaction genomic bins represent almost half of the most abundant bins in all three samples (9/20 in Kalø Vig, and 8/20 in Løgten and enrichment). Furthermore, the first six most abundant bins in the enrichment are shortlisted. Figs. 4 to 6 display the abundances of the shortlisted bins and highlight the bins which have an abundance > 65.

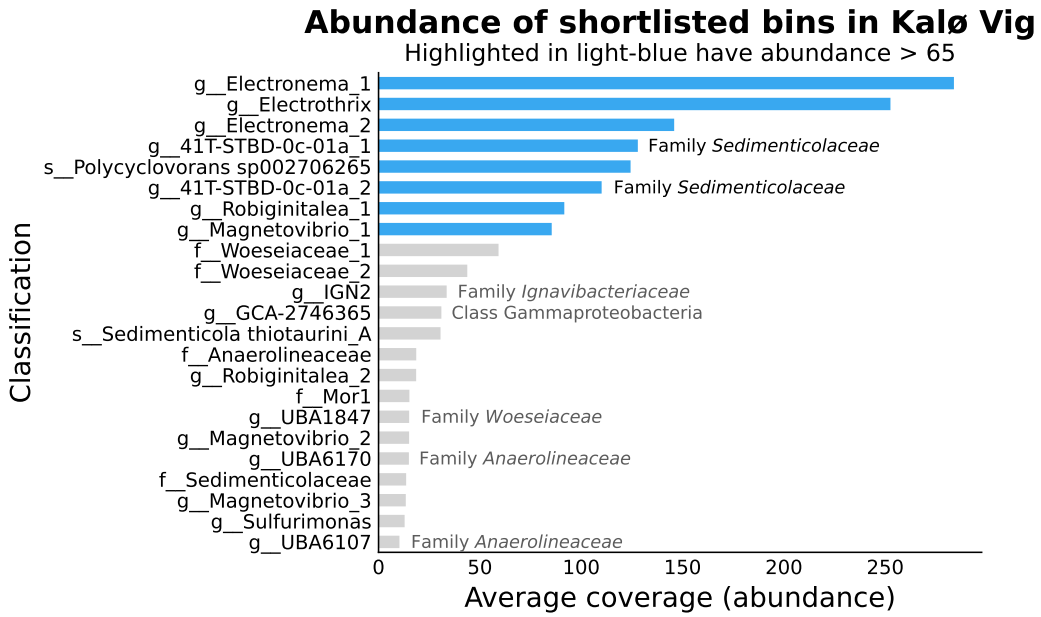


Figure 4: Abundance of shortlisted bins in Kalø Vig.

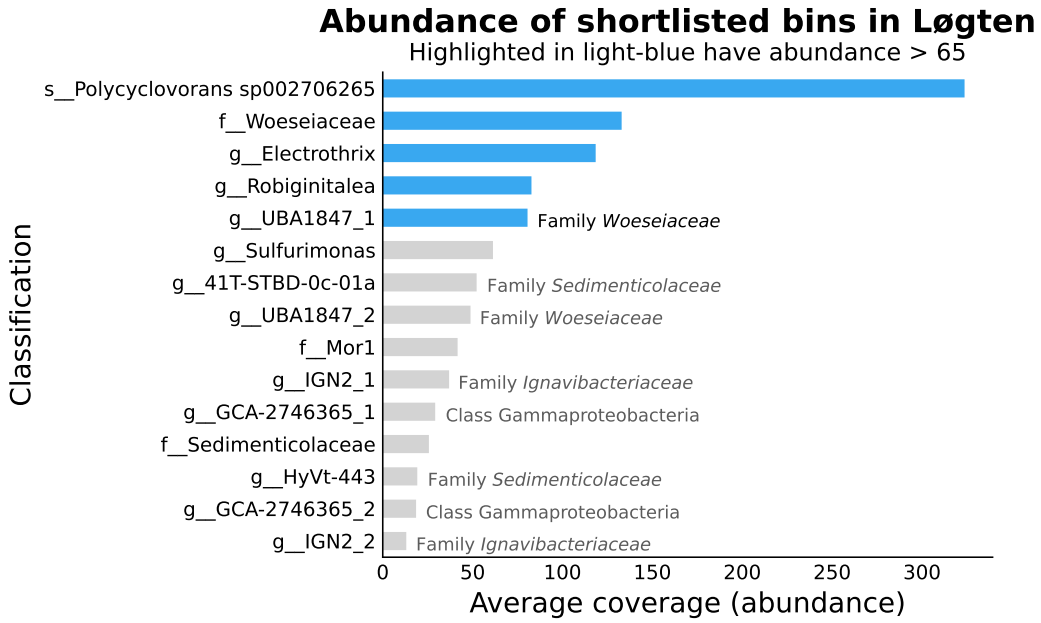


Figure 5: Abundance of shortlisted bins in Løgten.

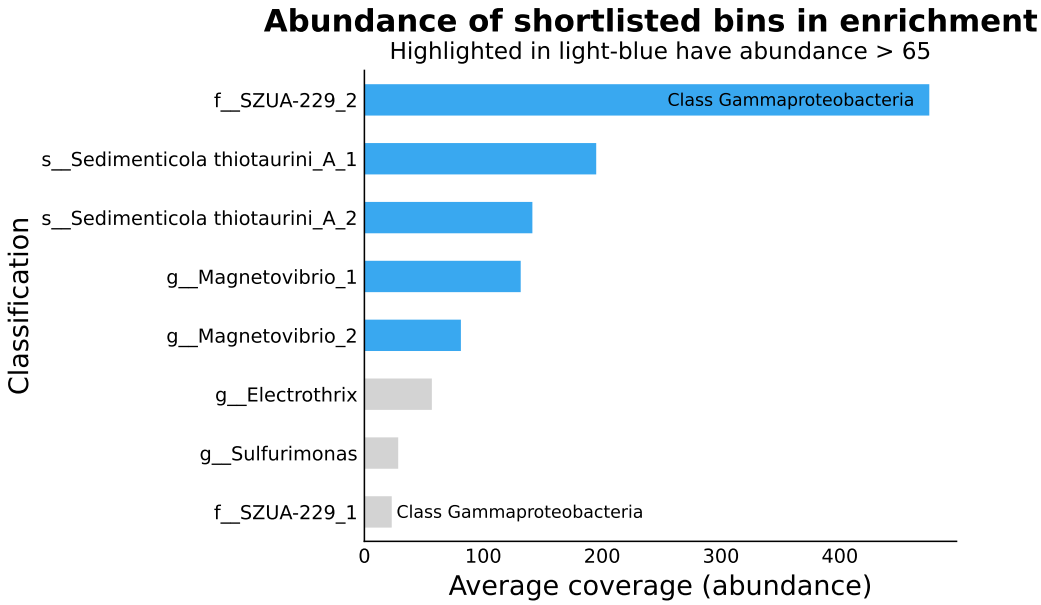


Figure 6: Abundance of shortlisted bins in the *Candidatus* Electrothrix communis RB enrichment.

Figs. 7 to 9 show twenty most abundant genomic bins in each of three samples and highlight the bins that were shortlisted for the interaction with cable bacteria.

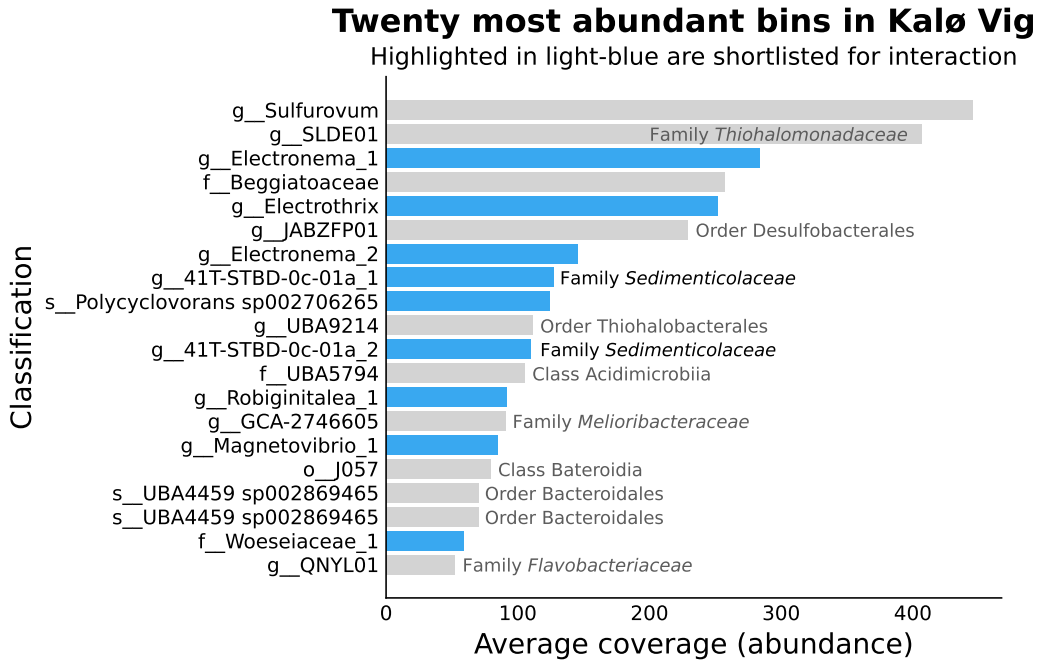


Figure 7: Twenty most abundant bins in Kalø Vig.

