Imperial College London

Study of Extracellular Electron Transfer In Engineered Methanogenic Systems

Theo Y.C. Lam

Feb 2025

Department of Civil and Environmental Engineering
Imperial College London

Thesis submitted for the degree of Doctor of Philosophy

Supervisors: Dr Po-Heng (Henry) LEE, Dr Giin-Yu Amy Tan

Abstract

Anaerobic digestion of low-strength wastewater is frequently constrained by slow hydrolysis, accumulation of fermentation intermediates, and limited energetic efficiency. While extracellular electron transfer (EET) and direct interspecies electron transfer (DIET) are now recognised as important for methanogenesis, their broader functional roles within complex anaerobic treatment systems remain poorly resolved, particularly at pilot scale.

This thesis investigates the role of EET in a staged anaerobic fluidised-bed membrane bioreactor treating domestic wastewater, integrating reactor performance data with metagenomic, metatranscriptomic, and pangenomic analyses. Conductive granular activated carbon was used to promote biofilm-associated electron exchange, and functional inference was performed using a rule-based gene-centric framework, enabling resolution of EET mechanisms in predominantly uncultured microbial communities.

The results demonstrate that EET is not restricted to methanogenesis but is associated with enhanced upstream carbon conversion. Distinct functional guilds were identified, with *Bacteroidota* enriched in multiheme cytochromes linked to hydrolysis and acidogenesis, and *Desulfobacterota* encoded electrically conductive pili associated with syntrophic oxidation of propionate. Under sulfate-limited conditions, sulfate-reducing bacteria were observed to engage in DIET with *Methanothrix*, indicating a cooperative rather than competitive interaction and challenging conventional models of sulfate–methane competition.

Pangenomic analysis of *Methanothrix* identified a previously uncharacterised MspA-like amyloid sheath protein, proposed as a putative electron-uptake module. The development of a bespoke amyloid protein prediction framework was necessary because existing databases lack close homologues, highlighting the limitations of conventional annotation approaches for EET-related functions.

No conclusive evidence for horizontal gene transfer of canonical EET genes was found; however, phage-associated transfer of antibiotic resistance genes was detected, suggesting that biofilms may facilitate incidental genetic exchange.

Overall, this work extends the conceptual scope of EET from a methanogenic shortcut to a system-level mechanism supporting hydrolysis, stability, and carbon conversion, with implications for the design and interpretation of advanced anaerobic treatment processes.

Statement of Originality

The entirety of this thesis is my original work. Any quotes, descriptions, or references to the work of others have been thoroughly acknowledged, citing both published and unpublished sources.

Material from Chapter 2 was published in Uali, A. S., **Lam, T. Y.**, Huang, X., Wu, Z., Shih, H. J., Tan, G. Y., & Lee, P. H. (2024). Role and potential of the semi-classical/-quantum mechanism of the extracellular environment and cell envelope in Direct Interspecies Electron Transfer (DIET)-driven biomethanation. *Critical Reviews in Environmental Science and Technology*, 54(7), 581-601.

Material from Chapters 4, 5, and 6 is currently prepared for publication in Lam, T. Y., Shin, C., Tan, G. Y., Lee, P. H. [Extracellular Electron Transfer in Staged Anaerobic Digestion: Sheath Proteins and Electron Flow in the Carbon Cycle]. *Environmental Science & Technology*, In preparation

Material from Chapter 7 was published in Yu, R., Huang, Z., **Lam, T. Y.**, & Sun, Y. (2024). Utilizing profile hidden Markov model databases for discovering viruses from metagenomic data: A comprehensive review. Briefings in Bioinformatics, 25(4), bbae292.

Copyright Statement

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Acknowledgement

I would like to thank my supervisors, Henry Lee and Amy Tan, whose steady support and good humor have made this journey both challenging and rewarding. Their guidance has not only enriched my research but also helped me understand my strengths and weaknesses—and even how to mitigate the weakness of my strength.

I'm also grateful to Stephen Smith for his thoughtful comments along the way, which have provided useful insights into my progress.

A warm thank you goes to Geoff Fowler, Ángel Nievas Pino, and Celia Lima for their practical help in the lab, ensuring that the experiments went as smoothly as possible.

I would also like to acknowledge the enjoyable and inspiring moments shared with my colleagues Emma Berins, Michael Ho, Henrique Dornelles, Jonathan Hungerland, and Aitolkyn Uali. Our intellectual exchanges and fun discussions have added a memorable lightness to this challenging period.

Lastly, my heartfelt thanks to my family, partners, and friends for being there when the balance between family commitments and academic work was tested. Your support kept me going even when I was torn in two.

Contents

1 Introduction **22**

1.1 Background . . . 22

1.1.1 Extracellular electron transfer promotes methane production . . . 23

1.1.2 Pilot-scale staged anaerobic fluidised-bed membrane bioreactors as a study ground for direct interspecies electron transfer application . . . 24

1.2 Research gaps . . . 26

1.3 Aims and objectives . . . 27

1.3.1 Aims . . . 27

1.3.2 Objectives . . . 27

1.4 Research framework . . . 28

1.5 Thesis structure . . . 29

2 Literature Review **31**

2.1 Anaerobic digestion . . . 31

2.1.1 AD via mediated interspecies electron transfer . . . 31

2.1.2 Anaerobic digestion via direct interspecies electron transfer . . . 33

2.1.3 Conductive materials synergies with DIET . . . 34

2.1.4 Potential reason for DIET's stabilising effect . . . 35

2.1.5 DIET effect on organic waste degradation . . . 36

2.1.6 Literature-based non-conductive reference communities for comparison to GAC-associated microbiomes . . . 37

2.1.7 Validation of DIET occurrence . . . 38

2.1.8 Theoretical considerations in EET . . . 39

2.2 Biological processes of EET . . . 40

2.2.1 Brief history of EET study . . . 40

2.2.2 Diverse electron import mechanisms in methanogens . . . 41

2.2.3 Electron shuttles . . . 44

2.3 Proteins related to extracellular electron transfer . . . 45

2.3.1 Multi-heme c-type cytochrome and porin-cytochrome conduits . . . 45

2.3.2 Electrically conductive pili . . . 47

2.3.3 Amyloid protein . . . 49

2.3.4 Adaptation in EET capability . . . 49

2.4 Horizontal gene transfer . . . 50
2.4.1 Scope and mechanistic focus of this thesis . . . 51

3 Materials and Methods **52**
3.1 Staged anaerobic fluidised membrane bioreactors . . . 52
3.1.1 SAF-MBR 1.0 . . . 53
3.1.2 SAF-MBR 2.0 . . . 55
3.1.3 Operational performance and cross-system normalisation . . . 56
3.1.4 Biomass sampling rationale and representativeness of reactor states . . . 59
3.1.5 Biomass collection, preparation, and nucleic acid extraction . . . 59
3.2 Genomic data processing . . . 61
3.2.1 Methodological rationale . . . 61
3.2.2 DNA and RNA sequencing . . . 61
3.2.3 Sequencing denoising . . . 62
3.2.4 Metagenome assembly . . . 62
3.2.5 Metagenome binning . . . 63
3.2.6 MAGs refinement . . . 64
3.2.7 MAGs clustering and dereplication . . . 64
3.3 Taxonomy profiling via read-based approach . . . 65
3.4 Phylogenetic analysis . . . 66
3.4.1 Assign taxonomy for metagenome-assembled genomes . . . 66
3.4.2 MAGs phylogeny . . . 66
3.4.3 Protein phylogeny . . . 67
3.5 Functional annotation . . . 67
3.5.1 Community-based functional annotation . . . 67
3.5.2 Genome-based Functional Annotation . . . 68
3.6 EET gene annotation . . . 69
3.6.1 Remote homologues sequence detection via hidden Markov model . . . 70
3.6.2 MHC detection . . . 71
3.6.3 E-pili detection . . . 72
3.7 Expression level analysis . . . 73
3.8 Pangenomic analysis . . . 73
3.8.1 Pangenomic analysis using Anvi'o . . . 74
3.8.2 Homology detection and definition in Pangenomes . . . 74
3.8.3 Functional enrichment analysis . . . 75
3.8.4 Amyloid protein prediction . . . 76
3.9 Horizontal gene transfer analysis . . . 76
3.9.1 Genome-wide detection of putative HGT . . . 76
3.9.2 Identification of mobile genetic elements . . . 77
3.9.3 Gene flow annotation . . . 78
3.10 Statistical analysis . . . 78

4 Diversity of Genes Related to Extracellular Electron Transfer and Their Significance in Carbon Degradation 79
4.1 Introduction . . . 79
4.2 Materials and methods . . . 80
4.2.1 Sample collection and data pre-processing . . . 80
4.2.2 Metagenome-assembled genome refinement and post-processing . . . 80
4.2.3 Gene prediction . . . 81
4.2.4 Phylogenetic tree construction . . . 81
4.2.5 Statistical analysis . . . 81
4.3 Results . . . 82
4.3.1 Metagenomic reconstruction and EET-related genes prediction . . . 82
4.3.2 Correlations between the Distribution of EET Genes and SAF-MBR Operating Conditions . . . 94
4.3.3 Correlation between EET genes and metabolic capability . . . 95
4.4 Discussion . . . 97
4.4.1 Enhanced detection of EET genes through rule-based approaches . . . 97
4.4.2 Diversity and prevalence of EET-Capable microbial communities in SAF-MBR 98
4.4.3 Metabolic associations of EET genes with carbon metabolism pathways . 99
4.4.4 Implications for anaerobic wastewater treatment . . . 99

5 Direct Interspecies Electron Transfer Driven Syntropy between Sulfate-reducing Bacteria and *Methanothrix* in Pilot-scale Reactor 101
5.1 Introduction . . . 101
5.2 Materials and methods . . . 102
5.2.1 Bioreactor operation, sample collection, preparation, and sequencing . . . 102
5.2.2 Metagenomic and metatranscriptomic analysis . . . 102
5.2.3 Data availability . . . 103
5.3 Results and discussion . . . 103
5.3.1 Divergence in metagenomic and metatranscriptomic taxonomy profiles . . 103
5.3.2 Metabolic function transcription in community . . . 105
5.3.3 Metagenome-assembled genomes and gene transcription . . . 107
5.3.4 GAC-facilitated DIET in sulfate-reducing bacteria and promoted symbiotic relationships . . . 109

6 Identification of Extracellular Electron Transfer Sheath Proteins in *Methanothrix* Using Pangenomic Analysis 114
6.1 Introduction . . . 114
6.2 Materials and methods . . . 116
6.2.1 Genome collection . . . 116
6.2.2 Pangenome analysis . . . 116
6.2.3 Functional enrichment analysis in pangenome . . . 117

6.2.4 MspA and Spq prediction . . . 117
6.2.5 Motif analysis . . . 118
6.2.6 3D protein structure prediciton . . . 118
6.3 Results . . . 118
6.3.1 *Methanothrix*'s pangenome . . . 118
6.3.2 Prediction of major sheath protein in *Methanothrix* MAGs . . . 120
6.4 Discussion . . . 122
6.4.1 The EET pathway in *Methanothrix* . . . 122
6.4.2 Comparing to other methanogens . . . 128
6.4.3 Role of structural proteins in EET . . . 130
6.4.4 Metabolic requirements and flexibility . . . 130
6.4.5 Implications for biotechnological applications . . . 130
6.4.6 Critical evaluation of current models . . . 131

7 Horizontal Gene Transfer of Antibiotic Resistance Gene in SAF-MBRs 132
7.1 Introduction . . . 132
7.2 Materials and Methods . . . 133
7.2.1 Global horizontal gene transfer event identification . . . 133
7.2.2 Plasmid and phage detection . . . 133
7.2.3 Antibiotic resistance gene (ARG) detection . . . 134
7.2.4 Quantifying ARGs . . . 134
7.2.5 Plasmid and phage abundance . . . 134
7.3 Results and Discussion . . . 135
7.3.1 Overview . . . 135
7.3.2 ARG distribution across samples . . . 135
7.3.3 HGT of sulfonamide resistance genes . . . 135
7.3.4 Phage-host interactions in the SAF-MBR . . . 139

8 General Discussion 141
8.1 Genomic evidence for extracellular electron transfer in engineered anaerobic biofilms 141
8.2 Microhabitat-structured organisation of EET gene inventories . . . 142
8.3 Limits of taxonomic inference and the necessity of gene-centric approaches . . . 143
8.4 EET-associated functions beyond canonical methanogenesis . . . 144
8.5 *Methanothrix* as a focal taxon: pangenomic evidence for putative surface-associated EET modules . . . 145
8.6 Distribution and transfer of EET-related and accessory genes in SAF-MBR microbiomes . . . 147
8.7 Implications for interpreting EET in engineered anaerobic systems . . . 148

9 Conclusions 150
9.1 Key findings and contributions . . . 150
9.1.1 EET gene diversity and system-level carbon conversion . . . 150

9.1.2 DIET-driven syntrophy between sulfate-reducing bacteria and methanogens 151
9.1.3 Molecular basis of extracellular electron uptake in *Methanothrix* 151
9.1.4 Constraints on EET gene transfer and accessory gene dynamics 151
9.2 Implications and potential applications 152
9.3 Limitations of the study 153
9.4 Future works 154
9.4.1 Physical examination of the surface proteins using state-of-the-art Bio-SECM 154
9.4.2 Transposon sequencing for *De Novo* functional annotation 154
9.4.3 Pure and co-culture experiments 154

10 References **156**

A Supplementary Notes **189**
A.1 Active biomass estimation in SAF-MBR 189
A.2 Refinement of Metagenome-Assembled Genomes 190
A.2.1 Four Modes of Refinement 190
A.3 Phylogenetic trees oddity 192
A.4 E-pili identification 192

B Supplementary Tables **198**

C Supplementary Figures **215**

List of Tables

2.1 Summary of genus-level 16S rRNA trends reported in PVDF-tagged AFMBR studies used as non-conductive (or paired-carrier) reference conditions for community comparison. Reported values are those explicitly stated in the cited manuscripts. 39
2.2 The naturally occurring modes of EET 45

3.1 Integrated operating conditions, sulfate content, COD removal performance, and CH_4 yield normalised per litre of wastewater treated for SAF-MBR systems and the PVDF AFMBR reference. 58
3.2 Biomass samples collected from SAF-MBR 1.0 and SAF-MBR 2.0, with sequencing availability. For ease of reference, samples were assigned a simple prefix. 60

4.1 Comparison of MHC-encoding and non-MHC *Bacteroidota* MAGs based on the `Genes` annotation column. Values represent normalised abundance and mean total abundance per sample in bulk and GAC fractions. 88
4.2 Top 10 genera contributing to the GAC-associated EET+ dereplicated MAG pool. Values are expressed as percentages. Columns “MHC”, “e-pili”, and “Both” indicate the within-genus distribution of EET gene inventories. 92
4.3 Spearman correlation between EET gene relative abundance and sample types. . 94

B.1 DNA quality of the biomass samples 199
B.2 RNA quality of the biomass sample 200
B.3 General statistics of the metagenomes collected in this study. % Dups is the percentage of duplicated sequences. % GC is the percentage of GC content, which refers to the percentage of nitrogenous bases in a DNA or RNA molecule that are either guanine (G) or cytosine (C). M seq is sequence in million. 201
B.4 List of Type C Multiheme Cytochrome Genes for Iron Reduction and Oxidation. This set of genes was used for HMM database training. 202
B.5 List of e-pili genes identified in literature review. This set of genes was used for HMM database training. 204
B.6 Generalisation of cell membrane structure according to the latest version of Bergey’s Manual of Systematics of Archaea and Bacteria [318] 206
B.7 Shannon diversity index and richness across bioreactor samples 207
B.8 ANOVA Results and Tukey’s HSD Test for Shannon Index 207
B.9 Electron Donors and EET Mechanisms of Microbes 207

B.10 Quality control and phylogenetic classification of the MAGs. 210
B.11 The pili sequences' aromaticity and largest gap between neighbouring aromatic amino acids identified in bin 142. Tyrosines (Tyr), Phenylalanines (Phe), Histidines (His), and Tryptophans (Trp). The pili genes with aromaticity higher than 9.8% are bolded and considered to be putative e-pili. 211
B.12 Overview of Antibiotic Classes, Their Mechanisms of Action, and Common Resistance Mechanisms 212
B.13 Removal Efficiency of Various Substances Under Different Conditions in AFMBR . 214

List of Figures

1.1 An overview of the research objectives based on observations from the SAF-MBR 28

2.1 Simplified view of the major metabolic pathways in AD 32
2.2 Illustrations of EET pathways in anaerobic treatments 34
2.3 The proposed mechanism for augmenting anaerobic digestion of complex organic waste via DIET is illustrated, with the red arrow indicating the DIET pathway (adapted from [12]). 37
2.4 proposed EET models in methanogens 42
2.5 Mechanisms of EET in methanogens. (a) Direct transfer occurs via redox proteins and (b) microbial pili or nanowires, while (c) indirect transfer occurs through intermediary electron carriers. Additionally, the process involves electron transfer with (d) the aid of extracellular polymeric substances and cell wall polymers. 43
2.6 The process of electron exchange using MHC conduit. The conventional pathway for electron export, associated with anodic activities, is represented in light blue, whereas the hypothetical pathway for cathodic electron import is depicted in pink. Solid lines indicate established electron transfer routes, whereas hypothesised routes are illustrated with dashed lines. Image is reproduced from [143] 46
2.7 Expressing *pilA* pilin monomer genes from external sources in *G. sulfurreducens* results in pili exhibiting significantly varied conductivity, with differences reaching up to 10^6-fold compared to individual native wires. These genetically modified pili maintain the same diameter as the natural wild-type pili at 3 nm, differing from the 4 nm diameter of OmcS filaments. Such alterations in pili conductivity notably degrade the efficiency of extracellular electron transport, adversely affecting related phenotypes. This finding aligns with similar outcomes from a range of studies on heterologous expression, all of which corroborate the substantial impact of conductivity modifications on electron transport capabilities. Image is adapted from Lovley and Walker [163]. 48

3.1 (a) Schematic diagram and (b) a photo of the pilot-scale SAF-MBR system 1.0 [31] 54
3.2 The SAF-MBR 2.0 55
3.3 Steps for obtaining assembly-graph refined MAGs 65
3.4 Databases used by different bioinformatics pipelines (DRAM, METABOLIC 4.0, and eggNOG-mapper v2) 69

3.5 MHC identification workflow 72
3.6 E-pili identification workflow 73

4.1 E-pili, MHC, oxidative (OxiEET), and reductive (RedEET) EET protein abundances across the phylogenetic diversity of bioreactor microbial communities. The tree included 421 MAGs featuring at least one aforementioned EET gene category . Phylogenetic reconstruction utilised RAxML (v8.2.12) for maximum likelihood estimation [237]. Taxonomic classifications were conducted via GTDB-tk [227], with visualisation facilitated by tvBOT [312]. 83
4.2 MHC gene distribution among phylum, media, and reaction direction. More than half of the predicted gene (n=639) were unannotated and 439 matched with the experimentally verified EET MHC gene. 84
4.3 Redox potential range of the MHC genes identified in this study. The colour legends present reaction direction: orange is reductive, blue is oxidative, and green is bidirectional. 85
4.4 Scatter plot illustrating the relationship between percent aromatics and the largest gap within pilin sequences. Data points are categorised into three groups: hmm-search e-pili (grey dots); Putative bacterial e-pili (red dots), denoting sequences with a percentage aromatics $\geq$ 9.8% and the largest gap $\leq$ 22; and Putative archaeal e-pili (green dots), indicating sequences identified as Mhun_3140 with percent aromatics $\geq$ 9.8%. This classification aims to distinguish potential e-pili sequences based on their aromatic content and gap characteristics, highlighting the variability and specific traits that may influence their function and assembly. 86
4.5 Comparison of mean RPKM expression levels of Epili and MHC across various phyla. Error bars represent the standard deviation of expression levels across different conditions. Phyla with no detectable expression in one or both datasets are represented with empty bars. 86
4.6 Sankey diagram illustrating the flow of normalised coverage from MAGs clustered at phylum level to EET-gene categories and subsequently to the media type. ”Not Found” refers to no EET-gene detected, and both refer to the detection of MHC and e-pili within a MAG. (consider to update the label) 87
4.7 Comparison of alpha diversity between GAC and Bulk microbial communities. Boxplots show the Shannon Diversity Index (left) and Simpson Diversity Index (right) for both sample types. GAC samples exhibit significantly higher Shannon (p = 0.032) and Simpson (p = 0.016) indices. Significance was tested using the Kruskal-Wallis test. 89
4.8 Dendrogram showing hierarchical clustering of samples based on Bray-Curtis dissimilarity, representing beta diversity. Samples are categorised into Bulk, GAC, and PVDF, indicated by different colours. The vertical axis reflects Bray-Curtis distances, illustrating the similarity or dissimilarity in microbial community composition. 90

4.9 Clustered heatmap representing the relative abundance of microbial communities at the phylum level across different sample media types. Rows correspond to different microbial phyla, while columns represent individual samples. The samples are hierarchically clustered, with dendrograms at the top indicating clustering patterns based on microbial community composition. Colour intensity within the heatmap reflects the abundance of each phylum, with a scale ranging from 0 (deep purple) to 50 (bright yellow). Columns are colour-coded by sample media type: green represents Bulk, blue represents GAC, and orange represents PVDF.Only phylum with at least one sample higher than 1% is shown here. 91

4.10 Scatter plots of EETRatio (%) versus reactor operating conditions with LOWESS smoothing. Each point represents a pilot-scale sample, coloured by biomass type (GAC or bulk). The plots show weak monotonic associations without threshold-like transitions, and the data primarily cluster according to discrete operational modes rather than continuous parametric responses. 95

4.11 Heatmap of Pearson correlation coefficients among operational parameters and EET gene distributions. Variables include temperature, HRT, OLR, COD concentrations, COD removal efficiency, and ratios of EET, e-pili, and MHC genes to MAGs. . . . 96

5.1 Taxonomic profile at genus level 104

5.2 Metabolic weight scores of metabolic pathways in the metatranscriptome, with the corresponding contribution percentages from MAGs aggregated at the phylum level. Old and new phylum names are provided for consistency. 105

5.3 Transcription distribution of methanogenesis predicted using HUMAnN 3.0 against Uniref 50 database. CO_2 reduction pathway accounted for 77% of the total methanogenesis transcribed, of which 18% had been identified to come from *Methanothrix* species. 106

5.4 (a) The predicted 3D protein structure of pili from SRB bin 142 is illustrated, with aromatic residues highlighted in green. (b) A colour gradient indicates the error estimate. Yellow represents the highest error margin, while purple denotes the lowest. (c) An accompanying plot of the error estimate. 110

5.5 Proposed SRB MAG's metabolism and transcription level 113

6.1 The pangenome of genus *Methanothrix*. Each gene cluster contains at least one gene contributed by at least one genome. The pangenome is organised by gene cluster frequency, and gene clusters are divided into 4 groups based on their presence/absence patterns. The core gene cluster contains 1,740 gene clusters which all genomes within the pangenome possess and share. Conversely, the two accessory gene clusters (ACC_1 & ACC_2), which contain 4,291 and 3,079 gene clusters, pertain to sequences unique to a subset of strains within the pangenome. In comparison, singletons (17,430 gene clusters) are genomic sequences exclusive to individual strains. The arrangement of items follows a presence-absence criterion, utilising D (Euclidean distance) and L (Ward's method) for clustering. 119

6.2 Phylogenetic analysis and distribution of *MspA*-like protein sequences. (a) Phylogenetic tree of *MspA*-like protein sequences, where internal homologs (percentage identity $>$ 35%) are color-labeled in orange. (b) Expression levels ($\log_2$RPKM) of *MspA*-like genes in the metatranscriptomes of three bioreactor GAC samples. (c) Distribution of the MAGs by origin environment, with bubble size representing the number of genomes. (d) Total number of genomes encoding the corresponding gene. 121

6.3 N-terminal motif encoded in MspA-like genes, annotated using MEME 122

6.4 Predicted amyloidogenic regions in transcribed mspA-like genes from *Methanothrix* MAGs. 123

6.5 Heatmap showing the presence (1) or absence (0) of gene clusters across *Methanothrix* MAGs reconstructed from this study's metagenome. 124

6.6 Heatmap of fold-difference in MspA-like gene expression relative to the median RPKM across *Methanothrix* MAGs from this study's metagenomes. Fold-differences were calculated within the same sample by comparing metatranscriptome data to its corresponding MAG. 125

6.7 3D protein structure prediction on MspA-like genes recovered from the MAGs with significant expression in the metatranscriptomes. Interface accuracy was assessed using the interface predicted template modeling score (ipTM). Prediction was done by using AlphaFold 3. 126

6.8 Structural comparison of MspA-like proteins. The top row displays previously annotated MspA proteins [41], predicted using AlphaFold 2. These models represent interactions between viral SH-like proteins (Vir197, Vir071, and PBV238) and their corresponding host SH proteins from *Methanotrichales*, *Methanomicrobiales*, and *Syntropharchaeia*. Model confidence is indicated by the predicted local distance difference test (pLDDT), and interface accuracy is assessed using the ipTM. The bottom row shows the GC_00000231, which is identified from the core gene cluster in this study, predicted using AlphaFold 3. This MspA-like protein structurally resembled the Aromatic residues, which are highlighted in orange, with only residues having a nearest neighbour within 20 Ådisplayed. 127

6.9 Proposed *Methanothrix* EET methanogenesis pathway. Arrows are coloured according to each methanogenesis pathway. Acs, acetyl-CoA synthase; CoB-SH, coenzyme B; Cdh, Acetyl-CoA decarbonylase/synthase complex; CoMSH, coenzyme M; Fae, formaldehyde activating enzyme; Fd, ferredoxin; Fpo, $F_{420}H_2$ dehydrogenase; Ftr, formylmethanofuran–H_4MPT formyltransferase; Fwd, formylmethanofuran dehydrogenase; H4MPT, tetrahydromethanopterin; HCHO, formaldehyde; Hdr, heterodisulfide reductase; mch, methenyl-H_4MPT cyclohydrolase; Mcr, methyl-coenzyme M reductase; Mer, F_{420}-dependent methylene-H_4MPT reductase; MP, methanophenazine; MspA, major sheath protein; Mtr, Na^+-translocating methyl-H_4MPT-coenzyme-M methyltransferase; PQQ, pyrroloquinoline quinone; RHP, reductive hexulose phosphate; Sqp, surface quinoprotein. The image was created using BioRender. 129

7.1 ARG quantification grouped by the sample type. 136

7.2 A stacked barplot of drug class abundance across different samples. 136

7.3 Phylogenetic distribution and putative transfer mechanisms of *sul* genes. *sul4* clusters are highlighted in yellow. 138

7.4 Correlation between phage and MAG abundance in the SAF-MBR. 139

7.5 Stacked bar plot illustrating the distribution of phage- and plasmid-mediated HGT events across different samples. 140

A.1 Scattered boxplots of the refined MAGs' completeness and redundancy. 192

A.2 Venn diagram of MHC and e-pili genes identified in the MAGs. 193

A.3 Phylogenetic tree constructed using MAGs that contain EET genes. The taxonomy (phylum level) was assigned based on the result done by (a) PhyloPhlAn and (b) GTDB-TK. 194

C.1 Sequence quality histograms of the metagenomes collected in this study accessed using FastQC [207] and summarised using MultiQC [428]. A phred score above 30 thus all samples are high quality. 215

C.2 Top ten prevalent phyla by mean relative abundance based on the percentage reads assigned to EET MAGs. 216

C.3 e-pili and MHC gene count across various phyla. Each bar represents the count of genes identified within each phylum. Blue coloured bar represent e-pili genes and red is for MHC genes. For visualisation purpose, only phylum with gene count more than 5 was shown. 216

C.4 Heatmap showing the correlation between MHC genes and metabolic capacity. Only statistically significant Pearson correlation coefficients (p-value < 0.05, determined using Student's *t*-test) with a random forest feature importance greater than 0.01 are displayed. 217

C.5 Heatmap showing the correlation between e-pili genes and metabolic capacity. Only statistically significant Pearson correlation coefficients (p-value < 0.05, determined using Student's *t*-test) with a random forest feature importance greater than 0.01 are displayed. 218
C.6 Taxonomy profile classified via Kaiju. (a) search against metagenome at phylum level. (b) search against metagenome at family level. (c) search against metatranscriptome at phylum level. (d) search against metatranscriptome at family level. The cutoff is at 0.01% relative abundance. 219
C.7 Transcription level of genes involved in Wood-Ljungdahl acetate oxidation pathway. 220
C.8 Transcription level of genes involved in methylmalonyl-CoA propionate degradation pathway. 221
C.9 Transcription level of genes involved in formate/H_2 MIET pathways. 222
C.10 Transcription level of genes involved in electron transfer in SRB. 223
C.11 First quarter of the MspA amyloidogenic region analysis. 224
C.12 Second quarter of the MspA amyloidogenic region analysis. 225
C.13 Third quarter of the MspA amyloidogenic region analysis. 226
C.14 Fourth quarter of the MspA amyloidogenic region analysis. 227
C.15 MspA contains evenly distributed amyloidogenic regions. Analysis of the MspA and MspA-like sequences using AmyloGram [429] identified multiple regions with predicted high amyloidogenic potential (indicated above the horizontal lines). The sequences analysed include: (i) 33 expressed MspA-like genes predicted from the *Methanothrix* pangenome, (ii) WP_003891919, WP_011727317, WP_011730563, and WP_162139622 from *Mycobacterium smegmatis*, (iii) WP_011449234.1 from *Methanospirillum hungatei*, and (iv) ABK14853.1 from *Methanothrix thermoacetophila* PT. 227

Glossary

M. barkeri *Methanosarcina barkeri*. 18
M. mazei *Methanosarcina mazei*. 18
M. soehngenii *Methanothrix soehngenii*. 18

AD anaerobic digester. 18, 22
AFBR anaerobic fluidised-bed bioreactor. 18, 53, 80
AFMBR anaerobic fluidised-bed membrane bioreactor. 18, 53
AMG auxiliary metabolic gene. 18, 133, 135
AMR anti microbial resistance. 18
ANI average nucleotide identity. 18, 64, 117
ARG antibiotic resistance genes. 18, 26, 132, 133, 135

BES bioelectrochemical system. 18, 41
BGI Beijing Genomics Institute. 18
BOD biochemical oxygen demand. 18

CARD Comprehensive Antibiotic Resistance Database. 18
CDS coding sequence. 18
CH_4 methane. 11, 18, 22–24, 31, 33, 35, 36, 39, 41, 42, 46, 51, 53, 56–59, 79, 97, 100, 101, 105, 106, 114, 115, 128, 130, 131, 144, 145, 148, 151, 152, 189, 190
COD chemical oxygen demand. 18
COG Clusters of Orthologous Genes. 18, 116

DIET direct interspecies electron transfer. 18, 23, 101, 102

e-pili electrically conductive pili. 18, 25, 79, 97
EAG environmentally accessory genes. 18
EBNet Environmental Biotechnology Network. 18
ECG environmentally core genes. 18
EET extracellular electron transfer. 18, 23, 31, 73, 74, 79, 80, 97, 101, 132
EPS extracellular polymeric substances. 18, 43

Fpo proton-translocating F_{420}:methanophenazine oxidoreductase. 18

GAC granular activated carbon. 18, 23, 41, 79, 80, 97, 101, 132

H_2-MET H_2-mediated electron transfer. 18

HGT horizontal gene transfer. 18, 26, 77, 132, 133
HMM hidden Markov Model. 18
HRT hydraulic retention time. 18, 53

KEGG Kyoto Encyclopedia of Genes and Genomes. 18

MAG metagenome-assembled genome. 18, 63, 103, 116
MCL Markov Cluster Algorithm. 18, 75, 117
MePA methylparaben. 18
MET mediated electron transfer. 18, 39
MGE mobile genetic element. 18, 50
MHC multiheme c-type cytochrome. 18, 25, 34, 79, 97
MK menaquinone. 18
MP methanophenazine. 18, 42, 44, 128
MS mass spectrometry. 18
MSA multiple sequence alignment. 18, 66, 70
MspA major sheath protein. 18, 25
MW Metabolic Weight. 18

NCBI National Center for Biotechnology Information. 18, 116
NDMA N-nitrosodimethylamine. 18, 132

OLR organic loading rate. 18

PANI polyaniline. 18
PCET proton-coupled electron transfer. 18
PET polyethylene terephthalate. 18
PFOR pyruvate:ferredoxin oxidoreductase. 18
pHMM profile hidden Markov Model. 18, 81
PVDF polyvinylidene difluoride. 18, 53

RGI Resistance Gene Identifier. 18
RPH reductive hexulose phosphate. 18, 120

SAF-MBR staged anaerobic fluidised-bed membrane bioreactor. 18, 24, 29, 79, 80, 101, 102, 116, 132, 133
SCGs single-copy genes. 18, 63
SMX sulfamethoxazole. 18, 132, 133
SNP single-nucleotide polymorphism. 18
SRB sulfate-reducing bacteria. 18, 29, 101, 102
SRT solids retention time. 18

TCA tricarboxylic acid. 18

UASB up-flow anaerobic sludge blanket. 18, 23

UniProt Universal Protein resource. 18
UPLC ultra-performance liquid chromatography. 18
UQ ubiquinone. 18

VF virulence factors. 18
VFA volatile fatty acid. 18
VFDB Virulence Factor Database. 18

WHO World Health Organisation. 18
WWTP wastewater treatment plant. 18

Chapter 1

Introduction

1.1 Background

Global sustainability efforts underscore the urgent need to develop and adopt renewable, carbon-neutral energy sources [1]. Anaerobic digestion (AD) is widely regarded as a viable technology in this domain, as it converts organic matter into biogas under oxygen-free conditions. Frequently implemented worldwide, including in the UK, this process uses various feedstocks (e.g. food waste, sewage sludge, animal manure, and crops) to generate energy [2].

Anaerobic digestion aligns with several United Nations Sustainable Development Goals (SDGs) by addressing poverty and economic development (SDGs 1, 8, and 10), promoting clean and affordable energy (SDG 7), and fostering social progress and well-being (SDGs 2, 3, 4, 5, 9, 16, and 17) [3]. According to the International Energy Agency's *World Energy Outlook* (2020), biogas remains a niche but fast-growing component of modern bioenergy through to 2040. Under the Stated Policies Scenario, direct-use biogas is projected to reach approximately 75 Mtoe, while near-pure methane (CH_4), known as biomethane, reaches approximately 80 Mtoe. In the Sustainable Development Scenario, both expand substantially, with biomethane reaching approximately 240 Mtoe and biogas approximately 150 Mtoe (corresponding to a 40:60 ratio) [4].

In the United Kingdom, the anaerobic digestion sector produced 1.021 Mtoe in 2020. By 2023, its output of 3.5 TWh accounted for approximately 1% of total electricity generation and approximately 3% of renewable electricity production [5]. Although it still represents a relatively small share of the national energy mix, anaerobic digestion offers several advantages over intermittent renewable sources, including year-round dispatchability, diversion of food waste from landfills, and the production of nutrient-rich digestate for agricultural use. The growth potential remains substantial. An IEA geospatial analysis estimates that sustainably available residues—including crop straw, manures, municipal waste, wastewater sludge, and forestry by-products—could yield nearly 1,000 billion cubic metres per year of biogas, equivalent to approximately 730 Mtoe, sufficient to meet around one-quarter of current global natural gas demand [4].

Within wastewater treatment, anaerobic digestion has long been established as a core technology for sludge stabilisation, primarily for odour and pathogen control, volatile solids reduction,

and volume minimisation to facilitate downstream handling and agricultural reuse, with the treated digestate suitable for use as a fertiliser and soil improver [6]. More recently, anaerobic digestion technologies have gained additional traction for bioenergy recovery, contributing to renewable energy targets and decarbonisation strategies within the water sector. Despite these advantages, key challenges remain, particularly regarding bioenergetic constraints that limit conversion efficiency. Typically, CH_4 comprises 50 to 70% of the resultant biogas, with carbon dioxide (CO_2) accounting for the balance [7].

The performance of anaerobic processes largely depends on syntrophic associations among hydrolytic, fermentative, syntrophic, and methanogenic microorganisms. These groups exchange reducing equivalents via diffusible electron donors, most notably hydrogen and volatile fatty acids such as acetate. If these electron donors accumulate, the Gibbs free energy change (ΔG) for their oxidation becomes less favourable, inhibiting syntrophic metabolism [8]. Because these intermediates are acidic, their accumulation simultaneously lowers reactor pH, further suppressing methanogenic activity and, under severe imbalance, leading to digester souring [9, 10]. Consequently, electron donor accumulation constrains organic loading rates and biogas yields.

Moreover, upstream steps, such as hydrolysis and acidogenesis, may be compromised when electron donor turnover is kinetically or thermodynamically limited. This highlights the need to explore strategies that enhance electron-sharing mechanisms across multiple stages of anaerobic digestion.

1.1.1 Extracellular electron transfer promotes methane production

A promising strategy to address these limitations involves increasing extracellular electron transfer (EET) in microbial consortia. EET refers to the capacity of microorganisms to exchange electrons beyond their cell boundaries. Direct interspecies electron transfer (DIET), a specialised form of EET, enables cells to transfer electrons directly to neighbouring microbes without relying on intermediates such as hydrogen or formate. By providing a more direct mode of electron exchange, DIET can improve biogas yields and enhance the stability of microbial communities [11–13].

The first report of DIET in an anaerobic wastewater treatment process emerged from an up-flow anaerobic sludge blanket (UASB) reactor in 2010 [14]. In that system, *Geobacter* species were shown to transfer electrons directly to *Methanothrix*, augmenting methanogenesis. Subsequently, various studies demonstrated that conductive additives (e.g. granular activated carbon, GAC; magnetite; biochar; carbon cloth) can further promote DIET between exoelectrogens (e.g. *Geobacter*) and methanogens (e.g. *Methanothrix*, *Methanosarcina*), with CH_4 production increases averaging around 90% and peaking at 300% [11, 12, 15–21]. In many cases, methanogens can reduce CO_2 directly to CH_4 via EET-dependent electron uptake, effectively bypassing hydrogenotrophic and acetoclastic routes [22–24].

EET-Dependent CO_2 Reduction Pathway (EET-Methanogenesis):

$$\mathrm{CO_2} + 8\,\mathrm{H^+} + 8\,\mathrm{e^-} \longrightarrow \mathrm{CH_4} + 2\,\mathrm{H_2O} \quad \Delta G' \approx -150 \text{ kJ/mol } \mathrm{CH_4}\,^{[a]} \tag{1.1}$$

Hydrogenotrophic Methanogenesis:

$$\mathrm{CO_2} + 4\,\mathrm{H_2} \longrightarrow \mathrm{CH_4} + 2\,\mathrm{H_2O} \quad \Delta G^\circ \approx -131 \text{ kJ/mol } \mathrm{CH_4}\,^{[b]} \tag{1.2}$$

Acetoclastic Methanogenesis:

$$\mathrm{CH_3COOH} \longrightarrow \mathrm{CH_4} + \mathrm{CO_2} \quad \Delta G^\circ \approx -31 \text{ kJ/mol } \mathrm{CH_4} \tag{1.3}$$

Annotations:

[a] Estimated under *in situ* electroactive biofilm conditions assuming an electron donor potential of $E_{\text{donor}} = -0.30$ V vs SHE and a CO_2/CH_4 acceptor potential of $E_{\text{acceptor}} = -0.24$ V, corresponding to $\Delta G' = -nF(E_{\text{acceptor}} - E_{\text{donor}}) \approx -150$ kJ/mol CH_4.

[b] Calculated under standard biochemical conditions ($25\ ^\circ$C, 1 atm, pH 7, $p_{\mathrm{H_2}} = 1$ atm).

Because EET circumvents the accumulation of dissolved intermediates, it provides a kinetic advantage over hydrogen-mediated electron transfer: DIET can achieve an order-of-magnitude higher rate of electron transport [25]. Additionally, by avoiding hydrogen as an obligate intermediate and operating at higher effective redox potentials (typically -0.30 V vs SHE rather than the H^+/H_2 couple at -0.414 V), EET-based pathways can yield a more thermodynamically favourable route to CH_4 production under *in situ* conditions, with a reaction free energy approaching -150 kJ mol^{-1} (as compared to approximately -131 kJ mol^{-1} for hydrogenotrophic and -31 kJ mol^{-1} for acetoclastic methanogenesis) [26–28].

1.1.2 Pilot-scale staged anaerobic fluidised-bed membrane bioreactors as a study ground for direct interspecies electron transfer application

Although laboratory studies have illustrated DIET's potential to enhance methanogenesis, scaling up remains challenging. Incorporating conductive media can be cost-prohibitive and prone to material losses during long-term operation. For example, Zhang et al. [29] reported a 40% CH_4 production increase in a pilot-scale digester amended with GAC, but encountered daily GAC losses of 0.345 kg d^{-1}. Replacing lost GAC at 0.12–0.25 £ $gVSS^{-1}$ day^{-1} consumed 38.3–81.8% of the CH_4's potential revenue (approx. 0.28 pounds $gVSS^{-1}$ day^{-1}).

By contrast, pilot-scale staged anaerobic fluidised-bed membrane bioreactors (SAF-MBR) have demonstrated promising energy yields from low-strength wastewater [30–32]. Shin et al. [32] evaluated a SAF-MBR 2.0 system treating municipal wastewater. Here, SAF-MBR 2.0 denotes a second-generation staged configuration in which GAC was retained within an anaerobic fluidised-bed membrane zone, allowing sustained biofilm growth on a conductive carrier. The membrane prevented GAC washout, thereby maintaining high biomass retention and prolonged microbe-carrier contact, conditions favourable for syntrophic metabolism and methanogenesis.

Under this configuration, accelerated degradation kinetics were observed for acetate, propionate, glucose, and primary sludge. The reactor achieved significantly lower half-saturation

coefficients (3.05 mg COD L^{-1} for acetoclastic methanogenesis and 2.7 mg COD L^{-1} for propionate degradation) relative to conventional high-strength AD systems [33]. These reduced *K* values correspond to improved substrate uptake rates and undetectable levels of effluent acetate and propionate, substantially exceeding earlier performance benchmarks [31, 34–36].

Such rapid kinetics likely reflect DIET facilitated by GAC. In a prototype SAF-MBR fed with acetate and propionate, *Methanothrix*, *Geobacter*, and syntrophic bacteria (*Syntrophobacter*, *Smithella*) were all implicated in accelerating the methanogenic phase [24]. Importantly, DIET-driven benefits may extend beyond methanogenesis to upstream stages, including hydrolysis, acidogenesis, and acetogenesis [12, 37]. This is significant because hydrolysis is widely considered the primary rate-limiting step in anaerobic digestion, particularly for complex and particulate substrates such as primary sludge. Enhancement of electron transfer during these earlier stages may therefore alleviate kinetic bottlenecks that constrain overall process throughput.