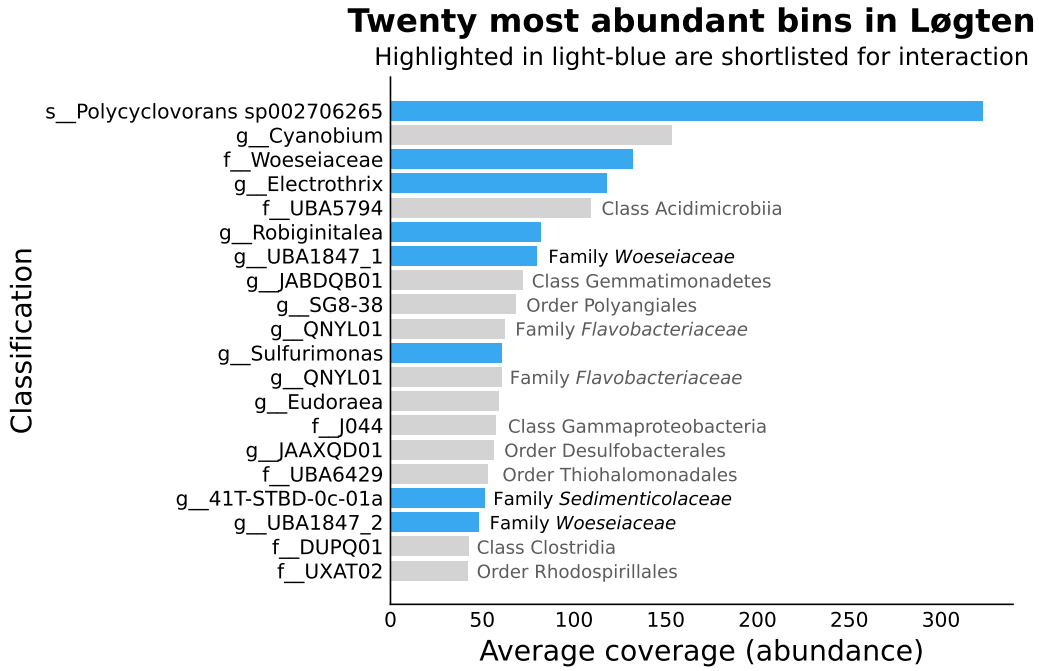


Figure 8: Twenty most abundant bins in Løgten.

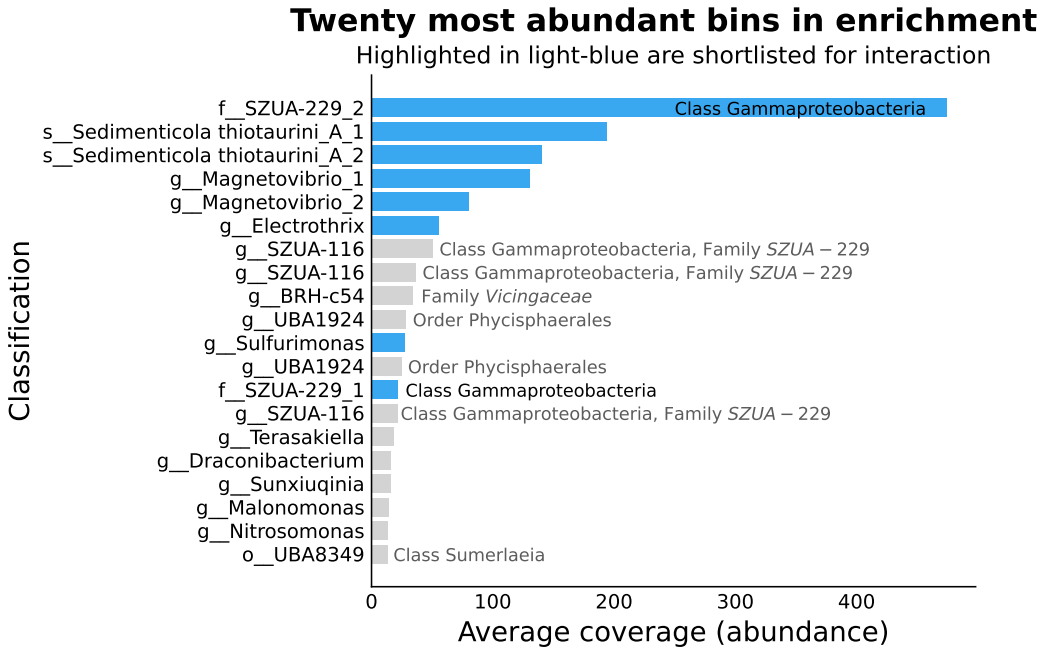


Figure 9: Twenty most abundant bins in the *Candidatus* Electrothrix communis RB enrichment.

Ten families (*Cyclobacteriaceae*, *DSM-19610*, *Desulfobulbaceae*, *GCA-2748-055*, *Magnetovibrionacea*, *Pontiellaceae*, *Rhodobacteraceae*, *SZUA-229*, *Sedimenticolaceae*, and *Sulfurimonadaceae*), and five genera (*Electrothrix*, *GCA-2748055*, *Magnetovibrio*, *Pontiella*, *Sulfurimonas*) appear in all three samples, while no identical species are present in them. This is probably due to the fact that GTDB-Tk is not powerful enough to classify many taxa up to the species level. This tool uses bin ANIs to reference genomes in the GT database for classification, therefore, its inability to more granular classification may be due to the incomplete database or incomplete genomic bins. However, the *DSM-19610* family refers to the *Thiogranum longum* species in the German Collection of Microorganisms and Cell Cultures.

If only the twenty most abundant species are used, the list shrinks to only two families: *Desulfobulbaceae* which is the family of cable bacteria, and *Sedimenticolaceae*, whereas only *Electrothrix* was found in all three samples.

Interestingly, the species *Polycyclovorans sp002706265* stays at the top of the Løgten list and is the ninth species in the Kalø Vig sample. Similarly, the family *Woeseiaceae* is the third species in Løgten and the nineteenth in Kalø Vig. However, none of these taxa could be found in the enrichment of cable bacteria.

Finally, a new species of *Electronema* was found in the Kalø Vig sample.

3.3 Phylogenetic tree

The phylogenetic tree of the selected species (Fig. 10) demonstrates a split into two groups: a smaller one belonging to the Chloroflexi phylum (most belonging to the family *Anaerolineaceae*) and a bigger one comprising eight

clear clusters. Similar taxa from different samples cluster together, and no sample-specific groupings can be observed.

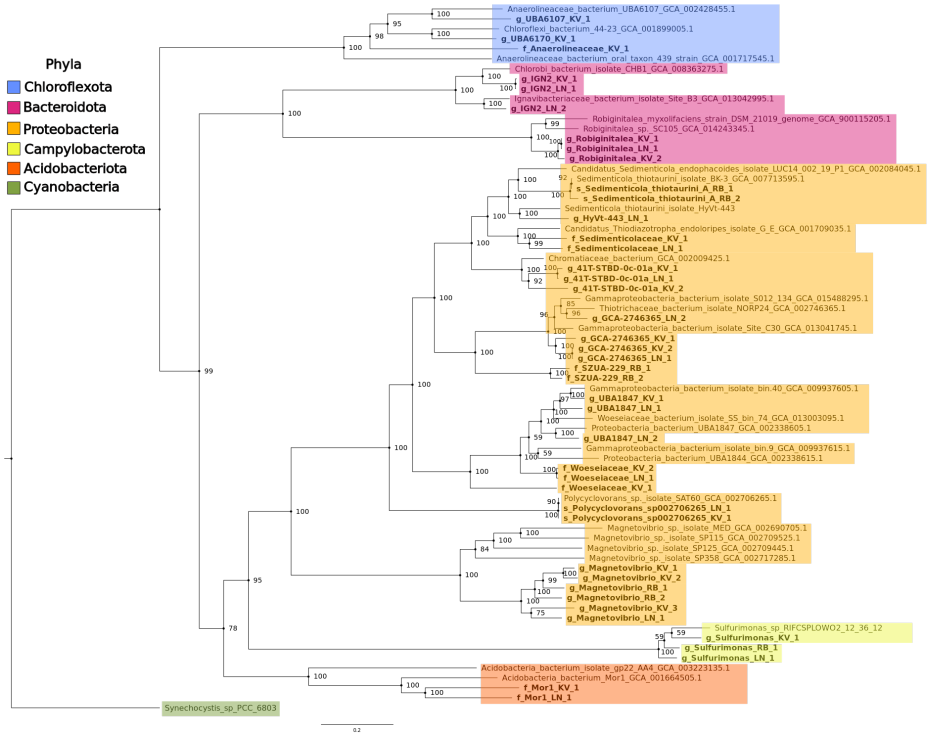


Figure 10: Maximum likelihood phylogenetic tree of shortlisted taxonomic groups with their closest and most complete relatives from GTDB. The tree was built with IQ-TREE 2.2.0.3 with 1000 bootstrap replicates to evaluate the reliability of the tree. The acronym KV corresponds to Kalø Vig, LN to Løgten, and RB to *Candidatus* Electrothrix communis RB. All groups are contained within five phyla (Chloroflexota, Bacteroidota, Proteobacteria, Campylobacterota, Acidobacteriota), with the majority in Proteobacteria. Cyanobacterium *Synechocystis* is an outgroup. In bold are shortlisted species, while the rest are related species from GTDB.

3.4 Functional profiling

3.4.1 BLAST

Six different proteins (OmcS, OmcB, OmcC, OmcZ, MtrC, PioA), involved in extracellular electron transfer were searched for in the bins of the three samples. At present, two main organisms are used for studies of this process: *Geobacter sulfurreducens* and *Shewanella oneidensis*, thus most published scientific papers in the field of EET revolved around these two species. Genomic bins were transformed into protein sequences to allow a more pinpointed search. However, they will still be referred to as bins in the rest of the thesis.

OmcS plays an important role in electron transfer to insoluble Fe(III) oxides and Mn(IV) oxide [12] and is involved in DIET in *G. sulfurreducens* [13]. OmcB in *G. sulfurreducens* is involved in the electron transfer to Fe(III) [14]. OmcC is a part of the OmcCB complex but is probably not involved in the

dissimilatory reduction of Fe(III) as previously thought [14]. OmcZ may constitute nanowires in *G. sulfurreducens* [15]. MtrC is a part of the MtrCAB complex responsible for the transfer of electrons across the outer membrane [16] and is involved in the reduction of Fe(III) [17]. Finally, PioA is necessary for iron oxidation in *Rhodopseudomonas palustris* [18].

In summary, OmcB and OmcC are very similar proteins [14], but their hits overlap only in a minor fashion (in the enriched sample, there are no overlaps). Interestingly, OmcS and PioA tend to appear together in a few taxonomic groups (i.e., classes Krumholzibacteria and Mor1), although these groups are different in the enrichment. Similarly, MtrC tends to appear with PioA in, for example, family *Woeseiaceae*. OmcZ does not have many hits (and no hits in the enrichment). This protein appears only once in the *Mor1* family, but it does not appear in the same family in the Kalø Vig sample, probably due to the lower completeness of the latter in the Kalø Vig sample (87.93% in Kalø Vig against 94.83% in Løgten).

The heatmaps of protein hits (Figs. 11 to 13) visually demonstrate the above summary, and tables Table 5 and **????????** contain raw data of the heatmaps. If certain genomic bins were classified identically (i.e., have the same genus, family, etc.), they were given consecutive numbers after one underscore (i.e., _1, _2, etc.). These numbers correspond to the numbers given to different bins in the phylogenetic tree. Cable bacteria are not included in the phylogenetic tree; however, they are present in the heatmap. These, if duplicated, also have the consecutive numbers which correspond to the order (by abundance) they appear in the tables. Finally, if the classification was unique, the numbers are omitted.

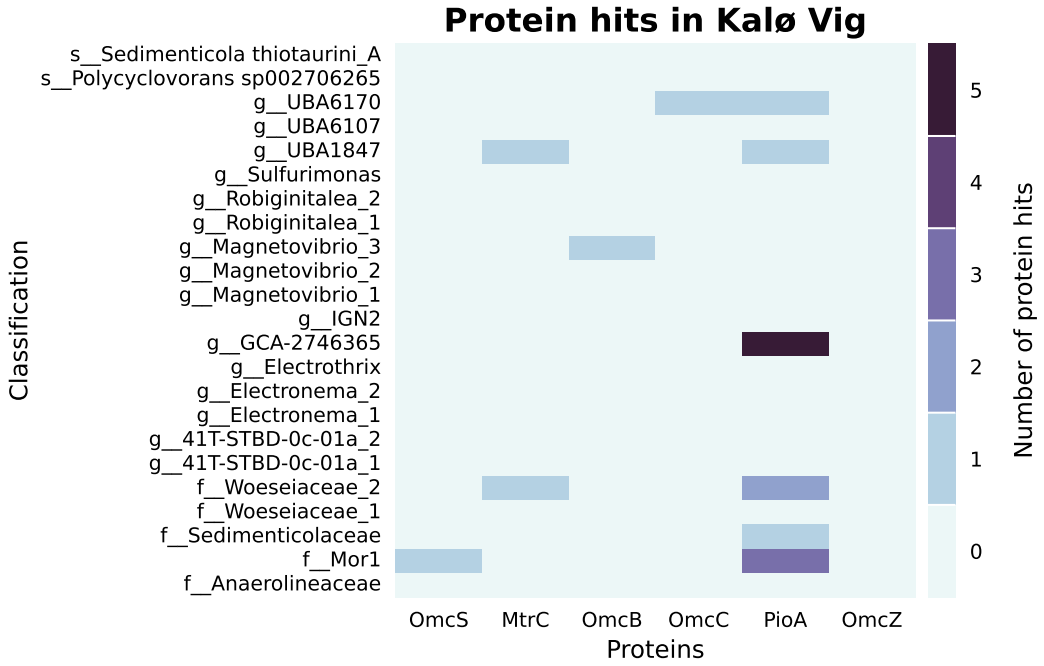


Figure 11: Protein hits in Kalø Vig.

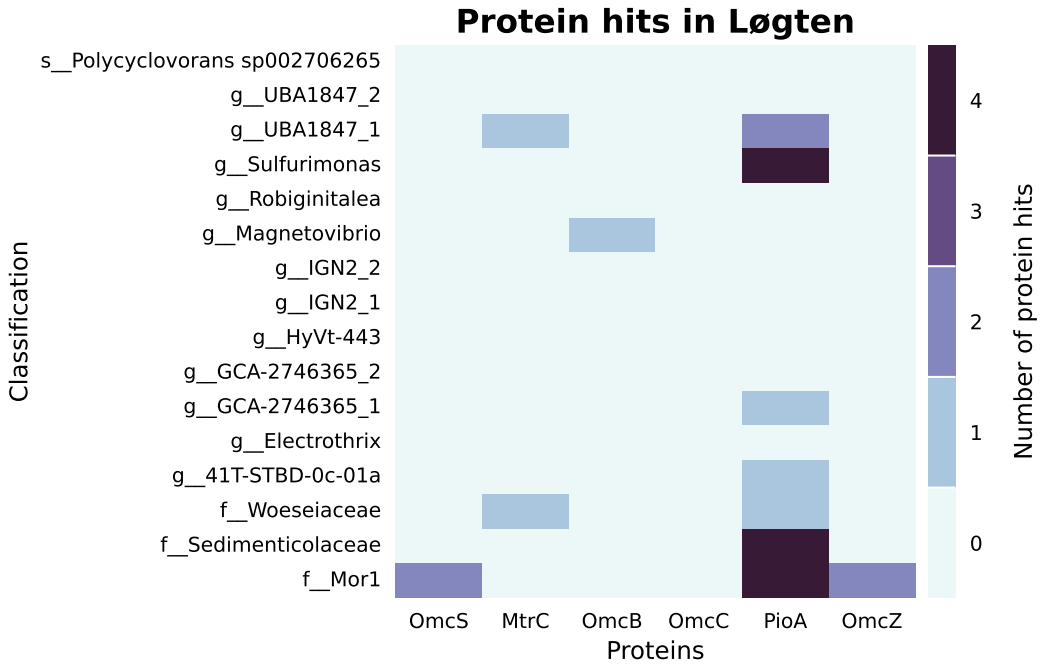


Figure 12: Protein hits in Løgten.

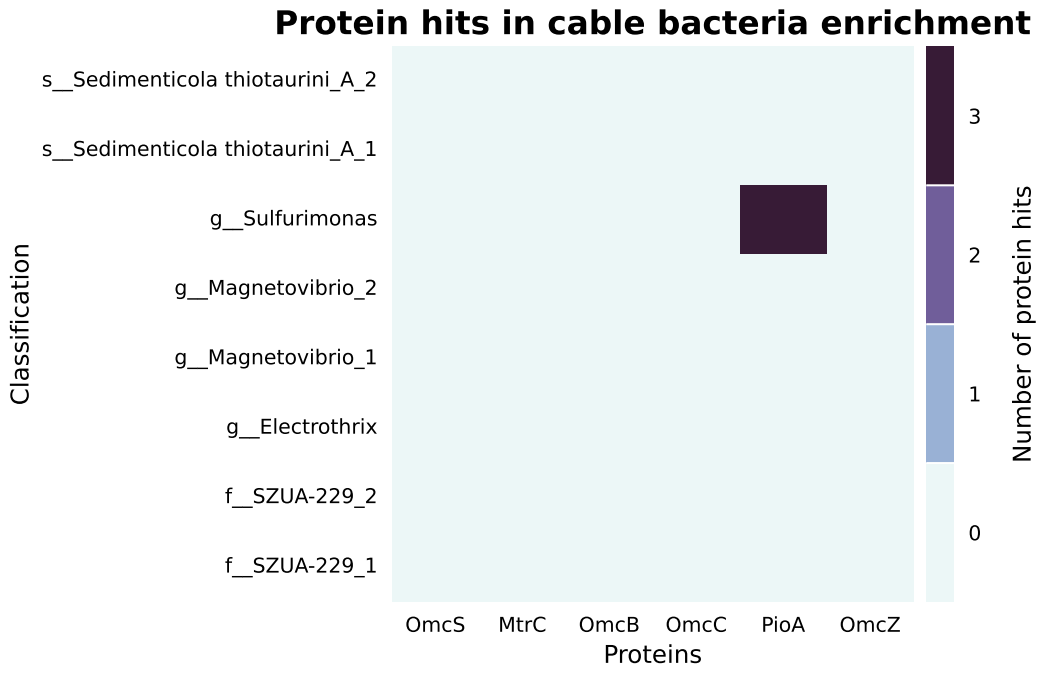


Figure 13: Protein hits in *Candidatus* Electrothrix communis RB enrichment.

3.4.2 FeGenie

The total number of protein hits related to iron oxidation was 66 (1.63%) and 55 (1.64%) in Kalø Vig and Løgten, respectively. Hits related to iron reduction

accounted for a substantially higher number of hits: 1005 (24.87%) and 845 (25.27%). Tables 4, 7 and 10 show the numbers of found genes related to iron reduction and oxidation in the shortlisted bins.

3.4.3 InterProScan

In total, only 14 (5.6%) and 10 bins (5.4%) contain either DsrA or DsrB proteins in the Kalø Vig and Løgten samples, respectively. DsrA and DsrB proteins were found in all cable bacteria species across the two samples. Among the shortlisted bacteria, five bins contained Dsr proteins in Kalø Vig, and four in Løgten. The *Sedimenticolaceae* family appeared in both samples (two bins in each of them). Interestingly, this family appeared also in the *Ca.* Electrothrix communis RB sample two times, both times as the *Sedimenticola thiotaurini_A* species which also appears (one time) in the Kalø Vig sample.

3.5 Fluorescence in situ hybridization

FISH generated good results. For example, cable bacteria can clearly be seen in big amounts in all six slides used in the experiment. The filaments resulted in being more than one millimeter long. However, only one slide was certain to contain Epsilonproteobacteria next to cables because the signals for Epsiloproteobacteria on the other slides did not appear under the examination of the universal bacterial probe. Moreover, there appeared many other bright and large yellow signals that were not consistent with a single species of bacteria; one such spot is visible in Fig. 14. Fig. 14 shows possible signals for *Sulfurimonas* sitting on cable bacteria filaments.

4 Discussion

4.1 Integrative approach improves genome binning

It was not known how MetaDecoder performs compared to the other binning tools. Moreover, it was also not known if the different combinations of the tools would substantially change the DAS Tool results. Thus, it was decided to separately run DAS Tool on the bins generated by the tools (including and excluding MetaDecoder) to evaluate its impact and the different tool combinations on the final results.

The comparison between these two approaches did not reveal a clear difference that would allow clearly differentiating between a better and a worse one. Even though, at first glance, the results of data integration with the inclusion of MetaDecoder, demonstrated better results (for example, the number of bins in both samples was higher, and the average number of contigs was lower), a more detailed investigation revealed that in reality, some bins were of higher quality in the setup without MetaDecoder.