Moreover, the discovery of electrically conductive pili (e-pili) across diverse bacterial taxa underscores the possibility of widespread DIET activity in both natural and engineered systems [38]. Beyond *Geobacter*, several other bacteria (e.g. *Acetobacterium*, *Thermincola*, *Syntrophus*, *Desulfovibrio*) may participate in electromethanogenic or DIET-enabled metabolisms, expanding the functional relevance of DIET beyond the classical methanogen-*Geobacter* paradigm [39, 40]. While the exact electron uptaking mechanism of *Methanothrix* remains to be explored, major sheath proteins (MspA) are a strong candidate that has been shown to be electrically conductive [41, 42].

Overall, these findings suggest a broader functional role for EET beyond methanogenesis, potentially supporting hydrolysis and acidogenesis under varying loading and environmental conditions. This expansion of EET's role warrants further study, particularly in industrial-scale reactors where process stability and substrate diversity are paramount.

Bioinformatics serves as a cost-effective tool for detecting EET-related activities in such systems. Low-cost 16S rRNA gene amplicon sequencing can reveal the presence of putative EET-capable taxa [43, 44]. At the same time, metagenomics and metatranscriptomics enable functional inference by linking genes to established EET pathways [23, 24, 45–47]. However, reference databases such as NCBI [48], KEGG [49], and InterPro [50] provide incomplete coverage of membrane-bound EET proteins, which are inherently difficult to purify and experimentally characterise. Moreover, functional convergence among EET proteins often occurs in the absence of strong sequence homology, limiting the effectiveness of conventional reference-based annotation.

To overcome these constraints, this study employs complementary rule-based and motif-driven bioinformatics strategies to infer EET potential from conserved biophysical and structural features rather than relying solely on sequence similarity. Specifically, EET candidates were identified using (i) motif-based searches with custom profile hidden Markov models (pHMMs) [51, 52], and (ii) rule-based frameworks grounded in established physical criteria for multiheme c-type cytochrome (MHC), electrically conductive pili, and MspA-like sheath proteins. This approach enables systematic discovery of putative EET genes beyond the limits of existing annotations and forms a central methodological question addressed in this thesis.

1.2 Research gaps

Research gap 1: Effective bioinformatics methods for identifying EET in AD microbiomes No standardised methodology exists for detecting EET or DIET in microbial communities. Discrepancies in bioinformatics workflows can drastically alter annotation outcomes. Further, the diversity of EET genes in methanogens and other bacteria remains underexplored [53]. Establishing a customised bioinformatics pipeline for *de novo* analysis would enable comprehensive identification of EET in AD microbiomes, particularly where reference sequences are scarce.

Research gap 2: EET-driven organic waste degradation across AD stages Although most EET research has focused on methanogenesis, evidence suggests that EET can also accelerate upstream processes such as hydrolysis, acidogenesis, and acetogenesis by mitigating the accumulation of inhibitory intermediates. Understanding whether EET-driven electron transfer can enhance the breakdown of complex organic substrates (e.g. primary sludge) would support the development of more robust and efficient AD systems.

Research gap 3: Genomic basis of EET capacity in methanogenic lineages While methanogens are increasingly implicated as key electron acceptors in EET- and DIET-enabled anaerobic systems, the genetic basis underpinning this capacity remains poorly resolved. In particular, it is unclear which EET-associated genes are consistently encoded within dominant methanogenic lineages, how these genes are distributed across strains, and whether they constitute core, lineage-defining traits or sporadic accessory features.

Putative EET-associated proteins such as MHC and MspA-like sheath proteins have been proposed in methanogens, yet their prevalence, conservation, and genomic organisation within ecologically relevant taxa such as *Methanothrix* remain insufficiently characterised. Resolving the pangenomic structure of EET capacity in methanogens is essential for linking observed reactor-scale performance to microbial genetic potential.

Research gap 4: Evolutionary stability versus mobility of EET genes in engineered anaerobic systems Biofilm-dominated anaerobic reactors, such as SAF-MBRs, promote dense microbial assemblages and frequent phage–host interactions, thereby creating conditions conducive to horizontal gene transfer (HGT). In such environments, mobile gene classes such as antibiotic resistance genes (ARGs) are readily disseminated via plasmids and phages. However, it remains unclear whether EET-related genes in methanogens exhibit similar evolutionary mobility or instead represent stable, lineage-conserved traits.

Direct comparisons between the evolutionary behaviour of EET genes and demonstrably mobile gene classes within the same engineered system are scarce. Determining whether enrichment of EET-capable methanogens arises through horizontal gene transfer or through ecological selection of pre-adapted lineages is critical for assessing the robustness and predictability of EET-enhanced anaerobic digestion.

Overall, advancing our understanding of EET in anaerobic wastewater treatment can improve system design and operation, with potential benefits for waste management, bioremediation, and bioenergy production. Such optimisation of EET processes contributes to global sustainability goals, aligning with efforts toward cleaner energy (SDG7), economic growth (SDG8), and responsible consumption (SDG 12).

1.3 Aims and objectives

1.3.1 Aims

The aim of this research is to demonstrate how EET enhances the performance and stability of anaerobic digestion by establishing the genetic basis of EET capacity in methanogenic communities, characterising its ecological associations beyond methanogenesis, and comparing the evolutionary behaviour of EET-related genes with other mobile gene classes in a pilot-scale staged anaerobic fluidised-bed membrane bioreactor.

The overarching aim of this research is to determine how EET structures carbon and electron flow in engineered methanogenic systems, and to establish the extent to which EET contributes to enhanced substrate conversion efficiency and functional stability in anaerobic wastewater treatment.

Specifically, this thesis aims to:

1. establish whether EET-enabled microbial interactions extend beyond methanogenesis to influence upstream processes of hydrolysis, acidogenesis, and syntrophic oxidation in low-strength anaerobic systems;

2. determine the dominant microbial guilds and electron transfer mechanisms responsible for EET under pilot-scale operating conditions, with particular emphasis on biofilm-associated communities;

3. elucidate the genomic and structural basis of electron uptake in *Methanothrix*, including the identification and functional interpretation of previously uncharacterised sheath-associated proteins;

4. assess the implications of EET-driven metabolism for reactor performance, stability, and design, beyond taxonomic enrichment-based interpretations.

1.3.2 Objectives

Figure 1.1 illustrates the research questions and objectives of this study.

1. Develop and apply a rule-based bioinformatics framework for identifying EET genes in anaerobic digestion microbiomes by integrating motif-based searches and physically informed

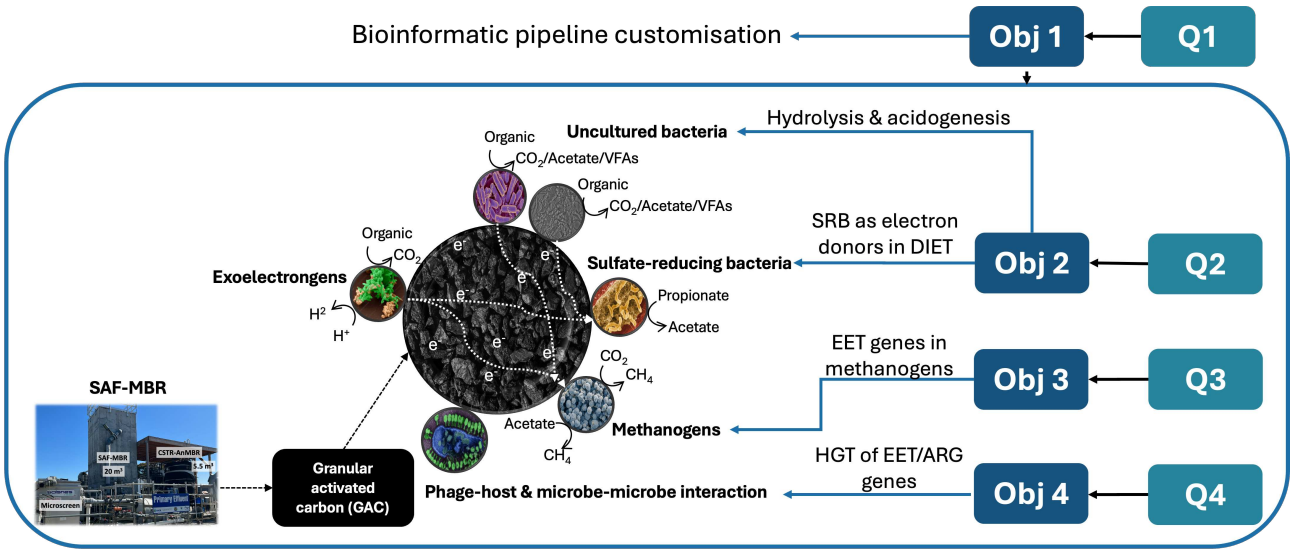


Figure 1.1: An overview of the research objectives based on observations from the SAF-MBR.

gene-inventory rules to detect MHC, e-pili, and MspA-like sheath proteins beyond the limits of reference-based annotation.

2. Characterise the functional roles and ecological distribution of EET-capable microbial guilds across multiple stages of anaerobic digestion by linking EET gene enrichment to uncultured fermentative bacteria, sulfate-reducing bacteria acting as electron donors, and methanogens acting as electron sinks within a pilot-scale SAF-MBR.

3. Resolve the genomic basis of EET capacity in dominant methanogenic lineages by constructing pangenomes (e.g. of *Methanothrix*) and mapping the distribution of EET-associated genes, including MspA-like sheath proteins, to determine whether EET capacity constitutes a lineage-conserved core trait or an accessory genomic feature.

4. Assess the evolutionary stability of EET-related genes relative to antibiotic resistance genes by analysing plasmid- and phage-associated sequences within the SAF-MBR to contrast lineage-conserved EET traits with demonstrably mobile gene classes under identical ecological and phage-mediated selective pressures.

1.4 Research framework

This study comprises three main components: (i) a comprehensive dataset, (ii) a customised computational pipeline, and (iii) targeted data analyses. The dataset combines metagenomic and metatranscriptomic data from the SAF-MBR and other digesters supplemented by publicly available databases. The computational pipeline includes data preprocessing, metagenomic/metatranscriptomic analyses, and investigations of phages and plasmids, aiming to identify DIET-related genes, antibiotic resistance factors, and microbe-phage interactions. Finally, the analyses address the specific objectives of identifying EET genes, characterising their roles in AD, and elucidating mechanisms of gene transfer.

1.5 Thesis structure

This thesis is organised into seven main chapters and several appendices, as follows:

- **Chapter 2: Literature Review** Provides a detailed examination of the fundamental principles and previous findings in anaerobic digestion. It covers the mechanisms of mediated and direct interspecies electron transfer, the role of conductive materials, and possible reasons for DIET's stabilising effects. Additional sections delve into biological processes of EET, including electron-export mechanisms in exoelectrogens and electron-import pathways in methanogens, as well as key proteins (MHC, e-pili, and amyloids) involved in EET. The chapter concludes by addressing HGT dynamics, particularly with respect to EET genes and phages.

- **Chapter 3: Materials and Methods** Describes the technical details of the SAF-MBR systems used in this study (SAF-MBR 1.0 and SAF-MBR 2.0), including reactor configurations and sampling protocols. The chapter also presents methodologies for processing genomic and transcriptomic data, including metagenome assembly and binning, refinement of metagenome-assembled genomes, taxonomic profiling, phylogenetic analyses, functional annotation, and identification of EET genes. Finally, it introduces pangenomic and phage analyses, explaining how HGT events and antibiotic resistance genes were investigated.

- **Chapter 4: Diversity of Genes Related to Extracellular Electron Transfer and Their Significance in Carbon Degradation** Explores the distribution of EET-related genes in various microbial communities reconstructed from metagenomic data, focusing on SAF-MBR samples. The chapter evaluates correlations between EET gene abundance and reactor operating conditions and examines how these genes may influence carbon-degradation pathways. It also discusses the effectiveness of rule-based approaches for improving the detection and annotation of EET genes.

- **Chapter 5: Direct Interspecies Electron Transfer Driven Syntropy between Sulfate-reducing Bacteria and *Methanothrix* in Pilot-scale Reactor** Investigates how DIET contributes to syntrophic interactions in pilot-scale systems, particularly those involving sulfate-reducing bacteria (SRB) and *Methanothrix*. Metagenomic and metatranscriptomic data elucidate the metabolic and ecological dynamics of these consortia, highlighting the role of GAC in facilitating electron transfer and stabilising reactor performance.

- **Chapter 6: Identification of Extracellular Electron Transfer Sheath Proteins in *Methanothrix* Using Pangenomic Analysis** Examines core and accessory genes in *Methanothrix* via pangenomic analysis, pinpointing potential major sheath proteins involved in EET. The chapter discusses structural proteins that could form electrically conductive sheaths and addresses how these findings expand our understanding of the EET pathways and their applications in biotechnological processes.

- **Chapter 7: Horizontal Gene Transfer of Antibiotic Resistance Gene in SAF-MBRs** Focuses on mobile genetic elements, particularly phages and plasmids, in the SAF-MBR microbiome. While initial hypotheses targeted EET-related HGT, the results instead highlight ARG flows, including the *sul4* gene, which confers resistance to sulfamethoxazole. This chapter discusses the prevalence, distribution, and potential transfer of ARGs, as well as their broader ecological implications for wastewater treatment systems.
- **Chapter 8: General Discussion** Integrates the experimental and multi-omics findings across reactor performance, microbial community structure, and genomic analyses to develop a system-level interpretation of extracellular electron transfer in engineered methanogenic environments.
- **Chapter 9: Conclusions** Summarises the principal scientific contributions of the thesis in relation to the stated aims, and identifies key limitations and future research directions relevant to the design and operation of EET-enabled anaerobic treatment systems.
- **Appendices** Contains supplementary information, extended data sets, and detailed protocols.
  - **Appendix A** includes supplementary notes on kinetic modelling and metagenome refinement.
  - **Appendices B & C** provide additional tables and figures supporting data interpretations throughout the thesis.

Chapter 2

Literature Review

2.1 Anaerobic digestion

Anaerobic digestion, a process that converts organic waste i into biogas containing approximately 50–70% CH_4, is a widely adopted bioenergy strategy with a history dating back centuries [54–56]. However, in full-scale operation the fraction of substrate carbon ultimately recovered as CH_4 remains well below its theoretical biochemical-CH_4 potential because hydrolysis and acidogenesis are slow and readily inhibited [10, 55, 57]. To enhance efficiency, strategies such as optimising environmental conditions within digesters, adjusting temperature and pH, and using microbial additives [58, 59], as well as influent pre-treatments [58, 60, 61], have been implemented.

Recent innovations include integrating conductive materials to promote and stabilise methanogenesis [54, 55]. These materials facilitate EET, altering microbial metabolism and improving organic degradation.

2.1.1 AD via mediated interspecies electron transfer

AD is a biodegradation process driven by a diverse group of Archaea and Bacteria that converts complex organic materials into CH_4 via a four-stage pathway: hydrolysis, acidogenesis (fermentation), acetogenesis, and methanogenesis. Figure 2.1 shows a simplified schematic of the main metabolic pathways in AD.

Polymeric organic compounds are hydrolysed into simpler monomers for methanogenesis. In conventional anaerobic digestion systems, acetoclastic methanogens primarily metabolise mono-methyl compounds and acetate. Acetate is both electron donor and acceptor in this pathway. By splitting acetate into methyl and carboxyl groups, electrons from the carboxyl group (COO-) was used to produce CH_4 through decarboxylation. Enzymes like carbon monoxide dehydrogenase/acetyl-CoA synthase (CODH/ACS) catalyse this electron transfer, which is essential for converting the methyl group to CH_4. Conversely, hydrogenotrophic methanogenesis reduces CO_2 to CH_4 using electrons from molecular hydrogen (H_2). This pathway involves converting CO_2 to formate and subsequently to CH_4 through a series of reductions mediated by coenzymes such as F_{420}.

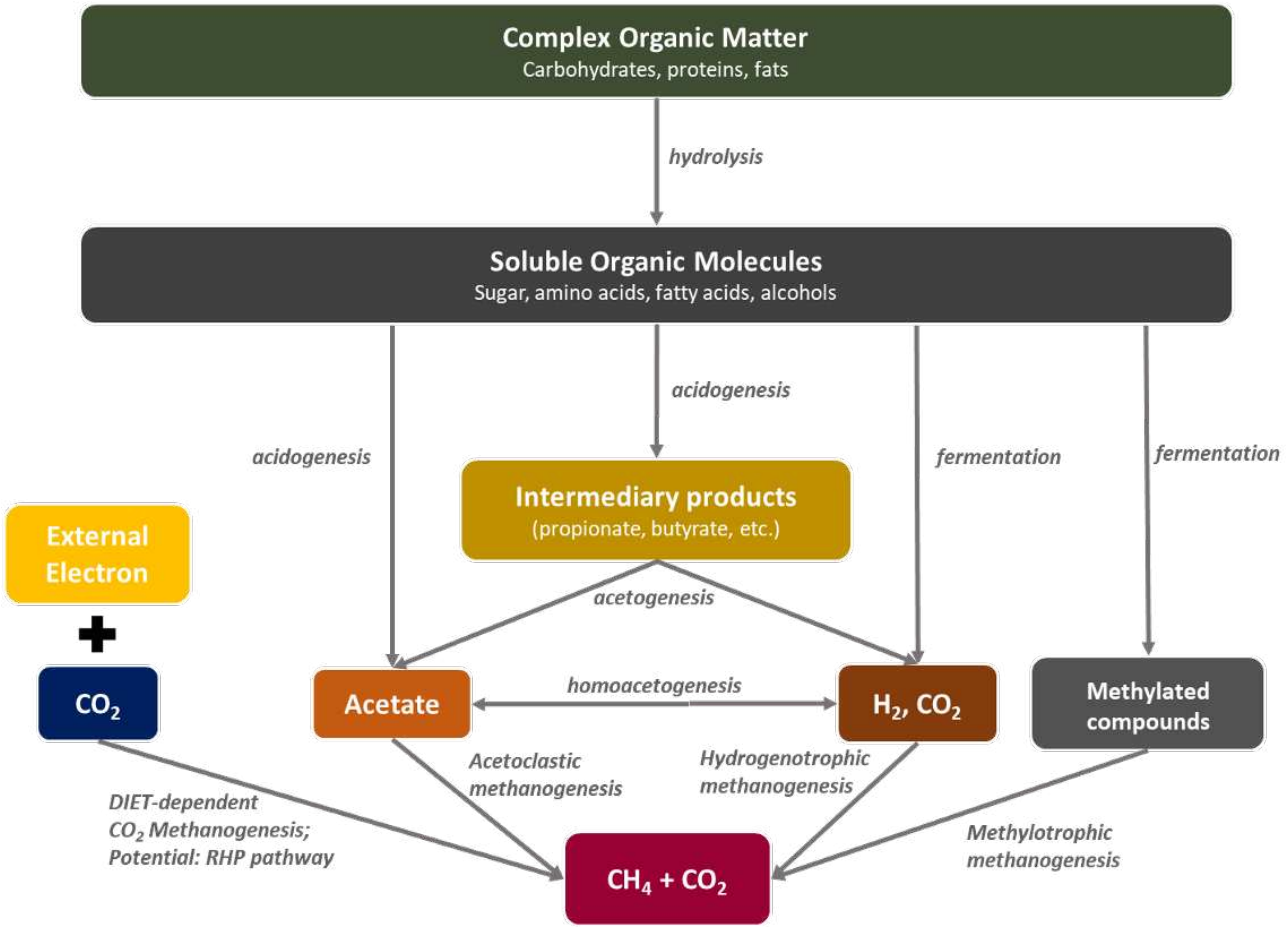


Figure 2.1: Simplified view of the major metabolic pathways in AD

Interspecies hydrogen transfer, the process by which bacterial H_2 production from organic substrates and its consumption by methanogens facilitate electron transfer, has been a part of the fundamental understanding of methanogenic systems for over five decades [8, 62, 63]. Research has predominantly focused on the metabolism of non-fermentable substances, such as fatty acids, alcohols, and aromatic compounds. This focus is due to the thermodynamic feasibility of H_2 production from these materials only when hydrogenotrophic methanogens efficiently consume H_2 to keep its concentration low ($<$10 Pa) [8].

Systems reliant on interspecies hydrogen transfer are prone to disruption due to the inherently slow electron transfer via H_2 diffusion and low solubility. When H_2 production outpaces consumption, volatile-fatty-acid (VFA) accumulation leads to acidification of the bulk phase, creating conditions that inhibit *methanogens* and precipitate system failure. Currently, no engineering solutions exist to expedite interspecies hydrogen transfer. The role of conductive particles in enhancing CH_4 production remains unclear, particularly concerning their influence on interspecies hydrogen transfer dynamics.

2.1.2 Anaerobic digestion via direct interspecies electron transfer

Extracellular electron transfer is a bidirectional mechanism that facilitates the movement of electrons both into and out of a cell across its membrane. This process is essential for electroactive bacteria, also known as exoelectrogens. Exoelectrogens can transfer electrons from internal metabolic processes to insoluble oxides in their external environment, such as Fe(III) and Mn(III&IV), or accept electrons from external sources. This capability offers a versatile alternative to traditional diffusion-based electron transfer methods, such as those mediated by H_2 or formate [64]. EET is particularly significant in enhancing the efficiency of AD processes.

The primary advantage of EET lies in its efficiency: methanogens can directly uptake extracellular electrons from exoelectrogens to transform CO_2 into CH_4 (Reaction 1), bypassing the need for chemical electron donors like H_2 (Reaction 2) or acetate (Reaction 3) [23, 24]. This process is not only more direct but also significantly faster in electron transfer. The CO_2 reduction methanogenesis induced by DIET is independent of H_2, unlike hydrogenotrophic methanogenesis. Control experiments using gene deletions of membrane-bound hydrogenase in methanogen mutants (*Methanococcus maripaludis*: *fru*, *frc*, *hmd*, *vhu*, *vhc*, and *ehb*; *Methanosarcina barkeri*: *ech*, *vhu*, and *fhr*) did not impede methanogenesis, suggesting electromethanogenesis is independent of H_2 [65, 66].

Reaction 1. EET-dependent CO_2 reduction pathway (electromethanogenesis):

$$CO_2 + 8H^+ + 8e^- \rightarrow CH_4 + 2H_2O$$

Reaction 2. Hydrogenotrophic methanogenesis:

$$CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O$$

Reaction 3. Acetoclastic methanogenesis:

$$2CH_3COOH \rightarrow 2CH_4 + 2CO_2$$

Electromethanogenesis may occur in the anaerobic suspended aggregate composed of

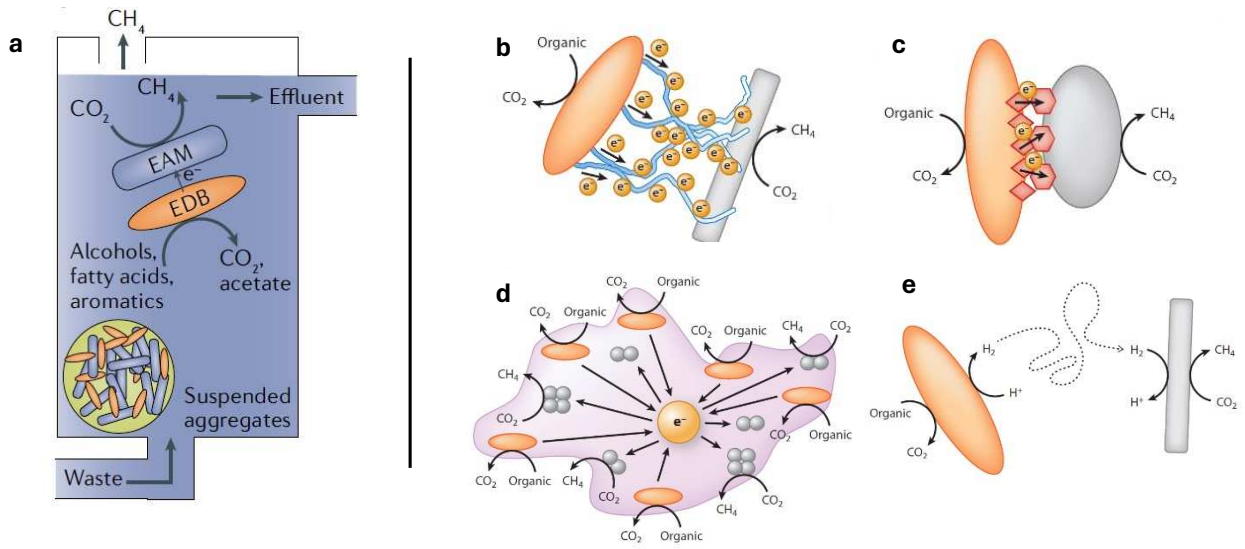


Figure 2.2: Illustrations of electron transfer modes of CO2 reduction pathway in anaerobic treatment. (a) the EET between the electron-accepting methanogens (EAM) and electron-donating bacteria (EDB). (b-e) Orange-coloured cells are EDB, and grey-coloured cells are EAM.; (b) EET via conductive pili (blue); (c) EET via direct contact of EET outer membrane proteins (red); (d) EET via conductive materials mediation (purple); (e) H_2-MET. Images are adapted from [69] and [44]

methanogens and exoelectrogens, which can be commonly observed in up-flow anaerobic sludge blanket systems (Fig. 2.2a) [14]. Electrons in aggregates are extracellularly transferred through conductive pili like PilA (Fig. 2.2b) or EET outer membrane proteins (Fig. 2.2c) like MHC OsmC in *Geobacter*. In the presence of conductive materials, however, aggregate is not formed. Instead, the methanogens and exoelectrogens preferentially form biofilm on the materials, performing EET without direct cell-to-cell contact (Fig. 2.2d) [17, 67, 68].

Conductive materials act as the conductive conduit and significantly promote the occurrence of EET-methanogenesis in anaerobic treatment. In such mediated treatment, EET is likely the predominant electron transfer mode for the CO_2 reduction pathway, as it is more kinetically (8-fold of electron transfer rate) and energetically (1.5-fold) advantageous than H_2-MET (Fig. 2.2e) [25].

2.1.3 Conductive materials synergies with DIET

Over the past decade, the inclusion of conductive materials has been extensively tested with various feed substrates, ranging from domestic to industrial wastewater, in laboratory-scale anaerobic treatment experiments [15, 16, 30, 70].

Conductive materials such as GAC, biochar, and magnetite have significantly enhanced biogas production, with increases up to 300% and an average improvement of 92.5 observed across more than 90 studies [15, 16, 30, 70]. Performance gains were achieved through accelerated methanogenesis, which permits higher organic loading and, in some cases, greater volatile-solids destruction, indicating more complete substrate conversion. [71] summarised the effect of various non-biological conductive materials on DIET in AD. Among the twenty-one conductive materials reviewed, GAC was considered to be the most viable due to low cost and the greatest mean

improvement in CH_4 production. Specifically, the use of GAC in a pilot-scale SAF-MBR led to CH_4 yield increases of 33 to 53% (0.2 to 0.23 $L\,CH_4\,gCOD^{-1}$), a substantial improvement over conventional treatments (0.15 $L\,CH_4\,gCOD^{-1}$) [30, 72]. This suggests that the benefits of conductive materials in enhancing biogas production are scalable. The pilot-scale SAF-MBR involved in this study was functional during the winter in South Korea (13°C) with 76% of COD removal rate [31].

In addition, conductive materials have proven effective in overcoming various environmental challenges that typically hinder CH_4 production. They are particularly beneficial in low pH conditions, a common obstacle in CH_4 generation. These materials increase CH_4 production in acidic environments and accelerate CH_4 generation after acidity levels are neutralised [73–75]. High ammonium concentrations, often resulting from high protein influent, which negatively impact anaerobic digestion, are also mitigated by the presence of conductive materials [76–79]. Furthermore, conductive materials provide a buffer against the inhibitory effects of high organic loading, which can lead to the accumulation of organic acids and ammonia, thus protecting the anaerobic digestion process from these metabolic imbalances [21, 80–82]. One possible explanation is that the biofilm formation enabled by the conductive materials provided a protective layer to the microbes against the ammonium stress [83].

In addition, extensive laboratory-scale anaerobic treatment experiments with different feed substrates (from domestic to industrial wastewaters) and conductive biocarriers (e.g., GAC, biochar, magnetite) and our pilot-scale operation had demonstrated substantial improvement in CH_4 production through electromethanogenesis (up to 3-fold) [15, 16, 30, 31, 70].

2.1.4 Potential reason for DIET's stabilising effect

Since the mechanisms of electron transfer routes with DIET remain relatively undefined, the reasoning for the effects of acceleration and stabilisation in anaerobic digestion obtained with DIET remains somewhat speculative. The potential energy yield from the overall reaction of organic substrates to CH_4 is equal to DIET, as well as when H_2/formate participates as electron carriers. The enhanced energy conservation available to DIET partners is likely due to the more direct electron transfer through DIET, which involves fewer steps of electron transfer with fewer energetic losses [84]. However, primary comparative studies on energy conservation and cell yield that support this speculation have not yet been conducted [44]. Efficient energy conservation is not a sufficient condition for DIET adoption [84], as DIET partners must invest energy in establishing extracellular electrical connections through protein synthesis. The cost of manufacturing the extracellular proteins can be reduced by adding conductive materials [85].

Once interspecies electrical connections have been established, electron transfer through DIET should be faster than diffusive electron exchange via soluble electron shuttles. This would also make improvements in the accumulation of H_2, possibly lowering its sensitivity [27, 80, 86–88]. The potential is particularly well-suited for DIET applications when organic loading rates are high in both wastewater and high solid system (OLR range of 2 to 6 $g\,VS\,L^{-1}\,d^{-1}$ [89]) and fermentable substrates are metabolised more rapidly, with the simultaneous production of high

levels of VFAs and H_2. Such benefits can be quantified using mathematical models that compare DIET performance with interspecies H_2 transfer under different environmental conditions and account for conductive material amendments, a valuable tool for anaerobic digester design once more data are available. However, there is currently a lack of data on fundamental parameters necessary for building a reliable/predictive model [44].

From a physical standpoint, the inhibitory effect of high partial hydrogen pressures (dissolved H_2) on VFA metabolism can be attributed to suppressing hydrogenase activity. High hydrogen concentrations induce feedback inhibition, diminishing enzyme efficiency by shifting the equilibrium of hydrogenase-catalysed reactions towards hydrogen formation, thereby impeding the metabolic utilisation of hydrogen [90]. Furthermore, increased hydrogen pressures make the methanogenesis thermodynamically unfavourable, inhibiting cellular energy production and the metabolic conversion of hydrogen into usable energy [91].

2.1.5 DIET effect on organic waste degradation

Conductive materials have been highlighted for assisting acetate methanogenesis through two major pathways [79], namely acetoclastic methanogenesis and syntrophic acetate oxidation. In acetoclastic methanogenesis, acetate is directly converted to CH_4 and CO_2 through acetate kinase and acetate-CoA ligase activities in organisms like *Methanosarcina* and *Methanothrix* [92]. The syntrophic oxidation pathway, conversely, consists of the conversion of acetate to H_2 and CO_2, thereafter to CH_4 with the synergistic action of acetate-oxidising bacteria and hydrogenotrophic methanogens. This interaction is notably enhanced in the presence of magnetite at thermophilic conditions [93], and a similar increase is observed in syntrophic methanogenesis via propionate and/or butyrate oxidation due to conductive materials [94, 95]. Thermodynamically, the oxidation of propionate and/or butyrate to acetate and H_2/formate is challenging, and this is made possible only when intermediates are maintained at low concentrations.

Conductive materials mediated DIET is hypothesised to enhance the degradation of complex organic wastes such as leachate and dog food, which showed rapid production and consumption of volatile fatty acids and efficient CH_4 production [96–98]. Furthermore, DIET in AD of real organic wastes is not solely mediated by *Geobacter* spp. A broader spectrum of exoelectrogenic bacteria participates in DIET [99]. This diversity highlights a more intricate DIET mechanism during the degradation of complex organics (Fig. 2.3). Interestingly, while fermentative bacteria are commonly used to describe bacteria that degrade complex organic waste anaerobically, EET is a type of cellular respiration, not fermentation. Given the prevalence of complex organics in practical AD processes, it is essential to investigate further how substrate composition influences the enrichment of various DIET-capable microorganisms.

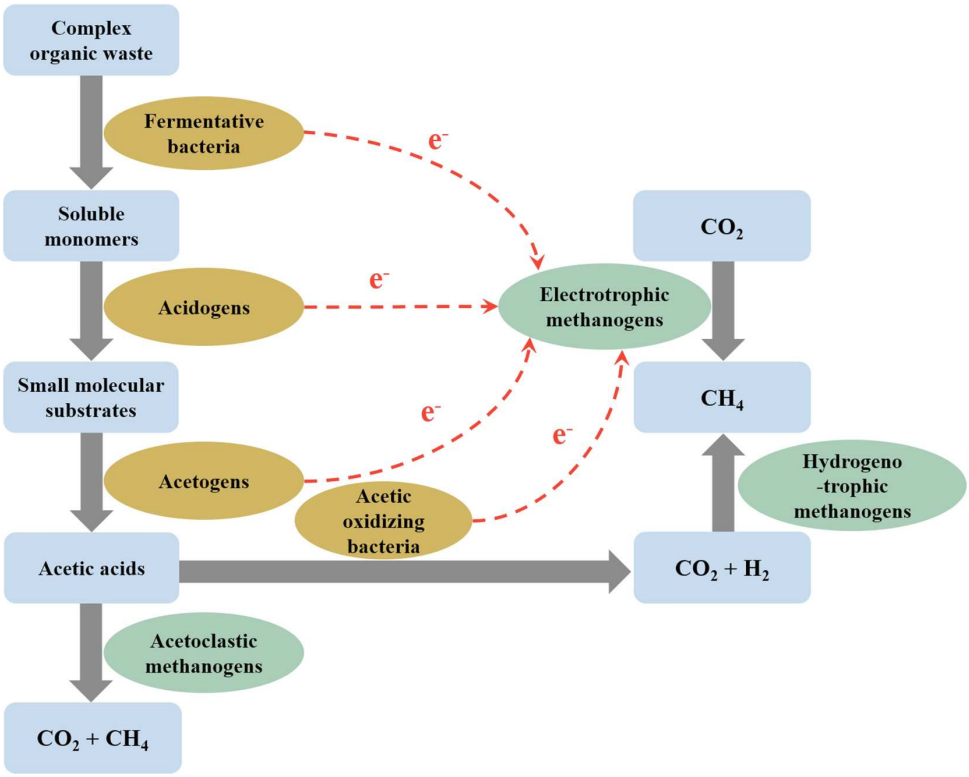


Figure 2.3: The proposed mechanism for augmenting anaerobic digestion of complex organic waste via DIET is illustrated, with the red arrow indicating the DIET pathway (adapted from [12]).

2.1.6 Literature-based non-conductive reference communities for comparison to GAC-associated microbiomes

A fully matched, pilot-scale control reactor without conductive media was not available during the SAF-MBR campaign. To provide context for assessing the influence of conductive carriers on microbial community structure, particularly on the enrichment of EET and DIET associated guilds, a literature-based reference set was assembled. This reference comprises three AFMBR studies that employed PVDF membrane modules in combination with non-conductive or low-conductivity fluidised carriers. Each study reported 16S rRNA gene-based community profiles during periods of stable reactor operation [100–102]. Together, these studies offer a pragmatic benchmark for evaluating whether conductive materials selectively enrich microbial taxa associated with syntrophic metabolism, acetoclastic methanogenesis, and putative DIET interactions—beyond the baseline established by non-conductive systems.

In one AFMBR operated with PVDF carriers, genus-level 16S rRNA data revealed the consistent presence of propionate-oxidising syntrophic taxa on the biofilm, with *Syntrophobacter* detected at 1.53% to 5.58% and sulfate-reducing *Desulfobulbus* observed at 3.21% to 4.51% on the PVDF media during later stages of operation [100]. A notable temporal shift was reported in the methanogenic community: early samples were dominated by hydrogenotrophic *Methanolinea* (>60% of archaeal reads). In contrast, acetoclastic *Methanothrix* accounted for over 70% at later time points, suggesting a transition toward acetate-driven methanogenesis as the reactor stabilised [100]. Importantly, *Geobacter*, a canonical exoelectrogen, was not detected, indicating

that its enrichment is not an inevitable feature of polymeric-carrier AFMBRs, even when conditions support stable syntrophic metabolism and methanogenesis.

A second AFMBR study directly compared conductive GAC to non-conductive acrylonitrile butadiene styrene (ABS) beads within the same reactor design, enabling a side-by-side evaluation of syntrophic and methanogenic taxa across carrier types [102]. In the ABS system, *Syntrophobacter* and *Geobacter* were detected at 2.01% and 0.65% relative abundance, respectively, substantially lower than in the GAC condition. The total syntrophic fraction in the ABS carrier remained modest (approximately 3.15%) [102]. Despite this, the archaeal community was dominated by *Methanothrix* (38.32%), confirming that acetoclastic methanogenesis can prevail even in the absence of conductive carriers. However, the reduced presence of syntroph/exoelectrogen-associated genera in the ABS condition was interpreted as indicative of fewer DIET-compatible interactions than in GAC. This within-study comparison offers a valuable internal control and suggests that conductive carriers can preferentially enrich microbial guilds associated with electron exchange and syntrophic metabolism under otherwise similar operating conditions.

A third study investigated community differences between polyaniline (PANI)-coated conductive media and uncoated polymeric carriers in an AFMBR treating low-strength wastewater [101]. Genus-level analysis revealed shifts in both bacterial and archaeal communities consistent with enhanced DIET selection in the conductive system. Across all conditions, *Methanothrix*, *Methanoregula*, and *Methanolinea* were dominant archaeal genera. Notably, *Methanothrix* was more abundant on the conductive-coated carriers (56.2% to 60.0%) compared to uncoated PVDF 37.1% on the PVDF-media and 31.5% in the PVDF-bulk [101]. Conversely, hydrogenotrophic *Methanoregula* exhibited higher relative abundance in the uncoated PVDF system (e.g., 49.4% on PVDF-media), suggesting a shift in methanogenic pathways. The bacterial community on conductive media was also enriched in propionate-oxidising syntrophs, such as *Cryptanaerobacter* (23% on the coated biofilm vs. 10.69% on the uncoated PVDF biofilm). Despite the conductive coating, *Geobacter* was not detected, reinforcing the notion that conductive materials can facilitate DIET-favourable conditions without necessarily promoting the growth of canonical exoelectrogens [101].

Taken together, these studies provide a robust literature-based reference for interpreting microbial community shifts observed in the SAF-MBR reactors. They support the use of a two-tiered control framework in this thesis, in which: (i) within-system comparisons between bulk and GAC-associated communities are used to assess localised enrichment patterns, and (ii) external references based on non-conductive carrier systems contextualise whether microbial taxa and functional guilds linked to syntrophy and acetoclastic methanogenesis are disproportionately enriched in conductive-carrier environments relative to polymeric baselines [100–102].

2.1.7 Validation of DIET occurrence

The application-oriented experiments were designed to evaluate performance enhancement achieved by the addition of conductive materials. Often, the observed efficiency improvements were attributed to DIET, as suggested by microbial community analysis. However, these analyses

Table 2.1: Summary of genus-level 16S rRNA trends reported in PVDF-tagged AFMBR studies used as non-conductive (or paired-carrier) reference conditions for community comparison. Reported values are those explicitly stated in the cited manuscripts.

Study / carrier comparison	**Bacterial guilds highlighted (genus level)**	**Archaeal trends highlighted (genus level)**
PVDF polymeric media AFMBR [100]	Propionate-oxidising syntrophs present; *Syntrophobacter* (1.53% to 5.58%); *Desulfobulbus* (3.21% to 4.51%); *Geobacter* negligible	Temporal shift from *Methanolinea* (>60%, earlier) to *Methanothrix* (>70%, later)
GAC (conductive) vs ABS beads (non-conductive) [102]	ABS: lower *Syntrophobacter* (2.01%) and *Geobacter* (0.65%) than GAC; total syntrophic fraction ∼3.15% in ABS	ABS dominated by *Methanothrix* (38.32%); interpreted as weaker DIET-favourable enrichment than GAC
PANI-coated vs uncoated polymeric media [101]	Coated biofilm enriched in syntroph-associated taxa; *Cryptanaerobacter* (23% coated vs 10.69% uncoated PVDF biofilm); *Geobacter* not detected	*Methanothrix* higher on coated media (56.21% to 60.0% reported) than uncoated PVDF (e.g., 37.1% PVDF-media); *Methanoregula* higher in uncoated PVDF (e.g., 49.4% PVDF-media)

typically struggle to distinguish between H_2/formate MET and DIET, primarily because there are no direct methods for measuring either process. The increased CH_4 production observed in the presence of conductive materials could suggest several mechanisms: 1) DIET between exoelectrogens and methanogens, 2) enhanced H_2/formate formation via EET, or 3) MET facilitation through an electroactive biofilm. Consequently, the term "DIET" is often used to describe syntrophic relationships among microbes via the EET route rather than to specify a particular electron transfer mechanism.

To confirm DIET, two main approaches can be employed: one involves using microbial fuel cells or bioelectrochemical systems to analyse current-time responses or cyclic voltammetry profiles [103]; the other relies on gene-deletion mutant experiments to rule out the possibility of MET [66, 104]. These validation methods are distinct from performance evaluation experiments.

2.1.8 Theoretical considerations in EET

Various approaches have been employed to characterise the physical attributes of EET and explain its seemingly accelerated reaction kinetics.

Cruz et al. (2014) [86] calculated theoretical electron transfer rates by comparing diffusive H_2 exchange with DIET mediated by magnetite. Using Fick's law and modelling, the study estimated the electron carrier flux for H_2 diffusion between acetogens and methanogens, yielding a maximum flux of approximately 2×10^{-8} nmol s^{-1}, about 4×10^{12} electrons per second. In contrast, electronic conduction via magnetite conduits exhibited a higher flux of approximately 3×10^{-5} A, representing a nearly 10^6-fold increase over H_2 diffusion.

In wastewater treatment, thermodynamic analysis and stoichiometry are used to determine Gibbs free energy changes, assess reaction favourability, and optimise microbial conversion of

substrates to biogas. These principles can be adapted to evaluate the impact of EET on metabolic efficiency.

Microbial communities in wastewater typically operate at the edge of energy conservation, where syntrophic metabolism yields energy close to the minimum required for ATP synthesis (-15 to -20 kJ mol^{-1}) [8]. [25] modelled electron transfer mechanisms in EET and interspecies electron transfer (IET), finding that EET (44.9×10^3 e^- cp^{-1} s^{-1}) exhibits an order of magnitude higher rate than H_2-mediated electron transfer (5.24×10^3 e^- cp^{-1} s^{-1}). This highlights EET as a more kinetically favourable pathway.