The metrics of contamination and completeness were adopted to further improve the quality of the final bins. The main scope of the project was the search for EET proteins, thus, the bins had to be as pure as possible to

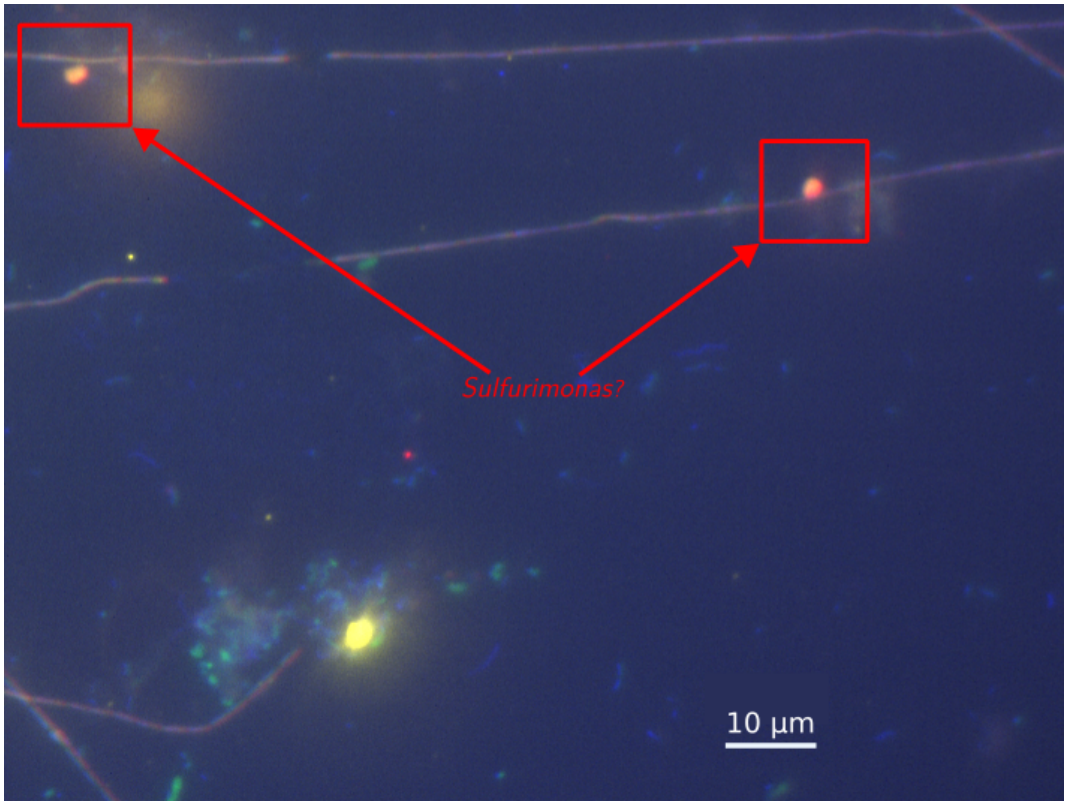


Figure 14: FISH slide. Epsilonproteobacteria close to cable bacteria are indicated by red square boxes. Cable bacteria are long filaments. A large yellow spot not consistent with a single species of bacteria is visible in the left bottom part of the figure.

reduce the number of false positives. At the same time, incomplete bins would lack sensitivity. Thus, the other metrics (N50, L50, total number of contigs, maximum contig length, and total contig length) were excluded as they would only increase the workload and would not provide additional information useful for the bin refinement with Python scripts.

The number of recovered bins was high, and their taxonomy was suspected to be imprecise and poorly described (as was revealed later); thus, the contamination metric was given priority over the completeness. The rationale behind this decision is that the bins in which EET proteins would be found should only point to the taxa which contain these proteins with a very high probability, thus reducing the BLAST error rate. Consequently, this choice would produce a shorter and more manageable list of selected taxa for a more in-depth analysis. The bin completeness can consistently be increased by using more advanced sequencing techniques, such as long-read sequencing in combination with a higher sequencing depth by more traditional approaches (i.e., Illumina short-read sequencing). However, in this case, for some genomic bins, the increased sequencing depth may not be useful as they already have very high sequencing depths (i.e., > 50X), after which bin completeness does not increase further, but this increase may be useful for genomic bins with lower depths. Nonetheless, the better solution here would be long-read sequencing

because longer reads span highly repeated regions, resulting in much better genome reconstruction.

It is crucial to note that, generally, completeness is overestimated because an incorrectly clustered contig may contain a single-copy gene used to estimate completeness, and thus it increases completeness, while, in reality, this is a sign of contamination. Moreover, these estimates are not completely correct because the knowledge of the microbial world is very limited, and only few single-copy genes are used to estimate completeness. Concurrently, the presence of all single-copy genes in a genomic bin does not mean that this genome is complete, which would require a confirmation that it is circular.

Nevertheless, this procedure resulted in a relatively high number of genomic bins, many of which were of high quality (30% in Kalø Vig and 32% in Løgten). Thus, an integrative approach proved to be highly efficient in genome binning.

However, it is not flexible enough to be applied to any metagenomic project as it requires a lot of parameter tuning and custom scripts. In addition, the choice of the tools to generate bins for DAS Tool (and potentially other integrative tools) is not straightforward, and at least in the context of this project, there was no "best" tool combination.

To make this procedure easily manageable by research groups (especially those not having bioinformaticians), a workflow generator tool, such as Snakemake, could be employed. Then, only configuration files should be customized to set tool parameters and input files for the project, as the rest of the job can be handled automatically by the workflow. Moreover, this workflow can also be made smoothly scalable by allowing users to select a flexible number of binning algorithms to run under the hood. In this project, only two combinations of tools were tested, and neither of them outperformed the other, thus it would be a good idea to comprehensively test different combinations using the same approach described by Yue et al. [1]. It is also worth noting that there may not be the best solution for every dataset as all environmental samples vary considerably, and the CAMI datasets, despite being a golden standard for the assessment of metagenomic tools, are artificial and may not (and were not intended to) capture the whole complexity of the microbial world.

4.2 Taxonomic and functional profiling revealed potential partners and competitors of cable bacteria

Apart from the BLAST procedure for protein screening, also FeGenie was used in order to find relevant proteins involved in iron oxidation and reduction and confirm the BLAST results in relation to these two categories. However, BLAST allowed a more zoomed-in view of the contents of the bins and was more specific in regards to this screening. Additionally, InterProScan was run to search for DsrA/B proteins to investigate which bacterial groups may be involved in the generation of the main substrate of cable bacteria.

Some of the proteins were not found consistently in the same taxa across samples; this may be attributed to the fact that some genomic bins were less complete than others, so the necessary sequences were not present in them.

4.2.1 *Magnetovibrio*

The *Magnetovibrio* genus (class Alphaproteobacteria) is constituted of magnetotactic species capable of so-called magnetotaxis, causing them to orient along the lines of a magnetic field. This genus was found in all three samples, and magnetotactic bacteria require ferrous iron, Fe(II), to form magnetosomes, while cable bacteria acidify the anoxic environment, which promotes the dissolution of iron sulfide, FeS, which releases Fe(II) and also contributes to the formation of hydrogen sulfide, H_2S, the main substrate of cable bacteria. Thus, *Magnetovibrio* can greatly benefit from the presence of cable bacteria in marine sediments. Moreover, a near-complete *Magnetovibrio* bin (in Løgten, completeness of 94.14%) contains the OmcB protein involved in the reduction of Fe(III). Furthermore, proteins for iron reduction were found with FeGenie in all three samples. For example, proteins of the Dmk group are located in the EET locus of the pathogen *Listeria monocytogene* [19]). This indicates a possibility that this genus transfers electrons to extracellular sinks, which can also be cable bacteria. Additionally, iron oxidation proteins were also found, although Fe(II) is not used as an electron donor [20]. *Magnetovibrio* were also found to use sulfide and thiosulfate (which is a result of the H_2S oxidation) [20] as electron donors (and either oxygen or nitrous oxide as electron acceptors), thus possibly competing or using the immediate products of the cable bacteria metabolism. Finally, in a study concerning the degradation of petroleum hydrocarbons through the use of snorkels and cable bacteria (by Marzocchi et al. [21]), this genus was considerably enriched in the treatments possessing cable bacteria compared to the treatments where their growth was inhibited. Nonetheless, it is not clear whether this taxon is able to use iron compounds as electron acceptors or donors, although it is possible as they use iron extensively to grow magnetosomes. In summary, *Magnetovibrio* are potentially able to use iron derivatives as electron acceptors and donors, and thus further investigation is required to understand this process as well as their potential to use cable bacteria as electron sources and sinks. For example, one can deplete their confirmed donors and acceptors and observe their growth with cable bacteria. However, an experiment should be carefully designed to eliminate the possibility of using iron for electron exchange as it is also vitally required for magnetosome production.

4.2.2 *Sedimenticolaceae*

Four, three, and two representatives of the *Sedimenticolaceae* family (class Gammaproteobacteria) were observed in the Kalø Vig, Løgten, and *Ca.* Electrothrix communis RB samples. This family was found to contain the *dsrAB* gene, necessary for the sulfite reduction [22], and thus involved in the sulfide (H_2S) production making this species beneficial for cable bacteria. This family also contains the *soeABC* gene, necessary for sulfite oxidation, potentially enabling it to take up electrons from cable bacteria. Moreover, in the cable bacteria enrichment, two microbes (*Sedimenticola thiotaurini_A*) of this family are the second and the third most abundant species, and it is also in the Kalø Vig sample in a relatively high abundance (37^{th} position out of 251 bins). In

addition, a species of Sedimenticola (*Sedimenticola selenatireducens*) oxidative protein type which reverses the traditional pathways of sulfite reduction [23], further suggesting that this family may be involved in sulfite oxidation.

Next, a protein for Fe(III) oxidation (PioA) is present in the *41T-STBD-0c-01a* genus of this family in the Løgten sample. However, this protein was observed neither in the less complete bins in the Kalø Vig samples nor in the identically complete (*Sedimenticola thiotaurini_A*) species in the same sample, making it wonder whether this protein appears specifically in the *41T-STBD-0c-01a* genus (as *Sedimenticola thiotaurini_A* belongs to the *Sedimenticola* genus). Finally, it is interesting to note that only proteins of iron reduction were reported (with the exception of one bin in the cable bacteria enrichment, which has very low completeness of around 10%) by FeGenie but not those of iron oxidation in all three samples.

4.2.3 *Anaerolineaceae*

Three members of the *Anaerolineaceae* family (phylum Chloroflexota) were observed in the Kalø Vig sample, and one of them is a nearly-complete bin (93.1%) contains a PioA protein as well as an OmcC protein. The latter is a part of the OmcCB complex, but OmcB was not observed. This complex is involved in the electron transfer to Fe(III), thus enabling it to transfer electrons extracellularly. Moreover, in [9] many positive interactions were found between Chloroflexi (the phylum of this family) and cable bacteria populations. Further confirmation for the presence of proteins contributing to electron transfer was obtained by FeGenie (as previously with *Magnetovibrio* Dmk proteins were found). Finally, in the study by Marzocchi et al. [21], this family was the second most enriched across all treatments performed in the research. However, the highest prevalence of *Anaerolineaceae* was observed in the "snorkel" treatment in which the growth of cable bacteria was inhibited. It is, therefore, not clear how this family may interact with cable bacteria because, despite it containing proteins for EET, it thrived in treatment without cable bacteria; therefore, the competition potential of this family may be studied by checking, for instance, its ability to oxidize sulfide.

4.2.4 *Mor1*

The family *Mor1* belonging to the Acidobacteriota phylum contains proteins OmcS, PioA, and OmcZ in both Kalø Vig and Løgten samples (even though OmcZ is not present in Kalø Vig probably due to lower completeness). The proteins OmcS and OmcZ enable this family to transfer electrons through nanowires as both these cytochromes generate them. OmcZ was found to be 1000-fold more conductive than OmcS, and its growth can also be stimulated by an electric field [15]. It would be interesting to study whether cable bacteria and *Mor1* cooperate with the former, producing an electrical field that stimulates the growth of nanowires and the latter donating electrons to cable bacteria. FeGenie found further confirmation as the bins of *Mor1* in Kalø Vig and Løgten contain the proteins OmcS and OmcZ as well as the proteins of the Dmk and Fmn groups involved in iron reduction. Next, Liu et al. [9] discov-

ered positive interactions of Acidobacteriota with cable bacteria (even though Mor1 was not specifically mentioned. Furthermore, many Acidobacteriota are acidophilic, aerobic bacteria (at least, in the class Acidobacteria) [24], thus they may benefit from the presence of cable bacteria in the anoxic zone of marine sediments, as they would simultaneously live in an acidic environment and be able to donate electrons in the absence of oxygen. However, the Mor1 class has not been studied extensively, and the literature on its properties could not be found.

4.2.5 *Woeseiaceae*

The family *Woeseiaceae* within the Gammaproteobacteria class occurs globally and seems to be at the core of most marine sediments. It is capable of employing a wide range of physiological mechanisms to adapt to diverse marine sediments environments [25]. This family was observed three times in the Kalø Vig and Løgten samples. Most of these observations are abundant (for example, in Løgten, this family is the third most abundant). This family contains the PioA and MtrC complexes (whose proteins appear together in Kalø Vig and Løgten), thus enabling it to oxidize and reduce iron. Moreover, the genus *Woeseia* was previously found (using co-occurrence networks) to co-occur with cable bacteria [2]. Even though the co-occurrence network methods are not very efficient in establishing microbial interactions [3], the combination with the genes found by BLAST involved in EET provides a good indication of possible interaction of this family with cable bacteria. Additionally, FeGenie found MtrA (periplasmic *c*-type cytochrome, homolog of PioA), and MtrB (integral outer-membrane β-barrel protein) genes in two of the bins (`metabat_bin.455_sub` and `metabat_bin.51`) in Kalø Vig, also involved in iron reduction as well as confirmed the presence of MtrC in the same bin indicating potential in electron transfer to cable bacteria from the species of this family. The same genes were also found in two bins of the Løgten sample (`metabat_bin.378` and `maxbin_bin.17`), further confirming the hypothesis of electron exchange.

4.2.6 *Polycyclovorans*

Polycyclovorans is a novel genus within the *Nevskiaceae* family (class Gammaproteobacteria), one species of which was isolated from a marine diatom *Skeletonema costatum* [26]. This genus was reported to participate in the degradation of PAHs. These hydrocarbons result from burning fossil fuels but also constitute up to 35% of crude oil. The degradation of petroleum hydrocarbons requires electron acceptors whose availability in marine environments is restricted. However, cable bacteria are able to connect the oxic and anoxic zones, thus stimulating the degradation [21]. Thus, it is not surprising that the species *Polycyclovorans sp002706265* thrives in the Kalø Vig and Løgten samples (ninth most abundant species in the former and the most abundant species in the latter). However, no *Polycyclovorans* was observed in the marine cable bacteria enrichment. In addition, EET proteins were not found by BLAST, but the proteins of the Dmk and FMN groups (involved in iron reduction) were detected by FeGenie in both samples. In conclusion, it is more

probable that this species greatly benefits from the presence of cable bacteria but does not perform any electron exchange with them.

4.2.7 *Sulfurimonas*

The *Sulfurimonas* genus within the Campylobacterota (or Epsilonproteobacteria) class is a widespread and physiologically diverse group of species. Its representatives were found mainly in sulfidic environments of marine sediments but also more extreme conditions such as deep-sea hydrothermal vents. This group is capable of using different electron donors such as sulfide, sulfite, and thiosulfate and electron acceptors (oxygen, nitrate, and nitrite), although not all substances can be used by all isolated species [27]. Furthermore, as oxygen and nitrate are absent deep in marine sediments, it is possible that this genus uses particulate manganese oxide, MnO_2, as an electron acceptor [28]. The capacity of using sulfide, nitrate, nitrite, and oxygen relates this group to cable bacteria, thus causing them to be potential competitors. At the same time, the use of thiosulfate may be beneficial for *Sulfurimonas* (similar to *Magnetovibrio*) as it is an immediate product of sulfide oxidation. Moreover, it is possible to think about cable bacteria as electron sinks of *Sulfurimonas* in the absence of their common electron acceptors as cable bacteria outcompete this genus in their usage. Marzocchi et al. [21] reported on finding an enrichment of *Sulfurimonas* in the Cable bacteria + Snorkel treatment, but it was nearly identical to the abundance (and lower) in the Snorkel treatment (which did not contain any cable bacteria). Furthermore, in a study of carbon uptake by Vasquez-Cardenas et al. [10] Epsilonprotebacteria (including Sulfurimonas were found to persist throughout oxic and suboxic zone and thrive in the presence of cable bacteria. As mentioned in the Introduction, it is plausible that these bacteria use cables as electron sinks to access oxygen indirectly. In that study, it was also observed that a cut below a certain depth halts the carbon uptake by the microbial community surrounding cable bacteria, providing more evidence to the hypothesis of oxygen access.

Two out of three samples (Løgten and *Ca.* Electrothrix communis RB) contain the PioA protein necessary for iron oxidation, but this protein was not observed in a nearly complete genome in the Kalø Vig sample, so its real presence is not fully confirmed. FeGenie observed the Dmk protein group (required for the iron reduction) in all samples, even though iron had not been reported as an electron acceptor for this taxon. Nevertheless, it may be speculated that cable bacteria play a role as their electron acceptor.

Finally, it was shown that one of the *Sulfurimonas* strains (CVO) strongly expresses type IV pilin-like proteins during the sulfur oxidation [29] which is often observed to contribute to surface adhesion and gliding mobility. As DIET requires a strong attachment of cells, this pilin may play a role in this process.

4.2.8 *SZUA-229*

The *SZUA-229* family belonging to the Gammaproteobacteria class is poorly described and was only recovered from a cold seep area in the South China Sea [30], and it was the most represented taxonomic group. Some representatives

of this family were found to possess some of the necessary for dissimilatory sulfate reduction, thiosulfate oxidation, and dissimilatory nitrate reduction KEGG ontologies such as dsrB (KEGG entry K11181), soxA (L-cysteine S-thiosulfotransferase, KEGG entry K17222), and narI (nitrate reductase gamma subunit, KEGG entry K00374).

PioA protein was found in the Kalø Vig sample in two out of three representatives in the Løgten sample (it was not found in the third one, probably due to its relatively low completeness of 77.43%). Additionally, this family resulted in the most abundant in the *Ca.* Electrothrix communis RB sample by a wide margin, although the PioA protein was not observed (similarly, due to lower completeness of 89.66%).

PioA is necessary for iron oxidation, thus it is worth noticing that one species in Kalø Vig possess the FoxY protein for this process. The same protein is also observed in two species in Løgten but not in the least complete one. Finally, the *Ca.* Electrothrix communis RB sample also has this protein. Thus, it is possible that this family is somehow involved in iron oxidation. Additionally, all species contain proteins for iron reduction (for example, those of the Dmk group), thus it may be speculated that they have an ability to also transfer electrons to other microorganisms.

Nonetheless, this taxonomic group is under-examined, therefore,it is very difficult to formulate any meaningful hypothesis, and a more comprehensive study of this group is required through the application of different gene-searching tools as well as data integration from other omic techniques (i.e., transcriptomics).

4.2.9 *Robiginitalea*

Little is known about the genus *Robiginitalea*, although a few species have been isolated, such as *Robiginitalea biformata* [31] and *Robiginitalea myxolifaciens* [32]. Thus, it is difficult to make any conclusions about their metabolism. They were found to be abundant in Kalø Vig (one bin is 13^th) and Løgten (the bin is sixth). Even though the iron reduction proteins (DmkB and FmnB) were found in both genomic bins, it is early to say about their interaction patterns with cable bacteria, and more investigation should be conducted.

4.2.10 *IGN2*

The *IGN2* genus belongs to the family *Ignavibacteriaceae* which was found to play an important in the iron reduction within plant biomass, thus showing its potential for using iron compounds as electron acceptors [33]. This ability was confirmed in the present project as both representatives found in the Kalø Vig and Løgten samples possess proteins involved in iron reduction (Dmk group). Nonetheless, this group appears to not have other EET proteins even in the near-complete bin in Løgten (completeness of 98.28%). Furthermore, in another paper [34], this family was not described as iron-reducing but rather as nitrite-reducing raising questions about its metabolic adaptability to different environmental conditions.

4.2.11 Novel *Electronema* species

Notably, a new species of *Electronema* found in the Kalø Vig sample, is able to survive in the marine environment despite being denominated as "freshwater". A genome screening was performed and the Na(+)/H(+) antiporter, *NhaA*, was detected (manuscript in preparation). *NhaA* is possibly required to sustain higher osmotic pressure in marine environments.

4.2.12 Dsr protein-containing bacteria

As the primary substrate of cable bacteria, hydrogen sulfide, is mainly produced by dissimilatory sulfate reduction, it was interesting to investigate which bacteria are potentially able to generate sulfide. As DsrA and DsrB proteins are essential for this process, screening for them was performed in InterProScan datasets. Interestingly, many of the found bacteria are shortlisted for interaction, and three of them in the Kalø Vig sample belong to the family of *Sedimenticolaceae*.

In addition, representatives of the Desulfobacterales order were found in both samples; this order is known for its sulfate-reducing exemplars. The order Thiohalomonadales was found to perform sulfur oxidation [35], but no evidence for its sulfate-reducing properties have been observed. Nonetheless, a bin of this taxon in Løgten contains DsrA or DsrB. As pointed out in the *Sedimenticolaceae* section above, the dissimilatory sulfite reduction may be reversed, and in this case, it is possible that this reversion was observed.