Thermodynamic analyses have demonstrated that EET significantly enhances energy conservation compared to standard biological conditions (25 ℃, pH 7.0, 1 atm, dilute aqueous solution; i.e., $\Delta G^{\circ\prime}$ conditions). A team from Tsinghua University used thermodynamic modelling to determine the modified Gibbs free energy of syntrophic metabolism under DIET conditions [26, 28]. Moreover, EET enables microbes to generate more ATP during fermentation, making EET-fermentation hybridisation energetically favourable. This phenomenon was demonstrated in lactic acid bacteria, where *Lactiplantibacillus plantarum* under EET conditions produced 1.75 times more product at a rate 3.5 times faster than canonical fermentation and anaerobic respiration [37].

2.2 Biological processes of EET

2.2.1 Brief history of EET study

The history of EET in microorganisms, particularly those involving *Geobacter* and textit Shewanella species, has been a focal point of microbial electrochemistry and bioelectrochemical systems research. The discovery of these bacteria and their unique capabilities to transfer electrons to insoluble Fe(III) oxides revolutionised our understanding of microbial metabolism and its environmental implications [105, 106].

Geobacter and *Shewanella* species are known for their ability to reduce Fe(III) oxides by transferring electrons outside their cell membranes. This process, termed EET, allows these bacteria to use Fe(III) oxides as electron acceptors in their respiratory processes. The mechanism involves the use of bacterial nanowires and conductive protein structures that facilitate electron transfer to distant electron acceptors [107]. This ability plays a crucial role not only in the biogeochemical cycling of iron but also in the bioremediation of contaminated groundwater and the generation of electricity in microbial fuel cells (MFCs).

Applying EET processes in MFCs represents a significant advancement, where bacteria such as *Geobacter* and *Shewanella* generate electricity by oxidising organic compounds and transferring electrons to anodes. The integration of *Geobacter* and *Shewanella* species in co-cultures has been shown to enhance electricity production in MFCs [108]. *Shewanella oneidensis*, in particular, has been studied for its immediate current response in MFCs and the impact of oxygen exposure on its electrical output, further highlighting the intricate relationship between microbial metabolism and electron transfer processes [109].

Furthermore, the concept of DIET has expanded the scope of EET studies. *Geobacter* spp. can form conductive networks that facilitate electron transfer not only to abiotic electron acceptors but also directly between microbial species. This process has been observed in co-cultures of *Geobacter metallireducens* and other microorganisms, such as *Methanothrix harundinacea* and *Methanosarcina horonobensis*, where electrons are transferred directly between species to support growth and CH_4 production [53, 104]. OmcS and PilA in *Geobacter metallireducens* were found to be responsible for the faster growth as the Omcs-deletion mutation could not sustain growth under the co-culture condition, with the possibility of hydrogen interspecies electron transfer removed by using hydrogenase deletion mutants.

In anaerobic digesters, the role of EET and DIET becomes crucial for the efficient breakdown of organic waste and biogas production. [14] proposed the occurrence of DIET in AD treatment systems based on the high conductivity of the AD aggregates and the enriched *Methanothrix-Geobacter* community revealed by 16S rRNA gene analysis. The conductive properties of *Geobacter* biofilms and their interaction with other microorganisms enhance the conversion of organic wastes into CH_4, showcasing the potential of these processes in sustainable energy recovery and waste treatment applications [107].

Consecutively, Liu et al. (2012) [17] confirmed that GAC promotes DIET between *Geobacter* and *Methanosarcina* in a co-culture experiment and Rotaru et al. (2014) [22] mapped out part of the EET metabolism between the co-culture of *Geobacter* and *Methanothrix* using transcriptomic analysis. The concept of DIET-mediated methanogenesis, also known as electromethanogenesis, emerged from these studies.

2.2.2 Diverse electron import mechanisms in methanogens

Both *Methanothrix soehngenii* and *Methanosarcina mazei* contain genes for the CO_2 reduction pathway (i.e., *fwd*, *ftr*, *mch*, *mtd*, and *mer*). Such genes have been shown to be highly expressed during electromethanogenesis [24], although the electron uptake mechanism in methanogens has yet to be identified. The recent application of comparative transcriptomic analysis has revealed gene regulation in methanogens during electromethanogenesis in co-culture with exoelectrogens or BES; this technique may help elucidate the EET mechanisms in *M. soehngenii* and *M. mazei*. One of the indicators of EET capability is the genes with MHC structure, often denoted by the heme motif ”CXXCH”. Such genes were observed in the OmcS in *Geobacter* and Mtr in *Shewanella* [104, 110] and in some Methanosarcina [53].

Such an indication is valid for *M. acetivorans*, as the co-culture of the multiheme cytochrome A (MmcA) gene deletion mutants with *Geobacter metallireducens* produced only one-fifth of the CH_4 compared to the wild-type control. Even with GAC mediation, the mutant co-culture took about 140 days to activate the methanogenesis and about 100 days longer than the control. On the other hand, transcriptome analysis reveals that the MmcA of the wild type *M. acetivorans* was up-regulated by 4.6-fold in the co-culture [47]. Up-regulations were also shown in the membrane-bound proteins H^+/Na^+-translocating ferredoxin:NAD^+ oxidoreductase (Rnf), heterodisulfide reductase complexes (HdrED), and $F_{420}H_2$dehydrogenase (Fpo), indicating that these

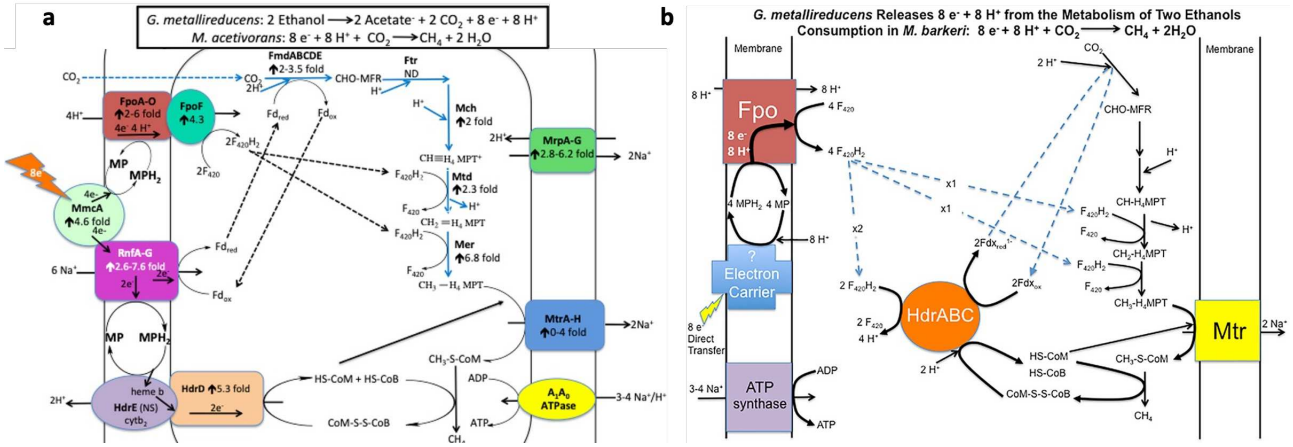


Figure 2.4: Models of electromethanogenesis in (a) *Methanosarcina acetivorans* and (b) *Methanosarcina barkeri*. Images are adapted from [47].

genes are highly involved in the electromethanogenesis in *M. acetivorans*. Two electron uptake pathways were proposed based on this transcriptomic annotation (Fig. 2.4a): 1) MmcA likely uptakes and transfers extracellular electrons to its neighbouring Rnf, which the Rnf then passes the electrons to HdrED complex via the redox reaction of methanophenazine (MP) for the subsequent intracellular electron transfer; 2) MmcA might also directly reduce MP to MPH_2, and the Fpo oxidises the reduced MPH_2 and regenerates coenzyme F_{420} for the subsequent intracellular electron transfer. Because *M. soehngenii* and *M. mazei* are phylogenetically related to *M. acetivorans*, they might share similar EET pathways as the genes for the Fpo and HdrED synthesis are also found in the reference genomes of *M. soehngenii* and *M. mazei*.

However, MHC is not essential for electromethanogenesis in some methanogens. For instance, the CH_4 production of MHC (MM_0633) gene deletion mutants of *M. mazei* had a similar CH_4 production as to the wild type in a BES experiment [111]. Furthermore, the *Methanosarcina barkeri* knockout mutant lacks MHC and yet up-took extracellular electrons through an unknown route (Fig. 2.4b) [112]. In co-culture with *G. metallireducens*, the oxidation of two moles of ethanol to acetate generated eight electrons and protons each. These electrons were conveyed to MP within the cell membrane and subsequently transferred to Fpo. This transfer facilitated proton translocation, driving the Fpo-mediated conversion of F_{420} to its reduced form, $F_{420}H_2$. Approximately half of this $F_{420}H_2$ acted as a reducing agent in the CO_2 reduction pathway, while the remainder serves as the electron source for HdrABC. HdrABC utilised this electron supply in a bifurcation reaction to reduce both ferredoxin and CoM-S-S-CoB [113]. Hydrogenotrophic methanogens such as *Methanoococcus* and *Methanobacterium* have also been observed to participate in EET without MHC [65, 114].

Methanospirillum hungatei is noted for synthesising an electrically conductive archaellum, which resembles type IV pili [115]. However, its capability for DIET remains unconfirmed. Conversely, *Methanothrix soehngenii*, lacking MHC and e-pili, has demonstrated DIET activity [19, 24, 46, 116].

Multiple experiments have suggested that the EET route in methanogens can be independent of MHC and e-pili. Thus, methanogens have likely adapted diverse strategies for electron uptake.

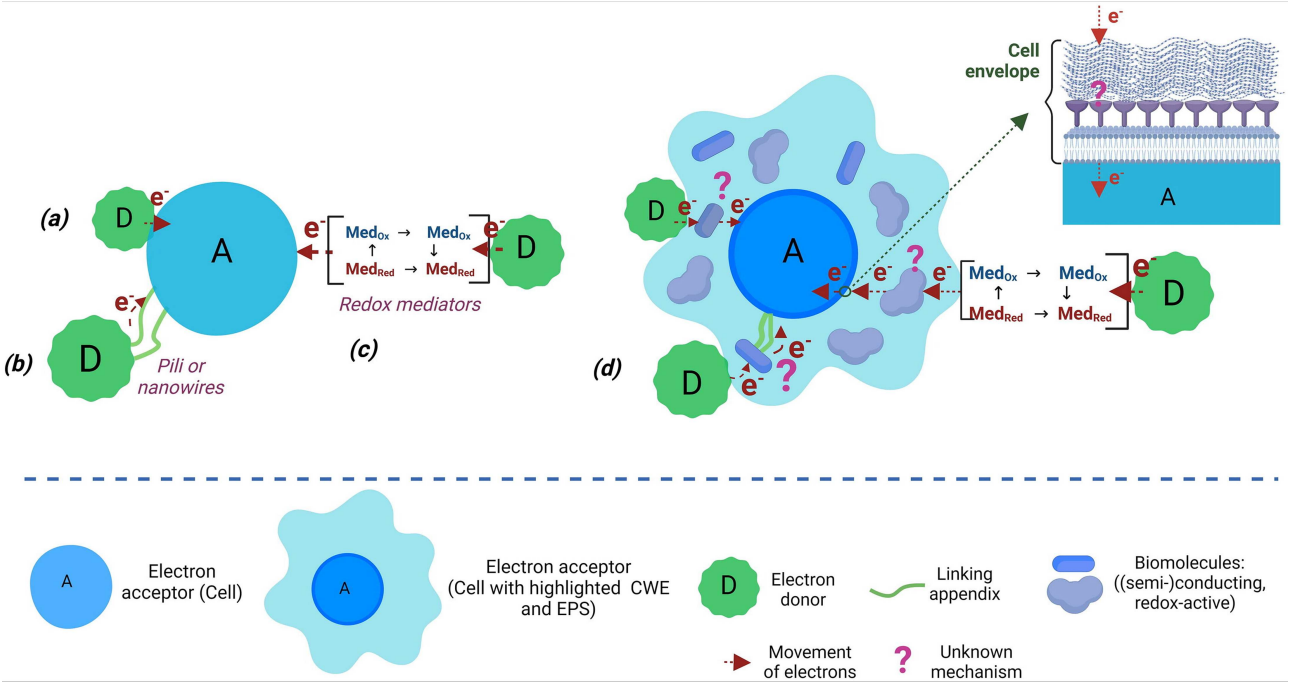


Figure 2.5: Mechanisms of EET in methanogens. (a) Direct transfer occurs via redox proteins and (b) microbial pili or nanowires, while (c) indirect transfer occurs through intermediary electron carriers. Additionally, the process involves electron transfer with (d) the aid of extracellular polymeric substances and cell wall polymers.

Since the cell surface is the first point of contact between the methanogen and an incoming electron, cell envelope and membrane-bound surface proteins could likely partake in EET.

The cell surface of *Methanosarcina* spp., including the extracellular polymeric matrix and cell envelope, contains charge-transferring molecules, facilitating the electroactivity of these methanogens [111]. This capability allows *M. barkeri* to reduce extracellular electron acceptors such as Fe(III) and anthraquinone-2,6-disulfonate [117]. Moreover, a study involving a *M. barkeri*–cadmium sulfide (CdS) biohybrid highlighted the critical role of membrane-bound proteins in converting CO_2 to CH_4 [118]. The degradation of these proteins through various concentrations of proteinase K (1.0, 2.5, and 5.0 U/mL) resulted in a diminished CH_4 production capacity of the treated *M. barkeri*–CdS biohybrids. Additionally, *M. mazei* demonstrated the ability to reduce vanadium ($V5^+$) ions under both growth and non-growth conditions, with reduction rates and extents varying with substrate and $V5^+$ concentrations. At a 10 mM concentration of vanadate, a 100% reduction of $V5^+$ was observed at rates ranging from 14 to 110 μM/h, with vanadium found exclusively in extracellular precipitates, indicating an extracellular bioreduction mechanism [119]. These findings point to two primary categories of substances in *Methanosarcina* spp. responsible for charge transport: secreted extracellular matrix components (Fig. 2.5) (e.g., extracellular polymeric substances (EPSs) and cell structure molecules (e.g., methanochondroitin (MC), the surface layer (SL), and cytoplasmic membrane (CM)) [120]. Further description of molecular structure and postulation on semi-classical/quantum chemistry of these components detailed in a review article I co-authored with [121].

2.2.3 Electron shuttles

Transferring electrons to the outer surface constitutes the first step in EET. However, the efficiency of electron transfer to solid-phase acceptors or other cells can be enhanced by additional electrical transport mechanisms. Notably, soluble, redox-active molecules can be electron shuttles between cells and particulate or distant electron acceptors. In environments such as soil, where direct contact between cells and Fe(III) oxide is minimal, the addition of redox-active humic substances—natural components of soils and sediments—can substantially increase the reduction rate of Fe(III) oxide in both defined cultures and organic-deficient subsurface soils [122, 123]. The quinone groups in these shuttles are reduced at the cell surface, and the reduced shuttle moves until it encounters and reduces Fe(III) oxide, thereby replenishing the shuttle's oxidised state. This mechanism allows for significant effects with only catalytic quantities of the electron shuttle, with humic substances and analogous natural shuttles being abundant enough to support Fe(III) oxide reduction across a variety of anaerobic environments [123, 124].

Certain microorganisms have evolved to secrete their own electron shuttles to facilitate EET, a phenomenon initially observed in *Geothrix* and *Shewanella* species [123]. These shuttles can include compounds such as flavins, quinones, and phenazines [125–127], with flavin-mediated shuttle transfer through the porin–cytochrome system accounting for up to 75% of *S. oneidensis*' EET capability towards solid-phase electron acceptors [128]. Enhancing *S. oneidensis* with a synthetic flavin biosynthesis module to increase flavin production resulted in more than a tenfold increase in EET to electrodes [129]. The production of self-generated electron shuttles is particularly beneficial in confined, organic-rich settings, aligning well with the characteristics of anaerobic wastewater treatment facilities, especially those bioreactor designs that ensure prolonged retention of biomass/biofilm.

Similarly, a potential way to promote methanogenesis in *Methanosarcina* is via lipophilic electron carrier supply. Increase in the concentration of redox mediators, such as the natural-occurring MP [130], the addition of phenazine [130], or the solid-state shuttle nanoFe_3O_4 [131] promotes growth and methanogenesis.

The electron flux via MP in *Methanosarcina acetivorans* culminates in the generation of a scalar transmembrane ion gradient, crucial for ATP synthesis. This mechanism was notably more efficient than electron donation to cytoplasmic acceptors, highlighting the significance of MP in methanogenic efficiency. Furthermore, the relationship between cell density and H_2 diffusion is critical; at low cell densities, H_2 diffusion was compromised, necessitating higher MP concentrations within the membrane. Conversely, increased cell density enhanced H_2 diffusion efficiency, resulting in reduced membrane MP concentrations [130].

When MP concentration was sufficiently high(i.e., > 10 μM), isopotential electron tunnelling became possible and the cell membrane of *Methanosarcin* became quantised.

In brief, the naturally occurring EET mechanism can be summarised in four modes (Table 2.2).

Table 2.2: The naturally occurring modes of EET

Mechanism	Description	References
Short-range direct contact (in nm)	Microorganisms directly transfer electrons to electron sinks through cell surface proteins or conductive pili.	[106, 132]
Electron Shuttling	Soluble redox-active compounds mediate electron transfer from microorganisms to electron sinks.	[133, 134]
Long-range direct contact (in μm)	Extracellular appendages (i.e., e-pili or nanowire) facilitate long-range electron transfer to electron sinks.	[135, 136]
Conductive materials	Materials act as conduits for electron transfer between different microbial cells.	[137]

2.3 Proteins related to extracellular electron transfer

Electroactive microorganisms utilise various EET mechanisms to interact with their environment. These include MHCs for electron transport to cell surfaces, porin-cytochrome conduits for electron movement across membranes in diverse bacteria, e-pili acting as biological wires, multicellular intracellular electrical conduit of cable bacteria for long-range electron conduction in sediments, and electron shuttles that facilitate microbial electron exchange with external electron acceptors or donors [69]. Each mechanism demonstrates the diverse and complex ways these organisms manage electrical interactions with their surroundings.

2.3.1 Multi-heme c-type cytochrome and porin-cytochrome conduits

The widespread presence of MHC in EET across diverse bacteria and archaea suggests that their detection in metagenomes can predict the EET capabilities of uncultured microorganisms [138, 139]. Gram-negative species such as *Shewanella oneidensis* and *Geobacter sulfurreducens* have been demonstrated to sustain electron transport across their outer membranes via porin–cytochrome complexes. These complexes consist of a MHC located in the periplasm, associated with a transmembrane β-barrel protein, and another MHC on the cell's exterior surface (Fig. 2.6). Additionally, inner membrane-associated MHC such as CymA initiates the electron transfer from the quinone/quinol pool. The porin–cytochrome mechanism in *S. oneidensis* has been elucidated through extensive, multi-laboratory research over more than a decade, using genetic, biochemical, and functional methodologies previously reviewed [140–142]. Furthermore, evidence supports the role of porin-cytochrome conduits in *S. oneidensis* as essential components

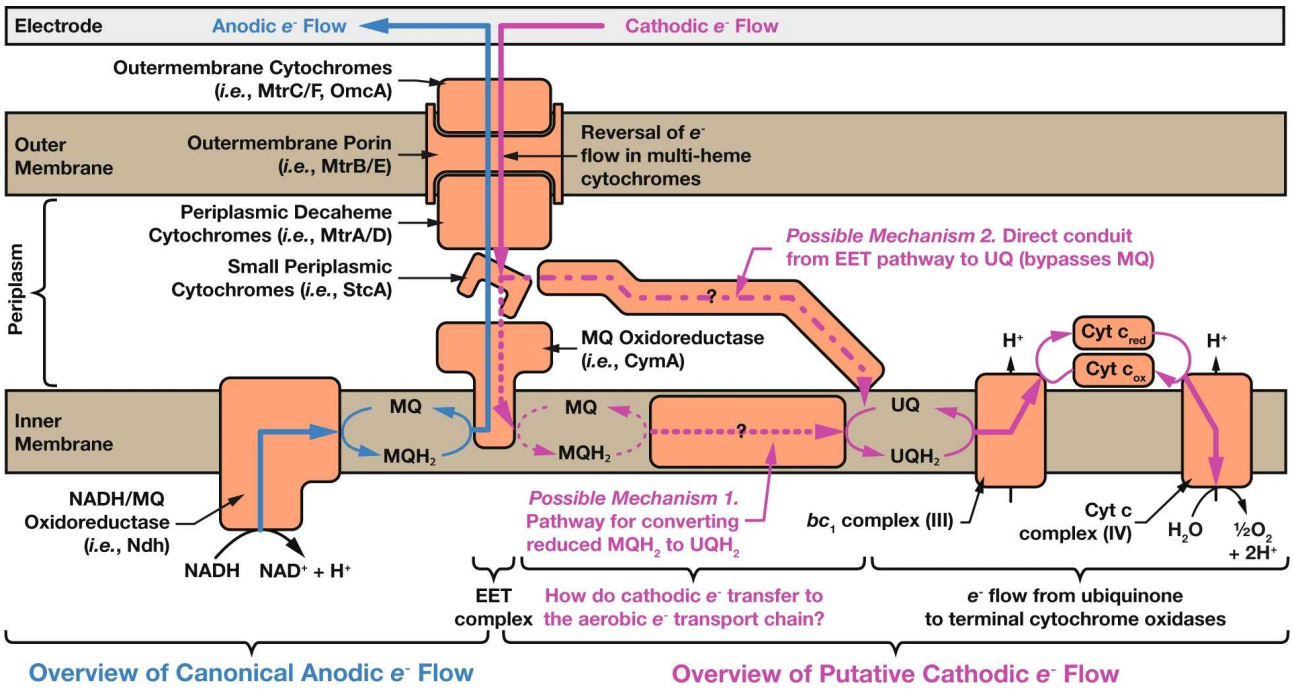


Figure 2.6: The process of electron exchange using MHC conduit. The conventional pathway for electron export, associated with anodic activities, is represented in light blue, whereas the hypothetical pathway for cathodic electron import is depicted in pink. Solid lines indicate established electron transfer routes, whereas hypothesised routes are illustrated with dashed lines. Image is reproduced from [143]

of electron uptake mechanisms, as demonstrated by transposon array (Fig. 2.6) [143].

Heme motif as a genetic signature aids in identifying probable exoelectrogenic organisms via genomic analysis [64, 139, 144]. However, while MHC is a signature of EET capability, it might not indicate DIET capability. For example, although *Shewanella* and *Geobacter* species exhibit these complexes, DIET has only been observed in *Geobacter* spp. Attempts to demonstrate *Shewanella* growth via DIET have been unsuccessful, with *S. oneidensis* failing to thrive under such conditions. Conversely, *Geobacter* species increase electrical connectivity by augmenting redox-active sites on the decaheme cytochrome OmcA, thereby expanding its electron transfer surface area.

As discussed in 2.2.2, emerging evidence suggested the role of cytochrome-based mechanisms in *Methanosarcina acetivorans* [47]. The presence of MHC in other EET-capable archaea, including hyperthermophiles [145, 146] and anaerobic CH_4 oxidisers [147–151], points to a common cytochrome-mediated EET pathway in archaea. Nonetheless, alternative EET mechanisms remain unidentified in certain archaea. Notably, *Methanothrix* and several *Methanosarcina* species that participate in DIET and highly efficient Fe(III)-reducing archaea like *Pyrobaculum* spp. [146, 152] do not possess MHC. Unravelling these alternative EET strategies, especially in the absence of c-type cytochromes, is crucial for a deeper understanding of the comparative benefits of cytochrome-dependent and independent EET pathways across diverse archaeal habitats.

Opposite to methanogenic behaviour, a study utilised anaerobic methanotrophic *Ca. Methanoperedens* in bioelectrochemical systems and observed CH_4-dependent current constituting 91-93% of the total current. The current production was associated with high enrichment of *Ca. Methanoperedens* on the anode, reaching up to 82% of the microbial community. This EET mechanism was

evident at various electrode potentials, possibly involving an uncharacterised short-range electron transport protein complex [153].

2.3.2 Electrically conductive pili

E-pili, or electrically conductive pili, are defined as any pili that are sufficiently electrically conductive along their length to facilitate long-range electron exchange between the microorganism expressing e-pili and its external environment [154]. E-pili is one of the observed routes for long-range EET between microbes and minerals[140], and also frequently observed in DIET [44, 154]. Their evolution, conductivity features, and significance in *Geobacter* species have been extensively studied [14, 104, 155, 156]. Research indicates that DIET can occur without e-pili in the electron-accepting partner but is hindered by poorly conductive pili in the electron-donating strain [157]. Notably, the addition of conductive material plays a regulatory role in the pili expression (Fig. **??**).

Interestingly, e-pili have evolved independently in various bacteria [38, 39], including *Syntrophus aciditrophicus*, a key player in anaerobic digestion and non-fermentable substrate metabolism through interspecies electron transfer. This organism, along with others capable of interspecies hydrogen transfer, exhibits genes for e-pili, suggesting a broad potential for DIET across diverse microbial communities.

The specific conductivity threshold that defines an e-pilus depends on its function. For example, the e-pili of *G. sulfurreducens*, with a conductivity at pH 7 of 51 $mS\,cm^{-1}$ [158], have been shown to effectively support growth under conditions that necessitate long-range electron transport, in line with theoretical expectations. It is suggested that e-pili with conductivity lower than 30-40 $mS\,cm^{-1}$ could not sustain EET [159, 160].

The e-pili of *G. sulfurreducens* is characterised by the *pilA* gene, encoding pilin monomers. In *Geobacter* species, *pilA* represents a modified version of the standard type IV pilin, likely adapted to improve conductivity [161]. This adaptation involves a truncation that allows for the closer arrangement of aromatic amino acids, thereby creating a conductive channel within the e-pili. Through the heterologous expression of PilA pilin monomers, studies have demonstrated the influence of aromatic amino acid content on electrical conductivity, revealing a direct correlation between the density of these amino acids and the conductivity of e-pili (Fig. 2.7) [160, 162].

Originally thought to be unique to a specific phylogenetic cluster, these e-pili genes are a hallmark of the *Desulfuromonadales* order within the *Deltaproteobacteria*. Nonetheless, more recent experimental and metagenomic studies have expanded our understanding of the phylogenetic range of e-pili, indicating a broader distribution than initially recognised [38, 164].

Other than *pilA* e-pili, OmcS nanowire found in *Geobacter* also contributes to EET. OmcS nanowires are not type IV pili but are polymerized filaments of the cytochrome OmcS, with continuous chains of parallel-stacked hemes optimising electron transport at short distances [165]. However, several observations support the secondary role of OmcS in electron transport. Deleting the gene for OmcS increases biofilm conductivity without impacting the production of high current densities [156, 166]. Strains like KN400, which express very little OmcS but abundant e-pili, generate biofilms with higher conductivity and current densities. Conversely, cells with

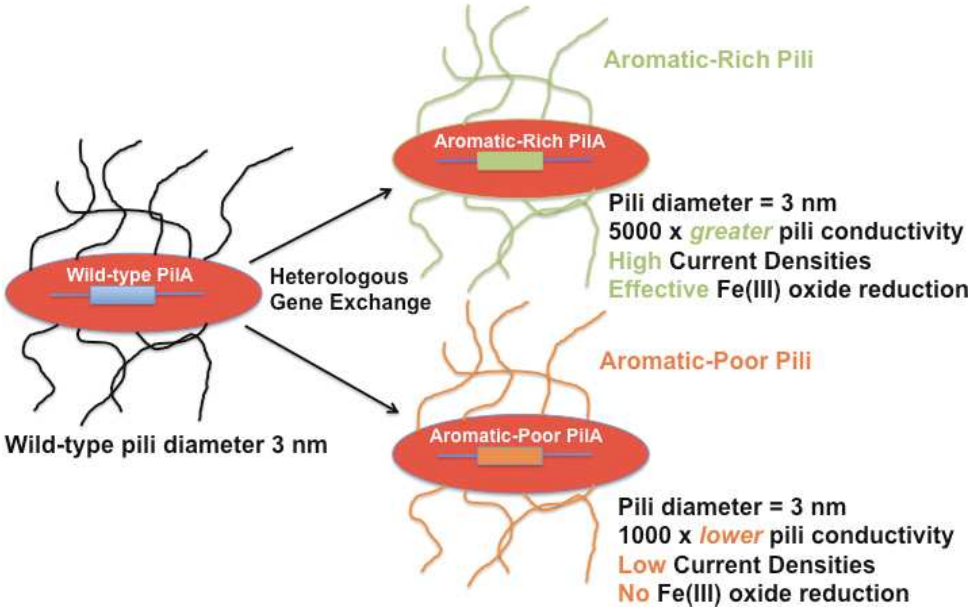


Figure 2.7: Expressing *pilA* pilin monomer genes from external sources in *G. sulfurreducens* results in pili exhibiting significantly varied conductivity, with differences reaching up to 10^6-fold compared to individual native wires. These genetically modified pili maintain the same diameter as the natural wild-type pili at 3 nm, differing from the 4 nm diameter of OmcS filaments. Such alterations in pili conductivity notably degrade the efficiency of extracellular electron transport, adversely affecting related phenotypes. This finding aligns with similar outcomes from a range of studies on heterologous expression, all of which corroborate the substantial impact of conductivity modifications on electron transport capabilities. Image is adapted from Lovley and Walker [163].

abundant OmcS but poorly conductive pili are ineffective in reducing Fe(III) oxides. Moreover, *G. uraniireducens*, which possesses an OmcS homolog but produces poorly conductive pili, relies on an electron shuttle for Fe(III) oxide reduction and produces low current densities. Thus, on ecological significance, *pilA*-like e-pili is of uttermost interest in studying EET, especially in biofilm, an argument concluded by [163].

Archaellum of *Methanospirillum hungatei*

Archaeal pili exhibit a significant diversity, with approximately 5,000 gene products associated with type IV pili in archaea, with over half being previously classified as hypothetical in public databases [167].

The archaellum, a primary locomotive structure in Archaea, has been identified as electrically conductive in *Methanospirillum hungatei*, drawing parallels with the e-pili found in bacteria [115]. This discovery employed conductive atomic force microscopy to assess the archaellum's conductivity, placing it on a conductive substrate with a conductive tip acting as a movable top electrode, demonstrating a conductive capacity surpassing that of the e-pili in *Geobacter sulfurreducens* through a linear response in current-voltage spectroscopy.

Moreover, the *M. hungatei* archaellum's role extends beyond mobility, acting as a medium for long-range EET, potentially aiding cellular adhesion by mitigating charge barriers and facilitating electrical intercellular communication.

Despite the similarities in the alpha-helix filament structures and assembly mechanisms between the archaellum and bacterial type IV pili, notable differences exist in their tertiary structures and subunit arrangements, with the archaellum lacking an inner channel [168]. This unique structural configuration suggests that the archaellum's conductive properties may operate through mechanisms distinct from those observed in *G. sulfurreducens*' e-pili and other bacterial analogues.

2.3.3 Amyloid protein

Methanogenic archaea, such as *Methanosarcina acetivorans* and *Methanospirillum hungatei*, have demonstrated the ability to EET via MHC and archaella [115, 169]. However, this mechanism is notably absent in *Methanothrix* species [53].

Despite the absence of MHC and archaella, *Methanothrix* species form distinctive paracrystalline sheaths composed of self-assembling amyloid fibres with a β-sheet structure [170, 171]. To date, these sheath fibres have been isolated from two methanogenic species: *Methanothrix thermophila* [171] and *Methanospirillum hungatei* [172]. The fibres are primarily composed of a MspA, which is characterised by a high aromatic amino acid content (exceeding 11%).

In co-culture experiments with *Methanothrix thermoacetophila* and *Geobacter metallireducens*, MspA was upregulated under DIET conditions, indicating that MspA may facilitate electron uptake through amyloid fibrils. Evidence from recent isolation experiments suggests these fibres are conductive, likely due to aromatic stacking within the amyloid proteins [173, 174].

Electron transfer in amyloid proteins occurs via hole hopping through stacked tyrosine residues within the β-sheet structure, enabling efficient transport over micrometre distances through multi-step hopping, rather than single-step tunnelling [174]. This process operates without redox-active metal cofactors or photosensitisers, driven by a proton-coupled electron transfer (PCET) mechanism, where nearby proton acceptors, such as glutamine, play a crucial role. Metal ions, such as Zn^{2+}, enhance conductivity by stabilising tyrosine stacking and the hydrogen-bonding network.

At the time of writing, the MspA protein (locus tag: Mthe_1069) identified in *Methanothrix thermophila* has no homologues (i.e., amino acid identity $>$ 25%) in the NCBI GenBank and RefSeq databases since its reporting in 2018 [172]. Consequently, predicting MspA in other *Methanothrix* species using metagenomic data necessitates prediction methods that rely on a rule-based method instead of an alignment-based method.

2.3.4 Adaptation in EET capability

The ability to perform EET can be gained or lost depending on the substrate environment. Zheng et al. (2020) [175] highlighted the selective pressure on EET capabilities by comparing the *Methanobacterium electrotrophus* strain YSL, identified from coastal river sediments [176], with *Methanobacterium formicicum* strain MF. While strain YSL is capable of engaging in DIET, strain MF, despite being its closest relative, failed to grow via DIET when co-cultured with *Geobacter metallireducens*, even in the presence of GAC [22, 177].

One possible explanation for this difference is that strain MF may have lost its DIET capability

due to mutations affecting key components necessary for EET. Strain MF was isolated in 1966 with H_2 as the electron donor [178] and has been maintained under these conditions for over 50 years. The loss of EET capability might have been evolutionarily advantageous, as it could eliminate unnecessary protein production in an H_2-based culture. This suggests that pure culture strains may exhibit behaviours different from those of their phylogenetic neighbours that thrive in environmental and engineering systems.

2.4 Horizontal gene transfer

Horizontal gene transfer facilitates rapid microbial evolution and enables organisms to adapt to diverse environments by directly exchanging genetic material, bypassing slower vertical inheritance [179, 180]. Mobile genetic elements (MGEs) have profoundly influenced microbial diversity by rapidly acquiring novel metabolic traits. For example, HGT underpins the dissemination of antibiotic resistance genes [181] and supports the exploitation of new nutrients and ecological niches [179]. This capacity for genetic exchange underscores a fundamental process driving microbial evolution.

Classic mechanisms of HGT include transformation (the uptake of free DNA), transduction (phage-mediated transfer), and conjugation (primarily via conjugative pili). Additional mechanisms, such as DNA transfer via outer membrane vesicles [182] or virus-like particles [183], underscore the complexity of HGT. Detecting MGEs in metagenomes commonly relies on gene markers in metagenomic-assembled genomes; database comprehensiveness critically affects identification [184, 185]. Meanwhile, *de novo* binning approaches facilitate discovery but can lead to false associations [186–188]. Recently, transformer-based methods have enhanced plasmid and phage detection by encoding contigs into protein-cluster ”sentence” and applying positional embeddings [189, 190]. Such advances allow the identification of previously unannotated mobile elements, supporting comparative genomic and phylogenetic analyses that elucidate HGT in complex microbial communities.

In methanogens, HGT has helped shape metabolic capacities and expand protein families [191, 192]. A prominent case is the acquisition of bacterial genes by *Methanosarcina*, enabling the use of varied substrates (e.g. acetate, methanol) in methanogenesis [193]. Specifically, genes encoding acetate kinase (*ack*) and phospho-acetyl-transferase (Pta) conferred an additional pathway for acetoclastic methanogenesis [194], promoting the colonisation of new ecological niches and influencing carbon cycling. Bacterial-derived proteins also assist stress mitigation and membrane transport in methanogens [195], underlining the importance of tracking HGT to understand their evolution and environmental adaptations.

Broader insights into electron transport illustrate the mobility of key functional genes. The MHC MtrCAB system, originally characterised in *Shewanella oneidensis*, has been detected in at least 148 bacterial species from environments spanning marine ecosystems to contaminated wastewater [196]. Phylogenetic incongruence suggests HGT as a main driver of MtrCAB dispersal across taxa with diverse metabolic strategies, including chemoorganotrophy and chemolithoautotrophy. Genomic regions containing *mtrCAB* often feature transposases, integrases, endonucleases, and

recombinases, indicating that mobile elements mediate insertion or loss of these genes. In certain genomes, *mtrCAB* disrupts syntenic regions,[1] while the presence of recombination enzymes in regions lacking *mtrCAB* points to gene loss events.

Phages also play a vital role in prokaryotic evolution and disseminating adaptive traits. They are abundant in raw sewage and may outnumber bacteria in wastewater treatment plants [197]. By transferring genetic material, phages shape microbial community composition and spread antibiotic resistance. Their effects in biofilm settings are particularly nuanced [198]. Phage-induced lysis of *Shewanella oneidensis* MR-1 can release MHCand extracellular DNA (eDNA). These compounds may act as electron shuttles and enhance EET in biofilms [199, 200]. In other scenarios, phages degrade the biofilm matrix of pathogenic bacteria, demonstrating that phage–host interactions can either facilitate or hinder biofilm development, depending on the ecological and metabolic context [201].

2.4.1 Scope and mechanistic focus of this thesis

The literature reviewed in this chapter highlights that extracellular and direct interspecies electron transfer involve a wide range of metabolic, cellular, and genetic mechanisms that operate across multiple biological scales. While these mechanisms provide crucial conceptual context, the present thesis does not seek to resolve all aspects of the EET or DIET experimentally.

Instead, this work focuses on three interconnected levels of investigation. First, the genetic potential for EET is examined through genome-resolved analyses, with emphasis on the distribution of genes encoding MHC and e-pili within metagenome-assembled genomes. Second, community-level metabolic organisation is assessed by linking the enrichment of putatively electroactive bacteria with shifts in methanogenic guilds, particularly acetoclastic methanogens, under conductive and non-conductive reactor conditions. Third, these microbial patterns are interpreted in the context of process-level reactor performance, including organic removal efficiency and CH_4 production.

Mechanistic processes occurring at finer spatial or biochemical resolutions, such as direct measurements of electron flux, protein-level electron transfer dynamics, or single-cell interactions, are beyond the scope of this thesis. These mechanisms are therefore discussed as interpretative frameworks rather than experimentally resolved phenomena. This scoped approach enables robust linkage between reactor-scale performance, community structure, and genomic potential for EET, while avoiding over-interpretation beyond the resolution of the available data.

[1] Syntenic regions are stretches of DNA in which genes are preserved in the same order and orientation; gene transfer can alter this arrangement, potentially influencing metabolic functions.

Chapter 3

Materials and Methods

3.1 Staged anaerobic fluidised membrane bioreactors

The staged anaerobic fluidised membrane bioreactors, designed and operated by research teams led by Dr Chungheon Shin (Research Director at the Codiga Resource Recovery Centre at Stanford University), represent an advanced approach to domestic wastewater treatment in temperate climates. The primary objective of SAF-MBR systems is to efficiently manage and treat low-strength domestic wastewater by harnessing the synergistic effects of anaerobic digestion and membrane filtration [30–32, 202].

The integration of GAC within the SAF-MBR enhances both performance and system stability. GAC acts as a biofilm carrier that promotes microbial growth, retains biomass, supports slow-growing species, enhances diversity, and protects microbes from local physicochemical stress within the extracellular polymeric substances matrix. Moreover, its scouring properties help reduce membrane fouling, extend its lifespan, and reduce maintenance needs. Significantly, GAC facilitates DIET, a crucial process in the anaerobic breakdown of complex organic compounds. As a result, the use of GAC in SAF-MBR systems increases treatment capacity and contributes to energy-efficient operation by minimising the energy required for membrane cleaning and maintenance. The pore size of the commercial-grade GAC is approximately 10 μm without any pretreatment. Each bioreactor contained 2028 kg of GAC and had a volume of 4500 L.

Pilot-scale studies are critical for exploring complex biological processes, such as DIET, in anaerobic digestion. Operating a SAF-MBR system at pilot scale enables detailed examination of microbial interactions involved in DIET under conditions that closely replicate real-world wastewater treatment scenarios. This scale of operation is essential for optimising key system parameters, including hydraulic retention time (HRT), organic loading rate (OLR), and GAC dosing, all of which can affect DIET efficiency. Furthermore, pilot-scale experiments provide an opportunity to monitor process dynamics in detail, offering valuable insights into the mechanisms driving DIET, the influence of operational conditions on microbial community structure, and the overall performance of the SAF-MBR system in domestic wastewater treatment.

In this project, biomass from two SAF-MBR systems was provided by Inha University, South Korea, and Stanford University, USA, for metagenomic and metatranscriptomic analysis. Dr

Chungheon Shin provided datasets for both pilot-scale systems, covering reactor operation, design, and chemical analysis. Detailed information on reactor design and operation is provided below.

3.1.1 SAF-MBR 1.0

Reactor design

The SAF-MBR system treated real wastewater, deployed at the Bucheon domestic wastewater treatment facility in the Republic of Korea, processed 2 mm screened primary effluent using a dual-reactor configuration, as depicted in Figure 3.1. The anaerobic fluidised bed reactor (AFBR), devoid of membranes, provided initial wastewater treatment. In contrast, the anaerobic fluidised membrane bioreactor (AFMBR) with membranes completed the process by retaining all suspended solids for subsequent handling or disposal. The AFBR, constructed from STS 304 stainless steel and measuring 3.0 m in height and 0.65 m in diameter, had a volume of 0.99 m^3. A settling tank atop the AFBR, equipped with a submerged weir, prevented GAC overflow. The reactor utilised 139 kg of GAC with a size range of 0.8–1 mm, maintaining fluidisation through effluent recirculation at a rate of 0.15 $m^3\,min^{-1}$, resulting in GAC expansion to 40% of the reactor height.

The AFMBR, rectangular and made of STS 304 stainless steel, had dimensions of 0.9 m by 0.4 m and a depth of 2 m, creating a working volume of 0.77 m^3. It was topped with a settler designed to prevent GAC loss, housing five membrane modules composed of polyvinylidene fluoride (PVDF) hollow fibre membranes with a 0.03 μm pore size and a total surface area of 39.5 m^2. With 264 kg of GAC occupying 60% of its volume at rest, the recirculation flow rate of 0.53 $m^3\,min^{-1}$ expanded the GAC to fill the reactor volume. This ensured comprehensive contact between the membrane surfaces and fluidised GAC, mitigating membrane fouling efficiently, a notable advantage over traditional gas-sparged systems.

Reactor operation

A SAF-MBR was continuously operated for 555 days without the need for membrane chemical cleaning and processing primary-settled domestic wastewater. The wastewater temperatures ranged from 8 to 30 ℃, and the total hydraulic retention time (HRT) varied between 4.6 and 6.8 hours. Initially, during the colder months with temperatures of 8–15 ℃, before acclimation was complete, the system achieved average removal efficiencies of chemical oxygen demand (COD) and biochemical oxygen demand (BOD_5) removal efficiencies of 81% and 85%, respectively. Following full acclimation, the reactor exhibited enhanced performance with COD removal efficiencies of 94% in summer and 90% in winter, and BOD_5 removal efficiencies of 98% in summer and 90% in winter. The average effluent concentrations were maintained below 23 mg/L for COD and 9 mg/L for BOD_5. The operational energy requirement was 0.23 $kWh\,m^{-3}$, which could be supplied by producing primary and secondary CH_4 and potentially reduced by optimising the hydraulic retention time. Biosolids production was consistent throughout all seasons, averaging 0.051 g of volatile suspended solids (VSS) per g COD removed. In this project, ten samples were collected

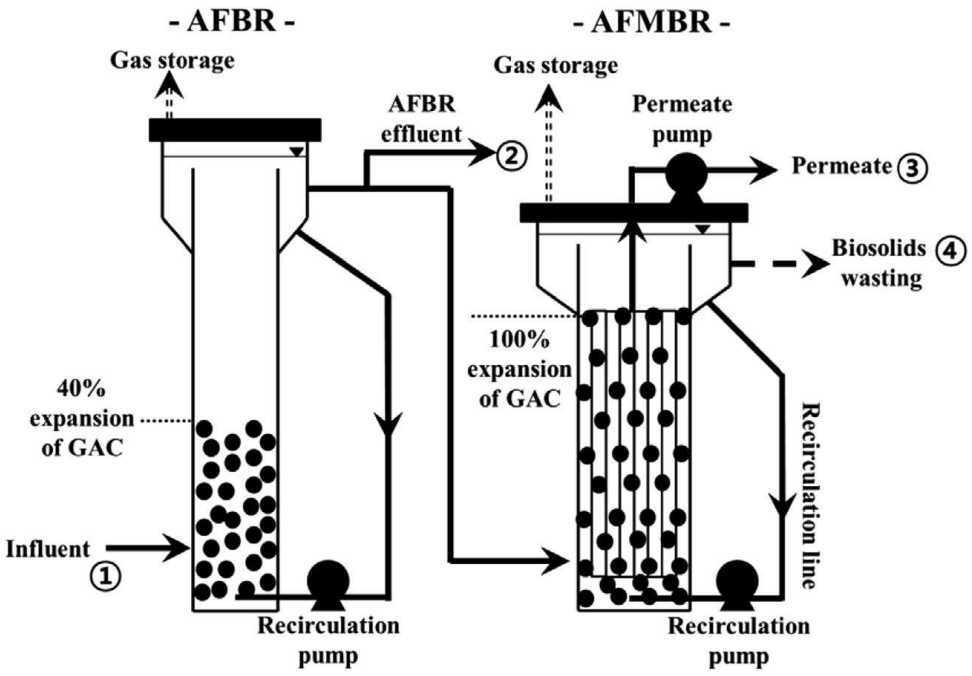




Figure 3.1: (a) Schematic diagram and (b) a photo of the pilot-scale SAF-MBR system 1.0 [31]



Figure 3.2: The SAF-MBR 2.0

during winter, spring, and summer, comprising both bulk and GAC samples, for metagenomic sequencing in SAF-MBR 1.0.

3.1.2 SAF-MBR 2.0

Reactor design

At the Codiga Resource Recovery Centre in Stanford, CA, a pilot-scale wastewater treatment setup was established, featuring a grit tank and micro-screen for initial treatment, followed by a SAF-MBR 2.0 system for advanced processing (Fig. 3.2). Unlike the single GAC-fluidised bed with an integrated membrane module in SAF-MBR 1.0, the 2.0 upgrade decouples these functions into two parallel GAC–AFBRs, which feed an independent, gas-sparged ultrafiltration tank with LEAP^{TM} diffusers, providing hydraulic redundancy, higher throughput, and lower specific energy demand. This setup treated municipal wastewater from the Serra Street sewer at Stanford University for over three years, employing a 0.3 mm micro-screen (MS-28, Hydro International) for primary screening.

The SAF-MBR 2.0 system comprised two parallel AFBRs (i.e., AFBR1 and AFBR2) with GAC biocarriers and a gas-sparged ultrafiltration membrane tank. The system maintained adequate mixing and solids control by recirculating fluid at twice the influent rate, with solids retention facilitated by constant waste removal. Gas phases across the tanks were interconnected to stabilise pressure, while biogas sparging ensured even gas distribution.

Operational specifics included GAC packing ratios of 60% for AFBR 1 and 57% for AFBR 2, with respective upflow velocities of 43.8 m h^{-1} and 27.1 m h^{-1}, leading to around 30% bed expansion for adequate mixing. The membrane tank used two ZeeWeed 500D modules (SUEZ) with 0.04 μm pores and LEAPmbr diffusers, having volumes of 1.06, 1.72, and 1.72 m^3 for AFBR 1, AFBR 2, and the membrane tank, respectively.

Reactor operation

The SAF-MBR 2.0 consistently demonstrated COD removal efficiencies exceeding 70%, with specific values of 84% during a 10.0 h HRT at 22.3 ℃, 83% during a 7.0 h HRT at 24.7 ℃, 79% during a 5.3 h HRT at 22.5 ℃, and 78% during a 5.3 h HRT at 18.4 ℃. Due to the attached growth on the GAC carrier, the mean cell residence time ranges from 550 to 1150 days. Operational stability was sustained over several months, during which H_2S concentrations in the biogas were monitored regularly. Initially detectable in trace amounts during the startup phase, H_2S dropped below 0.4 ppm once the reactor reached stable performance—coinciding with consistent COD removal and biogas output. Recirculated bulk COD was examined as an indicator of potential membrane fouling, along with mixed liquor suspended solids (MLSS), to gauge the biological process's concentration. The particulate COD (pCOD) fraction was analysed to determine mixed liquor volatile suspended solids (MLVSS). Empirical monitoring indicates a pCOD fraction of 90% in the recirculated bulk COD, with a COD/SS conversion ratio of 2.1 for pCOD to MLVSS and a dispersed growth active biomass to inert biomass ratio of 1.4 (160mgCOD/113mgVSS). The system produced 16 to 18 mL of biogas per litre of wastewater, with approximately 85% CH_4 content. This biogas production accounts for approximately 9In this project, one GAC sample was collected for metagenomic and metatranscriptomic sequencing in the SAF-MBR 2.0.

3.1.3 Operational performance and cross-system normalisation

Table 3.1 summarises the operating conditions, treatment performance, sulfate content, and CH_4 recovery metrics for the SAF-MBR systems investigated in this study, alongside a lab-scale AFMBR [101] reference operated under comparable low-strength wastewater conditions. For SAF-MBR 1.0, three biomass sampling days (Days 342, 436, and 555) were selected to represent distinct operational phases characterised by progressively higher organic loading rates and shorter hydraulic retention times. On each sampling day, influent characteristics, temperature, HRT, SRT, organic loading rate, influent and permeate COD concentrations, and COD removal efficiency were obtained directly from reactor monitoring data. Seasonal sulfate concentrations for winter, spring, and summer operations were assigned based on reported values for the same pilot SAF-MBR system, providing contextual information on the availability and removal of alternative electron acceptors during each sampling period.

Methane recovery for SAF-MBR 1.0 is reported as COD-equivalent CH_4 per litre of wastewater treated and was not measured directly on the biomass sampling days. Instead, CH_4 yield was conservatively estimated by applying a representative COD-to-CH_4 recovery factor (56.5%), derived as the midpoint of two independent COD mass balances reported for the same SAF-MBR configuration under stable operation. This approach avoids assuming complete conversion of removed COD to CH_4 and accounts for COD retained as biomass, diverted to sulfate reduction, or otherwise not recovered as biogas. In contrast, CH_4 yield for SAF-MBR 2.0 was calculated directly from measured biogas flow and composition on the sampling day and is therefore reported as a measured value. For SAF-MBR 2.0, sulfate concentrations in the influent and effluents were below

detection limits, indicating the absence of sulfate as a competing terminal electron acceptor during operation.

The AFBR PVDF reference represents a system treating synthetic wastewater with a soluble COD concentration of 500 mg L^{-1} at room temperature (25 °C) and an organic loading rate of 4.5 kg COD m^{-3} d^{-1}. Influent COD, permeate COD, and COD removal efficiency (approximately 90%) were taken directly from one of my published studies [101]. At the same time, CH_4 yield was reported on a treated wastewater volume basis and converted to COD-equivalent units for consistency with the SAF-MBR datasets. Across all systems, COD removal efficiency and CH_4 yield are reported together to provide a transparent basis for comparing treatment performance and energy recovery, while clearly distinguishing between measured and estimation-based values.

Table 3.1: Integrated operating conditions, sulfate content, COD removal performance, and CH_4 yield normalised per litre of wastewater treated for SAF-MBR systems and the PVDF AFMBR reference.

System	Timepoint / stage	Influent	T (°C)	HRT (h)	SRT (d)	OLR (kg m^{-3} d^{-1})	COD_{in} (mg/L)	COD_{perm} (mg/L)	COD removal (%)
SAF-MBR 1.0	Day 342 (Winter)	Raw sewage	13.8	7.3	487	0.9	286	69	76
SAF-MBR 1.0	Day 436 (Spring)	Raw sewage	24.9	4.8	320	1.0	203	32	84
SAF-MBR 1.0	Day 555 (Summer)[a]	Raw sewage*	25.7	3.2	128	2.7	364	63	83
SAF-MBR 2.0	Day 390	Raw sewage	20.3	5.2	518	2.6	512	86	83
AFMBR PVDF	Stage IV	Synthetic (SCOD)	25	8	n.a.	4.5	500	50	90

System	Timepoint / stage	Influent sulfate (mg/L)	AFBR effluent sulfate (mg/L)	AFMBR permeate sulfate (mg/L)	Methane yield (mg COD/L)
SAF-MBR 1.0	Day 342	39 ± 2	25 ± 11	8 ± 13	130 (est.)
SAF-MBR 1.0	Day 436	36 ± 1	5 ± 0	0 ± 0	102 (est.)
SAF-MBR 1.0	Day 555	34 ± 1	2 ± 0	0 ± 0	180 (est.)
SAF-MBR 2.0	Day 390	0	0	0	139 (meas.)
AFMBR PVDF	Stage IV	n.r.	n.r.	n.r.	310

[a] Milk was introduced to increase the COD load to test the OLR capacity.

Notes: COD removal is based on measured influent and permeate COD concentrations. Methane yield is reported as COD-equivalent per litre of wastewater treated. For SAF-MBR 1.0, CH_4 yield is estimated using a single representative COD-to-CH_4 recovery of 56.5%, taken as the midpoint of two independent COD mass balances reported for the same pilot SAF-MBR system (Shin et al., 2014). SAF-MBR 2.0 CH_4 yield is calculated directly from measured biogas flow and composition on the sampling day. AFMBR PVDF values are taken from literature at SCOD 500 mg L^{-1}, room temperature (25 °C), OLR 4.5 g COD L^{-1} d^{-1}, with approximately 90% COD removal, and CH_4 production reported per treated wastewater volume [101].

3.1.4 Biomass sampling rationale and representativeness of reactor states

Biomass sampling was conducted during periods of sustained and well-characterised reactor performance rather than at fixed calendar intervals, with the objective of capturing microbial communities representative of distinct and operationally relevant reactor regimes. For pilot-scale anaerobic systems, sampling was guided by the attainment of stable process indicators, including consistent COD removal efficiency, stable effluent quality, and reproducible biogas production, as commonly adopted in pilot-scale anaerobic digestion and SAF-MBR studies. For SAF-MBR 1.0, samples were collected during winter (Day 342), spring (Day 436), and summer (Day 555) under post-acclimation conditions [31], corresponding to stable performance windows spanning low-temperature operation, reduced HRT operation, and elevated organic loading, respectively. The final sampling event additionally coincided with a controlled increase in influent COD to probe system capacity under high-load conditions while maintaining stable effluent performance. For SAF-MBR 2.0, biomass was collected after more than three years of continuous operation during a period of long-term operational stability [32], characterised by consistent COD removal, CH_4 content, and low H_2 concentrations. In addition, a laboratory-scale AFBR employing non-conductive PVDF biocarriers was sampled under stable operating conditions [101] and used as a non-conductive reference system to contextualise microbial community structure and EET-related gene distributions observed in the GAC-based SAF-MBR systems.

3.1.5 Biomass collection, preparation, and nucleic acid extraction

Biomass samples were collected aseptically via the mid-height sampling port immediately before offline back-flush from two pilot-scale SAF-MBR systems: SAF-MBR 1.0 (bulk liquor and GAC) on days 342, 436 and 555; and SAF-MBR 2.0 (GAC only) on day 390. A control sample was obtained from a laboratory-scale AFBR employing PVDF biocarriers [101]. All samples were placed into pre-chilled, sterile 50 mL tubes and snap-frozen on dry ice within 5 minutes of collection. Fourteen specimens were collected: eleven in 50% ethanol (–20 °C, DNA extraction) and three in RNAlater® (–80 °C, RNA extraction).

Before processing, all glassware and plasticware (filtration units, grinding assemblies, scissors, tubes) were autoclaved. Samples were handled on ice or stored at –80 °C if not extracted immediately.

Bulk sample pre-treatment Bulk reactor liquor (1 L) was filtered through sterile 45 mm × 0.2 µm cellulose nitrate membranes to concentrate biomass. Filters were cut into strips for DNA extraction or transferred intact into lysis buffer for RNA extraction. Membranes not processed immediately were wrapped in sterile foil and stored at –80 °C ($\leq$ 1 month).

GAC sample pre-treatment GAC grains were rinsed twice in ice-cold 4% (w/v) NaCl solution (dialysed, autoclaved) and then sonicated in an ice bath with gentle grinding in sterile 200 mL conical flasks containing 100 mL of autoclaved water (DNA) or 1× PBS (RNA). The resulting

"black liquor" was collected in 50 mL tubes and centrifuged at 8,000 rpm for 10 min (4 ℃). This wash–sonicate–centrifuge cycle was repeated until sufficient pellet mass was obtained (typically six 50 mL tubes). Pellets were stored at –80 ℃ ($\leq$ 1 month) or processed immediately.

DNA extraction Total DNA was isolated with the PowerSoil™ DNA Isolation Kit (QIAGEN, Hilden, Germany), with two modifications: (i) 0.4 g wet-weight filter strips or GAC pellet per PowerBead™ tube; and (ii) proportional adjustment of Solutions C2–C4 when crude lysate exceeded 500 µL. DNA was eluted in 100 µL nuclease-free water in low-bind tubes and purified with AMPure® SPRI beads (Beckman Coulter, Brea, CA, USA).

RNA extraction Total RNA was extracted using the RNeasy® PowerSoil® Total RNA Kit (QIAGEN, Hilden, Germany). RNALater®-preserved bulk pellets were centrifuged (8,000 rpm, 5 min), washed twice in ice-cold PBS, then processed. GAC pellets were resuspended in lysis buffer, bead-beaten for 10 min (max speed), and processed per protocol. RNA was eluted in 50 μL RNase-free water and quantified by Qubit™ before DNase treatment.

The origin and preservation of each sample are summarised in Table 3.2.

Quality check for the extracted DNA and RNA samples was done using NanoDrop Spectrophotometer (NanoDrop 2000c, Thermo Scientific, MA, USA); a record can be found in Appendix B.1. DNA and RNA samples that met the quality thresholds (i.e., $A_{260}/A_{280} \geq 1.8$ and $A_{260}/A_{230} \geq 2.0$) were selected for metagenomics and metatranscriptomics sequencing.

Table 3.2: Biomass samples collected from SAF-MBR 1.0 and SAF-MBR 2.0, with sequencing availability. For ease of reference, samples were assigned a simple prefix.

Reactor	Prefix	Unit / location	Day	Metagenome	Metatranscriptome
SAF-MBR 1.0	inha_p1	afmbr_bulk	342	Y	N
SAF-MBR 1.0	inha_p2	afbr_bulk	436	Y	N
SAF-MBR 1.0	inha_p2	afmbr_bulk	436	Y	N
SAF-MBR 1.0	inha_p3	afmbr_bulk	555	Y	N
SAF-MBR 1.0	inha_p1	afbr_gac	342	Y	N
SAF-MBR 1.0	inha_p1	afmbr_gac	342	Y	N
SAF-MBR 1.0	inha_p2	afbr_gac	436	Y	N
SAF-MBR 1.0	inha_p2	afmbr_gac	436	Y	N
SAF-MBR 1.0	inha_p3	afbr_gac	555	Y	Y
SAF-MBR 1.0	inha_p3	afmbr_gac	555	Y	Y
SAF-MBR 2.0	sf_fp	safmbr_gac	390	Y	Y
AFMBR	ref	afmbr_pvdf	420	Y	N

3.2 Genomic data processing

3.2.1 Methodological rationale

The genomic data processing workflows employed in this study were selected to ensure analytical consistency, biological interpretability, and comparability with existing studies of anaerobic wastewater microbiomes. All tools applied represent established approaches in environmental metagenomics and were used here as components of an integrated analytical strategy rather than as objects of evaluation in their own right.

A central design principle of the analytical framework was the prioritisation of gene inventory–based functional classification over taxonomy-led inference. This approach reflects the study's objective, which was to examine the distribution of metabolic capabilities, particularly those associated with EET, across reactor compartments rather than to ascribe function to specific named taxa. Functional classification was therefore based on defined combinations of genes associated with electron transfer and syntrophic metabolism, synthesised from multiple strands of the literature.

Multiple analytical procedures were applied because different levels of resolution were required to address distinct research questions. Assembly-free profiling was used to characterise overall community structure, while genome-resolved analyses enabled interrogation of metabolic pathways and EET-related gene inventories. Dereplication was employed to standardise genome representation and to support relative comparisons across samples.

This framework enabled consistent comparison of functional guilds across bulk and carrier-associated communities while minimising dependence on reference taxonomy. Throughout, results are interpreted in terms of comparative patterns in functional potential across conditions, aligned with the scope of the research questions addressed.

3.2.2 DNA and RNA sequencing

DNA was sequenced on the Illumina HiSeq/NovaSeq platform (Illumina, CA, USA) to obtain a target metagenomics data size of 40 GB per sample. RNA was sequenced on the Illumina HiSeq/NovaSeq platform (Illumina, CA, USA) to obtain a target metatranscriptomics data size of 10 Gb per sample.

The sequencing depth in this study ($\geq$ 10 Gb, versus the $\sim$10 Gb often reported for shotgun surveys [203]) was chosen to achieve reliable genome recovery across the wastewater community. Such depth is costly in both budget and computation, but it is necessary because wastewater slurries harbour hundreds of low-abundance taxa spread across a chemically heterogeneous matrix. In simpler substrate systems, for example, monosaccharide-fed enrichment cultures or lab digesters dominated by a handful of specialist lineages, near-complete genomes can usually be reconstructed with $\leq$ 10 Gb, so additional sequencing is rarely justified.

Data availability Metagenome sequencing reads were submitted to NCBI's Sequence Read Archive (BioProject ID: PRJNA1085517).

3.2.3 Sequencing denoising

Metagenomics and metatranscriptomics data were processed using the Kneaddata pipeline v3 provided by Biobakery [204]. Initial processing involved quality control and trimming with Trimmomatic v0.39 [205], targeting the removal of low-quality bases and adapters. Criteria set for this step included a minimum read length of 50 bp and a base quality threshold of 20. Subsequently, DNA sequences were subjected to a filtration process to exclude non-prokaryotic sequences, utilising the hg37_human_contamination reference database [206]. A parallel filtration process was applied to RNA sequences, employing the human transcriptome database (hg38).

The quality of the resultant sequences was assessed via FastQC [207], with all sequences achieving a Phred score greater than 30. The mean quality scores were visualised using MultiQC, as illustrated in Fig. C.1, with detailed sequencing metrics presented in Appendix B.3. The final dataset comprised sequences with average lengths ranging from 49 to 138 bp and encompassed a total of 28.7 to 129.6 million paired reads.

3.2.4 Metagenome assembly

Metagenome assembly is a critical step in metagenomic analyses with the aim of reconstructing the composition of the microbial community from sequence reads. This process, analogous to whole-genome assembly, addresses the complexities inherent in metagenomic datasets. The de Bruijn graph methodology [208] serves as the foundation for most metagenome assembly approaches. In this method, sequencing reads are divided into overlapping k-mers, which form the vertices and edges of the graph. The primary objective of the assembler is to navigate this graph efficiently to reconstruct the genomes present in the sample accurately.

However, metagenome assembly presents distinct challenges. Variations in sequence coverage, arising from the differential abundance of genomes, complicate the assembly process. Additionally, closely related strains with nearly identical sequences can introduce ambiguity into the assembly graph, further complicating accurate genome reconstruction.

The effectiveness of different assembly tools remains a topic of debate within the scientific community [203]. This uncertainty is largely due to the variability in assembly outcomes, which are influenced by both the biological and technical characteristics of the samples. Comparative studies of assemblers, such as MEGAHIT and metaSPAdes, using mock and real community samples have underscored this variability, demonstrating that both the choice of assembler and the complexity of the sample significantly affect assembly results.

For this study, the metaSPAdes pipeline [209] within SPAdes (v3.15.5) [210] was employed to assemble the quality-filtered reads. The pipeline was executed using the default configuration settings. Post-assembly, the quality of the assembled genomes was assessed using Quast [211]. Detailed quality metrics and analysis results can be found in this GitHub repository.

3.2.5 Metagenome binning

The output of metagenome assembly typically consists of millions of contigs[1], providing a comprehensive dataset for further analysis. Genome binning is the process of categorising these contigs and associating them with their respective source genomes. Binning methods utilise various data features, such as sequence composition and coverage patterns, to organise contigs into bins, each representing a putative genome.

Traditional cultivation methods often fail to meet the growth requirements of many environmental microbes, limiting our understanding of their roles and behaviours. Genome binning, however, allows the study of these elusive microbes by analysing their genetic material directly from environmental samples [212, 213], enabling the exploration of microbes resistant to cultivation in laboratory settings.

The quality of metagenome-assembled genomes (MAGs) is evaluated primarily based on two key metrics: completeness and redundancy (or contamination)[2]. These metrics can be assessed using CheckM, which estimates these parameters based on the presence of SCGs.

In this study, medium-quality draft MAGs were selected for gene annotation to recover as many EET genes as possible. MAGs were classified as medium-quality if they had a genome completeness higher than 50% and a redundancy below 10%, according to the Genomic Standards Consortium's guideline [214].

While both completeness and redundancy are critical, the latter is crucial to ensure that MAGs do not exceed 10% redundancy. Over-splitting a valid MAG is less problematic than allowing a MAG to represent contigs from multiple genomes. MAGs exhibiting more than 10% redundancy are likely of poor quality, containing contigs from different genomes. In such cases, refining the bin or discarding it is advisable to avoid contaminating public databases with unreliable data.

Reducing MAG redundancy is a priority, especially in environmental samples containing a large proportion of unknown taxa. MetaBAT (v2.12.1) was reported to have the highest average purity (100%) and redundancy (0%) among hybrid binners, but the lowest average completeness (10.9%) [215]. This assessment, based on a human gut microbiome dataset (NCBI accession ERP002061), provides an acceptable approximation for wastewater and environmental microbiomes. Accordingly, assembled contigs in this study were binned using MetaBAT2 (v2.15) [216] to produce MAGs.

In this study, filtered reads were processed using Bowtie2 [217] and Samtools to facilitate efficient binning. Bowtie2, a sequence aligner, maps metagenomic reads to the metagenome assembly by first generating a mapping index and then mapping the filtered reads back to the assembly to obtain read coverage and contig linkage data. The mapping files were then sorted and compressed using Samtools [218] to improve computational efficiency. This mapping step is crucial for accurately grouping contigs into bins.

[1] Contigs, short for contiguous sequences, are the initial assembly outputs where overlapping short reads are merged to form longer sequences.

[2] The terms "redundancy" and "contamination" are often used interchangeably in bacterial genome contexts, but they have distinct meanings. Redundancy refers to the presence of multiple copies of single-copy genes (SCGs) in a genome bin, while contamination implies foreign genetic material. Contamination is not inferred without substantial evidence. This distinction was clarified by Jed Fuhrman.

The sorted mapping files were condensed into a depth file, which illustrated contig coverage and served as input for MetaBAT2 during binning. A minimum contig length of 1,000 bp was set for binning, as this is the empirical minimum for accurate binning; shorter contigs tend to produce noisy tetranucleotide frequency profiles and result in suboptimal binning performance [219]. The quality of the binned genomes was subsequently assessed by evaluating the completeness and redundancy of the MAGs using CheckM (v1.2.2) [220].

3.2.6 MAGs refinement

MetaBAT2 produces highly pure bins; however, this comes at the expense of completeness due to stringent cutoffs at sequence overlapping edges [221]. Bins with low completeness may struggle to capture functional elements located in conserved genomic regions. To address this, bins can either be refined to gain clearer insights into their contents or discarded.

To comprehensively recruit EET genes from MAGs, refinement was performed using the BinSPreader protocol [222], which aimed to enhance genome bin completeness for improved EET gene capture. The details of this protocol, with minor modifications, are provided in Section A.2.

The refinement process utilises an assembly graph to enhance bin completeness. An assembly graph captures the connectivity within an assembly, consisting of unitigs—longest contiguous sequences derived from clear k-mer[3] paths. It also includes links between unitigs, derived from read overlaps. While the assembler resolves repeat regions and generates contigs from non-branching paths, the assembly graph uniquely retains the connectivity information of the DNA, distinguishing it from contigs. Incorporating the short edges of the assembly graph into MAGs can improve bin completeness without introducing redundancy. The crucial steps of bin refinement are summarised in Figure 3.3, and the specific application and evaluation of this pipeline is detailed in Chapter 4.2.

3.2.7 MAGs clustering and dereplication

Dereplication of MAGs using dRep [223] was performed to address key aspects of this study. The primary objective was to ensure accurate identification and consolidation of MAGs collected at different sites and time points but originating from the same species, thereby maintaining consistency in species representation. Secondly, dereplication aimed to reduce redundancy in functional annotation, since genomes with high similarity are likely to perform the same functions. Finally, the process facilitated the greater distinction between genomes, enabling accurate mapping of annotated genes to the transcriptome. This approach prevented mis-mapping of short reads (typically around 150 bp) that could align equally well to nearly identical genomes, complicating the determination of their true origin [223].

The dereplication process involved selecting a representative set of Species-level Representative Genomes (SRGs) from a larger pool of genomes. This selection was guided by the Average Nucleotide Identity (ANI) threshold and aligned fraction, both crucial parameters in dereplication.

[3]A substring of length (k) within a biological sequence, composed of nucleotides (A, T, G, and C) in sequence analysis

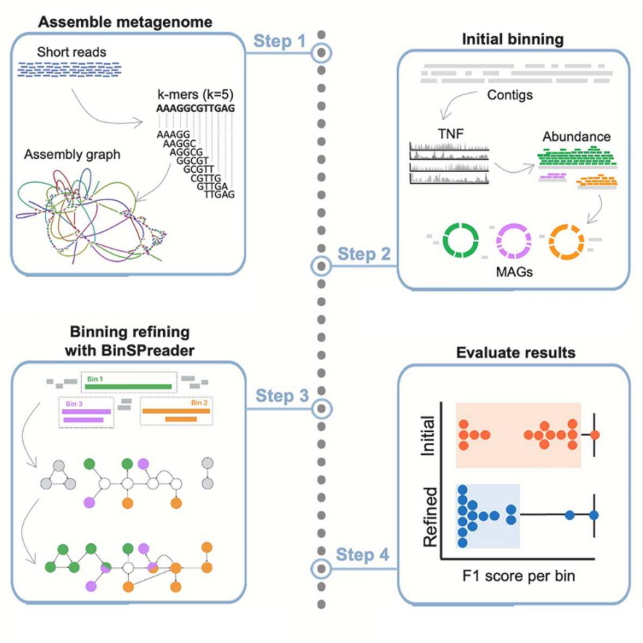


Figure 3.3: Steps for obtaining assembly-graph refined MAGs

ANI thresholds are context-dependent, and there is no universal standard defining when genomes are considered identical. In this study, ANI was determined using secondary clustering algorithms, with minimum secondary ANI and aligned fraction values set to ensure sufficient genome overlap for accurate ANI calculation. For species-level dereplication, a widely accepted 95% ANI threshold was employed, as supported by prior evaluations [224, 225].

3.3 Taxonomy profiling via read-based approach

The microbial profiles of the metagenome and metatranscriptome were characterised using a read-based taxonomy profiling technique, prioritising the inclusion of unassembled sequences to ensure a comprehensive analysis.

For the taxonomic assignment, filtered reads were analysed using Kaiju (v1.9.0) [226]. Nucleotide sequences were translated into amino acid sequences and aligned against the NCBI BLAST non-redundant (nr) protein database to identify exact matches. This approach improves sensitivity over nucleotide-based searches due to the redundancy of the genetic code, where 64 codons encode 20 amino acids, allowing for synonymous mutations that do not alter protein sequences. This, along with evolutionary conservation and statistical inference, enhances the method's effectiveness despite challenges like alignment gaps.

This read-based approach was instrumental in profiling the metagenome and metatranscriptome of the SAF-MBR 2.0, as discussed in Chapter 5.

3.4 Phylogenetic analysis

3.4.1 Assign taxonomy for metagenome-assembled genomes

Accurate taxonomic classification and phylogenetic analysis of MAGs are crucial for understanding microbial community composition and evolution. In this study, the taxonomic classifications and phylogenetic relationships of MAGs were determined using GTDB-Tk (v2.3.2) [227] and PhyloPhlAn (v3.0.3) [228]. A subsequent cross-validation process was conducted to ensure the accuracy of the results.

GTDB-Tk, specifically designed for the taxonomic categorisation of bacterial and archaeal genomes, utilises the Genome Taxonomy Database (GTDB) [229], which is updated biannually to reflect the latest genomic data from the NCBI Assembly database. GTDB-Tk employs Prodigal [230] for gene prediction through open reading frame () detection and HMMER to identify a specific set of marker genes—120 for bacteria and 122 for archaea. These marker genes are then aligned and concatenated into a MSA specific to each domain. The alignment is further refined using a GTDB-specific mask, facilitating the placement of genomes on a domain-specific reference tree through the pplacer tool [231].

On the other hand, PhyloPhlAn identifies ORFs corresponding to universal gene markers to integrate MAGs into a reference tree, utilising the species-level genome bins (SGBs) collection [232] as the reference framework.

Comparative analysis of the results indicated that GTDB-Tk produced more consistent and reliable outcomes for the genome set analysed in this study, as detailed in Appendix A.3. Consequently, the GTDB-based annotations were selected as the preferred taxonomic framework for this research.

3.4.2 MAGs phylogeny

To assess biodiversity and ensure the consistency of taxonomy with MAGs clustering, a phylogenetic tree was constructed using only the MAGs. This approach helps to illustrate the evolutionary distances among the MAGs.

Initially, BLAST Diamond (v0.9.24) was used for high-throughput DNA-to-protein alignment [233], facilitating the identification and alignment of sequences based on their evolutionary relationships.

Next, MSA was conducted using MAFFT (v7.505) [234]. To enhance the quality of the phylogenetic analysis, alignment trimming was performed with TrimAl (v1.4), which removed misaligned regions and gaps [235].

The phylogenetic tree was first constructed using FastTree (v2.1.11) through approximate maximum likelihood methods [236]. This tree was further refined using RAxML (v8.2.12) for maximum likelihood-based phylogenetic inference [237].

3.4.3 Protein phylogeny

To assess the diversity, evolutionary relationships, and potential functional profiles within each protein cluster, phylogenetic trees were constructed for the identified protein sequences, including putative EET genes such as MHC and e-pili.

MSA for these protein sequences were performed using T-coffee via EMBL-EBI for its high accuracy to divergent sequences, particularly suitable for small data set (i.e, less than 500 sequences) [238], which employs a three-stage algorithm specifically optimised for protein sequences. The alignment results were then examined through Mview [239], sequences with alignment coverage larger than 75% were retained. Following this, phylogenetic trees were generated using RAxML-NG [240] with model "LG+G+F0" that utilises amino acid replacement matrix developed by Le et al. [241], combined with machine-learning estimates for among-site rate heterogeneity and stationary frequencies.

3.5 Functional annotation

In metagenomic research, functional annotation is a critical step that provides insights into the metabolic potential and ecological roles of microbial communities. This study employs both genome-based and community-based annotation approaches to maximise the breadth and depth of functional insights. The rationale for integrating these methods is based on their complementary strengths, which are particularly relevant for the diverse and complex microbial communities under investigation.

3.5.1 Community-based functional annotation

The community-based approach, exemplified by HUMAnN 3.0 (HMP Unified Metabolic Analysis Network) [204], was selected for its ability to directly profile functional potential from metagenomic sequencing reads, bypassing the need for genome assembly. This method streamlines analysis by eliminating the computational demands of assembly and captures the functional potential across the entire microbial community, including rare or less abundant species. This is particularly useful in diverse, complex environments where genome assembly is challenging. HUMAnN provides pathway-level insights into community functions, offering a broad overview of metabolic capabilities. However, it has limitations, particularly in resolution and specificity, as it provides community-level information rather than organism-specific data. Additionally, reliance on reference databases may introduce biases, especially for uncharacterised genes or organisms, potentially leading to inaccurate or incomplete annotations.

In practice, the relative abundance of microbial taxa within each sample was estimated from the filtered reads. HUMAnN3 identified and annotated gene families using the UniProt UniRef50 database for higher sensitivity compared to UniRef90 [242], and modulated metabolic pathways through MetaCyc [243]. Clade-specific marker genes were sourced from the ChocoPhlAn database [244].

3.5.2 Genome-based Functional Annotation

To complement the broad insights provided by HUMAnN, a genome-based approach was used for the functional annotation of representative genomes obtained through dereplication. This annotation was performed using three specialised tools: DRAM (v1.5) [245], METABOLIC (v4.0) [246], and eggNOG-mapper (v2.0) [247]. By integrating these tools, which use various databases, the resulting annotations are more comprehensive (Fig. 3.4).

DRAM (Distilled and Refined Annotation of Metabolism) is tailored to annotate MAGs and VirSorter-identified viral contigs. It utilises multiple databases, including UniRef90, PFAM, dbCAN, RefSeq viral, VOGDB, and MEROPS peptidase, as well as custom user databases. DRAM operates in two stages: first, an annotation step assigns database identifiers to genes; second, a distillation step curates these annotations into functional categories, including electron transport chain and important functions relevant to anaerobic digestion. For viral contigs, DRAM also identifies potential Auxiliary Metabolic Genes (AMGs) by assigning auxiliary scores and flags that indicate the likelihood that a gene is both metabolic and viral.

METABOLIC (v4.0) was used to predict metabolic and biogeochemical functional traits in MAGs by integrating metagenomic and metatranscriptomic reads to assess pathway coverage within microbial communities. This enabled detailed views of community metabolism, the identification of active metabolic processes, and the calculation of genome coverage, with a particular focus on elemental and biogeochemical cycles such as nitrogen, carbon, and sulfur.

Metabolic functions were further examined through KEGG modules, which captured transcript coverage across core metabolic processes, and MetaCyc databases, which provided finer resolution of specific pathways. Transcript coverage was normalised to Reads Per Kilobase Million (RPKM), allowing quantification of activity levels across pathways. To generate a functional overview, the Metabolic Weight (MW) score was applied, reflecting the relative functional significance of metabolic activities within the community.

eggNOG-mapper (v2.0) provides rapid functional annotation of novel sequences using precomputed orthologous groups and phylogenies from the eggNOG database. Notably, eggNOG incorporates the Clusters of Orthologous Genes (COG) database, including archaeal COGs (arCOGs) crucial for AD microbiome annotation. The tool also integrates Pfam and KEGG for enhanced annotation capabilities, making it valuable for cross-checking with self-curated databases and validating novel EET gene annotations. By leveraging orthology predictions, eggNOG-mapper achieves higher precision in functional annotation compared to traditional homology-based methods like BLAST, as it transfers functional information from fine-grained orthologs, reducing issues related to functional divergence among close paralogs.

The collective use of DRAM, METABOLIC, and eggNOG mapper provides comprehensive information on the metabolic capabilities and potential ecological functions of the MAGs analysed. This genome-based approach, with its focus on high-resolution, detailed annotations, complements the community-level insights offered by HUMAnN. The integration of these tools results in a robust and nuanced understanding of microbial functions within the bioreactors.

	DRAM	METABOLIC 4.0	eggNOG-mapper v2
KEGG	Require license	No	No
Kofam	Yes	Yes	Yes
UniRef90	Yes	No	No
PFAM	Yes	No	Yes
dbCAN	Yes	Yes	No
CAZy	No	No	Yes
MEROPS	Yes	Yes	No
RefSeq viral	Yes	No	No
VOGDB	Yes	No	No
GTDB-Tk	No	Yes	No
eggNOG Orthology (v5)	No	No	Yes
Gene Ontology	No	No	Yes
EC numbers	No	No	Yes
BiGG reactions	No	No	Yes
COG	No	No	Yes
SMART	No	No	Yes
Custom databases	User-defined	Custom HMM profiles	No

Figure 3.4: Databases used by different bioinformatics pipelines (DRAM, METABOLIC 4.0, and eggNOG-mapper v2)

3.6 EET gene annotation

The annotation of novel genes is crucial for understanding their function, regulation, and evolutionary significance. Traditional sequence similarity approaches, such as BLASTp, have limitations when dealing with distantly related sequences or low sequence similarity. To overcome these challenges, more sophisticated methods, such as hidden Markov Models (HMMs), are often employed.

HMMs provide a robust alternative to sequence similarity approaches by modelling complex relationships between sequence features. Unlike traditional methods that rely on direct protein sequence comparisons, HMMs utilise statistical models to account for dependencies between adjacent sequence elements. This capability enables HMMs to identify conserved patterns and motifs indicative of gene function or regulation, even in cases of minimal sequence similarity or distantly related sequences. As a result, HMMs are particularly effective in capturing the underlying structure of biological sequences, such as genes [248, 249].

While sequence similarity methods excel at detecting homologous genes in closely related sequences, they often struggle with distantly related organisms where sequence similarity is low. Additionally, these methods do not consider the context or arrangement of sequence features, which are critical for understanding gene function and regulatory mechanisms. Consequently, their utility is reduced in the annotation of novel genes, particularly in evolutionary studies.

In contrast, HMMs offer a more flexible approach for gene annotation. By modelling the intricate relationships between sequence features and identifying conserved patterns in distantly related sequences, HMMs are invaluable for studying gene function, regulation, and evolution across diverse organisms. This methodology can provide more accurate gene annotations.

3.6.1 Remote homologues sequence detection via hidden Markov model

Hidden Markov Models (HMMs) have been widely applied in bioinformatics since their development in the 1970s, becoming one of the most prevalent models for biological sequence analysis due to their versatility and adaptability [250].

Profile hidden Markov model

The Profile Hidden Markov Model (pHMM) is a specialised variant of the HMM designed to represent a cluster of sequences by capturing evolutionary changes in protein sequences [251]. Compared to pairwise alignment-based methods, pHMMs demonstrate higher sensitivity in detecting remote homology among protein sequences [252].

The motivation for using pHMMs lies in their ability to effectively model evolutionarily conserved patterns within a family of homologous biological sequences, such as proteins or nucleic acids. By incorporating position-specific information from a multiple sequence alignment (MSA), pHMMs model both conserved and variable regions within a sequence family. This approach also considers the likelihood of insertions and deletions at each position, making pHMMs particularly adept at identifying distant homologs.

A pHMM is constructed by analysing multiple sequences within a family, significantly enhancing the model's coverage, sensitivity, specificity, and scope. The process involves capturing the variations and conserved motifs across the sequence family, including acceptable substitutions and patterns of insertions and deletions. This results in a model with greater depth and precision [251]. The multi-sequence approach provides a richer representation of the evolutionary depth of the sequence family compared to single-sequence models.

By incorporating evolutionary data from multiple alignments, pHMMs achieve heightened sensitivity and specificity. These enhanced models excel in identifying true homologs, even among unrelated sequences with superficial similarities [253]. The evolutionary insights embedded within these models are crucial for distinguishing closely related sequences from those that are not, thus enabling more precise and accurate sequence analysis.

Multi-sequence pHMMs are indispensable for complex analytical tasks, such as discovering distant homologs, delineating protein families, and annotating sequence domains [254]. The analytical depth and range provided by these models surpass the capabilities of single-sequence-based approaches, making them a critical tool in sequence analysis.

Search methods

HMMER is a free and widely used software package for running Hidden Markov Models (HMM) [255]. It offers two primary search methods, `hmmsearch` and `hmmscan`, each suited to different dataset types and research objectives. `hmmsearch` searches one or more HMM profiles against a sequence database, which is ideal for situations where a small number of profiles is compared to a large sequence database. Conversely, `hmmscan` is employed to query protein sequences—or DNA sequences in specific contexts—against large collections of protein profiles. This method

is more efficient when the profile database is significantly larger than the sequence set being queried, making it suitable for applications where the number of sequences is small but the library of profiles is extensive.

For large-scale metagenomic analyses involving millions of sequences, `hmmsearch` is typically preferred due to its computational efficiency, unless there are significantly more profiles than sequences [256]. Specifically, `hmmsearch` excels in identifying sequences in large databases that match a limited number of profiles, and in analysing specific protein families across diverse organisms.

HMMER provides per-target and per-domain outputs based on E-values. The per-target output reports sequences that meet an E-value threshold of ≤ 10, considering the entire sequence. The per-domain output highlights domains within these sequences that also meet this threshold, emphasising specific functional or structural regions.

Alignment quality in HMMER is assessed using E-value and bit-score metrics. E-value reflects the expected number of random hits with a score equal to or exceeding a threshold, with values ≤ 1 indicating biological significance. The bit-score, independent of database size, measures the likelihood of a match being model-based rather than random, with scores above 10 considered statistically significant.

For this study, significant gene matches were identified using an **E-value threshold of** ≤ 0.1 **and a bit-score threshold of** ≥ 10, ensuring biologically relevant results.

3.6.2 MHC detection

The primary objective of this pipeline is to detect MHCs (Fig. 3.5). MHCs participate in EET either as extracellular or outer membrane MHCs interacting with substrates (e.g., OmcE, OmcS, OmcZ), or as periplasmic MHCs incorporated into a porin to form a porin-cytochrome complex (PCC) (e.g., MtrABC, PioABC).