4.3 FISH gave indications of interaction between cable bacteria and *Sulfurimonas*

The use of a generic Epsilonproteobacteria probe was justified by the fact that the *Sulfurimonas* genus was the only one identified in the enrichment culture which was not subjected to filtering by completeness and contamination, so all genomic bins detected by MetaBAT 2 remained. Fig. 14 clearly demonstrates that the hypothesis of cable bacteria as electron sinks for this genus may be valid as two strong signals for Epsilonproteobacteria could be clearly observed in close proximity to the filaments of cable bacteria.

5 Conclusions and future research

This project shed more light on the interactions between bacteria of marine sediments and cable bacteria in the context of extracellular electron transfer. High-quality genomic bins were recovered by employing data integration of different binning tools and custom scripts to discover candidate species involved in this process. A more focused approach of searching only EET proteins was used to be more precise in species identification. The project yielded 10 candidate taxonomic groups that may exchange electrons with cable bacteria. The most interesting example is the *Sulfurimonas* genus, whose members may compete (by using common with cable bacteria electron acceptors and donors)

or cooperate with cable bacteria. Finally, FISH results confirmed that the members of this genus could be found in the close proximity of cable bacteria and thus have the potential to exchange electrons with cables. Fig. 15 summarized the hypotheses of cooperation and competition of Sulfurimonas with cable bacteria.

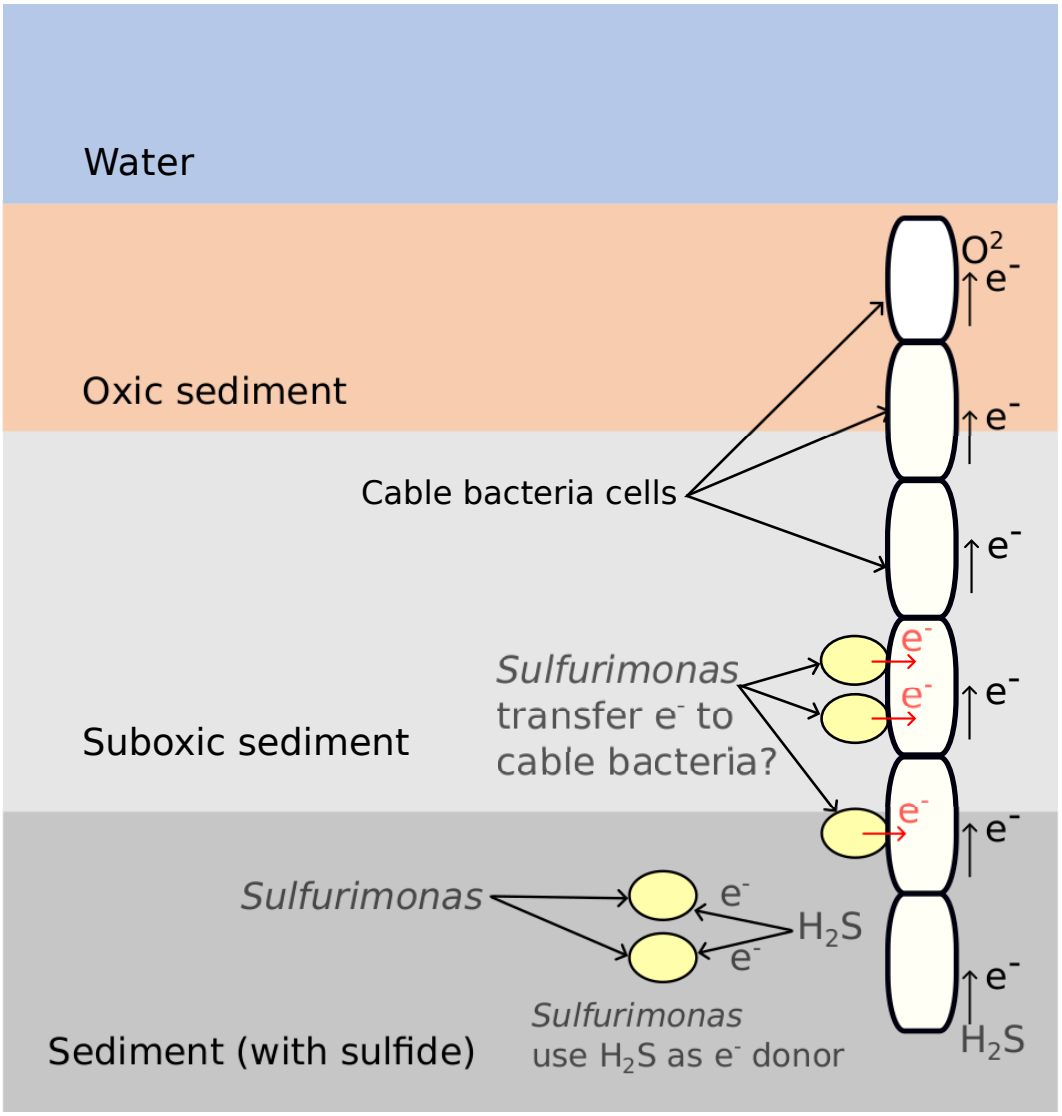


Figure 15: Conceptual diagram demonstrating possible interactions and competition of *Sulfurimonas* with cable bacteria.

Future research may include :

1. FISH on other shortlisted bacterial groups as well as using more focused probes which match exclusively the *Sulfurimonas* genus.
2. Addition of pure cultures of *Sulfurimonas* (for example, *Sulfurimonas gotlandica* into enrichments of cable bacteria and employment of live-cell labeling to see if they attach to cables.
3. When a pure culture of cable bacteria will be available, repetition of the above experiments and application of electron microscopy to observe if

any structure forms between cable bacteria and *Sulfurimonas* or other microbial groups.

4. Knockout of EET genes and observation if bacteria continue to collocate.

6 Code and data availability

Jupyter Notebooks and Python scripts used in this project are available on the project's GitHub page (https://github.com/artur-sannikov/aarhus-internship-code). The page also contains nontruncated unmodified versions of the tables from the Appendix, as well as more additional tables (for example, a list of all species included in the phylogenetic tree). Furthermore, to aid reproducible, the page contains `yml` files to recreate `conda` environments used to run the tools.

References

[1] Yi Yue, Hao Huang, Zhao Qi, Hui-Min Dou, Xin-Yi Liu, Tian-Fei Han, Yue Chen, Xiang-Jun Song, You-Hua Zhang, and Jian Tu. Evaluating metagenomics tools for genome binning with real metagenomic datasets and CAMI datasets. *BMC Bioinformatics*, 21(1):334, December 2020.

[2] Morten Hoppe. Finding cable bacteria associates using co-occurrence networks. Technical report, Aarhus University, June 2020.

[3] Hokuto Hirano and Kazuhiro Takemoto. Difficulty in inferring microbial community structure based on co-occurrence network approaches. *BMC Bioinformatics*, 20(1):329, December 2019.

[4] Christian Pfeffer, Steffen Larsen, Jie Song, Mingdong Dong, Flemming Besenbacher, Rikke Louise Meyer, Kasper Urup Kjeldsen, Lars Schreiber, Yuri A. Gorby, Mohamed Y. El-Naggar, Kar Man Leung, Andreas Schramm, Nils Risgaard-Petersen, and Lars Peter Nielsen. Filamentous bacteria transport electrons over centimetre distances. *Nature*, 491(7423):218–221, November 2012.

[5] M. P. Bryant, E. A. Wolin, M. J. Wolin, and R. S. Wolfe. Methanobacillus omelianskii, a symbiotic association of two species of bacteria. *Archiv fr Mikrobiologie*, 59(1-3):20–31, 1967.

[6] Amelia-Elena Rotaru, Pravin Malla Shrestha, Fanghua Liu, Beatrice Markovaite, Shanshan Chen, Kelly P. Nevin, and Derek R. Lovley. Direct Interspecies Electron Transfer between Geobacter metallireducens and Methanosarcina barkeri. *Applied and Environmental Microbiology*, 80(15):4599–4605, August 2014.

[7] Derek R. Lovley. Syntrophy Goes Electric: Direct Interspecies Electron Transfer. *Annual Review of Microbiology*, 71(1):643–664, September 2017.

[8] Filip J. R. Meysman, Rob Cornelissen, Stanislav Trashin, Robin Bonné, Silvia Hidalgo Martinez, Jasper van der Veen, Carsten J. Blom, Cheryl Karman, Ji-Ling Hou, Raghavendran Thiruvallur Eachambadi, Jeanine S. Geelhoed, Karolien De Wael, Hubertus J. E. Beaumont, Bart Cleuren, Roland Valcke, Herre S. J. van der Zant, Henricus T. S. Boschker, and Jean V. Manca. A highly conductive fibre network enables centimetre-scale electron transport in multicellular cable bacteria. *Nature Communications*, 10(1):4120, December 2019.

[9] Feifei Liu, Zhenyu Wang, Bo Wu, Jesper T. Bjerg, Wenzhe Hu, Xue Guo, Jun Guo, Lars Peter Nielsen, Rongliang Qiu, and Meiying Xu. Cable bacteria extend the impacts of elevated dissolved oxygen into anoxic sediments. *The ISME Journal*, 15(5):1551–1563, May 2021.

[10] Diana Vasquez-Cardenas, Jack van de Vossenberg, Lubos Polerecky, Sairah Y Malkin, Regina Schauer, Silvia Hidalgo-Martinez, Veronique

Confurius, Jack J Middelburg, Filip JR Meysman, and Henricus TS Boschker. Microbial carbon metabolism associated with electrogenic sulphur oxidation in coastal sediments. *The ISME Journal*, 9(9):1966–1978, September 2015.

[11] Morten Hoppe. Detecting microorganisms that interact with cable bacteria using cooccurrences and metagenomes. Master's thesis, Aarhus University, September 2021.

[12] T. Mehta, M. V. Coppi, S. E. Childers, and D. R. Lovley. Outer Membrane *c* -Type Cytochromes Required for Fe(III) and Mn(IV) Oxide Reduction in *Geobacter sulfurreducens*. *Applied and Environmental Microbiology*, 71(12):8634–8641, December 2005.