Genes were screened based on the number of heme motifs. Using regular expressions for heme binding motifs: [C·CH], [C··CH], [C···CH], and [C··········[! = C]··CH], amino acid sequences containing at least three heme motifs were considered potential MHCs.

PSORTm [257] predicted the subcellular localisation of potential MHCs, identifying if they function as extracellular or outer membrane proteins. Alternatively, they could be periplasmic proteins located within eight open reading frames of a beta-barrel outer membrane porin (BB-OMP) predicted using pHMM, or other extracellular or outer membrane-predicted MHC proteins.

For genes that passed the rule-based criteria, an HMM search was performed against custom pHMMs built from identified MHC genes [258, 259]. Each gene underwent a BlastP search [260] to collect amino acid sequences with over 35% identity and 70% coverage. These sequences were clustered using MMseqs2 [261] at a minimum identity of 70%. Subsequently, each protein cluster was aligned using MUSCLE v5 [262] and incorporated into a pHMM database, which was then compressed into a final database. A complete list of MHC-related pHMMs is found in Table B.4.

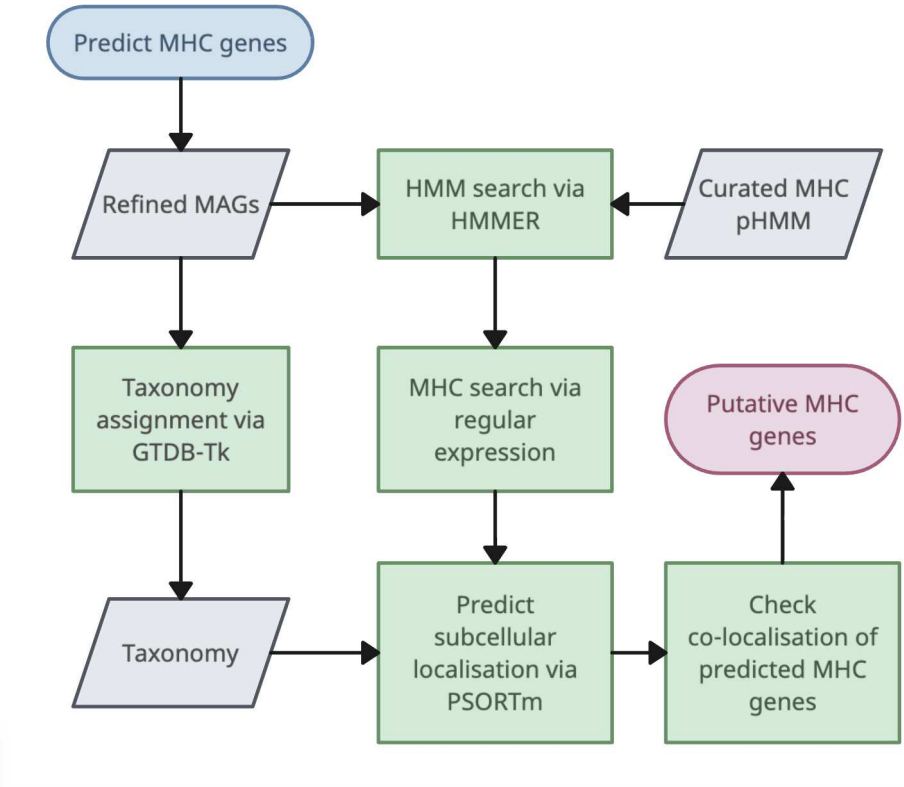


Figure 3.5: MHC identification workflow

3.6.3 E-pili detection

A HMM search was performed against a custom database curated specifically for this study, compiling from four studies [38, 115, 161, 164] for the most updated reference. This database included only genes identified as e-pili, either through experimental validation or genetic inference of conductivity, and these were compiled into a pHMM (Appendix B.5).

A HMM search was conducted against a custom database compiled from findings in four studies [38, 115, 161, 164]. Only genes identified as e-pili, either through experimental confirmation or genetic inference of conductivity, were included in this database and compiled into a pHMM (Appendix B.5).

The predicted genes were analysed using a rule-based approach. FlaFind [263] was employed to predict archaeal class III (type IV pilin-like) signal peptides and their corresponding prepilin peptidase cleavage sites. Similarly, PilFind [264] was used to predict bacterial type IV pilin-like signal peptides and their cleavage sites. After confirming the cleavage sites, only genes with aromatic amino acids content higher or equal to 9.8% and gaps of less than 22 and 35 amino acids between consecutive aromatic amino acids for bacteria and archaea, respectively (Fig. 3.6. Such thresholds were determined empirically based on the characteristics of experimentally verified e-pili [164].

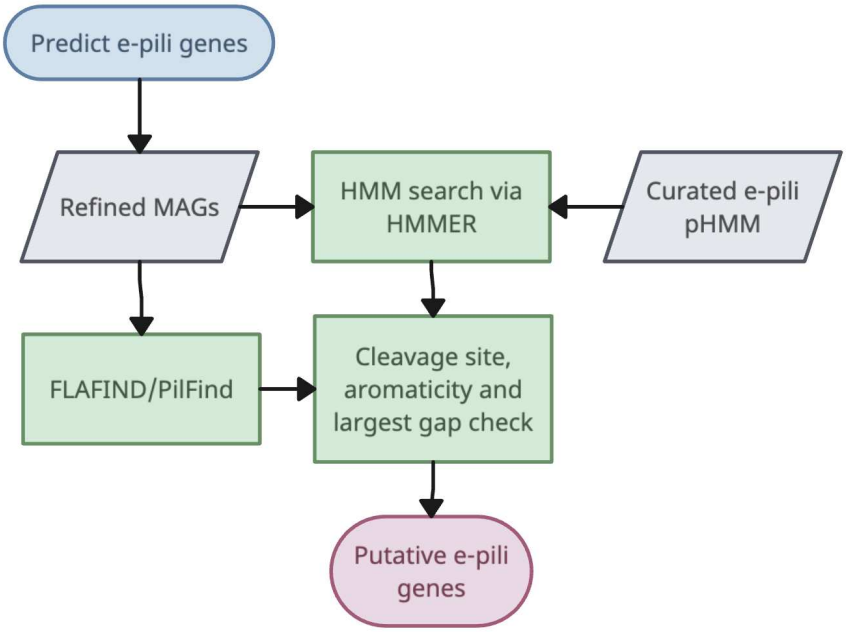


Figure 3.6: E-pili identification workflow

3.7 Expression level analysis

To account for the gene expression level from the MAGs, metatranscriptome assembly was performed to enhance the specificity of the gene mapping for longer reads.

The filtered RNA sequences were assembled using Trinity (v2.11.0) [265]. The quality of the metatranscriptome assembly was accessed using rnaQUAST (v2.2.2) [266].

Gene expression level was assessed by mapping the sequences of interest to the Trinity assembly using Seal in BBmap (v38.98) [267].

Seal can associate a single kmer with multiple references without alignment. This allows for the accurate assessment of gene expression levels in environments where sequences may be shared among different genes, isoforms, or organisms.

3.8 Pangenomic analysis

Pangenomes comprise the full complement of genes within a defined phylogenetic group and provide a more complete picture of genetic diversity than single reference genomes. A pangenome is constructed by comparing genome sequences across multiple individuals of a species to identify shared genes (”core genes”) as well as unique genes found only in certain strains (”accessory genes”) [268].

In the context of studying EET, pangenomic analysis can facilitate the functional annotation of unknown genes by leveraging the characterised genes in gene clusters. Given that only about 25% of gene clusters in methanogens were annotated with known functions [269], this approach can enhance our understanding of the genetic mechanisms of electron-accepting reaction sites

underlying EET. Specifically, this pangenomic study aims to determine whether the electron uptake genes of EET in methanogens are accessory or core genes.

By integrating metagenomics data, we can gain deeper insights into the functional aspects of methanogens and their fitness in their respective niches in EET. This method facilitates an understanding of how variations in EET genes contribute to the growth and efficiency of methanogens in diverse environmental conditions, particularly in anaerobic digestion ecosystems. By studying the presence or absence of specific gene clusters in varying environments, we can infer the significance of these genes in the ecological adaptability and metabolic efficiency of methanogens.

3.8.1 Pangenomic analysis using Anvi'o

Pangenome analysis was conducted using Anvi'o 8.0 [270] to investigate the participation of EET in *Methanothrix* by including the public genome available in the NCBI depository, as well as the MAGs generated from this study. Metadata for the public genome were collected using the `epost` command in Entrez Direct [271], serving as labels for subsequent pangenome analysis.

Medium-quality genomes (i.e., completeness $\geq$50%, redundancy $<$5% were subjected to a four-step pangenomic workflow, including the generation of Anvi'o database for each genome, identification of gene clusters across all genomes using `anvi-gen-genomes-storage`, computation of hierarchical clustering of genomes based on their gene cluster content using `anvi-pan-genome`, and visualisation and exploration of the pangenome using `anvi-display-pan`. Each genus was assigned to a pangenome. This preprocessing protocol followed a transparent workflow provided by Tom Delmont [272] and A. Murat Eren [273].

Following the manual allocation of gene clusters to core and accessory genomes based on gene presence/absence frequency, wherein clusters universally present in all or most genomes were categorised as the core genome, and the remainder were designated as the accessory genome, it was observed that the extent of the accessory genome varies with phylogenetic divergence (i.e., branching and distance). Following the export of the core and accessory genomes, gene clusters were annotated functionally and phylogenetically using the approaches discussed in earlier sessions.

3.8.2 Homology detection and definition in Pangenomes

Accurate identification and classification of homologous genes are essential for pangenome assembly. In closely related genomes, homologous genes exhibit significant conservation in their amino acid sequences, which sharply contrasts with the divergent nature observed in non-homologous genes across and within genomes. This conservation enables relatively straightforward clustering of genes at the species or genus level within a pangenome.

However, ambiguities in homology determination and potential clustering errors must be considered. To address these concerns, two approaches were employed to assess the robustness and precision of homology determination: (1) evaluating the stability of the pangenome in response to variations in amino acid sequence similarity thresholds by adjusting the inflation value of the

Markov Cluster (MCL) Algorithm, and (2) examining the functional diversity within gene clusters by comparing their functional heterogeneity against the Clusters of Orthologous Groups (COGs).

Pangenome analysis involves clustering genes based on sequence homology, where the MCL inflation parameter influences the granularity of clustering and, consequently, the identification of core and accessory genes. Typically, an initial inflation value between 2 and 3 is employed, with adjustments made based on the specificity and sensitivity required by this study [272].

Different MCL inflation values were tested to optimise pangenome analysis, particularly for environmental microbes. The performance of each clustering was evaluated to ensure an accurate representation of gene families. Cluster stability, defined as the robustness of gene clusters against minor variations in the MCL inflation value, served as the primary metric. Ideally, stable clusters exhibit minimal fluctuations in the number of gene clusters across varying inflation values.

In this study, the robustness of the pangenome to the MCL inflation parameter was assessed by adjusting the inflation parameter by ± 2. The pangenome configurations remained qualitatively consistent, with the overall structure and relative proportions of the genus core, species cores, and accessory genome showing minimal changes.

Gene clusters were defined based on amino acid sequence similarity without inherently considering functional similarity. However, it was hypothesised that gene clusters would generally consist of genes with analogous functions. This hypothesis was tested by evaluating the proportion of gene clusters containing genes annotated with different COG functions.

The internal consistency of the gene clusters was further examined through functional annotation. Although gene clusters may be subdivided based on amino acid sequence differences, they can still encompass homologous genes performing identical functions. The analysis involved comparing the predicted functions of gene clusters unique to specific groups against those characteristic of other groups.

3.8.3 Functional enrichment analysis

Upon establishing a pangenome, a fundamental objective is to examine the functions associated with gene clusters, particularly those accessory to specific subsets within the pangenome. Functional enrichment analysis was performed to identify functions disproportionately represented within specific clades or subclades of the pangenome. The genomes were segregated into distinct groups based on assigned attributes, such as the source environment or, for the wastewater sample, the reactor operating condition. Functions with significant association with any of these groups were pinpointed, specifically identifying functions that are prevalent in genomes within a particular group but largely absent in those from other groups.

Since the functional annotation was performed at the gene level, and in most cases a gene cluster presents only one function, for annotation purposes, the gene cluster function was collapsed to the function with the highest frequency within the cluster. Therefore, it is necessary to emphasise that the functional enrichment analysis performed here refers to functional association and not functional annotation.

3.8.4 Amyloid protein prediction

Given that the targeted MspA gene in *Methanothrix* lacks homologs, a rule-based approach, focusing on the predicted chemical properties of protein sequences, was favoured for amyloid protein prediction. Bioinformatics tools offer a cost-effective, time-efficient, and robust alternative to traditional experimental methods, such as atomic force microscopy [274]. Combining multiple tools that employ various rules improves prediction confidence through cross-verification.

The prediction process begins with identifying a signal peptide, which suggests that the protein can be secreted to the outer membrane. SignalP-6.0 [275] was utilised to detect signal peptides in the protein sequences, determining their potential for secretion if a sequence encodes a signal peptide (SP I). To further verify subcellular location, the motif-based prediction tools Cello [276] and PSORTb [257] were used.

The amyloidogenic potential of the protein sequences was assessed using AMYPred-FRL [277]. To further validate this potential, PASTA 2.0 [278] was employed to calculate the free energy of aggregation and evaluate the β-sheet content of the sequences, adding a layer of biophysical validation to ensure that the predicted amyloidogenic regions have the propensity to form amyloid structures under physiological conditions. Using Mthe_1069 as a reference, sequences with a free energy value lower than -10 and a β-sheet content greater than 30% were considered MspA-like proteins.

The final step involved cross-checking the protein sequences for significant functional annotation using eggnog-mapper 2.0 [279]. This tool provides functional annotation and orthology assignments, ensuring that the predicted amyloid proteins are not already well-characterised, thereby reducing the likelihood of misannotation.

By systematically applying these tools and criteria, the protein sequences were rigorously analysed for their potential to form amyloid proteins. The combination of signal peptide detection, extracellular localisation, amyloidogenic properties, favourable free energy, and beta-sheet content, validated against the reference sequence Mthe_1069, offers a comprehensive approach to amyloid protein prediction.

3.9 Horizontal gene transfer analysis

3.9.1 Genome-wide detection of putative HGT

HGTector 2 [280] was employed to identify HGT events within the MAGs obtained in this study. Protein sequences from the MAGs were first extracted and then subjected to BLASTp searches against the NCBI non-redundant protein database using an E-value cutoff of 1×10^{-5}, a minimum sequence identity threshold of 30%, and a minimum alignment coverage of 50%.

HGTector 2 is designed to detect HGT by integrating a phylogenetic-distance-based classification system that distinguishes between self-, close-, and distal-homologs. HGTector 2 dynamically retrieves and utilises the NCBI Taxonomy database, allowing for a systematic and accurate classification of homologs. This strategy effectively reduces false positives by differentiating between

vertically inherited and horizontally acquired genes.

Following the BLASTp searches, the resulting homologous hits were classified into self, close, and distal groups based on taxonomic distances. To minimise false positives, stringent filtering criteria were applied; only genes exhibiting significantly anomalous phylogenetic distributions, as determined by a custom-defined taxonomic exclusion list, were retained as putative HGT candidates.

3.9.2 Identification of mobile genetic elements

Mobile genetic elements associated with (HGT) were identified to differentiate between the primary HGT mechanisms of conjugation and transduction. PLASMe was employed to identify plasmid contigs within the genome collections. This method integrates a transformer-based approach to differentiate plasmid-derived sequences from chromosomal sequences and combines alignment-based techniques for closely related plasmids with machine learning models to address diverged plasmids [189]. A minimum coverage threshold of 0.9 was applied and validated using BLASTN to ensure accuracy. PhaBOX was employed to identify phage contigs from the genome collection. This tool incorporates several established methods: PhaMer for phage identification [190], PhaTYP for lifestyle prediction [281], PhaGCN for taxonomic classification [282], and CHERRY for host prediction [283]. By combining these tools, PhaBOX enables efficient detection of phages, determination of their lifestyles (lysogenic or lytic), taxonomic classification, and bacterial host prediction. The analysis used a minimum contig length of 1,000 bp, with identity and coverage thresholds set at 0.9 to ensure reliable results.

Detection of identical or nearly identical ARGs in phylogenetically disparate MAGs linked by similar plasmid sequences provided evidence of conjugation-mediated HGT [284]. Similarly, phage sequences bearing host-derived ARGs or other host-like genes indicated potential transduction events [285]. These assignments are supported by analysing gene synteny, orthologous gene clusters, and the presence of MGE-associated motifs (e.g., integrase or transposase genes).

Unlike conjugation and transduction, transformation (the direct uptake of free DNA from the environment) does not involve a discrete MGE [286]. Consequently, attributing ARG acquisition to transformation from metagenomic data alone is challenging. While the detection of bacterial competence genes (e.g., com operons) can imply the capacity for transformation, definitive evidence of transformation-mediated HGT requires temporal and ecological context [287]. For instance, longitudinal sampling combined with population genomics may show a gene entering a lineage that previously lacked it, without any associated plasmid or phage signature. However, metagenomic snapshots are insufficient to confirm such events. Therefore, transformation is acknowledged as a plausible HGT route, but it is not directly inferred from the static metagenomic data within this study.

3.9.3 Gene flow annotation

Using BlastP with identity score $>$ 99% and query coverage $>$ 75%. For target screening, although it is a common practice to set the `max_target_seqs` parameter to 1 in an attempt to fetch the best hit, however, for the local Blast version, this parameter will return the first hit that passed the criteria set by the user [288]. In response to this implicit yet significant issue, NCBI suggested not to set `max_target_seqs` parameter less than 5 [289] and revised the code for the online version to truly do the best-hit matching [290]. Thus, the `max_target_seqs` parameter was set to 5 and best-hit was settled based on 1) identity score, 2) query coverage, 3) e-value, and 4) bit-score, in that order.

3.10 Statistical analysis

All statistical procedures were executed in Python 3.11 using the pandas [291], numpy [292], scipy [293], scikit-learn [294] and statsmodels [295] libraries; figures were generated using matplotlib [296], seaborn [297], and plotly [298]. A two-tailed significance level of $p < 0.05$ was adopted throughout.

Microbial community data are typically non-Gaussian, zero-inflated and compositional. Accordingly, alpha-diversity metrics (Shannon, Simpson and richness) were first screened for normality; when assumptions held, a one-way ANOVA with Tukey's HSD post-hoc test was used because it offers greater power under Gaussian error, whereas the distribution-free Kruskal–Wallis test was applied otherwise. This parametric–non-parametric pairing is standard in metagenomic studies, ensuring robustness while retaining interpretability [299].

For beta-diversity, Bray–Curtis dissimilarities were chosen because they are insensitive to joint absences and remain the de facto measure for ecological count data [300, 301]. Ordination employed principal coordinate analysis (PCoA), which provides a metric-space representation compatible with Bray–Curtis distances, facilitating visual comparison of sample clusters. Group separation on the ordination space was formally evaluated by ANOSIM with 9,999 permutations, a permutation-based analogue of traditional ANOVA that does not require multivariate normality [302].

Correlation analyses used Spearman's rank coefficient for monotonic relationships and Pearson's product–moment coefficient when normality and linearity assumptions were met. This dual approach mirrors common practice in microbial ecology, allowing the capture of both linear and monotonic associations without inflating Type I error [303].

A random forest classifier (500 trees, default hyperparameters) was employed to rank environmental and genomic predictors of EET gene distribution. Random forests were selected because they (i) handle high-dimensional feature spaces without explicit variable selection, (ii) accommodate mixed data types (continuous, categorical, ordinal), and (iii) provide intuitive variable-importance scores while remaining computationally efficient [304]. These properties make them a conventional yet powerful choice for multi-omics data sets, which often comprise thousands of correlated predictors.

Chapter 4

Diversity of Genes Related to Extracellular Electron Transfer and Their Significance in Carbon Degradation

4.1 Introduction

Staged anaerobic fluidised-bed membrane bioreactor technology delivers anaerobic secondary treatment directly digesting low-strength municipal sewage without a separate sludge-digestion stage, and has demonstrated net-energy-positive operation and compliance with secondary BOD/TSS standards[30–32]. A key feature of this system is the fluidised GAC, which both scours the membrane to mitigate fouling and serves as a high–surface-area biocarrier that enhances carbon removal and CH_4 production; by retaining biomass in the reactor, the GAC increases the effective cell retention time without sludge return, enables the enrichment of slow-growing organisms, and buffers the community against shock loads, overloading, and inhibitory substances.

A pilot-scale study of the SAF-MBR reported an elevated hydrolysis rate constant of $0.35\,\mathrm{d}^{-1}$ at 37 °C, and biomethanation testing indicated that 98.5% of the CH_4 originated from the active biomass on the GAC (App. A.1). Kinetic analyses using glucose as a model organic substrate revealed a half-saturation coefficient of $1\,\mathrm{mgCOD\,L}^{-1}$ [32], which is notably lower than the typical $127\,\mathrm{mgCOD\,L}^{-1}$ reported in conventional operations [33].

Introducing GAC can induce EET by acting as an electrical conduit [29, 305]. EET facilitates the breakdown of complex organic compounds in anaerobic environments lacking conventional electron acceptors such as oxygen but rich in extracellular acceptors like Fe(III) and Mn(IV) [306]. By transferring electrons to these minerals, EET supports biogeochemical reactions that conserve more energy than fermentation [306]. Exoelectrogenic bacteria, which enable EET, are frequently identified by the presence of MHC and e-pili [69].

In wastewater treatment, various studies have shown that EET enhances hydrogen-independent CO_2 reduction methanogenesis, in which methanogens act as extracellular electron acceptors for exoelectrogenic bacteria [17, 22, 24]. This phenomenon has attracted considerable attention for engineering applications aimed at increasing CH_4 yields [74, 102]. While the role of EET

in promoting methanogenesis is widely recognised, its influence on other anaerobic processes, including hydrolysis and acidogenesis, has received comparatively less attention [12].

This chapter examines the diversity of genes related to EET in SAF-MBR systems and their significance in carbon degradation, with the goal of elucidating how EET contributes to reactor performance. Metagenomic and metatranscriptomic analyses were conducted on GAC-associated biofilm and bulk liquid biomass collected from two pilot-scale configurations, SAF-MBR 1.0 [31] and SAF-MBR 2.0 [32], thereby encompassing distinct operational conditions and microbial communities. These analyses resolve the EET-related genetic potential and its active expression and relate them to microbial community structure, metabolic pathways for complex organic carbon degradation, and system performance. Specifically, the chapter identifies key EET-associated genes, determines their distribution between suspension and GAC-based biofilm, evaluates correlations between EET gene abundance and expression and performance metrics, and investigates how EET genes and their transcriptional activity are associated with complex organic matter degradation and the composition of the corresponding microbial communities.

4.2 Materials and methods

4.2.1 Sample collection and data pre-processing

Eleven biomass samples were collected from two pilot-scale SAF-MBR systems treating real sewage and one lab-scale AFBR as described in Chapter 3. Specifically, the samples included seven GAC samples and four bulk samples from SAF-MBRs, as well as one PVDF sample from the AFBR as a non-conductive counterpart to GAC. Chapter 3 also described the techniques for sequencing, sequence preprocessing, genome assembly, and genome binning.

4.2.2 Metagenome-assembled genome refinement and post-processing

The metagenome-assembled genomes were refined using the BinSPreader tool to improve genome completeness, aiding in the recovery of genes from shorter DNA fragments that were not initially incorporated into contigs [221]. Eight refinement modes were compared based on the average F1 score and a total number of genomes refined with completeness $>50\%$ and redundancy $<10\%$. Detailed mode selection and parameter considerations are available in Appendix A.2.

Genome coverage for each MAG was assessed using CoverM by mapping the genomes to the metagenomes at the nucleotide sequence level [307]. The ORFs in each refined MAG were identified and translated into amino acid sequences using Prodigal v2.6.3 [230]. Taxonomic assignment of each MAG was performed by aligning them to a backbone tree generated using GTDK-db [227].

4.2.3 Gene prediction

To detect MHC and e-pili genes, custom pHMMs were constructed from previously identified genes [38, 115, 161, 164, 258, 259].

For each gene, a BlastP search [260] was conducted, collecting amino acid sequences with over 35% identity and 70% coverage. These sequences were clustered using MMseq2 [261] with a minimum identity threshold of 70%. Then each protein cluster was aligned using MUSCLE v5 [262], and the aligned sequences were integrated into a pHMM database. The complete set of pHMMs was compressed into a final database to optimise the search process.

The `hmmsearch` tool was employed to search across the ORFs of each genome using the custom pHMMs as references, applying an e-value cutoff of 1e-10. The best-hit results were then subjected to specific selection criteria for MHC and e-pili genes.

For MHC gene prediction, the identified genes were scanned for heme motifs using predefined regular expressions (e.g., [C·CH], [C··CH], [C· · ·CH], [C· · · · · · · · · · · ·[! = C]··CH]). Genes containing three or more heme motifs were further analysed for subcellular localisation using PSORTm [308], with taxonomy assignments tailored accordingly. Genes were categorised as cytoplasmic, cytoplasmic membrane, periplasmic, or outer membrane proteins. Those classified as outer membrane proteins, or as periplasmic proteins located within eight ORFs of a beta-barrel outer membrane porin (BB-OMP), were identified as putative MHC genes.

For e-pili gene prediction, the identified genes were evaluated for cleavage sites, aromaticity, and the largest gaps between aromatic residues. Genes with aromaticity of 9.8% or higher and with largest gaps of less than 22 for bacteria or 35 for archaea were considered putative e-pili genes. The Python script used for this analysis is available in Appendix A.4.

The putative EET genes were then cross-referenced using eggNOG-mapper v2 [247] to account for potential misannotations. The cross-referencing utilised the eggNOG 5.0 database [279], which includes COG [309], KEGG [49] pathways or modules, SMART [310], Pfam [182] domain content, and Gene Ontology [311] annotations.

4.2.4 Phylogenetic tree construction

To visualise the phylogeny the the MAGs identified with EET genes, a phylogenetic tree was constructed using MAFFT (v7.505) [234] for MSA and RAxML (v8.2.12) for maximum likelihood-based phylogenetic inference [237]. Similarly, to infer the diversity of the putative MHC and e-pili genes, a phylogenetic tree was constructed using muscle v5 [262] for MSA and RAxML (v8.2.12) for phylogenetic inference based on maximum likelihood [237]. The ”best_tree” generated from RAxML was visualised and annotated using tvBOT [312].

4.2.5 Statistical analysis

All statistical analyses, including Pearson’s and Spearman’s correlations, principal coordinate analysis (PCoA), ANOSIM, Shannon diversity, and richness indices, were performed using `pandas`

[291], `numpy` [292], `scipy` [293], `scikit-learn` [294], and `statsmodels` [295]. A random forest classifier (`scikit-learn`) was used to assess whether EET gene categories were predictive of metabolic preferences.

The analyses in this chapter rely on genome-resolved features; therefore, conventional biological replication of DNA and RNA extractions does not apply. Replication is instead represented at the level of MAGs, with each MAG treated as an independent observation for gene distribution and metabolic potential. For community-level analyses, the biomass samples collected from SAF-MBR 1.0 and SAF-MBR 2.0 serve as the available sample-level replicates across reactor compartments. No technical replication was performed for sequencing.

Correlations between EET gene distributions and reactor operating conditions (e.g., HRT, SCOD, OLR, biogas production) are based on single measurements per sampling event and therefore lack temporal or biological replication. These correlations are presented as exploratory associations that provide mechanistic context rather than inferential statistical tests.

4.3 Results

4.3.1 Metagenomic reconstruction and EET-related genes prediction

In this study, metagenomic analysis of twelve biomass samples from two pilot-scale SAF-MBRs yielded 1796 MAGs. Post-refinement, the number of MAGs that met the medium quality standard (CheckM-estimated completeness $\geq$ 50% and redundancy $\leq$ 10%) increased by 10.93%, bringing the total to 796 MAGs eligible for further analysis. This refinement increased the proportion of discovered EET genes by 8.51%, from 23.05% to 31.56%, ensuring substantial integrity for gene discovery. Details of the refinement process and results are available in Appendix A.2.

Among these, 421 MAGs were predicted to contain at least one putative MHC or e-pili gene, representing over 50% of the medium-quality MAGs. These are referred to as EET-positive (EET+) MAGs, while those without EET genes are labelled as EET-negative (EET-) MAGs.

Phylogenetic distribution of the EET+ MAGs

To depict the phylogenetic distribution of the EET+ MAGs, a phylogenetic tree was constructed based on the 421 EET+ MAGs, with taxonomy assigned utilising GTDB[1] annotations [227] (Fig. 4.1).

The 421 EET+ MAGs were assigned to 196 taxa across 42 phyla[2], most of which were uncultured genomes. Among these, 188 taxa were previously unreported for their EET capacity; a list of these taxa is available in the GitHub repository. The distribution of MHC and e-pili genes

[1] Initial taxonomy assignments, conducted through PhyloPhlAn annotations [228] against MetaRefSGB [232], were found to be largely incompatible with genome clustering, leading to their abandonment. For further details, refer to Appendix A.3.

[2] Taxon names at the phylum level with an alphabetical suffix indicate groups that are either not monophyletic in the GTDB reference tree but have alternative evidence supporting their monophyly or have unstable placement across releases.

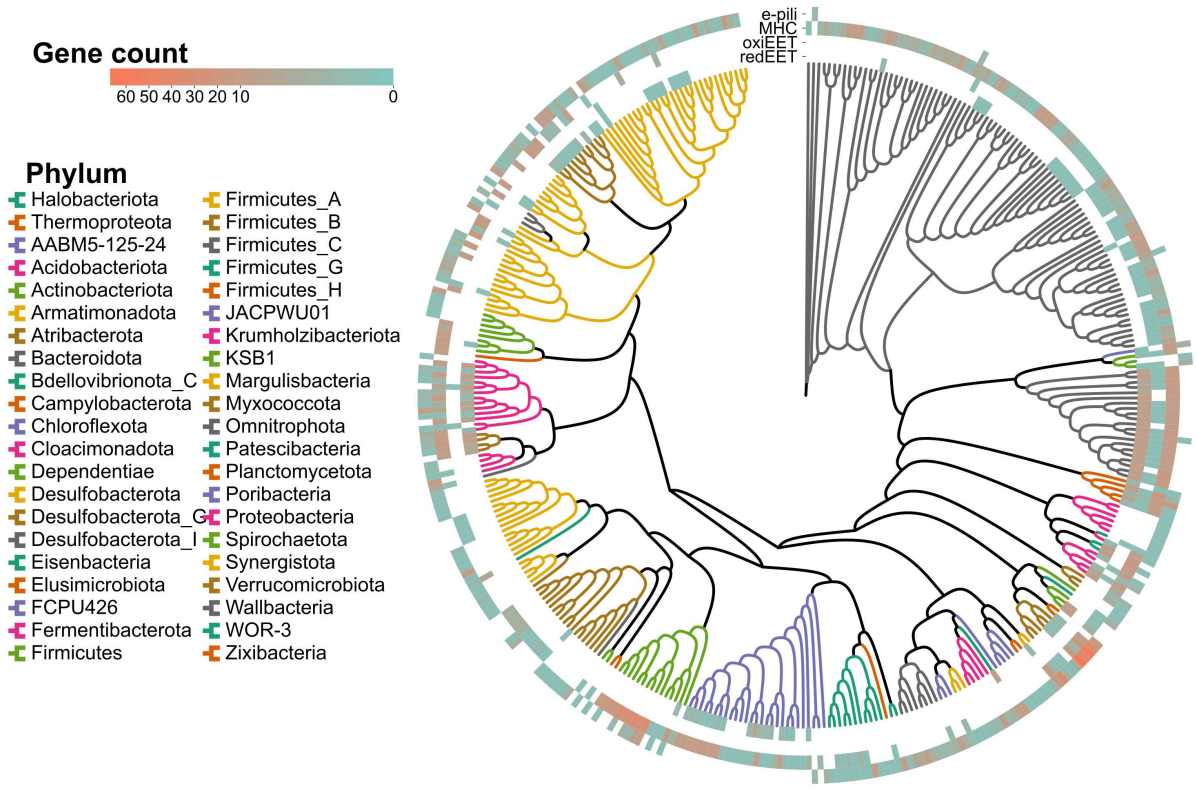


Figure 4.1: E-pili, MHC, oxidative (OxiEET), and reductive (RedEET) EET protein abundances across the phylogenetic diversity of bioreactor microbial communities. The tree included 421 MAGs featuring at least one aforementioned EET gene category . Phylogenetic reconstruction utilised RAxML (v8.2.12) for maximum likelihood estimation [237]. Taxonomic classifications were conducted via GTDB-tk [227], with visualisation facilitated by tvBOT [312].

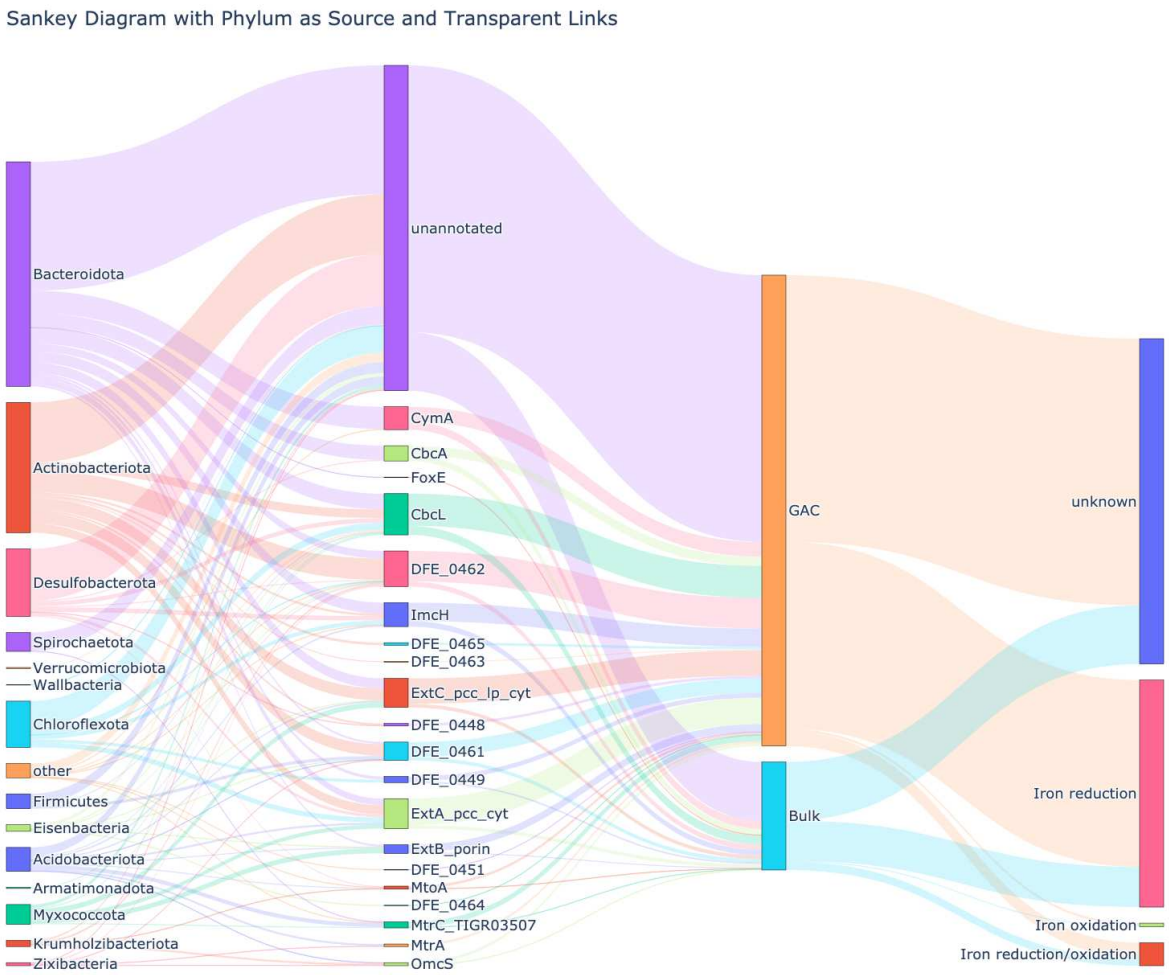


Figure 4.2: MHC gene distribution among phylum, media, and reaction direction. More than half of the predicted gene (n=639) were unannotated and 439 matched with the experimentally verified EET MHC gene.

across the phylogenetic tree suggested that the adaptation of EET strategies (i.e., short-range MHC and long-range e-pili) could be lineage-related.

Distribution of MHC and e-pili genes

MHC . In total 1747 MHC genes were predicted through HMM search, while the rule-based method identified 573 MHC genes, and 550 of them were not predicted by the HMM model. Since one of the premise of the pipeline is the presence of heme motif, this means that the majority of the MHC genes predicted by HMM method do not contain heme motif and therefore do not fit the physical requirement to be considered as MHC genes.

A total of 256 MAGs contained genes, with 154 unannotated and 28 identified as CymA, an inner membrane protein involved in both reduction and oxidation processes. Other notable genes included CbcA, CbcL, ExtC, and ImcH, all of which are associated with reduction at different potentials (Fig. 4.3). In the bulk sample, unannotated genes were again the most prevalent with 99 counts, followed by 16 counts each for CbcL and CymA.

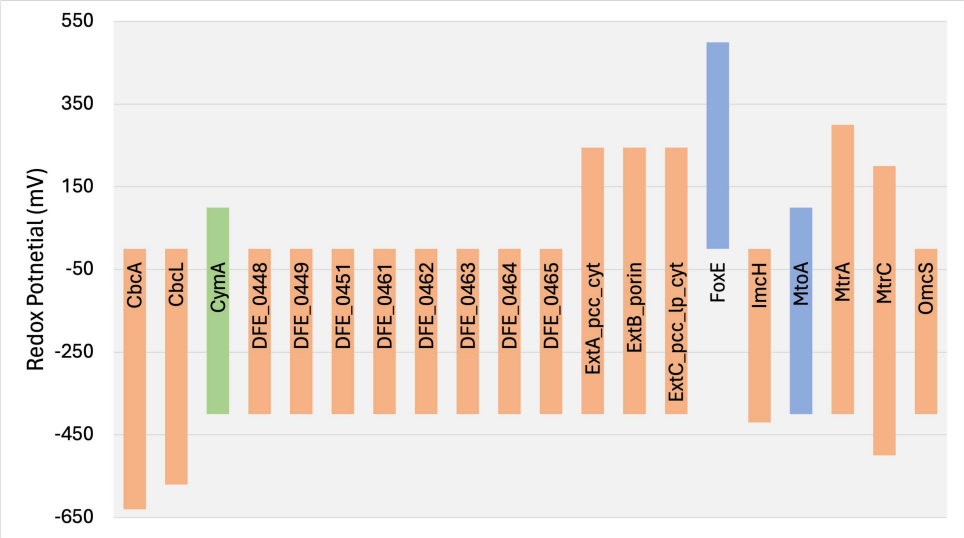


Figure 4.3: Redox potential range of the MHC genes identified in this study. The colour legends present reaction direction: orange is reductive, blue is oxidative, and green is bidirectional.

E-pili . A total of 1,444 pili gene sequences were predicted using *hmmsearch* against a custom e-pili pHMM (Fig. 4.4). Of these, 1,427 were predicted to undergo pilin maturation. Initially, the entire dataset was labelled as `hmmsearch e-pili`, reflecting default categorisation from the *hmmsearch* output. Further analysis refined the sequences based on two key parameters: the percentage of aromatic amino acids and the size of the largest sequence gap. These sequences that did not meet the criteria are likely to function as typical type IV pili, facilitating surface adhesion and other cellular processes.

Among them, 153 sequences were categorised as putative e-pili, characterised by at least 9.8% aromatic amino acids and a maximum gap size of 22 amino acids or fewer. Additionally, a distinct set of 10 sequences that matched Mhun_3140, which represents a conductive archaellum in *Methanospirillum hungatei* [115], were categorised as `putative archaeal e-pili` with aromatic-free gap less or equal to 35 amino acids.

Certain phyla were prevalent in the EET community in GAC

Analysis of the phylogenetic distribution of EET genotypes across different media types (Figs. 4.6 and C.3) revealed that *Bacteroidota* and *Desulfobacterota* had the highest coverage of EET-positive MAGs, with EET+ to EET- ratios of 6.8:1 for *Bacteroidota* and 1.7:1 for *Desulfobacterota*. These phyla also exhibited the highest mean relative abundances across samples (Fig. C.2). Specifically, *Bacteroidota* MAGs predominantly contributed to the prediction of MHC genes, accounting for 74.59% and 50.75% of total MHC genes in bulk and GAC samples, respectively. Conversely, *Desulfobacterota* contributed 16.56% and 45.06% of the e-pili genes in bulk and GAC samples, respectively.