[13] Zarath M. Summers, Heather E. Fogarty, Ching Leang, Ashley E. Franks, Nikhil S. Malvankar, and Derek R. Lovley. Direct Exchange of Electrons Within Aggregates of an Evolved Syntrophic Coculture of Anaerobic Bacteria. *Science*, 330(6009):1413–1415, December 2010.

[14] Ching Leang, M. V. Coppi, and D. R. Lovley. OmcB, a *c* -Type Polyheme Cytochrome, Involved in Fe(III) Reduction in *Geobacter sulfurreducens*. *Journal of Bacteriology*, 185(7):2096–2103, April 2003.

[15] Sibel Ebru Yalcin, J. Patrick O'Brien, Yangqi Gu, Krystle Reiss, Sophia M. Yi, Ruchi Jain, Vishok Srikanth, Peter J. Dahl, Winston Huynh, Dennis Vu, Atanu Acharya, Subhajyoti Chaudhuri, Tamas Varga, Victor S. Batista, and Nikhil S. Malvankar. Electric field stimulates production of highly conductive microbial OmcZ nanowires. *Nature Chemical Biology*, 16(10):1136–1142, October 2020.

[16] Marcus J. Edwards, Gaye F. White, Julea N. Butt, David J. Richardson, and Thomas A. Clarke. The Crystal Structure of a Biological Insulated Transmembrane Molecular Wire. *Cell*, 181(3):665–673.e10, April 2020.

[17] Brian H. Lower, Liang Shi, Ruchirej Yongsunthon, Timothy C. Droubay, David E. McCready, and Steven K. Lower. Specific Bonds between an Iron Oxide Surface and Outer Membrane Cytochromes MtrC and OmcA from *Shewanella oneidensis* MR-1. *Journal of Bacteriology*, 189(13):4944–4952, July 2007.

[18] Dinesh Gupta, Molly C. Sutherland, Karthikeyan Rengasamy, J. Mark Meacham, Robert G. Kranz, and Arpita Bose. Photoferrotrophs Produce a PioAB Electron Conduit for Extracellular Electron Uptake. *mBio*, 10(6):e02668–19, December 2019.

[19] Samuel H. Light, Lin Su, Rafael Rivera-Lugo, Jose A. Cornejo, Alexander Louie, Anthony T. Iavarone, Caroline M. Ajo-Franklin, and Daniel A. Portnoy. A flavin-based extracellular electron transfer mechanism in diverse Gram-positive bacteria. *Nature*, 562(7725):140–144, October 2018.

[20] Dennis A. Bazylinski, Timothy J. Williams, Christopher T. Lefèvre, Denis Trubitsyn, Jiasong Fang, Terrence J. Beveridge, Bruce M. Moskowitz, Bruce Ward, Sabrina Schübbe, Bradley L. Dubbels, and Brian Simpson. Magnetovibrio blakemorei gen. nov., sp. nov., a magnetotactic bacterium (Alphaproteobacteria : Rhodospirillaceae) isolated from a salt marsh. *International Journal of Systematic and Evolutionary Microbiology*, 63(Pt_-5):1824–1833, May 2013.

[21] Ugo Marzocchi, Enza Palma, Simona Rossetti, Federico Aulenta, and Alberto Scoma. Parallel artificial and biological electric circuits power petroleum decontamination: The case of snorkel and cable bacteria. *Water Research*, 173:115520, April 2020.

[22] Charlotte D. Vavourakis, Maliheh Mehrshad, Cherel Balkema, Rutger van Hall, Adrian-Ştefan Andrei, Rohit Ghai, Dimitry Y. Sorokin, and Gerard Muyzer. Metagenomes and metatranscriptomes shed new light on the microbial-mediated sulfur cycle in a Siberian soda lake. *BMC Biology*, 17(1):69, December 2019.

[23] Albert Leopold Müller, Kasper Urup Kjeldsen, Thomas Rattei, Michael Pester, and Alexander Loy. Phylogenetic and environmental diversity of DsrAB-type dissimilatory (bi)sulfite reductases. *The ISME Journal*, 9(5):1152–1165, May 2015.

[24] Svetlana N Dedysh and Jaap S Sinninghe Damsté. Acidobacteria. In *eLS*, pages 1–10. John Wiley & Sons, Ltd, 2018. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470015902.a0027685.

[25] Marc Mußmann, Petra Pjevac, Karen Krüger, and Stefan Dyksma. Genomic repertoire of the Woeseiaceae/JTB255, cosmopolitan and abundant core members of microbial communities in marine sediments. *The ISME Journal*, 11(5):1276–1281, May 2017.

[26] Tony Gutierrez, David H. Green, Peter D. Nichols, William B. Whitman, Kirk T. Semple, and Michael D. Aitken. Polycyclovorans algicola gen. nov., sp. nov., an Aromatic-Hydrocarbon-Degrading Marine Bacterium Found Associated with Laboratory Cultures of Marine Phytoplankton. *Applied and Environmental Microbiology*, 79(1):205–214, January 2013.

[27] Yuchen Han and Mirjam Perner. The globally widespread genus Sulfurimonas: versatile energy metabolisms and adaptations to redox clines. *Frontiers in Microbiology*, 6, September 2015.

[28] Jan V. Henkel, Angela Vogts, Johannes Werner, Thomas R. Neu, Cathrin Spröer, Boyke Bunk, and Heide N. Schulz-Vogt. Candidatus Sulfurimonas marisnigri sp. nov. and Candidatus Sulfurimonas baltica sp. nov., thiotrophic manganese oxide reducing chemolithoautotrophs of the class Campylobacteria isolated from the pelagic redoxclines of the Black Sea and the Baltic Sea. *Systematic and Applied Microbiology*, 44(1):126155, January 2021.

[29] Sven Lahme, Cameron M. Callbeck, Lucy E. Eland, Anil Wipat, Dennis Enning, Ian M. Head, and Casey R.J. Hubert. Comparison of sulfide-oxidizing *Sulfurimonas* strains reveals a new mode of thiosulfate formation in subsurface environments. *Environmental Microbiology*, 22(5):1784–1800, May 2020.

[30] Wen-Li Li, Xiyang Dong, Rui Lu, Ying-Li Zhou, Peng-Fei Zheng, Dong Feng, and Yong Wang. Microbial ecology of sulfur cycling near the sulfate–methane transition of deep-sea cold seep sediments. *Environmental Microbiology*, 23(11):6844–6858, November 2021.

[31] Jang-Cheon Cho and Stephen J. Giovannoni. Robiginitalea biformata gen. nov., sp. nov., a novel marine bacterium in the family Flavobacteriaceae with a higher G+C content. *International Journal of Systematic and Evolutionary Microbiology*, 54(4):1101–1106, July 2004.

[32] H. D. Manh, Y. Matsuo, A. Katsuta, S. Matsuda, Y. Shizuri, and H. Kasai. Robiginitalea myxolifaciens sp. nov., a novel myxol-producing bacterium isolated from marine sediment, and emended description of the genus Robiginitalea. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY*, 58(7):1660–1664, July 2008.

[33] Chang Su, Meiling Zhang, Luoying Lin, Guangwei Yu, Haitao Zhong, and Yunxiao Chong. Reduction of iron oxides and microbial community composition in iron-rich soils with different organic carbon as electron donors. *International Biodeterioration & Biodegradation*, 148:104881, March 2020.

[34] Younghyun Park, Jaecheul Yu, Van Khanh Nguyen, Seonghwan Park, Jeongmi Kim, and Taeho Lee. Understanding complete ammonium removal mechanism in single-chamber microbial fuel cells based on microbial ecology. *Science of The Total Environment*, 764:144231, April 2021.

[35] Yong Wang, Hong-Yu Bi, Hua-Guan Chen, Peng-Fei Zheng, Ying-Li Zhou, and Jiang-Tao Li. Metagenomics Reveals Dominant Unusual Sulfur Oxidizers Inhabiting Active Hydrothermal Chimneys From the Southwest Indian Ridge. *Frontiers in Microbiology*, 13:861795, May 2022.

Appendix

Table 3: Selected genomic bins (shortlisted species and cable bacteria) from Kalø Vig with their abundance, completeness, and contamination.

Bin id	Classification	Completeness (%)	Contamination (%)	Abundance
metabat_bin.189	g__Electronema_1	89.66	1.72	283.40
metabat_bin.479	g__Electrothrix	94.83	0	252.09
metabat_bin.603	g__Electronema_2	89.66	1.72	145.37
maxbin_bin.9	g__41T-STBD-0c-01a_1_	79.31	2.19	127.44
maxbin_bin.542	s__Polycyclovorans sp002706265	91.38	0.16	123.86
maxbin_bin.10	g__41T-STBD-0c-01a_2_	81.9	1.72	109.58
maxbin_bin.38_sub	g__Robiginitalea_1	95.4	5.17	91.17
maxbin_bin.654	g__Magnetovibrio_1	91.22	7.92	84.95
maxbin_bin.29	f__Woeseiaceae_1	96.55	6.9	58.71
metabat_bin.455_sub	f__Woeseiaceae_2	96.55	2.74	43.29
maxbin_bin.73_sub	g__IGN2_	68.97	3.95	33.12
metadecoder_kaloevig.metadecoder.700	g__GCA-2746365	93.1	3.45	30.51
maxbin_bin.657	s__Sedimenticola thiotaurini_A	98.28	0	30.10
metabat_bin.347	f__Anaerolineaceae	76.96	4.31	18.11
metabat_bin.402_sub	g__Robiginitalea_2	77.96	8.62	18.07
maxbin_bin.114	f__Mor1	87.93	5.33	14.74
metabat_bin.51	g__UBA1847	100	3.92	14.61
metabat_bin.184_sub	g__Magnetovibrio_2	78.46	3.61	14.59
metabat_bin.10	g__UBA6170	93.1	3.45	14.49
metadecoder_kaloevig.metadecoder.327_sub	f__Sedimenticolaceae	86.22	3.45	13.13
metadecoder_kaloevig.metadecoder.920_sub	g__Magnetovibrio_3	91.38	0	12.95
metabat_bin.152	g__Sulfurimonas	98.28	4.31	12.38
metadecoder_kaloevig.metadecoder.302	g__UBA6107_	60.19	3.62	9.82

Table 4: Selected genomic bins (shortlisted species and cable bacteria) from Kalø Vig and the number of iron-related genes in the categories of iron oxidation and iron reduction.