To resolve the functional heterogeneity within *Bacteroidota*, MAGs were further divided into MHC-encoding and non-MHC lineages using the `Genes` annotation column. As summarised in

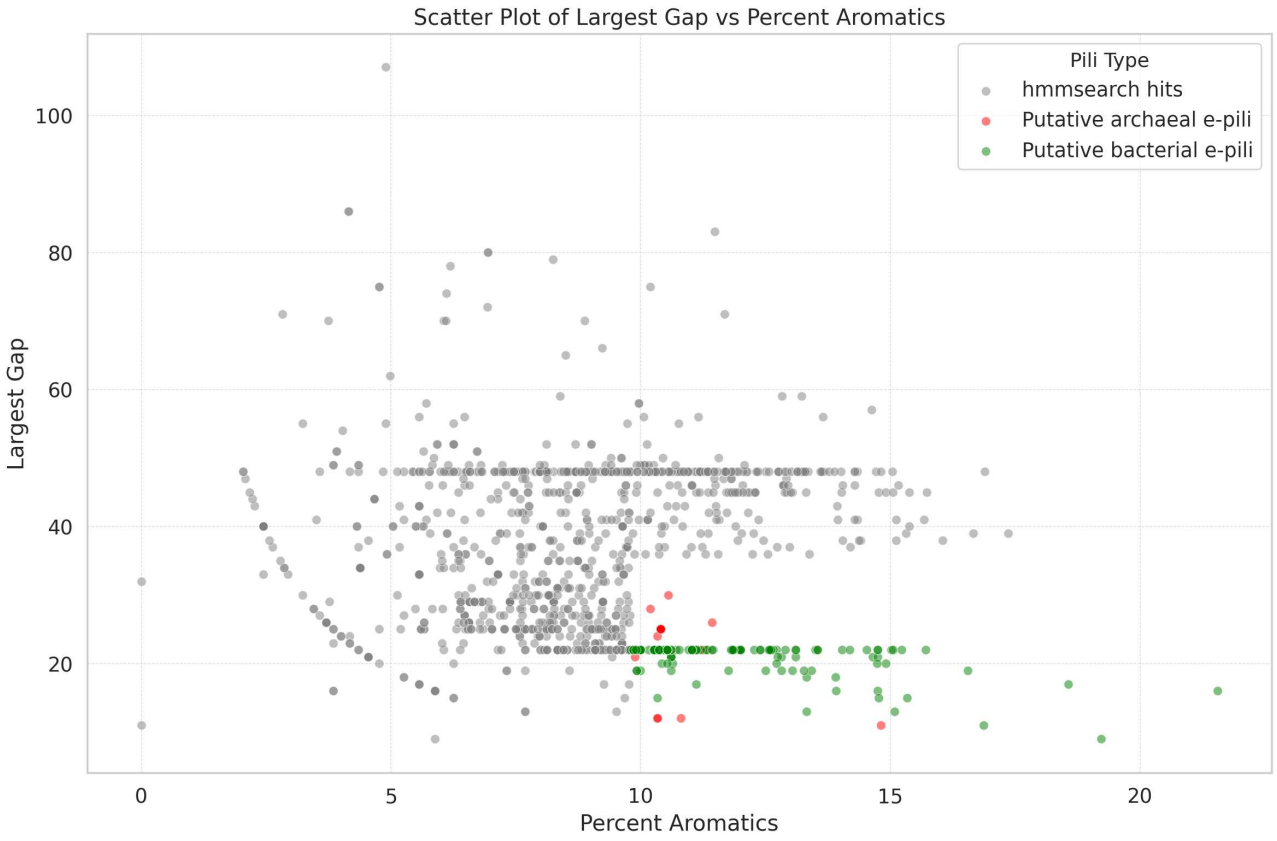


Figure 4.4: Scatter plot illustrating the relationship between percent aromatics and the largest gap within pilin sequences. Data points are categorised into three groups: hmmsearch e-pili (grey dots); Putative bacterial e-pili (red dots), denoting sequences with a percentage aromatics $\geq$ 9.8% and the largest gap $\leq$ 22; and Putative archaeal e-pili (green dots), indicating sequences identified as Mhun_3140 with percent aromatics $\geq$ 9.8%. This classification aims to distinguish potential e-pili sequences based on their aromatic content and gap characteristics, highlighting the variability and specific traits that may influence their function and assembly.

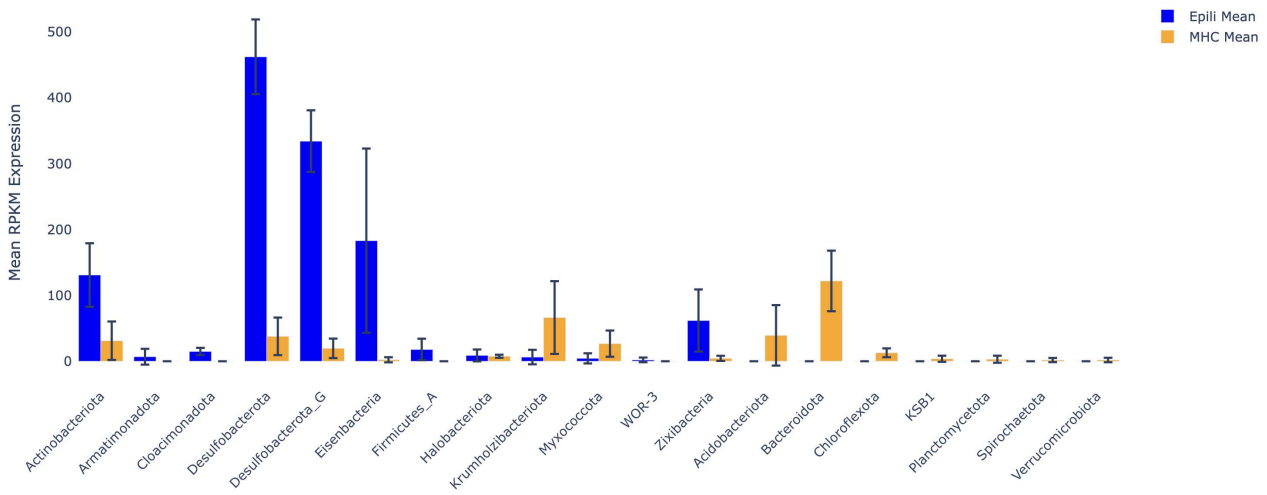


Figure 4.5: Comparison of mean RPKM expression levels of Epili and MHC across various phyla. Error bars represent the standard deviation of expression levels across different conditions. Phyla with no detectable expression in one or both datasets are represented with empty bars.

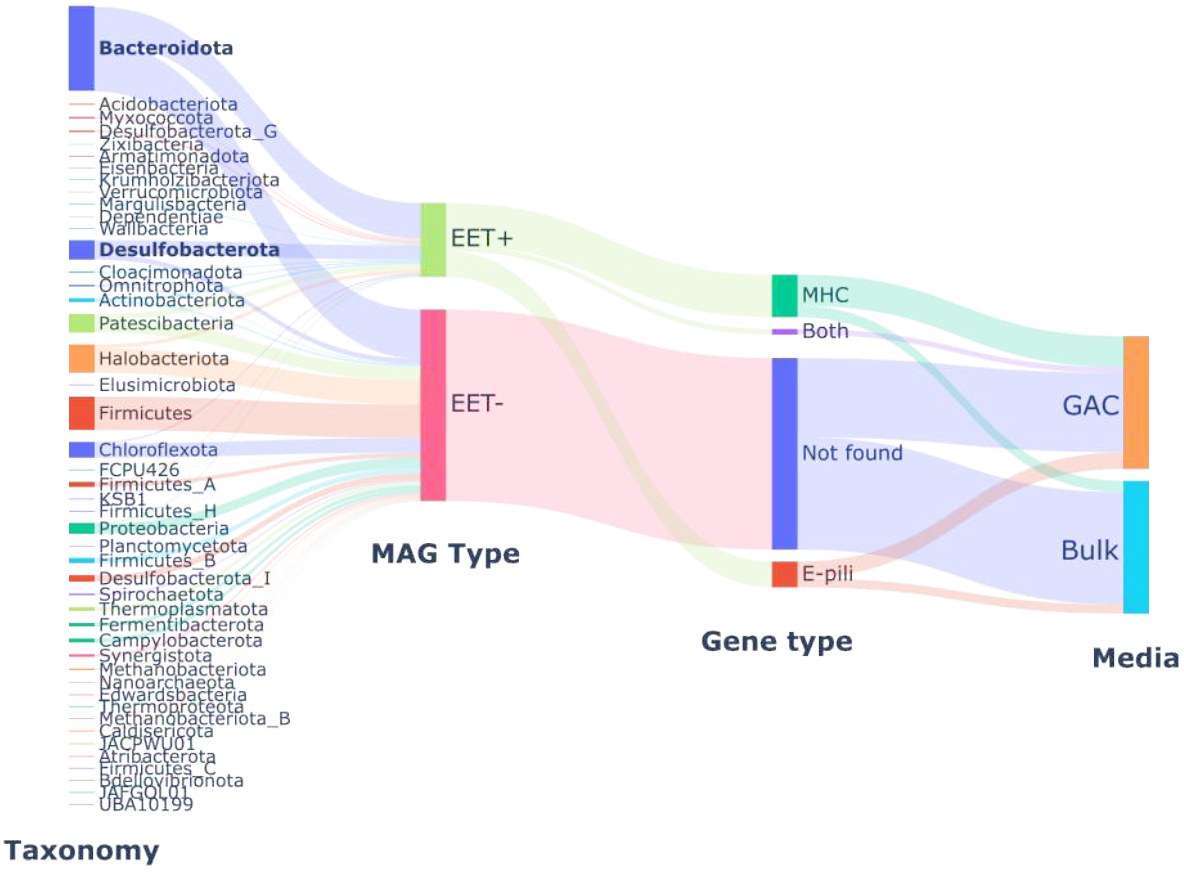


Figure 4.6: Sankey diagram illustrating the flow of normalised coverage from MAGs clustered at phylum level to EET-gene categories and subsequently to the media type. ”Not Found” refers to no EET-gene detected, and both refer to the detection of MHC and e-pili within a MAG. (consider to update the label)

Table 4.1, 47 of 88 *Bacteroidota* MAGs encoded multihaem cytochromes, while 41 lacked EET genes. These two groups showed distinct ecological distributions. MHC-encoding *Bacteroidota* were enriched on the GAC biofilm, exhibiting higher average normalised abundance (0.40 vs. 0.17 in bulk) and greater mean total abundance per sample (18.7 vs. 8.2). In contrast, non-MHC *Bacteroidota* dominated the bulk liquid, with substantially higher average abundance (0.80 vs. 0.10 in GAC) and higher mean total per sample (33.0 vs. 4.1). This partitioning demonstrates that only a subset of *Bacteroidota* participates in EET. These results also explain why *Bacteroidota* appear highly abundant in bulk while still strongly contributing to MHC-associated EET pathways in the Sankey diagram.

Bacteroidota (formerly known as *Bacteroidetes*[3]) are facultative anaerobes found in various environments such as soil, sediments, seawater, and animal guts. They are known for their hydrolytic capabilities, including polysaccharides (e.g., cellulose) and proteins, fermenting them into VFAs such as propionate and acetate [313]. Given their strong hydrolytic capabilities, *Bacteroidota* could also contribute to the accelerated hydrolysis of complex substrates, thereby enhancing breakdown rates within the system.

Table 4.1: Comparison of MHC-encoding and non-MHC *Bacteroidota* MAGs based on the `Genes` annotation column. Values represent normalised abundance and mean total abundance per sample in bulk and GAC fractions.

Metric	**MHC-encoding**	**Non-MHC**
Number of MAGs	47	41
Average normalised abundance		
GAC	0.40	0.10
Bulk	0.17	0.80
Mean total abundance per sample		
GAC	18.7	4.1
Bulk	8.2	33.0

In contrast, the phyla *Desulfobacterota* and *Patescibacteria* exhibited a higher prevalence of e-pili genes. *Desulfobacterota* (formerly known as δ-*Proteobacteria*) encompasses a range of taxa common in anaerobic processes. In our samples, acetogenic families such as *Geobacteraceae*, *Smithellaceae*, and *Syntrophobacteraceae* were observed, along with sulphate-reducing families *Desulfobacteraceae*, *Desulfobulbaceae*, and *Desulfovibrionaceae*. These families significantly contribute to the degradation of fermentative products like VFAs and ethanol into acetate and CO_2 [314]. *Patescibacteria*, also known as Candidate Phyla Radiation (CPR), is a superphylum comprising numerous bacterial lineages with highly reduced genomes. These microorganisms are characterised by their small size and lack many metabolic pathways found in other bacteria, suggesting a symbiotic or parasitic lifestyle [315].

Alpha diversity test (Fig. 4.7) indicated higher microbial diversity in GAC samples than in Bulk samples (Simpson Index , p = 0.016; Shannon index, p = 0.032). Notably, the GAC samples had a higher Shannon Index, with a mean difference of 0.3614 (Table. B.7 & B.8).

[3]The taxonomy nomenclature has been debated in recent years without a unified consensus.

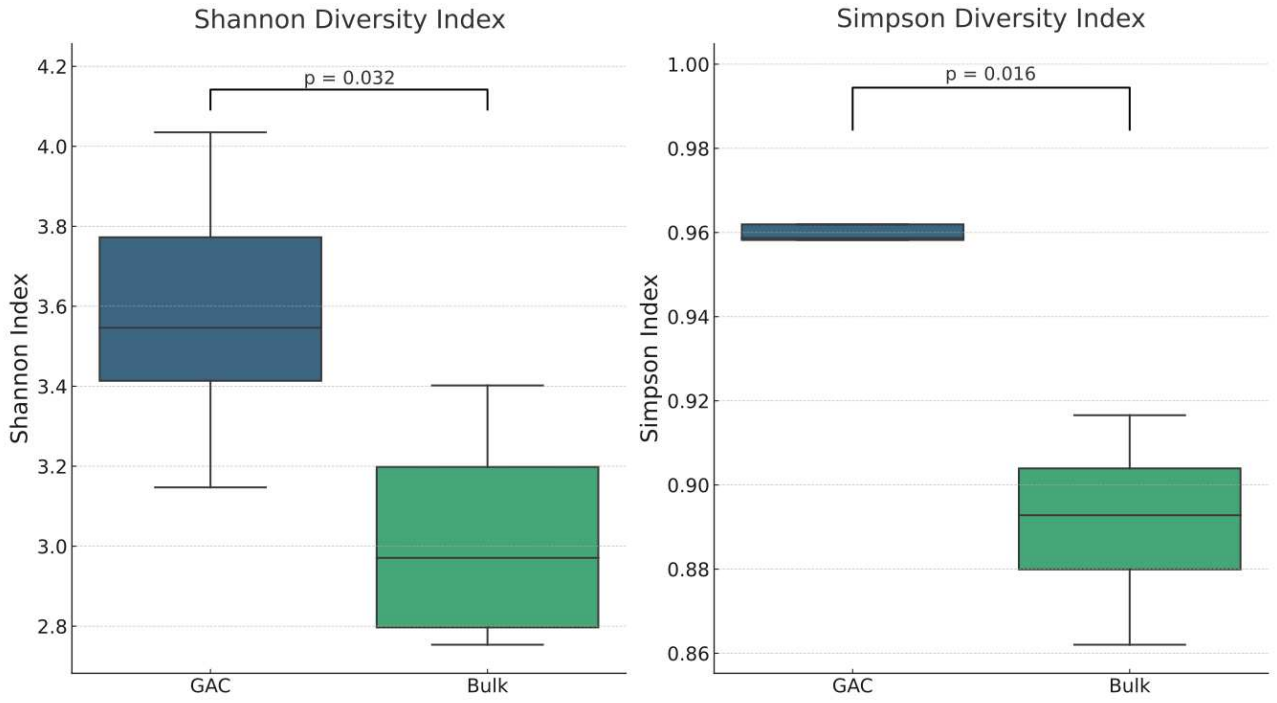


Figure 4.7: Comparison of alpha diversity between GAC and Bulk microbial communities. Boxplots show the Shannon Diversity Index (left) and Simpson Diversity Index (right) for both sample types. GAC samples exhibit significantly higher Shannon (p = 0.032) and Simpson (p = 0.016) indices. Significance was tested using the Kruskal-Wallis test.

Beta diversity analysis using Bray-Curtis dissimilarity (Fig. 4.8) further confirmed distinct community composition between media types, with bulk, GAC, and PVDF samples forming separate clusters. To aid interpretation, the phyla driving these differences are highlighted here. *Bacteroidota* were consistently enriched in bulk samples, reflecting their hydrolytic role. In contrast, several syntrophic and methanogenic groups were enriched on GAC, including *Desulfobacterota* and *Methanothrix*, consistent with their attachment-based lifestyles and associations with conductive materials. *Firmicutes* and *Chloroflexota* were also more abundant on GAC, in line with the presence of fermentative and syntrophic taxa within these phyla that preferentially grow in biofilms. Conversely, bulk samples contained higher proportions of *Proteobacteria* and *Patescibacteria*, which are commonly associated with suspended or fast-growing populations. These phylum-level differences corroborate the beta diversity results and clearly demonstrate that the separation between bulk and GAC communities is driven by the enrichment of hydrolytic taxa in bulk and syntrophic, methanogenic, and biofilm-forming taxa on GAC.

The clustering patterns observed in the Bray–Curtis beta diversity dendrogram were closely mirrored by those in the phylum-level clustered heatmap (Fig. 4.9). In the heatmap, GAC samples formed a distinct block characterised by elevated relative abundances of *Desulfobacterota*, *Methanothrix*, *Firmicutes*, and *Chloroflexota*, whereas bulk samples clustered separately and were enriched in *Bacteroidota*, *Proteobacteria*, and *Patescibacteria*. This clear correspondence between the dendrogram structure and heatmap colour patterns indicates that the separation between bulk and GAC communities is driven by pronounced shifts in a small number of metabolically important

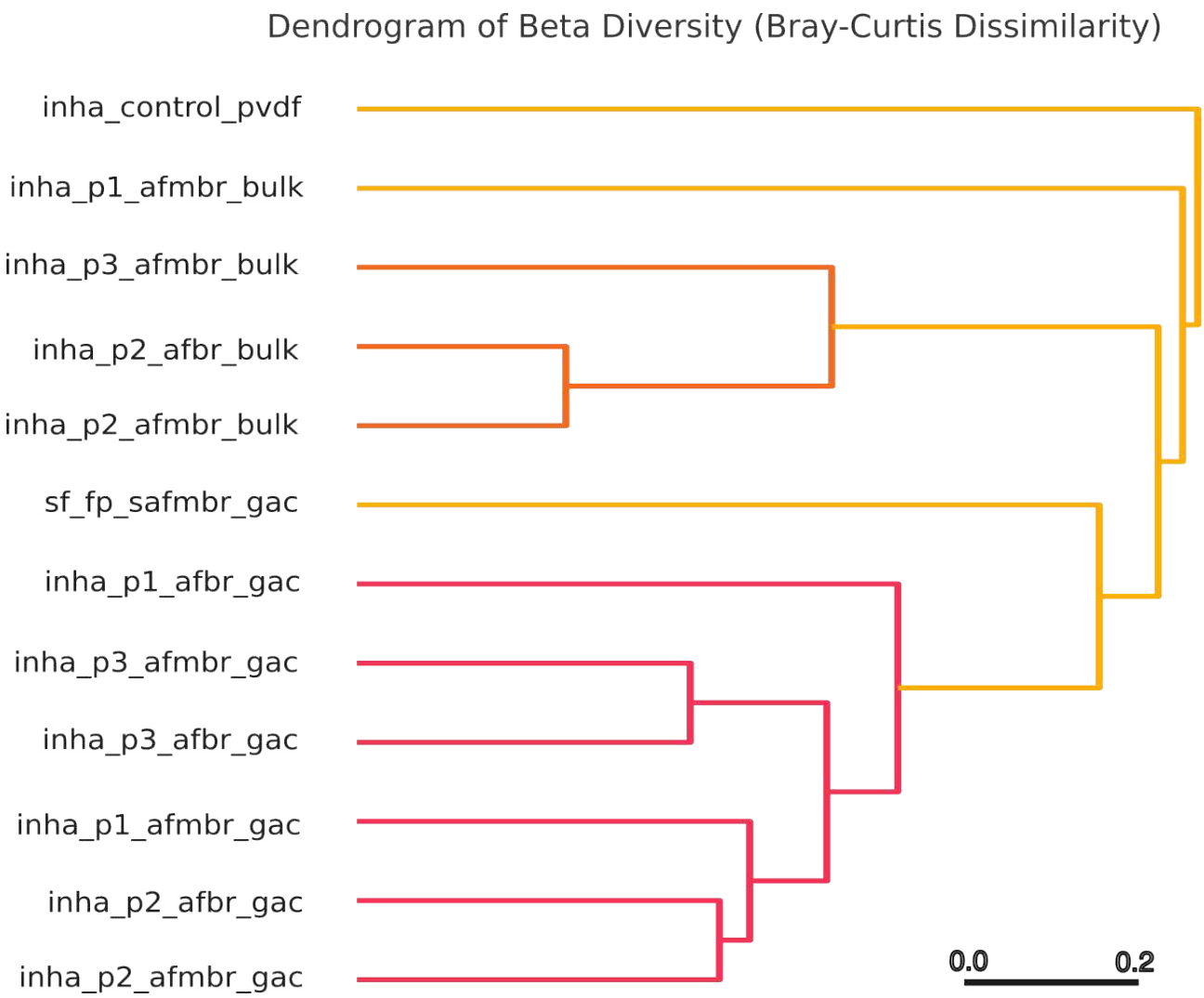


Figure 4.8: Dendrogram showing hierarchical clustering of samples based on Bray-Curtis dissimilarity, representing beta diversity. Samples are categorised into Bulk, GAC, and PVDF, indicated by different colours. The vertical axis reflects Bray-Curtis distances, illustrating the similarity or dissimilarity in microbial community composition.

phyla, rather than by subtle variation across many taxa.

Genus-level distribution of EET gene inventories in GAC-associated MAGs

To resolve the phylum-level patterns at finer taxonomic resolution, EET gene possession was summarised at the genus level using dereplicated MAG annotations. ”GAC-associated” EET+ MAGs were defined as dereplicated genomes classified as EET-positive (column `EET`) with non-zero mean abundance in GAC (column `GAC (mean)`). Genus-level assignments were taken from GTDB taxonomy (column `genus`), and EET gene inventories were classified as MHC-only, e-pili-only, or dual (column `Genes`).

At the genus level, the GAC-associated EET+ pool was phylogenetically dispersed, with a small number of genera contributing a substantial fraction of the EET+ genomes and a long tail of low-frequency genera. The top 10 genera accounted for 50.0% of GAC-associated EET+ dereplicated MAGs (Table 4.2), while most genera were represented by a single dereplicated EET+ genome

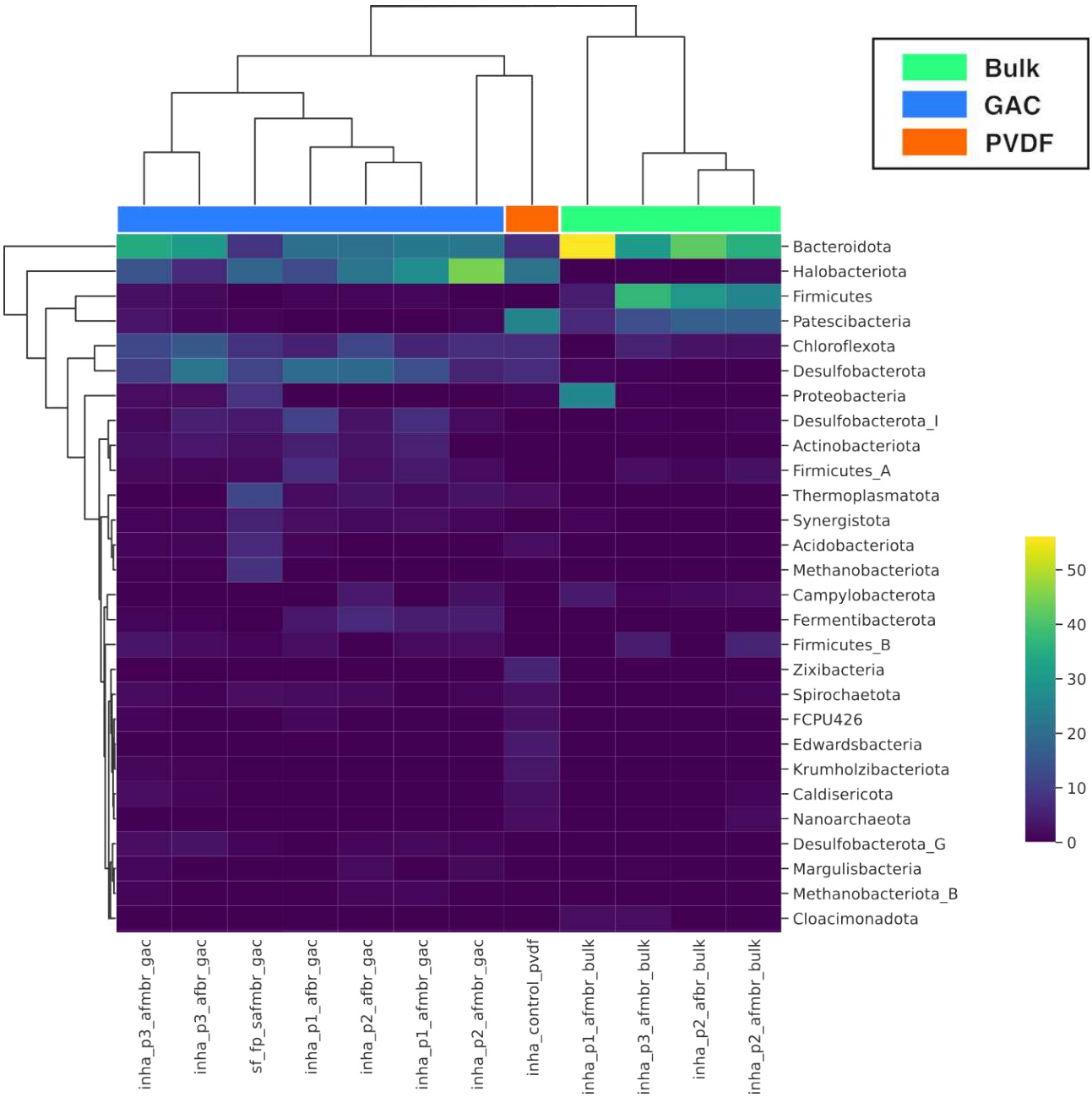


Figure 4.9: Clustered heatmap representing the relative abundance of microbial communities at the phylum level across different sample media types. Rows correspond to different microbial phyla, while columns represent individual samples. The samples are hierarchically clustered, with dendrograms at the top indicating clustering patterns based on microbial community composition. Colour intensity within the heatmap reflects the abundance of each phylum, with a scale ranging from 0 (deep purple) to 50 (bright yellow). Columns are colour-coded by sample media type: green represents Bulk, blue represents GAC, and orange represents PVDF.Only phylum with at least one sample higher than 1% is shown here.

(i.e., $> 70\%$ of genera were singletons). The largest single bin comprised genomes without a resolved GTDB genus label (reported as `unclassified`), representing 19.8% of GAC-associated EET+ MAGs, consistent with incomplete taxonomic resolution even at genus rank.

Across the GAC-associated EET+ set, MHC-only and e-pili-only strategies were both prevalent, while dual inventories were less common. Specifically, MHC-only inventories comprised 46.5% of GAC-associated EET+ MAGs, e-pili-only inventories comprised 40.7%, and dual inventories comprised 12.8%. This partitioning was also evident within the dominant genera: several high-contributing genera were exclusively MHC-only (e.g., `LD21`, `UBA4417`, `CAIKZJ01`), whereas others were exclusively e-pili-only (e.g., `UBA8904`, `RPRC01`). In contrast, *Holophaga* included mixed inventories, with half of its EET+ MAGs encoding both MHC and e-pili markers.

Importantly, the dominant genera were largely assigned to uncultured lineages represented by GTDB placeholder names (e.g., `UBA4417`, `CAIKZJ01`). In GTDB, such alphanumeric labels (including strain-identifier-derived genus names) are explicitly used as placeholders when no existing genus name is available and/or no binomially named representative genome exists, and are intended to be replaced as taxonomy stabilises and nomenclature catches up with genome discovery [227]. Consequently, while genus-level summaries help describe community-level patterns, the strong representation of placeholder genera indicates that a taxonomy-first approach is intrinsically limited for identifying EET+ microorganisms in engineered systems. The observation that EET+ potential was distributed across many low-frequency, predominantly uncultured genera further suggests that functional screening (gene-centric classification) provides a more robust basis than taxonomic labels alone for defining the EET+ fraction of the GAC-associated microbiome.

Table 4.2: Top 10 genera contributing to the GAC-associated EET+ dereplicated MAG pool. Values are expressed as percentages. Columns “MHC”, “e-pili”, and “Both” indicate the within-genus distribution of EET gene inventories.

GTDB Genus	Share of GAC-Associated EET+ MAGs (%)	MHC (%)	e-pili (%)	Both (%)
`unclassified`	19.8	17.6	70.6	11.8
`LD21`	7.0	100.0	0.0	0.0
`Holophaga`	4.7	25.0	25.0	50.0
`UBA4417`	4.7	100.0	0.0	0.0
`CAIKZJ01`	2.3	100.0	0.0	0.0
`UBA8904`	2.3	0.0	100.0	0.0
`T78`	2.3	100.0	0.0	0.0
`PNON01`	2.3	100.0	0.0	0.0
`RPRC01`	2.3	0.0	100.0	0.0
`SR-FBR-E99`	2.3	100.0	0.0	0.0

Beyond the dominant genera, two anaerobic guilds commonly invoked in engineered systems, sulfate-reducing bacteria (SRB) and syntrophs, were examined to test whether EET gene possession could be inferred reliably from genus labels alone. Because SRB functionality is not consistently captured by taxonomy (particularly for uncultured lineages), SRB potential was

operationally defined using DRAM module completeness for dissimilatory sulfate reduction ("Sulfur metabolism: dissimilatory sulfate reduction (and oxidation) sulfate $\Rightarrow$ sulfide") (3.5.2 and applying a conservative threshold (module completeness ≥ 0.5). Under this definition, putative SRB constituted 8.0% of dereplicated genomes detected in GAC (i.e., non-zero `GAC (mean)`). Notably, EET markers were frequent within this functional guild: 61.9% of putative SRB genomes were classified as EET+. Among the EET+ SRB subset, the inferred EET strategy was dominated by e-pili-only inventories (69.2%), with smaller contributions from MHC-only (15.4%) and dual inventories (15.4%). Syntrophs showed a similarly weak mapping between taxonomic labels and EET potential: syntroph-labelled genera (GTDB genera containing *Syntroph-*) were rare in the GAC-associated dereplicated set (1.1% of GAC-detected genomes), and only one-third of these syntroph-labelled genomes were EET+ (33.3%). The sole EET+ syntroph-labelled genome (*Syntrophorhabdus*) encoded a dual inventory (MHC + e-pili), whereas the remaining syntroph-labelled genera detected in GAC (*Syntrophobacter* and *Syntrophomonas*) were EET-. Together, these patterns reinforce that genus-level taxonomy is an unreliable proxy for identifying EET+ microorganisms, even within guilds classically associated with anaerobic carbon flux (syntrophy) or electron-acceptor competition (sulfate reduction), and particularly where a substantial fraction of relevant genomes fall within uncultured placeholder taxa.

Finally, the genus-level distributions observed here contrast with the literature-based non-conductive reference communities assembled from PVDF (or low-conductivity carrier) AFMBR studies (Section 2.1.6). In those reference systems, genus-level 16S rRNA profiles during stable operation consistently highlighted named syntrophic and sulfate-reducing genera, most notably *Syntrophobacter* (typically $\sim$1.5% to 5.6%) and *Desulfobulbus* ($\sim$3.2–4.5%) on polymeric carriers, and reported archaeal dominance by *Methanothrix* under multiple non-conductive (or paired-carrier) configurations [100–102]. By contrast, the GAC-associated EET+ MAG pool was not dominated by these canonical genus labels, but instead by a mixture of `unclassified` genomes and GTDB placeholder genera (Table 4.2), while syntroph-labelled genera were rare and largely EET-negative in this dataset. At the same time, functionally defined SRB in GAC exhibited high EET+ prevalence (predominantly e-pili-only), indicating that electron-transfer potential may be widespread among sulfate reducers without being recoverable from genus names alone. Although direct quantitative alignment is constrained by methodological differences (MAG-derived genome inventories versus 16S rRNA relative abundance profiles), the qualitative comparison suggests that conductive-carrier environments may redistribute "DIET-relevant" potential away from a small set of well-known genera and toward a broader set of uncultured lineages, such that enrichment is better captured by gene-centric EET screening than by tracking a limited list of canonical syntroph/SRB genera from non-conductive baselines [100–102].

Correlation analysis of MHC and e-pili genes across media types

Spearman rank correlation analysis evaluated the monotonic relationships of e-pili and MHC genes across different media types (Table 4.3). The relative abundance of e-pili and MHC genes was calculated by dividing the MAGs-mapped reads by the corresponding metagenome reads and

normalising to the percentage of MAGs reads.

Distinct gene expression patterns were identified across sample types. MHC genes displayed strong, statistically significant positive correlations with GAC samples, whereas e-pili genes showed moderate positive correlations that were not statistically significant. Bulk samples exhibited strong, significant negative correlations with both e-pili and MHC gene abundance. Correlations with PVDF samples were weaker: e-pili genes showed moderate, non-significant positive correlations, and MHC genes showed no significant correlation. These results suggest that GAC is associated with a higher abundance of MHC and other EET-related genes, while bulk samples show lower abundance.

The difference in abundance of MHC and e-pili in GAC and PVDF materials may be attributed to microbial adaptability to conductive and non-conductive environments. The higher abundance of the MHC gene in GAC samples likely resulted from GAC's conductive properties, which facilitate short-range extracellular electron transfer [69]. Conversely, the absence of a statistically significant difference of e-pili genes in GAC and PVDF samples may reflect microbial adaptability through e-pili formation, enabling long-range electron transfer and potentially compensating for the lack of conductivity in the case of PVDF [316]. Additionally, e-pili may function similarly to type IV pili, aiding in surface adhesion and biofilm formation in PVDF samples [317].

These observations suggested that biocarrier selection significantly influences microbial electron transfer strategies, with GAC's conductive properties promoting the expression of both MHC and e-pili genes.

Table 4.3: Spearman correlation between EET gene relative abundance and sample types.

Gene	GAC Correlation	GAC P-Value	Bulk Correlation	Bulk P-Value	PVDF Correlation	PVDF P-Value
E-pili	0.514	0.087	-0.768	0.004	0.393	0.206
MHC	0.808	0.001	-0.819	0.001	-0.044	0.893

4.3.2 Correlations between the Distribution of EET Genes and SAF-MBR Operating Conditions

To investigate how reactor operation influences the distribution of EET+ genes in the SAF-MBR systems, Pearson correlation analysis was performed using two complementary metrics of EET gene abundance: the EET ratio (percentage of EET+ MAGs relative to total MAGs) and EET/MAGs (EET gene counts normalised by MAG number). Genes encoding multiheme cytochromes were further separated into oxidative (oxiEET) and reductive (redEET) categories. These metrics were correlated with operational parameters including temperature, hydraulic retention time (HRT), solids retention time (SRT), organic loading rate (OLR), and influent and permeate COD.

The correlation matrix (Fig. 4.11) showed a negative association between EET metrics and HRT, indicating that shorter hydraulic retention times did not suppress the enrichment of EET+ MAGs. Positive correlations with OLR and COD removal efficiency suggested that EET+ populations were

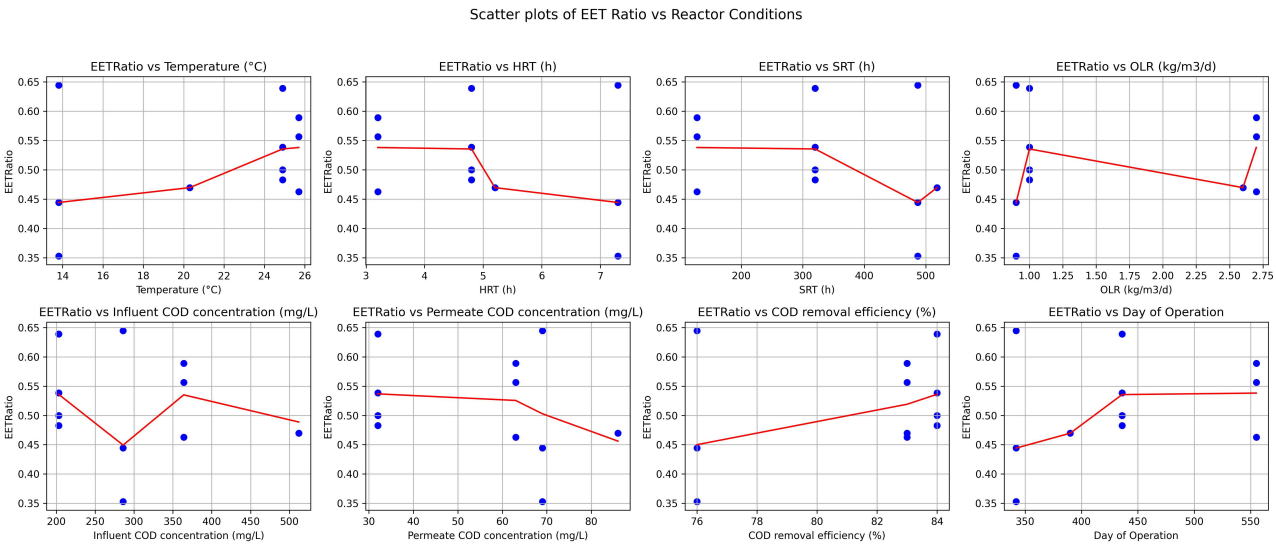


Figure 4.10: Scatter plots of EETRatio (%) versus reactor operating conditions with LOWESS smoothing. Each point represents a pilot-scale sample, coloured by biomass type (GAC or bulk). The plots show weak monotonic associations without threshold-like transitions, and the data primarily cluster according to discrete operational modes rather than continuous parametric responses.

more abundant under higher organic throughput and improved carbon conversion performance. These relationships align with the role of EET-capable microbes in accelerating redox turnover and sustaining metabolic flux under elevated substrate input. The SAF-MBR configuration, which decouples SRT from HRT through membrane retention, further supports the development of dense biofilms on GAC, contributing to the stability of EET+ communities.

To complement the coefficient-based analysis, scatter plots with LOWESS smoothing were used to examine the form of these relationships (Fig. 4.10). Across all operational parameters, EETRatio displayed weak monotonic trends and no evidence of sharp inflexion points or threshold-like transitions. Instead, samples formed distinct clusters that reflected the discrete operational modes of the pilot systems (e.g., Inha Phases 1–3 and Seoul Facility operation). This clustering indicates that variations in EET gene abundance were driven primarily by broader changes in reactor regime and biomass configuration, such as GAC-associated biofilm development versus bulk suspension, rather than by continuous gradients of individual environmental variables. The combined evidence supports a functional interpretation in which EET+ microbes are favoured under high-rate, biofilm-dominated conditions where redox cycling and carbon degradation are intensified.

4.3.3 Correlation between EET genes and metabolic capability

To evaluate potential metabolic linkages between EET+ MAGs and carbon transformation processes beyond methanogenesis, Pearson correlation coefficients (r) and Random Forest importance scores (I) were used to quantify associations between MHC and e-pili genes and a range of carbon metabolism pathways (Fig. C.4; Fig. C.5).

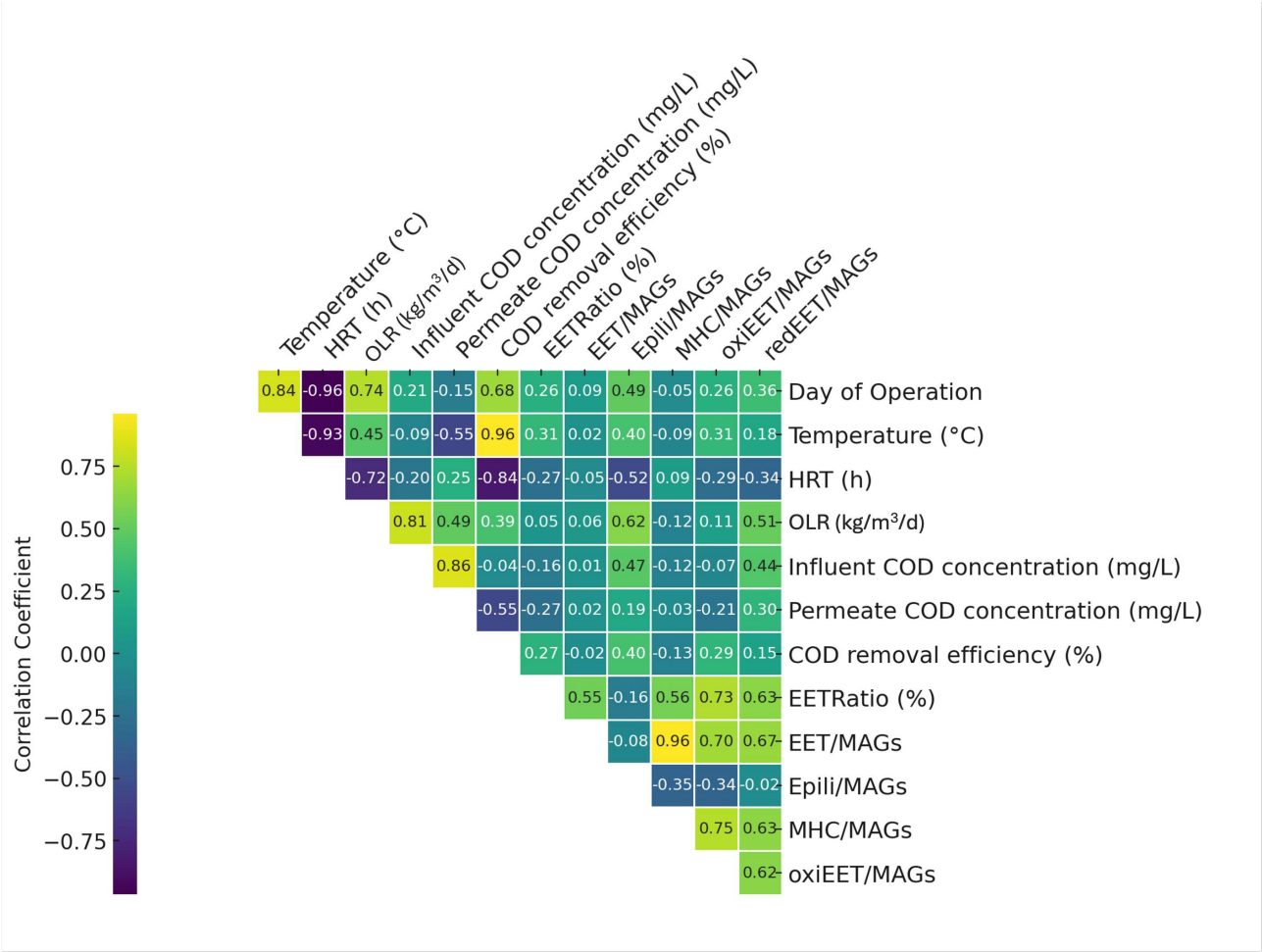


Figure 4.11: Heatmap of Pearson correlation coefficients among operational parameters and EET gene distributions. Variables include temperature, HRT, OLR, COD concentrations, COD removal efficiency, and ratios of EET, e-pili, and MHC genes to MAGs.

MHC-encoded MAGs demonstrated moderate positive correlations with primary acidogenic and early acetogenic pathways, including glycolysis ($r = 0.17$, $I = 0.0277$) and the citrate cycle ($r = 0.38$, $I = 0.042$). E-pili-encoded MAGs exhibited similar but slightly weaker correlations with these pathways ($r = 0.15$, $I = 0.041$ for glycolysis; $r = 0.13$, $I = 0.041$ for the citrate cycle). These associations indicate that EET genes are not confined to taxa involved solely in methanogenesis but extend to microorganisms responsible for carbohydrate fermentation and the subsequent conversion of fermentation intermediates into acetyl-CoA precursors.

Correlations with alternative or peripheral carbon pathways—including the glyoxylate cycle, Entner–Doudoroff pathway, pentose phosphate pathway, hydroxypropionate–hydroxybutyrate cycle, and dicarboxylate–hydroxybutyrate cycle—were consistently lower for both MHC and e-pili genes. This pattern suggests that EET+ MAGs are more strongly coupled to the core fermentative and acetogenic routes central to anaerobic digestion.

In addition, MHC-encoded MAGs displayed moderate associations with hydrolytic enzyme systems involved in the degradation of complex polysaccharides, including amorphous cellulose ($r = 0.11$, $I = 0.011$), xyloglucan ($r = 0.088$, $I = 0.026$), arabinan ($r = 0.12$, $I = 0.029$), and mixed-linkage glucans ($r = 0.21$, $I = 0.007$). The presence of these correlations indicates that EET genes may also be linked to microorganisms involved in the upstream hydrolysis of plant-derived polymers, suggesting that the influence of EET may extend to the earliest stages of carbon breakdown in the SAF-MBR community.

Across both MHC- and e-pili-encoded MAGs, strong associations were observed with energy-conserving electron transport complexes, including complex I (NADH:quinone oxidoreductase), complex II (succinate dehydrogenase), complex IV (low-affinity cytochrome c oxidase), and complex V (F-type ATPase). These complexes ranked among the highest Random Forest features ($I = 0.02$–0.06), indicating that EET gene presence aligns with taxa possessing robust electron-transport and oxidative phosphorylation capacity. Together, these results illustrate coordinated associations between EET genes and multiple stages of carbon transformation, spanning hydrolysis, acidogenesis, acetogenesis, and energy conservation.

4.4 Discussion

The metagenomic results revealed an extensive range of EET-related genes (e.g., MHC and e-pili) across 42 microbial phyla. Comparative analyses of bulk liquid and GAC samples indicated a distinct microbial community on the GAC, where EET mechanisms were prevalent. These findings highlight the potential of EET-mediated processes to enhance anaerobic treatment performance for primary wastewater treatment, demonstrating both improved treatability and CH_4 production.

4.4.1 Enhanced detection of EET genes through rule-based approaches

Predicting the occurrence of DIET in bioreactor studies typically involves the taxonomic classification of microbial communities using 16S rRNA amplicon sequencing, an economically viable

option [11]. When metagenomic sequencing data is available, the prediction of EET biomarkers, such as MHC and e-pili, becomes feasible. Due to the unique physical properties of MHC and e-pili, specific functional motifs (e.g., heme motifs in MHC) and amino acid arrangements (e.g., aromaticity and gap distances in e-pili) are essential for a protein to perform EET theoretically. Consequently, rule-based approaches have been adapted and refined for de novo discovery [164, 259].

Using the rule-based results as the standard, this study identified 1,682 (96.2%) and 1,291 (89.4%) false-positive predictions for MHC and e-pili, respectively. For MHC genes predicted by BlastP and HMMsearch, the majority contained fewer than 3 heme motifs, with some lacking any heme motifs altogether. In the case of e-pili, low aromatic density and extended gap distances predominantly ruled out predictions. Additionally, the rule-based approach supplemented 508 novel MHC gene predictions that were previously unannotated. This demonstrates that alignment-based methods can drastically overestimate the number of EET biomarkers, making them less ideal for researchers aiming to identify the significance of EET communities within their systems.

4.4.2 Diversity and prevalence of EET-Capable microbial communities in SAF-MBR

The abundance of EET-capable microbial communities in the SAF-MBR with conductive GAC was notably high. This study identified new phyla exhibiting EET capabilities, demonstrated high microbial diversity, and observed a greater prevalence of EET-associated taxa, compared to the previous survey of global anaerobic digesters' microbiome composition [259, 314]. Specifically, EET microbes from the phylum *Bacteroidota* remain poorly understood, with no physiological tests conducted on environmental cultivars [69]. This phylum was more frequently associated with MHC. Conversely, the phylum *Desulfobacterota* comprises numerous syntrophic and sulfate-reducing bacteria commonly observed in anaerobic digestion processes [314, 318], and was more closely associated with e-pili.

The diversity of MHC genes among *Geobacter* species is well-documented. The genome of *Geobacter sulfurreducens* alone encodes over 100 cytochrome c genes [319], of which 78 are MHC with more than three heme motifs per gene and an average of 7.5 hemes per gene [320]. Not all MHC genes are well conserved across *Geobacter* species. Similarly, another EET model organism, *Shewanella oneidensis*, encodes 30 cytochrome c genes. Decades of research have characterised some MHC genes in these organisms genetically and electrochemically; however, more than half of the identified MHC genes remain unannotated [321]. Consistent with this diversity, 57.13% of the MHC genes predicted by motif search in this study were unannotated, highlighting the extensive genetic diversity and the potential for novel EET mechanisms within the SAF-MBR microbial communities.

4.4.3 Metabolic associations of EET genes with carbon metabolism pathways

Correlation analyses revealed previously unrecognised associations between EET genes and the early stages of carbon degradation. In anaerobic digestion research, EET is almost exclusively discussed in relation to syntrophic VFA oxidation and methanogenesis. In contrast, its potential influence on upstream steps—particularly hydrolysis and acidogenesis—has rarely been examined, and, to our knowledge, has not been demonstrated in genome-resolved datasets. Here, both Pearson correlations (r) and Random Forest importance scores (I) showed that MAGs encoding MHC were positively associated not only with electron transport complexes but also with metabolic pathways that support hydrolytic and primary fermentative processes, including glycolysis and the truncated TCA cycle. These patterns suggest that EET-capable bacteria may modulate intracellular redox states to facilitate the rapid turnover of hydrolysis-derived sugars and intermediates.

Under anaerobic conditions, the conversion of glucose to pyruvate and, subsequently, to acetate generates excess NADH via glycolysis and the partial TCA cycle. Without efficient electron disposal, NADH accumulation can restrict carbon flux and slow hydrolysis-derived fermentation. The strong correlations observed between EET+ MAGs and NADH-linked pathways indicate that these organisms may alleviate redox pressure by exporting electrons through MHC-mediated EET [37]. This mechanism aligns with the observed associations between EET genes and cellulolytic pathways, implying that electron export may enable sustained hydrolysis of polymeric substrates by maintaining a favourable intracellular redox balance.

E-pili genes exhibited complementary associations, being enriched in MAGs linked to acidogenesis and acetogenesis, including the Wood–Ljungdahl pathway. Together, these findings show that EET-capable microbes are not limited to downstream syntrophic roles but are also coupled to the earliest stages of carbon processing. This represents an important and previously underappreciated dimension of EET in anaerobic digestion, highlighting its potential influence on the rate-limiting hydrolytic steps that govern overall process efficiency.

4.4.4 Implications for anaerobic wastewater treatment

In the context of anaerobic wastewater treatment, these findings indicate that EET-capable microbes can enhance carbon degradation and energy recovery by preventing NADH accumulation and maintaining high catabolic flux rates. Compared with classical syntrophic VFA oxidation, which depends on interspecies electron transfer via hydrogen or formate and is often constrained by thermodynamic and kinetic limitations [8], EET-driven respiration of fermentable substrates and intermediates is less dependent on dissolved interspecies carriers. By using extracellular electron acceptors to re-oxidise NADH and related redox cofactors, EET-capable microbes can maintain glycolysis and truncated TCA cycle activity, supporting higher acetate production rates and more efficient energy conservation [37].

Coupling this EET-driven respiration with methanogenesis provides a mechanistic basis for designing high-rate anaerobic systems in which hydrolysis, acidogenesis and methanogenesis

are more tightly integrated. In biofilm-based reactors containing conductive media (e.g. GAC, conductive polymers or semi-conductive plastics), electroactive bacteria can respire to the carrier surface, while electrotrophic methanogens (e.g., *Methanothrix*) receive electrons via DIET. This configuration can increase the fraction of carbon channelled to CH_4 rather than accumulating as VFAs, expand the tolerable organic loading rate, and reduce the risk of process failure driven by redox imbalance.

At the process scale, engineering reactors to select for EET+ consortia by choosing conductive media, staged fluidised-bed or membrane configurations, and controlling redox conditions could enable stable treatment of low-strength municipal wastewater at short HRT, with low biosolids yields and net-positive energy balances. Integration of EET-enriched fluidised-bed or membrane bioreactors with anaerobic digestion would therefore offer a route to compact, energy-efficient mainstream treatment, improved CH_4 recovery, and potentially lower greenhouse gas emissions from wastewater infrastructure.

Chapter 5

Direct Interspecies Electron Transfer Driven Syntropy between Sulfate-reducing Bacteria and *Methanothrix* in Pilot-scale Reactor

5.1 Introduction

Building on the previous chapter's demonstration that EET can enhance organic removal in SAF-MBR, this chapter investigates the role of DIET in accelerating CH_4 production in SAF-MBR 2.0, which treated low-strength and low-sulfate real-life sewage. In DIET, microorganisms exchange electrons directly, circumventing conventional soluble intermediates such as hydrogen, formate, or acetate [67, 70, 322]. This mechanism can couple exoelectrogenic bacteria to methanogens and potentially enhance overall reaction kinetics [22, 46].

Recent evidence suggests that DIET within GAC-based biofilms may explain the effective substrate utilisation and high CH_4 yields observed in pilot-scale SAF-MBR systems [12, 16]. Low-strength sewage provides an environment conducive to DIET by preventing dominant growth of fast-replicating bacteria, while the biofilm retains slow-growing or syntrophic populations. The enrichment of specific microbial consortia capable of DIET has thus emerged as a key element in improving the energy balance of anaerobic treatment [24].

In earlier work, syntrophic interactions among *Methanothrix*, *Geobacter*, and other exoelectrogens were identified in reactors fed with simple substrates (e.g., acetate and propionate) [24, 176]. However, relatively few studies have examined whether similar DIET-driven processes underpin the rapid conversion of complex organics in real wastewater [12, 15]. Notably, while our previous 16S rRNA gene survey on the SAF-MBR 2.0 treating domestic sewage revealed a substantial population of SRB with *Desulfovibrio* and *Desulfobulbus* identified as abundant genera, H_2S production was negligible [323]. In chapter 4, *Desulfobacterota* was found to encode the most e-pili genes among all phyla, which contains *Desulfovibrio* and *Desulfobulbus*. Although abundant SRB populations are often used as an indicator of poor digester performance due to competition with

methanogens for acetate and hydrogen sulphide toxicity that inhibits methanogenesis, promotes corrosion, odour, and biogas contamination [324, 325], recent evidence suggests a more nuanced role.

Moreover, some SRB have recently been implicated in possible DIET pathways [38]. This observation contrasts with the conventional view that SRB hinder methanogenesis by completing methanogens on the carbon source (typically acetate), generating H_2S which inhibits methanogenic archaea [326], corrode pipes, cause odour, and contaminate biogas [324, 325]. In conductive methanogenic aggregates from coastal sediments, *Desulfovibrio* and *Methanobacterium* dominated and were metabolically active; *Desulfovibrio* expressed e-pili genes, and the isolate from the aggregate (*Desulfovibrio* sp. JY) directly transferred electrons to *Methanothrix harundinacea* (which cannot use H_2 or formate) and to *Methanobacterium* strain YSL in co-cultures [40].

In SAF-MBR 2.0, the coexistence of a substantial SRB population with minimal H_2S is an interesting pilot-scale observation consistent with a possible DIET-mediated association between SRB and methanogens.

Accordingly, this chapter is to quantify the role of SRB-mediated DIET in SAF-MBR 2.0 with respect to sustaining methanogenesis while minimising sulphide-related drawbacks. Metagenomic and metatranscriptomic analyses are employed to identify microbial consortia, functional genes, and metabolic pathways that may underlie these unusual microbial interactions. In doing so, the work seeks to clarify whether SRB can indeed establish a beneficial partnership with methanogens when electron transfer occurs directly, thereby offering new perspectives on managing and harnessing SRB populations in anaerobic systems.

5.2 Materials and methods

5.2.1 Bioreactor operation, sample collection, preparation, and sequencing

Granular activated carbon samples were obtained from a pilot-scale SAF-MBR located at the Codiga Resource Recovery Center, Stanford University, California, on day 390 of operation, post-attainment of a steady-state condition. This SAF-MBR system has been operational for over three years, treating municipal wastewater sourced from the Serra Street sewer of the university campus. Detailed descriptions of the operational conditions are documented in Shin et al. (2021) [32]. On the sampling day, COD measurements indicated an influent concentration of 500 mg L^{-1} and an effluent concentration of 100 mg L^{-1}, achieving an 83% removal efficiency. Sulfate concentrations were below 5 mg L^{-1} in the influent and under 4 mg L^{-1} in the effluent, with sulfide concentrations in the effluent maintained below 0.5 mg L^{-1}. Samples were then subject to DNA and RNA extraction and sequencing as described previously in Section 3.1.5.

5.2.2 Metagenomic and metatranscriptomic analysis

Raw DNA/RNA reads were quality-trimmed and adapter-filtered, then profiled at the community level for taxonomy (Kaiju against NCBI nr) and function (HUMAnN3 to MetaCyc pathways via

UniRef50) (see Section 3.3). To resolve population genomes, metagenomes were assembled and binned to generate MAGs; bins were assigned phylogeny and quality-checked prior to pathway reconstruction (Section 3.2.5). Genome- and transcript-level evidence for carbon, nitrogen, and sulfur metabolisms was integrated using a standardised module framework (Section 3.5).

The functional genes of the MAGs were annotated via Prokka [327] and Protein BLAST [260] against the RefSeq database and custom database. The custom database was a collection of *actP*, *apr*, *bbs*, *pilQ*, *fsqr*, *gct*, *hnd*, *mmut*, *omcA*, *pfl*, *pflB*, *sat*, *seq*, *tfp*, sulfate metabolism from *Desulfovibrio* spp. The Protein BLAST result was filtered based on the following criteria: 1) query coverage $\geq$ 75%; 2) identity $\geq$ 35%; 3) e-value $\leq$ 0.1; 4) bitscore-to-query length $\geq$ 0.7. To access gene transcription level, the metatranscriptome was assembled using Trinity [265], with functional genes mapped to the assembly via BBmap [267].

DIET-relevant features were evaluated by annotating MHC, e-pili, hydrogenases, and related markers using curated references and custom searches; expression levels were quantified by mapping metatranscriptomes to the assemblies (Section 3.7). Subcellular localisation of *c*-type cytochromes was inferred with PSORTb [257], while pilus genes were screened with PilFind [264] and HMMER [254] respect to reported e-pili genes [38, 115, 161, 164] and assessed for maturation motifs and aromaticity; representative candidates were structurally modelled using RoseTTAFold [328].

5.2.3 Data availability

The 3D protein structure in PDB format is available at Github repository.

5.3 Results and discussion

5.3.1 Divergence in metagenomic and metatranscriptomic taxonomy profiles

The taxonomic profiling of the metagenome and metatranscriptome, conducted via Kaiju, revealed a dominance of specific microbial phyla within the metagenome (Fig. C.6), including *Desulfobacterota*, *Halobacteriota*, *Methanobacteriota*, *Bacteroidota*, *Bacillota*, and *Chloroflexota*. At the family level, *Methanotrichaceae*, *Geobacteraceae*, and *Syntrophaceae* emerged as prevalent, aligning with prior research on membrane bioreactors employing GAC [24, 102] and PVDF [100, 101] biocarriers.

In conventional operation, fermentative bacteria such as in *Bacteroidota*, *Bacillota*, and *Chloroflexota* initiate the degradation of complex organic materials into simpler compounds. These are further processed by syntrophic (*Desulfobacterota*/*Syntrophaceae*) and exoelectrogenic bacteria (*Geobacteraceae*), producing substrates for methanogenesis by *Halobacteriota*/*Methanotrichaceae* methanogens [314].

Further genus-level analysis categorised the predominant microbiota into fermenters, syn-

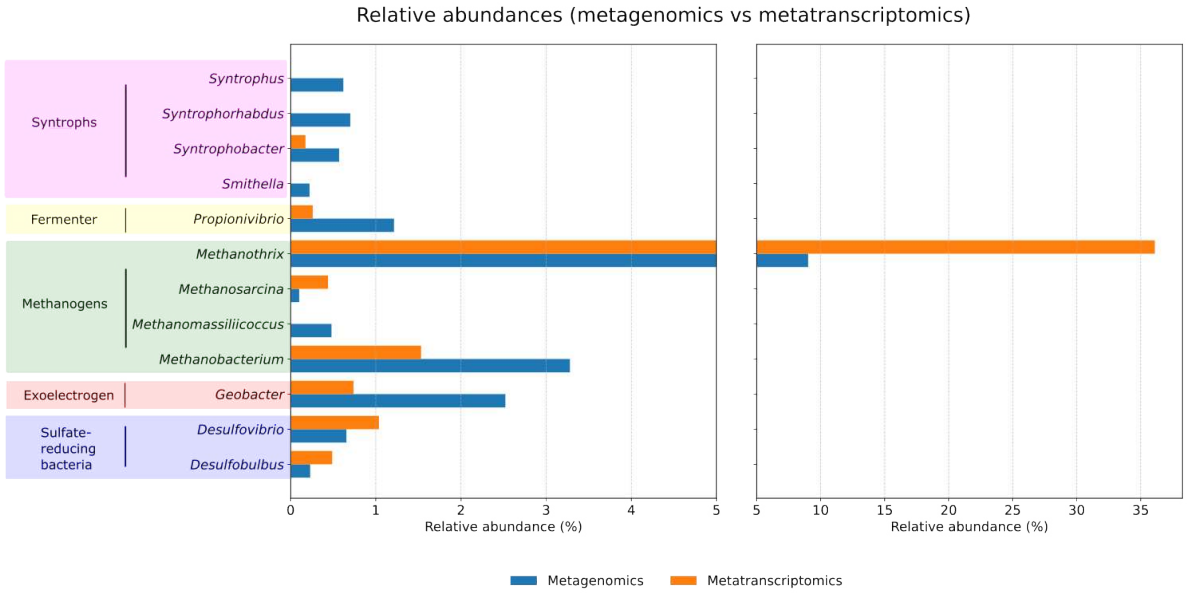


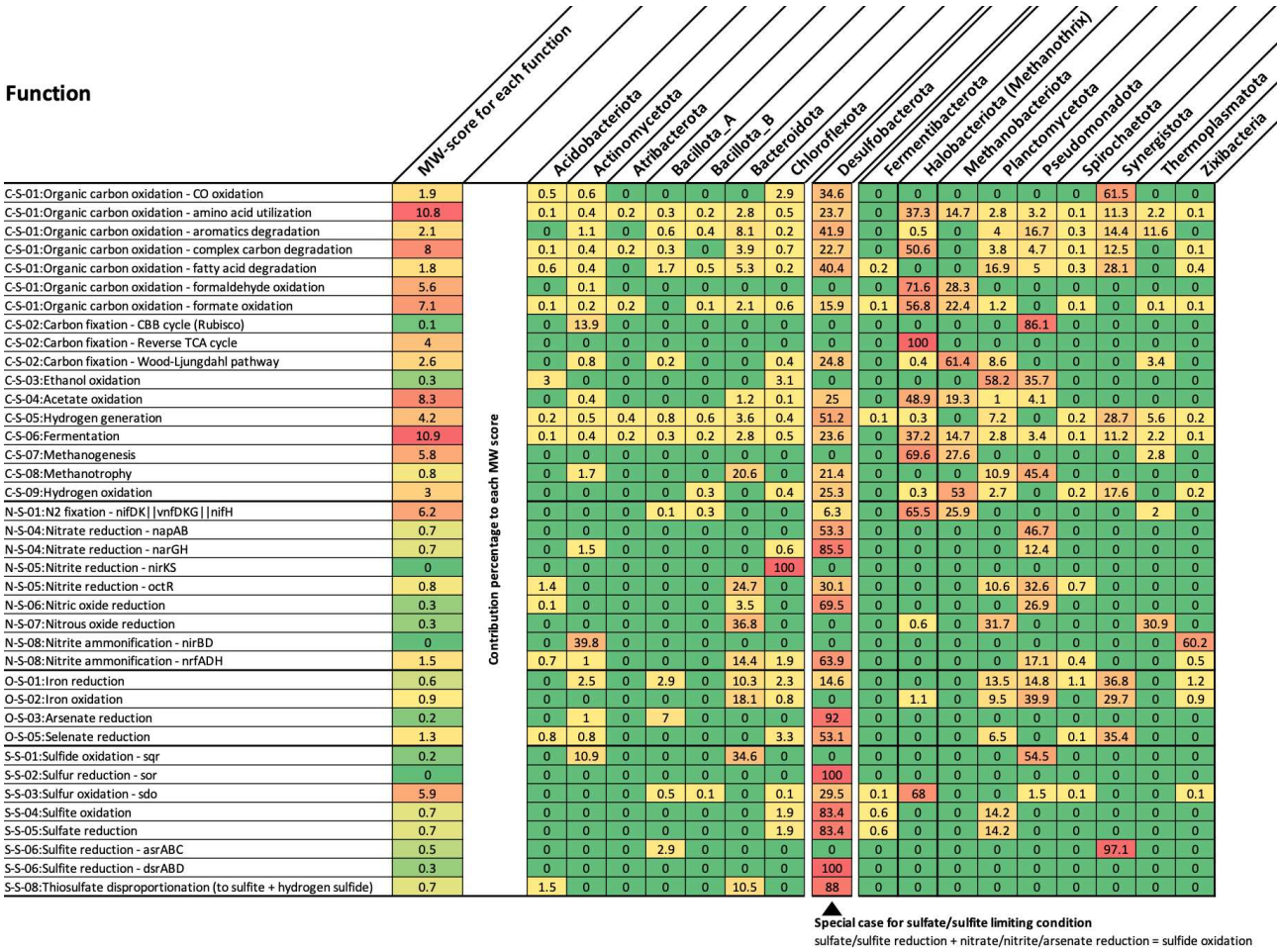
Figure 5.1: Top taxon (>0.1%) at genus level from metagenomics (blue coloured bar) and metatranscriptomic (orange coloured bar) classified using Kaiju against NCBI NR database.

trophs, exoelectrogens, SRB, and methanogens (Fig. 5.1). *Propionivibrio*, accounting for 1.21%, was identified as the leading fermenter, producing acetate and propionate. Syntrophs, such as *Smithella* (0.22%), *Syntrophobacter* (0.57%), and *Syntrophus* (0.62%), alongside exoelectrogens *Syntrophorhabdus* (0.70%) and *Geobacter* (2.52%), facilitated the breakdown of these compounds into CO_2 and electron shuttles for methanogens like *Methanothrix* (9.01%), *Methanobacterium* (3.28%), *Methanomassiliicoccus* (1.21%), and *Methanosarcina* (0.10%) to utilise in methanogenesis.

Intriguingly, the metatranscriptomic analysis unveiled a distinct pattern, with *Methanothrix* (36.1%) remaining highly transcribed, unlike the limited transcription observed for syntrophs and exoelectrogens, where only *Syntrophobacter* (0.17%) and *Geobacter* (0.74%) were notably active. Additionally, the transcriptional levels of SRB, specifically *Desulfobulbus* (0.49%) and *Desulfovibrio* (1.04%), were elevated compared to their metagenomic presence, indicating a higher functional activity in the metatranscriptome.

Conventionally, SRB were thought to compete with methanogens for electron donors like acetate, thereby inhibiting methanogenesis through this competition [329, 330]. In contrast, the concentration of H_2S in biogas dropped below detectable levels (<0.4 ppm) once the bioreactor achieved stable operation, indicating a mature wastewater microbiome. Significantly, the concentrations of common electron acceptors for SRB, including sulfate and sulfide, were minimal in both influent and effluent, not exceeding 0.5 mg L^{-1}, which is much lower than the common range of 20 and 500 mg L^{-1} [331].

This observation suggests that SRBs might utilise alternative electron acceptors under sulfate-scarce conditions. The absence of sulfate might prompt SRBs to transfer electrons to conductive materials such as GAC or directly to *Methanothrix*, indicating a shift towards DIET partnerships rather than competition. The establishment of DIET connections, which requires time as the



Function	MW-score for each function	Acidobacteriota	Actinomycetota	Atribacterota	Bacillota_A	Bacillota_B	Bacteroidota	Chloroflexota	Desulfobacterota	Fermentibacterota	Halobacteriota (Methanothrix)	Methanobacteriota	Planctomycetota	Pseudomonadota	Spirochaetota	Synergistota	Thermoplasmatota	Zixibacteria
C-S-01:Organic carbon oxidation - CO oxidation	1.9	0.5	0.6	0	0	0	0	2.9	34.6	0	0	0	0	0	0	61.5	0	0
C-S-01:Organic carbon oxidation - amino acid utilization	10.8	0.1	0.4	0.2	0.3	0.2	2.8	0.5	23.7	0	37.3	14.7	2.8	3.2	0.1	11.3	2.2	0.1
C-S-01:Organic carbon oxidation - aromatics degradation	2.1	0	1.1	0	0.6	0.4	8.1	0.2	41.9	0	0.5	0	4	16.7	0.3	14.4	11.6	0
C-S-01:Organic carbon oxidation - complex carbon degradation	8	0.1	0.4	0.2	0.3	0	3.9	0.7	22.7	0	50.6	0	3.8	4.7	0.1	12.5	0	0.1
C-S-01:Organic carbon oxidation - fatty acid degradation	1.8	0.6	0.4	0	1.7	0.5	5.3	0.2	40.4	0.2	0	0	16.9	5	0.3	28.1	0	0.4
C-S-01:Organic carbon oxidation - formaldehyde oxidation	5.6	0	0.1	0	0	0	0	0	0	0	71.6	28.3	0	0	0	0	0	0
C-S-01:Organic carbon oxidation - formate oxidation	7.1	0.1	0.2	0.2	0	0.1	2.1	0.6	15.9	0.1	56.8	22.4	1.2	0	0.1	0	0.1	0.1
C-S-02:Carbon fixation - CBB cycle (Rubisco)	0.1	0	13.9	0	0	0	0	0	0	0	0	0	0	86.1	0	0	0	0
C-S-02:Carbon fixation - Reverse TCA cycle	4	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0
C-S-02:Carbon fixation - Wood-Ljungdahl pathway	2.6	0	0.8	0	0.2	0	0	0.4	24.8	0	0.4	61.4	8.6	0	0	0	3.4	0
C-S-03:Ethanol oxidation	0.3	3	0	0	0	0	0	3.1	0	0	0	0	58.2	35.7	0	0	0	0
C-S-04:Acetate oxidation	8.3	0	0.4	0	0	0	1.2	0.1	25	0	48.9	19.3	1	4.1	0	0	0	0
C-S-05:Hydrogen generation	4.2	0.2	0.5	0.4	0.8	0.6	3.6	0.4	51.2	0.1	0.3	0	7.2	0	0.2	28.7	5.6	0.2
C-S-06:Fermentation	10.9	0.1	0.4	0.2	0.3	0.2	2.8	0.5	23.6	0	37.2	14.7	2.8	3.4	0.1	11.2	2.2	0.1
C-S-07:Methanogenesis	5.8	0	0	0	0	0	0	0	0	0	69.6	27.6	0	0	0	0	2.8	0
C-S-08:Methanotrophy	0.8	0	1.7	0	0	0	20.6	0	21.4	0	0	0	10.9	45.4	0	0	0	0
C-S-09:Hydrogen oxidation	3	0	0	0	0	0.3	0	0.4	25.3	0	0.3	53	2.7	0	0.2	17.6	0	0.2
N-S-01:N2 fixation - nifDK||vnfDKG||nifH	6.2	0	0	0	0.1	0.3	0	0	6.3	0	65.5	25.9	0	0	0	0	2	0
N-S-04:Nitrate reduction - napAB	0.7	0	0	0	0	0	0	0	53.3	0	0	0	0	46.7	0	0	0	0
N-S-04:Nitrate reduction - narGH	0.7	0	1.5	0	0	0	0	0.6	85.5	0	0	0	0	12.4	0	0	0	0
N-S-05:Nitrite reduction - nirKS	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0
N-S-05:Nitrite reduction - octR	0.8	1.4	0	0	0	0	24.7	0	30.1	0	0	0	10.6	32.6	0.7	0	0	0
N-S-06:Nitric oxide reduction	0.3	0.1	0	0	0	0	3.5	0	69.5	0	0	0	0	26.9	0	0	0	0
N-S-07:Nitrous oxide reduction	0.3	0	0	0	0	0	36.8	0	0	0	0.6	0	31.7	0	0	0	30.9	0
N-S-08:Nitrite ammonification - nirBD	0	0	39.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60.2
N-S-08:Nitrite ammonification - nrfADH	1.5	0.7	1	0	0	0	14.4	1.9	63.9	0	0	0	0	17.1	0.4	0	0	0.5
O-S-01:Iron reduction	0.6	0	2.5	0	2.9	0	10.3	2.3	14.6	0	0	0	13.5	14.8	1.1	36.8	0	1.2
O-S-02:Iron oxidation	0.9	0	0	0	0	0	18.1	0.8	0	0	1.1	0	9.5	39.9	0	29.7	0	0.9
O-S-03:Arsenate reduction	0.2	0	1	0	7	0	0	0	92	0	0	0	0	0	0	0	0	0
O-S-05:Selenate reduction	1.3	0.8	0.8	0	0	0	0	3.3	53.1	0	0	0	6.5	0	0.1	35.4	0	0
S-S-01:Sulfide oxidation - sqr	0.2	0	10.9	0	0	0	34.6	0	0	0	0	0	0	54.5	0	0	0	0
S-S-02:Sulfur reduction - sor	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0
S-S-03:Sulfur oxidation - sdo	5.9	0	0	0	0.5	0.1	0	0.1	29.5	0.1	68	0	0	1.5	0.1	0	0	0.1
S-S-04:Sulfite oxidation	0.7	0	0	0	0	0	0	1.9	83.4	0.6	0	0	14.2	0	0	0	0	0
S-S-05:Sulfate reduction	0.7	0	0	0	0	0	0	1.9	83.4	0.6	0	0	14.2	0	0	0	0	0
S-S-06:Sulfite reduction - asrABC	0.5	0	0	0	2.9	0	0	0	0	0	0	0	0	0	0	97.1	0	0
S-S-06:Sulfite reduction - dsrABD	0.3	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0
S-S-08:Thiosulfate disproportionation (to sulfite + hydrogen sulfide)	0.7	1.5	0	0	0	0	10.5	0	88	0	0	0	0	0	0	0	0	0

Figure 5.2: Metabolic weight scores of metabolic pathways in the metatranscriptome, with the corresponding contribution percentages from MAGs aggregated at the phylum level. Old and new phylum names are provided for consistency.

biofilm community matures [44], could explain the gradual reduction of H_2S levels during reactor operation.

Desulfovibrio, in particular, is hypothesised to engage in DIET with hydrogenotrophic methanogens like *Methanobacterium* for CO_2 reduction to CH_4 [40]. This interaction is plausible in the SAF-MBR, given the negligible influent sulfite levels (below the detection limit of 4 mg L^{-1}) (Section 3.1.3).

5.3.2 Metabolic function transcription in community

The functional overview based on the KEGG module showed that carbon metabolism constituted 77.3% of the total metabolic activity (Fig. 5.2). Predominant carbon metabolisers included *Proteobacteria* (*Desulfobacterota* and *Pseudomonadota*) and *Euryarchaeota* (*Halobacteriota* and *Methanobacteriota*), which facilitated a diverse range of carbon metabolic processes.

Within *Desulfobacterota*, SRBs, syntrophs, and *Geobacter* were notably active, according the previous taxonomy assignment, contributing to carbon fixation via the Wood-Ljungdahl pathway and acetate oxidation, accounting for 24.8% and 25% of the MW-scores, respectively. Analysis of community scores revealed a combined C2 to C1 pathways score of 10.9%, higher than methanogenesis at 5.8%. This suggests acetate may not be the primary methanogenic sub-

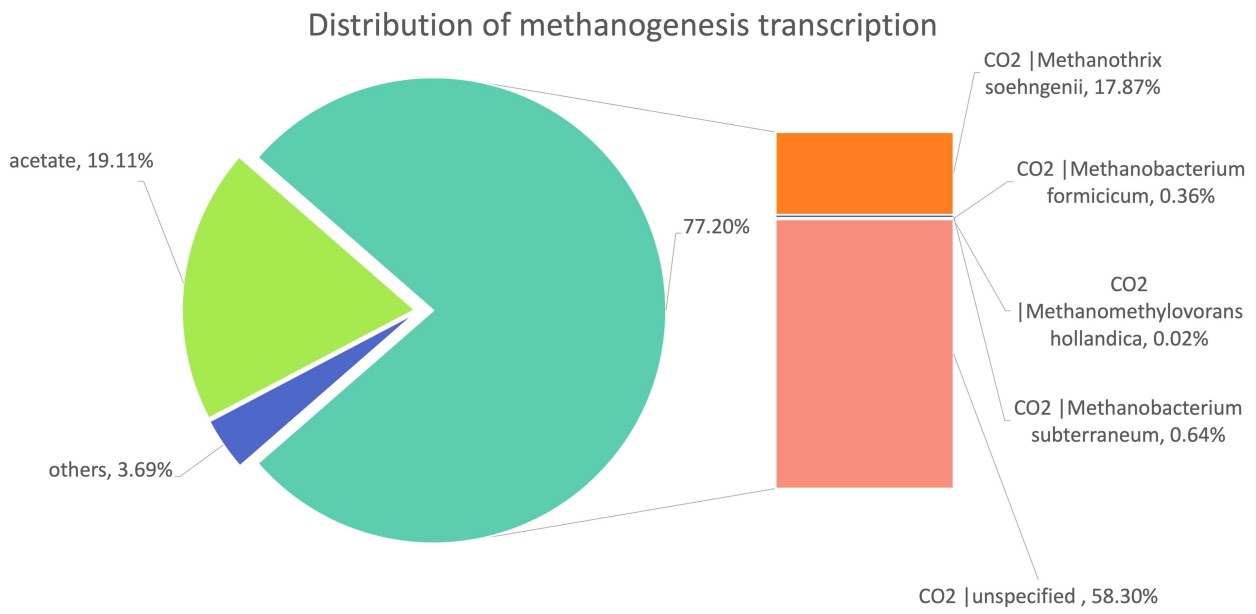


Figure 5.3: Transcription distribution of methanogenesis predicted using HUMAnN 3.0 against Uniref 50 database. CO_2 reduction pathway accounted for 77% of the total methanogenesis transcribed, of which 18% had been identified to come from *Methanothrix* species.

strate. Additionally, the detection of the TCA cycle V (226.3 copies per million reads, cpm) was predominantly attributed to *Geobacter*, as per MetaCyc pathways.

Euryarchaeota played a critical role in carbon metabolism, primarily through methanogenesis and formaldehyde oxidation. *Halobacteriota*, containing *Methanothrix*, contributed 71.6% to the MW-score for formaldehyde oxidation. This score aligned with methanogenesis composition from MetaCyc pathway reconstruction, indicating that CO_2 reduction pathways represented 77.20% of total methanogenesis, equating to 2.26% of the overall metabolism (Fig. 5.3). *Methanothrix* was identified as the leading contributor at 17.87%. Formaldehyde, an intermediate in CO_2 consumption for electromethanogenesis via the RHP pathway, was processed by *Methanothrix* (61.93 cpm). The absence of membrane-bound hydrogenase in *Methanothrix* suggests DIET as the likely electron donor for CO_2 reduction to CH_4, consistent with the observed methanogenesis patterns [24, 46].

Sulfur metabolism was particularly pronounced in *Desulfobacterota*, notably the sulfite and sulfate reduction. Interestingly, alongside these, certain microbial communities have demonstrated the ability to transcribe nitrate and nitrate reduction pathways [332–334]. This co-transcription suggests a potential coupling between sulfide oxidation and nitrate reduction pathways, an interplay that is pivotal for understanding sulfur cycling dynamics [332–334].

The oxidation of sulfide to elemental sulfur, followed by its disproportionation or further oxidation through the reverse sulfate reduction pathway, signifies a critical juncture in sulfur metabolism. Although explicit pathway modulation was not delineated, the operation of the SAF-MBR system under conditions of low sulfite/sulfate influent allowed for the inference of such mechanisms. The involvement of the dissimilatory bisulfite reductase (Dsr) enzyme underscores the intricate

relationship between sulfide oxidation and sulfate reduction pathways, illustrating the genomic capacity for sulfate reducers to facilitate sulfide oxidation by coupling with alternative electron acceptors.

Besides canonical nitrate/nitrite reduction, the transcript profile also suggested the coupling of sulfide oxidation with arsenate reduction, with *Desulfobacterota* contributed 92% of arsenate reduction transcription. These genes, which typically mediate sulfite reduction, operate in reverse under sulfite-limiting conditions [334]. These patterns indicate a flexible microbial sulfur metabolism that can utilise alternative electron acceptors when available. Because concentrations of nitrate, nitrite, sulfur species, and arsenate were not measured, we cannot assess their in situ availability; accordingly, the relative contribution of these pathways is inferred solely from transcriptional data and remains small compared to the overall metabolic score.

5.3.3 Metagenome-assembled genomes and gene transcription

Based on the functional profile, CO_2 appears to be the dominant methanogenic substrate. Since the KEGG module does not assign a specific module for propionate metabolism, and MetaCyc pathways currently cover only the propionate-to-acetate pathway during lactate metabolism, propionate degradation pathways were reconstructed by identifying hallmark genes in MAGs (e.g., propionyl-CoA carboxylase and methylmalonyl-CoA mutase).

Ten high-quality MAGs were identified (completeness $\geq$ 90% and redundancy $\leq$ 10%) of which seven were assigned to *Desulfobulbus* (bin 142), *Desulfovibrio* (bin 172,177), and *Desulfovibrionaceae* (bin 35,61,112,195), and three were assigned to syntrophs *Syntrophobacter* (bin 7), *Syntrophomonas* (bin 92), and *Smithella* sp F21 (bin 97) (Table B.10). These showed essential metabolic pathways: propionate degradation via the methylmalonyl-CoA pathway and acetate oxidation through the Wood-Ljungdahl pathway, notably in *Desulfobulbus* (bin 142).

Although MAGs assigned to *Geobacter* and *Methanothrix* with 64.89% to 84.86% completeness (Table B.10), these mid-quality bins offer limited metabolic pathway insights.

Propionate and acetate metabolism in SRB and syntrophs

A methylmalonyl-CoA dependent propionate degradation pathway, previously identified in *Desulfotomaculum kuznetsovii* under anaerobic conditions [335], was well represented in bin 142. All genes in this pathway, or their functional equivalents, showed homologous identities of over 94% and e-values of at least 3e-22. Relevant genes, including glutaconate CoA-transferase, methylmalonyl-CoA epimerase, methylmalonyl-CoA mutase, succinate-CoA ligase, succinate dehydrogenase, fumarate hydratase, malate dehydrogenase, pyruvate carboxykinase, and pyruvate:ferredoxin oxidoreductase, all with transcription from 7.12 to 11.28 $\log_2$RKPM (Fig. C.8).

In the absence of sulfate, *D. propionicus* can ferment three moles of pyruvate to two moles of acetate and one mole of propionate stoichiometrically via the methylmalonyl-CoA pathway [336–340]. This pyruvate degradation pathway can be coupled with Fe(III) reduction and independent of sulfate reduction, as demonstrated with sulfate reduction inhibitor molybdate (1 to 10 mM) [341].

Since the SAF-MBR was operated at low sulfate influent, the result suggested that Fe(III) reduction or DIET could substitute sulfate as an electron sink.

H_2/formate mediated interspecies electron transfer (MIET)

Several potential pathways were identified for electron transfer in bin 142 (Fig. C.9). The first involves indirect electron transfer in which H_2 and formate are oxidised by an NADP-reducing hydrogenase and formate dehydrogenase, respectively. Three NADP-reducing hydrogenase subunits (*hndA*, *hndC* and *hndD*) were transcribed (7.35 to 12.90 log2 RPKM), as well as formate dehydrogenase (8.53 log2 RPKM). The H_2 and formate were likely to be synthesised by fermenters, syntrophs, and the SRB itself present in the bioreactor before being transported into the periplasm of the SRB via formate transporter (Foc).

Another possibility is that the SRB could utilise H_2, the byproduct of acetoclastic methanogenesis, from *Methanothrix*; such coexistence has been reported [342]. The electrons produced during oxidation were then donated to c-type cytochromes. Relevant proteins, including cytochrome c3 family protein and cytochrome *c* biogenesis proteins (Ccm), were transcribed. The cytochromes shuttled the electrons down the periplasm to the quinone-reductase (Qrc) complex in the inner membrane, which operated in a redox loop with an adjacent complex (Qmo) complex. The Qmo complex participated in the dissimilatory sulphate reductase pathway, together with enzymes adenylyltransferase (Sat), adenylylsulfate reductase (Apr), and dissimilatory sulfite reductase (Dsr) to reduce sulfate to hydrogen sulfide. All three subunits of the QmoABC complex were identified with e values $\geq$ 3e-24. Notably, the described pathway (from H_2/formate reduction to sulfate reduction) has previously been identified as the basis for sulfate respiration in deltaproteobacterial SRB [343].

Pilus-mediated direct interspecies electron transfer (DIET)

Type C multiheme cytochromes (MHCs) is biomarkers for DIET, distinguished by their ability to harbor multiple heme groups, enabling efficient electron transport across microbial membranes, coupling between syntrophic partners in microbial consortia [68, 344]. Cytochrome *c* biogenesis proteins were identified in all SRB MAGs, notably in bins 35, 112 and 142, which demonstrated active transcription. Despite that, motif search categorised these proteins as mono- and di-cytochrome *c* variants. Subcellular localisation predictions suggested their placement in the periplasm, making them unlikely to function as extracellular electron transfer mediators. Furthermore, among the 30 hypothetical proteins exhibiting patterns akin to multiheme cytochrome *c*, their localisation was confined to cytoplasmic, periplasmic, or inner membrane regions.

Another potential pathway could occur through transmembrane type-IV pili that could export electrons to the external environment, reduce Fe(III) and/or participate DIET. Such a pathway has been proposed in co-culture experiments [40, 345]. For instance, *Desulfovibrio ferrophilus* transferred electrons extracellularly via aromatic-amino-acid-rich e-pili. The transcription of genes encoding an aromatics-rich putative pilin was upregulated during growth on Fe(III) oxide, sug-

gesting the involvement of this protein in facilitating reduction reactions with Fe(III). Additionally, *D. ferrophilus* reduced Fe(III) oxides occluded within porous alginate beads at a slightly slower rate than cultures amended with the artificial electron shuttle anthraquinone–2,6-disulfonate. This indicated that SRB could employ multiple strategy to electron export, including MIET for Fe(III) reduction, in which genetic manipulation on MHC deletion and pili inhibition [346] further suggested this possibility.