Bin id	Classification	Iron oxidation	Iron reduction
metabat_bin.189	g__Electronema_1	0	3
metabat_bin.479	g__Electrothrix	0	3
metabat_bin.603	g__Electronema_2	1	1
maxbin_bin.9	g__41T-STBD-0c-01a_1_	0	6
maxbin_bin.542	s__Polycyclovorans sp002706265	0	8
maxbin_bin.10	g__41T-STBD-0c-01a_2_	0	5
maxbin_bin.38_sub	g__Robiginitalea_1	0	4
maxbin_bin.654	g__Magnetovibrio_1	5	5
maxbin_bin.29	f__Woeseiaceae_1	1	6
metabat_bin.455_sub	f__Woeseiaceae_2	0	8
maxbin_bin.73_sub	g__IGN2_	0	3
metadecoder_kaloevig.metadecoder.700	g__GCA-2746365	1	5
maxbin_bin.657	s__Sedimenticola thiotaurini_A	0	4
metabat_bin.347	f__Anaerolineaceae	0	1
metabat_bin.402_sub	g__Robiginitalea_2	0	3
maxbin_bin.114	f__Mor1	0	5
metabat_bin.51	g__UBA1847	0	6
metabat_bin.184_sub	g__Magnetovibrio_2	5	4
metabat_bin.10	g__UBA6170	0	6
metadecoder_kaloevig.metadecoder.327_sub	f__Sedimenticolaceae	0	3
metadecoder_kaloevig.metadecoder.920_sub	g__Magnetovibrio_3	5	6
metabat_bin.152	g__Sulfurimonas	0	3
metadecoder_kaloevig.metadecoder.302	g__UBA6107_	0	2

Table 5: Selected genomic bins (shortlisted species and cable bacteria) from Kalø Vig and the number of protein hits in each bin for EET proteins.

Bin id	Classification	OmcS	PioA	OmcZ	OmcC	OmcB	MtrC
metabat_bin.189	g__Electronema_1	0	0	0	0	0	0
metabat_bin.479	g__Electrothrix	0	0	0	0	0	0
metabat_bin.603	g__Electronema_2	0	0	0	0	0	0
maxbin_bin.9	g__41T-STBD-0c-01a_1_	0	0	0	0	0	0
maxbin_bin.542	s__Polycyclovorans sp002706265	0	0	0	0	0	0
maxbin_bin.10	g__41T-STBD-0c-01a_2_	0	0	0	0	0	0
maxbin_bin.38_sub	g__Robiginitalea_1	0	0	0	0	0	0
maxbin_bin.654	g__Magnetovibrio_1	0	0	0	0	0	0
maxbin_bin.29	f__Woeseiaceae_1	0	0	0	0	0	0
metabat_bin.455_sub	f__Woeseiaceae_2	0	2	0	0	0	1
maxbin_bin.73_sub	g__IGN2_	0	0	0	0	0	0
metadecoder_kaloevig.metadecoder.700	g__GCA-2746365	0	5	0	0	0	0
maxbin_bin.657	s__Sedimenticola thiotaurini_A	0	0	0	0	0	0
metabat_bin.347	f__Anaerolineaceae	0	0	0	0	0	0
metabat_bin.402_sub	g__Robiginitalea_2	0	0	0	0	0	0
maxbin_bin.114	f__Mor1	1	3	0	0	0	0
metabat_bin.51	g__UBA1847	0	1	0	0	0	1
metabat_bin.184_sub	g__Magnetovibrio_2	0	0	0	0	0	0
metabat_bin.10	g__UBA6170	0	1	0	1	0	0
metadecoder_kaloevig.metadecoder.327_sub	f__Sedimenticolaceae	0	1	0	0	0	0
metadecoder_kaloevig.metadecoder.920_sub	g__Magnetovibrio_3	0	0	0	0	1	0
metabat_bin.152	g__Sulfurimonas	0	0	0	0	0	0
metadecoder_kaloevig.metadecoder.302	g__UBA6107_	0	0	0	0	0	0

Table 6: Selected genomic bins (shortlisted species and cable bacteria) from Løgten with their abundance, completeness, and contamination.

Bin id	classification	Completeness (%)	Contamination (%)	Abundance
maxbin_bin.466	s__Polycyclovorans sp002706265	100	2.51	323.11
maxbin_bin.17	f__Woeseiaceae_	100	0.31	132.34
maxbin_bin.450	g__Electrothrix	82.18	1.02	117.90
maxbin_bin.29	g__Robiginitalea;s__	96.55	0	82.18
metabat_bin.378	g__UBA1847_1	97.41	1.36	80.00
metabat_bin.124	g__Sulfurimonas	96.21	2.59	60.76
metabat_bin.351	g__41T-STBD-0c-01a	98.28	6.9	51.67
metabat_bin.103	g__UBA1847_2	96.55	3.45	48.27
metabat_bin.312_sub	f__Mor1	94.83	8.62	41.09
maxbin_bin.85_sub	g__IGN2_1	74.14	2.76	36.36
metabat_bin.399	g__GCA-2746365_1	98.28	1.72	28.62
metabat_bin.57	f__Sedimenticolaceae	100	1.72	25.12
metadecoder_loegten.metadecoder.439	g__HyVt-443	96.55	0	18.66
metabat_bin.403	g__GCA-2746365_2	77.43	5.49	17.99
metadecoder_loegten.metadecoder.470	g__IGN2_2	98.28	1.72	12.54
metabat_bin.341_sub	g__Magnetovibrio	94.14	3.29	10.72

Table 7: Selected genomic bins (shortlisted species and cable bacteria) from Løgten and the number of iron-related genes in the categories of iron oxidation and iron reduction.

Bin id	Classification	Iron oxidation	Iron reduction
maxbin_bin.466	s__Polycyclovorans sp002706265	0	9
maxbin_bin.17	f__Woeseiaceae_	0	7
maxbin_bin.450	g__Electrothrix	0	4
maxbin_bin.29	g__Robiginitalea;s__	0	3
metabat_bin.378	g__UBA1847_1	0	7
metabat_bin.124	g__Sulfurimonas	1	5
metabat_bin.351	g__41T-STBD-0c-01a	0	7
metabat_bin.103	g__UBA1847_2	0	5
metabat_bin.312_sub	f__Mor1	0	11
maxbin_bin.85_sub	g__IGN2_1	0	3
metabat_bin.399	g__GCA-2746365_1	1	5
metabat_bin.57	f__Sedimenticolaceae	0	6
metadecoder_loegten.metadecoder.439	g__HyVt-443	0	6
metabat_bin.403	g__GCA-2746365_2	0	4
metadecoder_loegten.metadecoder.470	g__IGN2_2	0	6
metabat_bin.341_sub	g__Magnetovibrio	1	5

Table 8: Selected genomic bins (shortlisted species and cable bacteria) from Løgten and the number of protein hits in each bin for EET proteins.

Bin id	Classification	OmcS	PioA	OmcZ	OmcC	OmcB	MtrC
maxbin_bin.466	s__Polycyclovorans sp002706265	0	0	0	0	0	0
maxbin_bin.17	f__Woeseiaceae_	0	1	0	0	0	1
maxbin_bin.450	g__Electrothrix	0	0	0	0	0	0
maxbin_bin.29	g__Robiginitalea;s__	0	0	0	0	0	0
metabat_bin.378	g__UBA1847_1	0	2	0	0	0	1
metabat_bin.124	g__Sulfurimonas	0	4	0	0	0	0
metabat_bin.351	g__41T-STBD-0c-01a	0	1	0	0	0	0
metabat_bin.103	g__UBA1847_2	0	0	0	0	0	0
metabat_bin.312_sub	f__Mor1	2	4	2	0	0	0
maxbin_bin.85_sub	g__IGN2_1	0	0	0	0	0	0
metabat_bin.399	g__GCA-2746365_1	0	1	0	0	0	0
metabat_bin.57	f__Sedimenticolaceae	0	4	0	0	0	0
metadecoder_loegten.metadecoder.439	g__HyVt-443	0	0	0	0	0	0
metabat_bin.403	g__GCA-2746365_2	0	0	0	0	0	0
metadecoder_loegten.metadecoder.470	g__IGN2_2	0	0	0	0	0	0
metabat_bin.341_sub	g__Magnetovibrio	0	0	0	0	1	0

Table 9: Selected genomic bins (shortlisted species and cable bacteria) from the *Ca.* Electrothrix communis RB enrichment with their abundance, completeness, and contamination.

Bin id	Classification	Completeness (%)	Contamination (%)	Abundance
bin.11	f__SZUA-229_2	89.66	13.79	474.45
bin.5	s__Sedimenticola thiotaurini_A_1	74.14	0	194.22
bin.1	s__Sedimenticola thiotaurini_A_2	10.34	0	140.52
bin.9	g__Magnetovibrio_1	89.66	0	130.69
bin.17	g__Magnetovibrio_2	89.66	0	80.38
bin.37	g__Electrothrix	98.28	0	55.95
bin.16	g__Sulfurimonas	57.84	0	27.62
bin.44	f__SZUA-229_1	91.38	57.29	22.19

Table 10: Selected genomic bins (shortlisted species and cable bacteria) from the *Ca.* Electrothrix communis RB enrichment and the number of iron-related genes in the categories of iron oxidation and iron reduction.

Bin id	Classification	Iron oxidation	Iron reduction
bin.11	f__SZUA-229_2	1	6
bin.5	s__Sedimenticola thiotaurini_A_1	0	4
bin.1	s__Sedimenticola thiotaurini_A_2	0	0
bin.9	g__Magnetovibrio_1	5	4
bin.17	g__Magnetovibrio_2	4	4
bin.37	g__Electrothrix	0	4
bin.16	g__Sulfurimonas	1	5
bin.44	f__SZUA-229_1	4	15

Table 11: Selected genomic bins (shortlisted species and cable bacteria) from the *Ca.* Electrothrix communis RB enrichment and the number of protein hits in each bin for EET proteins.

Bin id	Classification	OmcB	OmcS	PioA	OmcZ	OmcC	MtrC
bin.11	f__SZUA-229_2	0	0	0	0	0	0
bin.5	s__Sedimenticola thiotaurini_A_1	0	0	0	0	0	0
bin.1	s__Sedimenticola thiotaurini_A_2	0	0	0	0	0	0
bin.9	g__Magnetovibrio_1	0	0	0	0	0	0
bin.17	g__Magnetovibrio_2	0	0	0	0	0	0
bin.37	g__Electrothrix	0	0	0	0	0	0
bin.16	g__Sulfurimonas	0	0	3	0	0	0
bin.44	f__SZUA-229_1	0	0	0	0	0	0