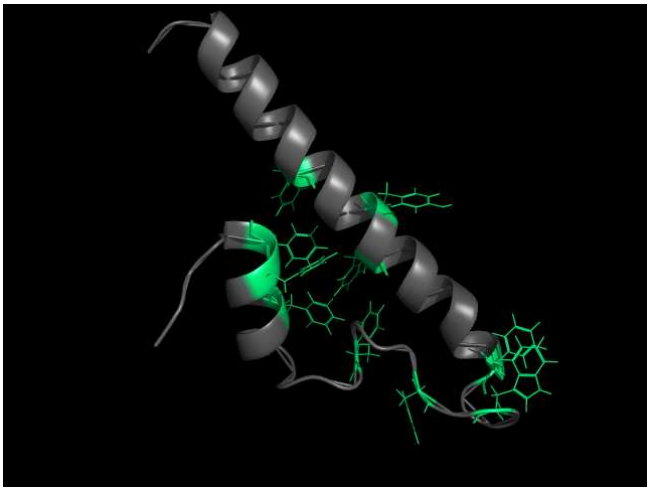
Type IV pili, particularly a subset known as e-pili, are capable of electron conduction through pi-pi stacking. In *G. sulfurreducens*, these pili contain 9.8% aromatic amino acids, spaced 3.2 Å apart. This structure, critical for conductivity, was verified in other bacterial homologs using conducting-probe atomic force microscopy. In contrast, pilus-deficient *G. sulfurreducens* mutants were unable to reduce Fe(III) oxides, highlighting the role of aromaticity in conductivity, as mutants with fewer aromatic residues exhibited reduced conductivity. However, this pili have been associated only with electron export, with no import activity observed.

Through searching against reported e-pili genes [38, 115, 161, 164], 8, 1, 4, and 19 putative pili genes were identified in bins 7, 94, 97, and 142, corresponding to syntrophs and *Desulfobulbus* (Table. B.11). Further analysis of a type IV pili (sequence ID: FHFJCJBA_03220) in bin 142 revealed an aromaticity of 22.58%, with the largest gap being seven amino acids (Table B.11). Effective electron transport in these pili requires consistent spacing of aromatic amino acids within 19 Å [347], ideally between 3-4 Å [348]. However, such regular sequences were not evident in the primary structures of the pili gene candidates and should be informed by tertiary structures. To investigate this, the 3D protein structure of the pili was predicted using RoseTTAFold (Fig. 5.4) with a 0.79 confidence. Spacings identified between aromatic amino acids ranged from 0.35 to 4.93 Å, suggesting the potential for electron hopping in this pili protein.

The transcriptomic evidence, together with the fast propionate degradation kinetics reported in the performance evaluation of the SAF-MBR [32], hinted that DIET might mediate propionate and acetate degradation by the SRB, thus benefiting methanogenesis (Fig. 5.5). The direct electron transfer mechanisms highlighted previously can be mediated by conductive GAC, reducing or bypassing the requirement for outer-membrane cytochromes or pili, thereby diminishing the necessity for conductive proteins [140, 163].

5.3.4 GAC-facilitated DIET in sulfate-reducing bacteria and promoted symbiotic relationships

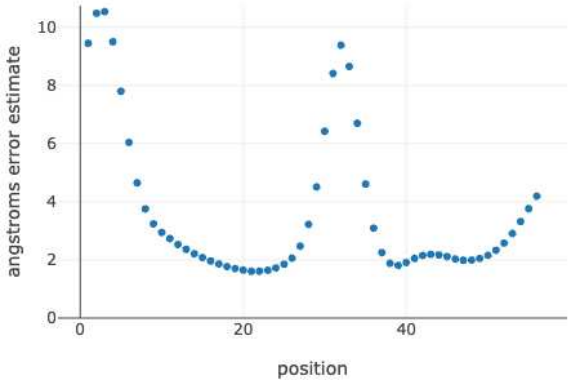
To evaluate whether the microbial community associated with conductive GAC in SAF-MBR 2.0 exhibited features beyond those expected under non-conductive carrier conditions, the community profile was interpreted in the context of literature-based reference AFMBR systems employing polymeric PVDF or ABS carriers (Section 2.1.6). This comparison was introduced to provide a defensible control framework in the absence of a fully matched, parallel pilot-scale non-conductive reactor. The reference systems represent lab- and pilot-scale AFMBRs operated under stable conditions with non-conductive media, for which genus-level 16S rRNA gene-based community



(a)



(b)



(c)

Figure 5.4: (a) The predicted 3D protein structure of pili from SRB bin 142 is illustrated, with aromatic residues highlighted in green. (b) A colour gradient indicates the error estimate. Yellow represents the highest error margin, while purple denotes the lowest. (c) An accompanying plot of the error estimate.

compositions have been reported.

Across PVDF- and ABS-based AFMBRs, syntrophic bacteria and acetoclastic methanogens were consistently detected but typically at modest relative abundances. For example, in a PVDF-based AFMBR, *Syntrophobacter* accounted for 1.53–5.58% of reads and sulfate-reducing *Desulfobulbus* for 3.21–4.51%, while canonical exoelectrogens such as *Geobacter* were not detected [100]. In paired-carrier comparisons, non-conductive ABS beads supported lower relative abundances of *Syntrophobacter* (2.01%) and *Geobacter* (0.65%) than conductive GAC operated under otherwise comparable conditions, with the total syntrophic fraction in ABS remaining at approximately 3.15% [102]. Despite this, acetoclastic methanogenesis was sustained in non-conductive systems, with *Methanothrix* comprising 38.32% of archaeal reads in ABS carriers [102] and exceeding 70% at later stages in a PVDF-based AFMBR [100]. Collectively, these studies demonstrate that non-conductive AFMBRs can maintain stable syntrophic and methanogenic communities, but typically without strong enrichment of exoelectrogen-associated taxa.

In contrast, the GAC-associated community in SAF-MBR 2.0 displayed a distinct quantitative profile. At the DNA level, *Geobacter* was detected at 2.52%, exceeding abundances reported for ABS carriers and absent PVDF systems, while multiple syntrophic genera (*Syntrophobacter* 0.57%, *Smithella* 0.22%, *Syntrophus* 0.62%, *Syntrophorhabdus* 0.70%) co-occurred within the same biofilm. Methanogens were dominated by *Methanothrix*, accounting for 9.02% of metagenomic reads but increasing to 36.10% in the metatranscriptome, indicating strong functional prioritisation under conductive-carrier conditions. Although transcriptomic data are not available for the PVDF and ABS reference systems, this comparison indicates that GAC supports the co-enrichment of exoelectrogen-linked taxa and highly active acetoclastic methanogens beyond levels typically observed in non-conductive AFMBRs.

Beyond community composition, SAF-MBR 2.0 provided insight into the role of SRB under long-term exposure to conductive media. Previous studies have shown that GAC promotes the formation of conductive biofilms and microbial aggregates, enhancing interspecies electron exchange efficiency [305, 349]. In SAF-MBR 2.0, rapid propionate degradation was observed despite low influent sulfate and sulfite concentrations [32]. Metatranscriptomic analysis revealed the persistence of SRB alongside syntrophs, *Geobacter*, and *Methanothrix*, with expression profiles indicating metabolic flexibility and the ability to utilise multiple electron acceptors, including conductive surfaces, sulfate, nitrate, and nitrite. The continued activity of SRB under sulfate-limited conditions suggests that electron flow was primarily sustained via DIET rather than conventional sulfate respiration.

Genome-resolved transcriptional analysis further indicated syntrophic cooperation between *Desulfobacterota* (*Deltaproteobacteria*) and *Halobacteriota* (*Euryarchaeota*). In particular, *Desulfobulbus* (bin 142) was implicated in propionate oxidation via the methylmalonyl-CoA pathway, generating acetate that was subsequently processed through acetoclastic methanogenesis by *Methanothrix* and through complementary carbon oxidation pathways by *Geobacter* and other syntrophs. This pattern supports a tightly coordinated division of metabolic labour facilitated by conductive biofilm architecture.

The dominance of *Methanothrix* in SAF-MBR 2.0 is consistent with the low-strength sewage influent and high biomass retention characteristic of membrane-coupled systems. Under conditions where bulk and biofilm-associated acetate concentrations approach near-zero, the lower substrate affinity constant of *Methanothrix* (K_s = 3.05 mgCOD L^{-1}) confers a competitive advantage over *Methanosarcina* (K_s = 221 mgCOD L^{-1}), despite the latter's higher maximum growth rate [350]. Accordingly, the observed community structure reflects substrate-limited selection rather than a direct consequence of conductive media alone.

This contrasts with studies reporting *Methanosarcina* dominance in GAC-amended systems operated at higher organic loads and elevated permeate COD concentrations (137–1013 mg L^{-1} compared to 32–86 mg L^{-1} in SAF-MBR 2.0) [351]. Under such conditions, the affinity advantage of *Methanothrix* diminishes, allowing *Methanosarcina* to proliferate. While several studies have demonstrated GAC-stimulated DIET between *Methanosarcina* and *Geobacter* [111, 177], a recent co-culture study reported no DIET enhancement between *Methanothrix thermoacetophila* and *Geobacter metallireducens* in the presence of GAC [116]. These discrepancies highlight the strong context dependence of GAC-mediated interactions and underscore the importance of complex biofilm structure and reactor-scale constraints in shaping DIET-enabled syntrophy.

Taken together, these observations directly address the central aim of Chapter 5 by demonstrating that, under sulfate-limited and low-strength wastewater conditions, SRB persist as active members of the carbon-converting consortium through DIET-enabled syntrophic associations rather than classical sulfate respiration. The enrichment and sustained transcriptional activity of SRB, *Geobacter*, and *Methanothrix* within GAC-associated biofilms indicate that conductive media restructure electron flow pathways in favour of tightly coupled, multi-partner syntrophy. This supports the interpretation that DIET represents a stabilising, functionally relevant electron transfer mode linking SRB metabolism to acetoclastic methanogenesis during pilot-scale SAF-MBR operation.

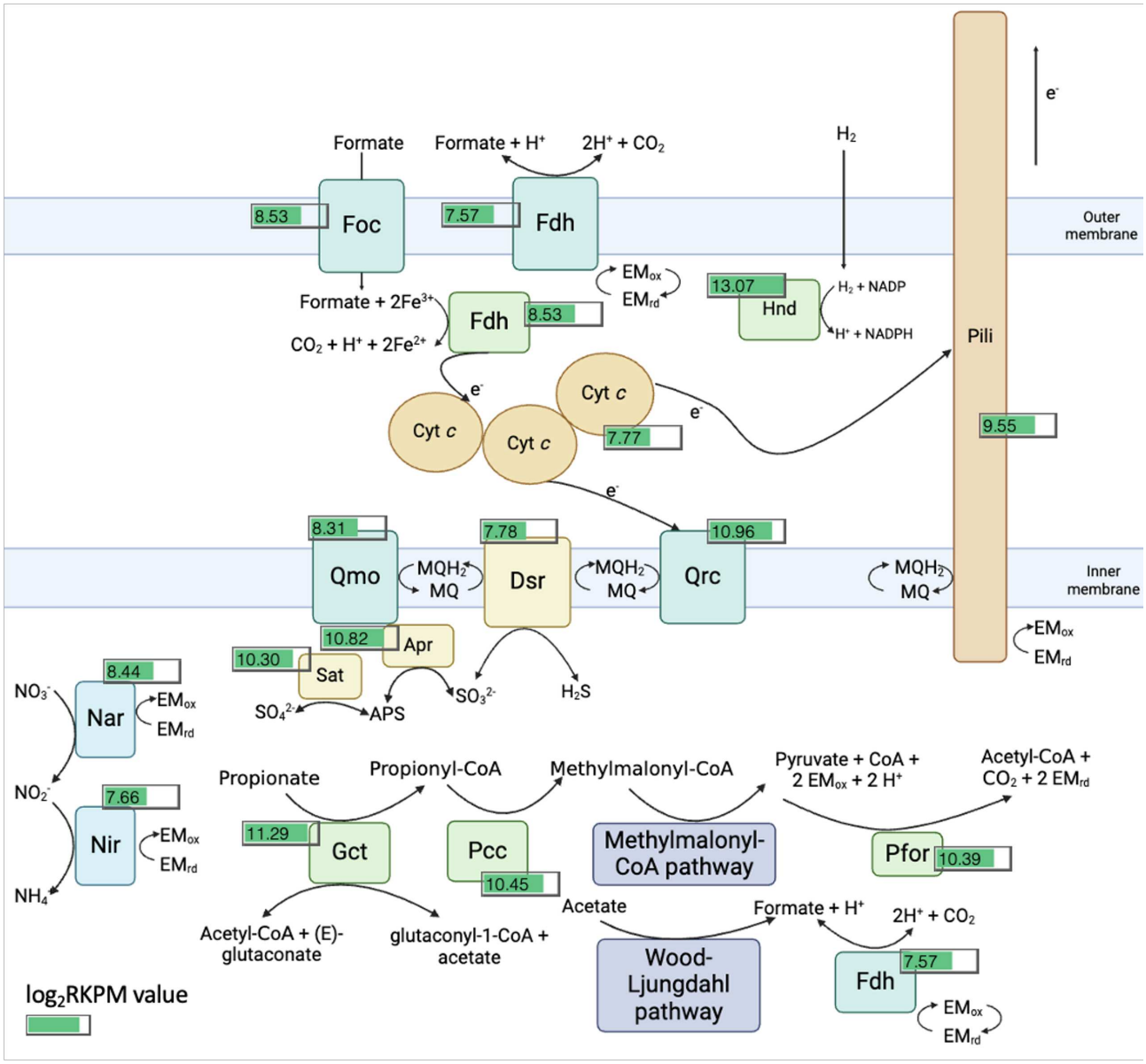


Figure 5.5: Proposed SRB *Desulfobulbus* (bin 142) metabolism and transcription level in $\log_2$ RKPM. These proposed pathways include pyruvate degradation, propionate to acetate oxidation, formate/H_2 mediated electron transfer, and the potential electron export via type IV pilus. Apr, Adenylylsulfate; Cyt c, Cytochrome c; Dsr, Sulfite reductase; Fdh, Formate dehydrogenase; Fdh, formate dehydrogenase; Foc, Formate transporter; Gct, Glutaconate coenzyme A transferase; Hnd, NADP-dependent hydrogenase; Pcc, Propionyl-CoA carboxylase; Pfor, Pyruvate:ferredoxin oxidoreductase; Pili, hypothetical e-pili (FHFJCJBA_03220); Qmo, Quinohaemoprotein methanol dehydrogenase; Qrc, Menaquinone-interacting cytochrome c reductase; TCA, Tricarboxylic acid cycle; Sat, Sulfate adenylyltransferase.

Chapter 6

Identification of Extracellular Electron Transfer Sheath Proteins in *Methanothrix* Using Pangenomic Analysis

6.1 Introduction

Methanothrix (previously known as *Methanosaeta*) is an obligate aceticlastic methanogen vital to anaerobic environments, including sediments and wastewater treatment systems. It converts acetate into CH_4, significantly contributing to global carbon cycling [352–354]. Its high affinity for acetate enables it to outcompete other methanogens under low-substrate conditions, with effective functioning observed at acetate concentrations as low as 25 mg COD L^{-1}. This trait renders *Methanothrix* prevalent in systems with low effluent COD, indicating efficient wastewater treatment [72, 314].

Methanothrix typically resides within biofilms formed by syntrophic bacteria. In natural environments, such as rice paddies, it engages in DIET with *Geobacter* [46]. Laboratory-scale bioreactors have also demonstrated DIET in *Methanothrix*-based systems, including SAF-MBR with GAC [24], AFMBR with PVDF [100], and microbial electrolysis cells employing graphite biocathodes [19]. These associations highlight the versatility of *Methanothrix* in facilitating efficient CH_4 production via DIET across engineered and natural systems.

Despite its slow growth rate and environmental sensitivity [92, 355–357], *Methanothrix* forms syntrophic associations with microbes such as *Geobacter* through EET [46, 116]. In this study, the acetate affinity of *Methanothrix* reached as low as 3.05 mg COD L^{-1} when mediated by GAC[32] (Section 1.1.2). While the EET mechanisms of various methanogens are relatively well characterised, those in *Methanothrix* remain elusive [53]. The major sheath protein MspA in *Methanothrix thermoacetophila* has garnered attention due to its aromatic-rich composition, resembling the electrically conductive pili observed in other methanogens [19]. Experimental studies have characterised MspA as an amyloid protein with evenly spaced β-sheet structures, a configuration hypothesised to facilitate electron transport. Furthermore, co-culture experiments involving *M. thermoacetophila* and *Geobacter* sp. revealed upregulation of both MspA and the

PQQ-binding quinoprotein SqpA under DIET conditions [116].

The PQQ-associated quinoprotein (SqpA) suggested a role in mediating electron transfer across the *Methanothrix* membrane. SqpA contains a high proportion of aromatic amino acids, enhancing its electron transfer capabilities [116]. This aligns with observations in *Methanosarcina*, where surface quinoproteins facilitate DIET-based electron uptake [358]. These findings support the involvement of both MspA and SqpA in EET, in which *Methanothrix* accepts electrons from syntrophic partners and contributes to CH_4 production in anoxic environments.

Generalising the involvement of MspA across the *Methanothrix* genus presents challenges. To date, only two MspA-like proteins have been confidently identified in methanogenic species: *Methanothrix thermoacetophila* PT and Mhun_2271 in *Methanospirillum hungatei* JF-1 [172]. The sole known MspA protein in *M. thermoacetophila* lacks homologs in publicly available databases [41, 172]. A 2024 study utilising Foldseek and UniProt50 reported no meaningful structural homologs for MspA (Meth_1069; ABK14853.1) [41], highlighting a significant knowledge gap in generalising the role of this amyloid sheath protein across the genus.

Pangenome analysis provides a robust method for elucidating the genetic diversity and functional capabilities within the genus *Methanothrix*, particularly for identifying and characterising outer-surface proteins [272, 359]. By differentiating the core genome shared by all species from accessory and singleton genes unique to specific strains, pangenomics addresses the current limitation of having only one *Methanothrix* species with an annotated surface protein. This approach facilitates systematic prediction of surface amyloid proteins across species and MAGs, leveraging rule-based tools to improve annotation accuracy. Additionally, evaluating aromaticity using pangenomic data helps determine the conductive properties of MspA homologs, thereby linking genetic variation to functional traits. Categorising genes into core, accessory, or singleton groups provides insights into their evolutionary significance and ecological roles, enhancing our understanding of *Methanothrix*'s adaptability and interactions within diverse environments [273, 360]. Pangenomic analysis thus serves as an entry point for investigating the molecular mechanisms underlying surface protein diversity in *Methanothrix*, advancing both genomic annotation and functional characterisation, and addressing the challenges in generalising EET mechanisms across the genus. In addition, building on the understanding that the *Methanothrix* sheath comprises amyloid proteins with regular β-sheet structures, rule-based tools proven effective in predicting amyloid domains were employed [274]

This study aims to determine whether *Methanothrix* genomes encode conserved outermost-layer amyloid proteins (MspA-like) that facilitate EET and to establish a rule-based framework for their reliable annotation. To achieve this, (i) a rule-based amyloid-screening workflow was developed and applied to newly reconstructed MAGs and publicly available *Methanothrix* genomes; (ii) pangenomic comparisons were performed to assess conservation, copy number, and genomic context of MspA-like candidates across the genus; (iii) sequence and structural features relevant to EET, including predicted β-sheet domains, signal peptides, repeats, and aromatic amino acid composition, were characterised and evaluated; and (iv) candidate occurrence was then linked to putative EET capacity by examining co-localised or associated EET genes and, where available,

transcriptional evidence. This integrated aim-and-objectives statement closes the Introduction by defining how the thesis identifies, compares, and functionally prioritises MspA-like surface-layer amyloids in *Methanothrix*, including cases lacking direct homologs to Meth_1069. By elucidating these molecular characteristics, this study provided an entry point to the mechanisms underlying EET in Methanothrix and establishes a framework for further investigation into the structural and functional diversity of electron transfer proteins in methanogenic archaea.

6.2 Materials and methods

6.2.1 Genome collection

High-quality genomes (completeness $>$ 90% and redundancy $<$ 10%) were obtained from metagenome-assembled genomes reconstructed from the staged anaerobic fluidised-bed membrane bioreactor (SAF-MBR). These MAGs were identified as belonging to *Methanothrix*. In addition, *Methanothrix* genomes were retrieved from the (NCBI) genome database, including reference genomes, annotated genomes, and MAGs with assembly levels ranging from complete to contigs. All genomes were assessed using CheckM [220], and those failing to meet the completeness ($>$ 90%) and redundancy ($<$ 10%) criteria were discarded.

To determine the environmental origins of the *Methanothrix* genomes downloaded from NCBI, a custom query was constructed using the BioProject and BioSample identifiers associated with each genome. Entrez Direct commands (`efetch` and `esummary`) [271] were used to retrieve information on culture collections, isolate sources, and related publications. The metadata was then manually curated and categorised into groups such as anaerobic digesters and wastewater treatment facilities. All curated metadata and quality assessment results are available on GitHub repository.

6.2.2 Pangenome analysis

Following the genome collection, pangenome construction was undertaken to explore the genomic diversity, core genes, and accessory elements of *Methanothrix*. Detailed methodology is provided in Section 3.8.

A minimum contig length of 1,000 bp was applied to all selected genomes. Open reading frames were identified using Prodigal [230] and then analysed as follows:

1. **Gene Profiling:** The s were profiled using an HMM database containing 76 archaeal core genes, 71 bacterial core genes [361], and 5S, 16S, and 23S rRNA genes (archaea and bacteria) from Barrnap (GitHub: Barrnap) to infer genome-level evolution and prepared the data for phylogenomic analysis.

2. **Functional Annotation:** The s were functionally annotated by mapping them to the (COGs) database [309] and Kofam [362].

3. **Horizontal Gene Transfer Prediction:** Each gene was evaluated for its probability of being transferred horizontally.

Genome dereplication. To take advantage of Anvi'o [270] for interactive and in-depth analyses, a soft limit of 100 genomes was set for constructing the pangenome. Because there were 184 *Methanothrix* genomes initially, dereplication was performed to reduce redundancy and maintain annotation quality. Genomes were clustered based on (ANI) thresholds of 90%, 95%, and 99%, and a centrality-based method selected the genome with the highest average similarity as the cluster representative. The 95% ANI threshold was ultimately chosen due to its balance between coverage and specificity.

Gene clustering and MCL inflation. The 98 *Methanothrix* representative genomes were used to construct the pangenome. Genome similarity was assessed via `pyani` [363]. Genes were then clustered using the MCL [364], which defines the sensitivity of cluster detection. Inflation factors of 2, 4, and 6 were tested; an inflation factor of 4 was selected for its balanced performance. Core genes or gene clusters were defined as those present in at least a quarter of the median coverage of the originating sample; all remaining genes or clusters were deemed accessory. The representative sequences from core gene clusters were mapped to the three metatranscriptomic assemblies via BBmap [267] to calculate the gene expression level in SAF-MBR.

6.2.3 Functional enrichment analysis in pangenome

Functional enrichment analysis was conducted using the approach described by Shaiber et al. [365] to identify functions disproportionately represented within specific clades or subclades of the pangenome. Genomes were grouped based on assigned attributes, such as source environment or reactor operating conditions for wastewater samples. Functions significantly associated with any group were identified, focusing on those prevalent in one group but largely absent in others. Gene-level functional annotations were consolidated at the gene cluster level, assigning each cluster the most frequent function. This analysis reflects functional associations rather than direct functional annotation.

6.2.4 MspA and Spq prediction

Protein prediction on two potential EET genes MspA and SpqA were first conducted using HMMsearch. Similar to the approach used in Chapter 4, for MspA (Meth_1069) and SpqA (Surface pyrroquinoline quinone (PQQ) protein), a BlastP search [260] was conducted, collecting amino acid sequences with over 35% identity and 70% coverage. These sequences were clustered using MMseq2 [261] with a minimum identity threshold of 70%. Each protein cluster was then aligned using MUSCLE v5 [262], and the resulting alignments were integrated into a pHMM database.

For MspA, rule-based approach leveraging multiple bioinformatics tools was implemented to identify MspA-like protein sequences in *Methanothrix* pangenomic gene clusters. Signal peptide

presence, indicative of potential extracellular localisation, was first assessed using SignalP-6.0 [275]. Sequences encoding a type I signal peptide (SP I) were further examined for extracellular localisation using the motif-based prediction tools Cello [276] and PSORTb [257]. The amyloidogenic potential was predicted using AMYPred-FRL [277], and aggregation propensity was quantified with PASTA 2.0 [278] by evaluating free energy and β-sheet content. Using Mthe_1069 as a reference, sequences exhibiting a free energy value below -10 and a β-sheet content exceeding 30% were grouped as MspA-like, using Mthe_1069 as a reference. Functional annotations were cross-verified using eggnog-mapper 2.0 [279] to exclude well-characterised proteins and minimise misannotations.

6.2.5 Motif analysis

Motif analysis was performed using Multiple Expectation Maximisation for Motif Elicitation (MEME) [366] to identify novel, ungapped motifs. Thirty-nine protein sequences (83–554 amino acids, mean: 221.8) were analysed in classic discovery mode, optimising motif information content based on the E-value. A zero-order background model was generated to refine motif significance with any number of motif occurrences per sequence.

6.2.6 3D protein structure prediciton

The three-dimensional structures of MspA-like proteins identified in this study were predicted using AlphaFold 3 [367]. The predicted structures were subsequently visualised, and the structural features were annotated in UCSF ChimeraX [368].

6.3 Results

6.3.1 *Methanothrix*'s pangenome

A pangenome analysis of 98 *Methanothrix* genomes yielded 26,540 gene clusters. Of these, 1,740 were identified as core genes, accounting for 6.66% of the pangenome (Fig. 6.1). These core genes encompass essential cellular functions, including DNA replication, transcription, translation, and energy metabolism. In contrast, the variable genome, constituting 93.34% of the pangenome, contained genes involved in diverse functions, including methanogenesis, nitrogen fixation, stress response, and horizontal gene transfer. This extensive variable component indicates a pronounced capacity for genomic diversity and adaptation, a pattern seen in microbial pangenome studies, where large accessory and unique gene sets are linked to metabolic versatility and niche-specific adaptation [369, 370] Notably, 2,718 out of 5,669 variable gene clusters were unique to individual strains, further highlighting the high level of genomic variability among *Methanothrix* strains.

All key genes associated with the acetoclastic pathway, except phospho-acetyl-transferase (*pta*), which is specific to *Methanosarcina*, were detected in the core genome. Additionally, genes encoding components of the hydrogen-independent CO_2 reduction pathway (a hydrogenotrophic

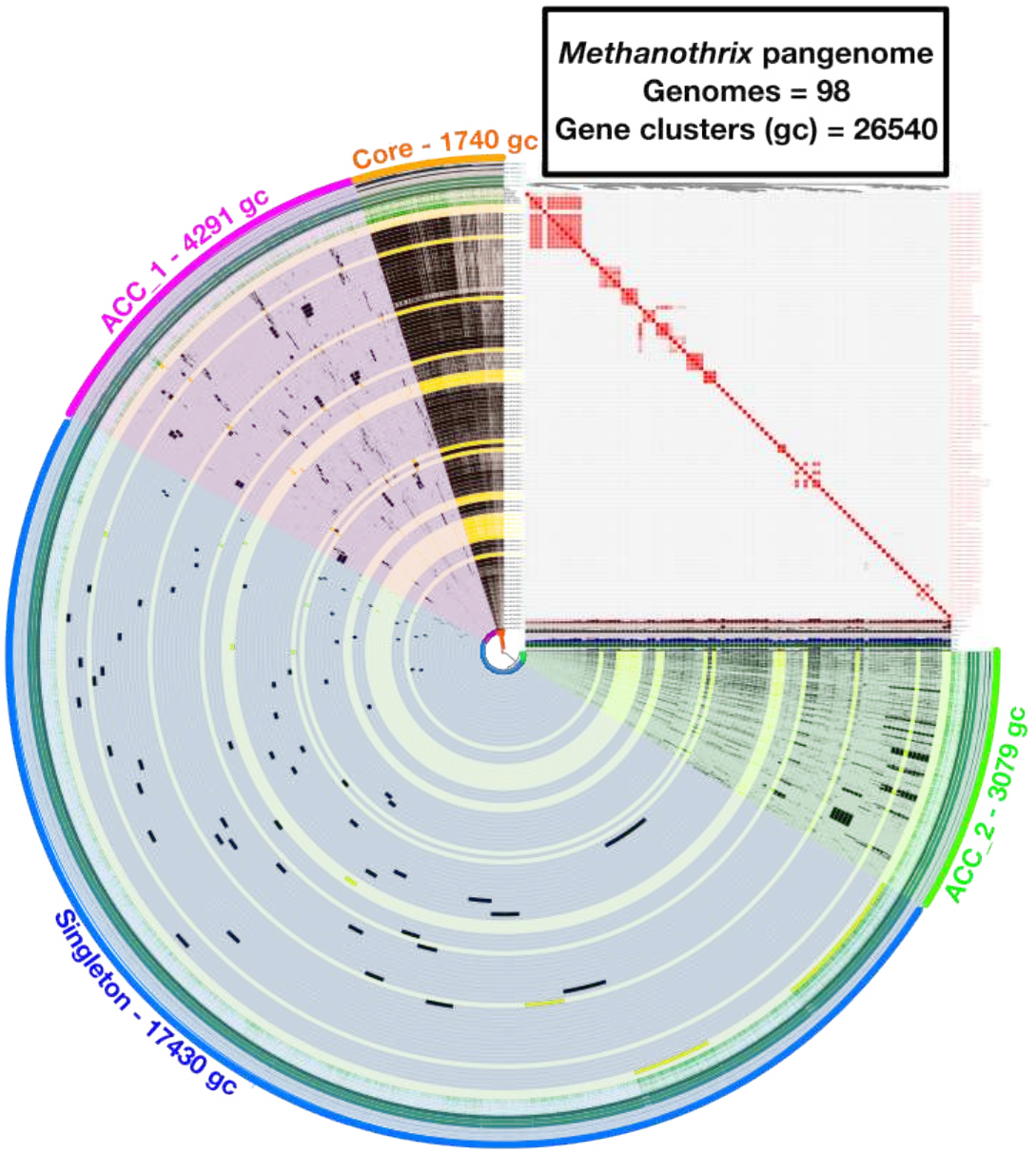


Figure 6.1: The pangenome of genus *Methanothrix*. Each gene cluster contains at least one gene contributed by at least one genome. The pangenome is organised by gene cluster frequency, and gene clusters are divided into 4 groups based on their presence/absence patterns. The core gene cluster contains 1,740 gene clusters which all genomes within the pangenome possess and share. Conversely, the two accessory gene clusters (ACC_1 & ACC_2), which contain 4,291 and 3,079 gene clusters, pertain to sequences unique to a subset of strains within the pangenome. In comparison, singletons (17,430 gene clusters) are genomic sequences exclusive to individual strains. The arrangement of items follows a presence-absence criterion, utilising D (Euclidean distance) and L (Ward's method) for clustering.

pathway lacking membrane-bound F_{420} hydrogenase [Frh] and F_{420}-non-reducing hydrogenase [Mvh]) were identified in the core set. These findings underscore the uniform presence of crucial methanogenic pathways within the *Methanothrix* lineage.

Essential genes for the reductive hexulose phosphate (RPH) pathway, implicated in carbon fixation for methanogenesis [24, 116], were also encoded in the core gene clusters. Enrichment analysis revealed that the non-oxidative phase converting fructose 6-phosphate (F6P) to ribulose 5-phosphate (Ru5P) was notably enriched in the MAGs derived from this study (enrichment score: 8.489, *p*-value: 0.004). It is worth noting that D-arabino-3-hexulose-6-phosphate (Hu6P), an uncommon intermediate typically associated with methanogenic archaea [371], lies between F6P and Ru5P. Hu6P can be converted into formaldehyde, subsequently forming methylene-tetra-hydro-methanopterin, which enters the methanogenesis pathway.

To evaluate the potential for EET, rule-based methods (Section 3.6 were used to screen for known EET-related elements. Neither multiheme cytochrome *c* nor electrically conductive pili were identified in the *Methanothrix* pangenome, aligning with previous findings that *Methanothrix* employs mechanisms distinct from well-characterised EET strategies [53, 69, 372]. Furthermore, no genes associated with biofilm formation were detected. This absence suggests that, within a bioreactor setting, the creation of biofilm matrices—often enriched in redox mediators that facilitate EET—may rely on other microbial community members rather than *Methanothrix* itself.

The pangenome of *Methanothrix* comprised a relatively small core but a substantially larger variable component, indicating great genetic variability. However, the specificity and ecological relevance of this variability, such as potential adaptation to environmental conditions, remain to be further investigated. Core gene clusters encode central metabolic pathways, including acetoclastic methanogenesis, hydrogen-independent CO_2 reduction, and RHP-linked carbon metabolism. While no multiheme cytochrome *c*, conductive pili, or biofilm formation genes were identified, these results are consistent with alternative EET mechanisms in *Methanothrix*.

6.3.2 Prediction of major sheath protein in *Methanothrix* MAGs

The HMM search did not yield any results, aligning with the current understanding that MspA (Meth_1069) lacks homologs in the *Methanothrix* MAGs constructed in this study. Conversely, 263 MspA-like genes were predicted using rule-based amyloid protein prediction tools, selecting genes with an aromaticity higher than 9.8%. These MspA-like genes were further verified using eggnog-mapper 2.0 to ensure they were not previously annotated. To confirm their activity in DIET-occurring environments, the genes were mapped to three metatranscriptomes sequenced from the GAC sample of the SAF-MBRs (details in Chapter 3). Among these, 33 genes were highly transcribed (mean: 15.70 $\pm$ 1.48 $\log_2$ RPKM) across the three metatranscriptomes (Fig. 6.2), indicating their active role in the SAF-MBR.

These highly expressed MspA-like genes were present in all four gene clusters. Specifically, GC_00001177, GC_00001249, and GC_00000231 from the core gene cluster were identified in over 60 genomes from various environments and were prevalent in the SAF-MBR group. While internal homologs existed within the *Methanothrix* MAGs, only nine sequences formed four groups

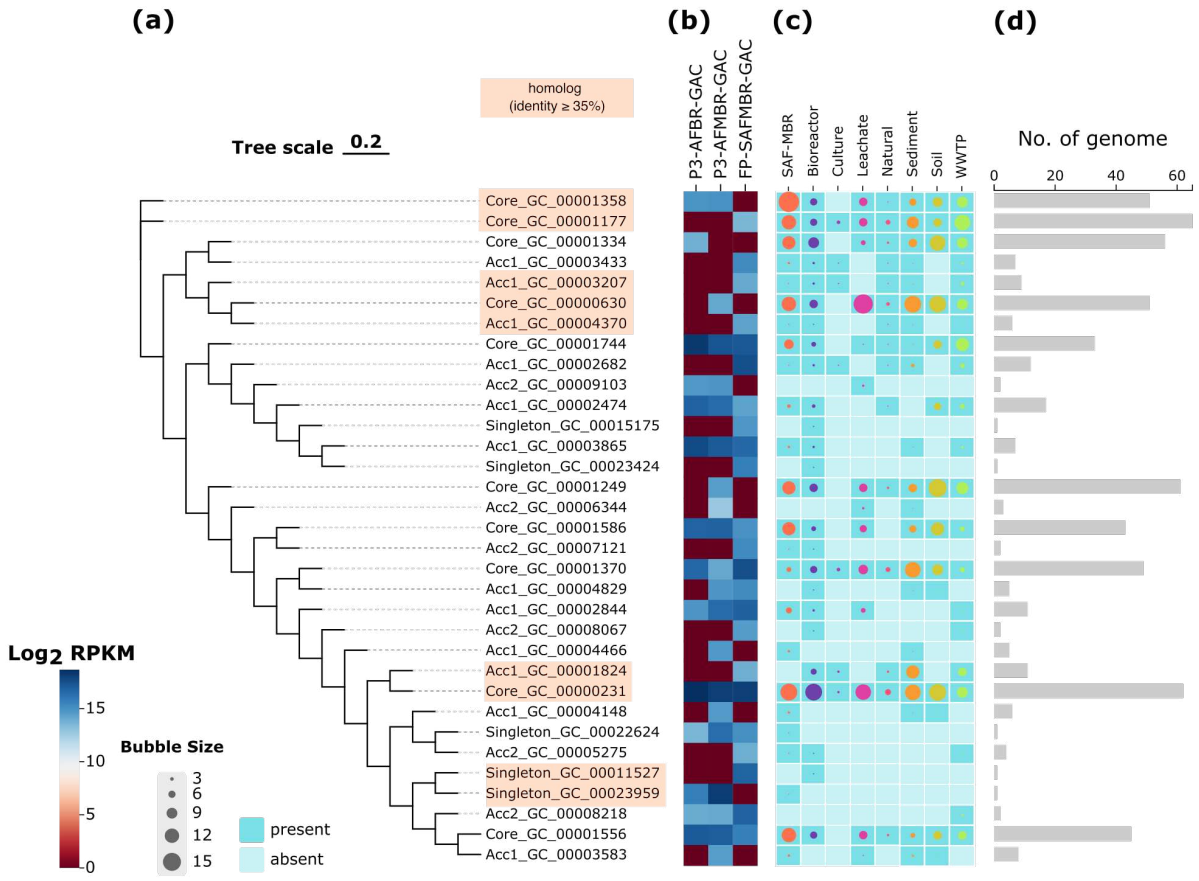


Figure 6.2: Phylogenetic analysis and distribution of *MspA*-like protein sequences. (a) Phylogenetic tree of *MspA*-like protein sequences, where internal homologs (percentage identity $>$ 35%) are color-labeled in orange. (b) Expression levels ($\log_2$RPKM) of *MspA*-like genes in the metatranscriptomes of three bioreactor GAC samples. (c) Distribution of the MAGs by origin environment, with bubble size representing the number of genomes. (d) Total number of genomes encoding the corresponding gene.

of homologs, contrasting with *Methanospirillum*, where several species shared highly similar MspA homologs [172].

The amyloidogenic regions of the transcribed MspA-like genes were regularly spaced (Fig. C.15), similar to the MspA (WP_011449234.1) identified in *Methanospirillum hungatei* JF-1, a distant homolog to Meth_1069. Motif analysis using MEME revealed the only common motif among all MspA-like and MspA genes in *Mycolicibacterium smegmatis*, *M. hungatei* JF-1, and *Methanothrix thermophila* PT is `MKAILLLLILLLLALAALA`, representing the N-terminal region of a signal peptide (Fig. 6.3). Notably, the motif from *Methanothrix* MAGs was more similar to *M. smegmatis* (e-value $\leq$10e-10) than to *M. hungatei* JF-1 (e-value $\geq$10e-6) (Fig. 6.4).

To validate the significance of the predicted MspA-like genes in the SAF-MBR, which serves as a proxy for DIET conditions in the bioreactor, the presence of these 33 genes was examined across eight *Methanothrix* MAGs reconstructed from the same sample as the metatranscriptomes. Genes in more than one genome were all from the core gene clusters, occasional genes from two

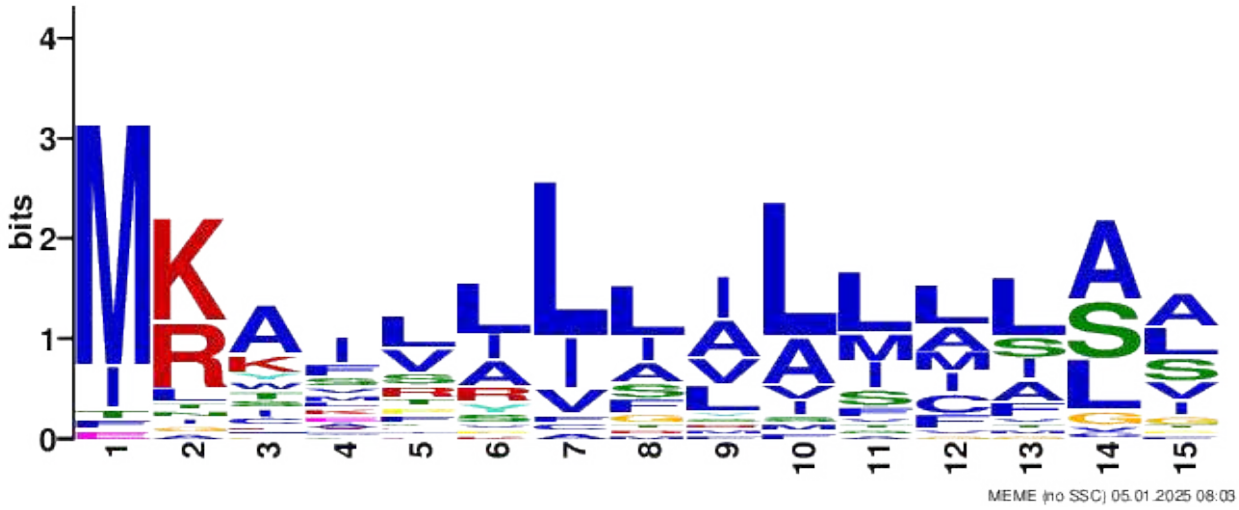


Figure 6.3: N-terminal motif encoded in MspA-like genes, annotated using MEME

accessory gene clusters, and none from the singleton gene cluster (Fig. 6.5). Furthermore, genes in more than half of the MAGs were compared to the median gene expression of each MAG. The MspA-like genes were expressed by 8 to 438 folds, highlighting their significance in total gene expression (Fig. 6.6).

To evaluate whether the aromatic residue placement in MspA-like genes allows electron hopping for EET, the 3D structures of six MspA-like genes were predicted using AlphaFold3 [367]. Among these, only GC_00000231 exhibited a clear β-sheet structure (Fig. 6.7), with inter-residue aromatic gaps ranging from 3.80 to 19.56 Å, spanning the entire protein. This structural arrangement closely resembles known MspA proteins found in archaea (Fig. 6.8).

Despite AlphaFold3's capabilities, discrepancies in predicted secondary structures were observed. Some MspA-like sequences expected to form β-sheets did not exhibit this feature. Several factors may account for these inconsistencies. AlphaFold3 struggles with proteins containing significant intrinsically disordered regions (IDRs), leading to unreliable folds [373]. Predictive accuracy also declines when homologous structures are absent in the database, as AlphaFold relies heavily on evolutionary conservation [374]. Additionally, errors such as sequencing gaps, frameshifts, or misannotations in metagenome-derived sequences can result in inaccurate secondary structure predictions [374].

6.4 Discussion

6.4.1 The EET pathway in *Methanothrix*

This study identified a highly expressed MspA-like sheath protein in *Methanothrix*, supported by 3D structural analysis indicative of its potential role in electron hopping. The proposed EET pathway in *Methanothrix* involves a coordinated series of redox-active components spanning from the extracellular sheath to the cytoplasmic membrane. The key elements of this pathway include the

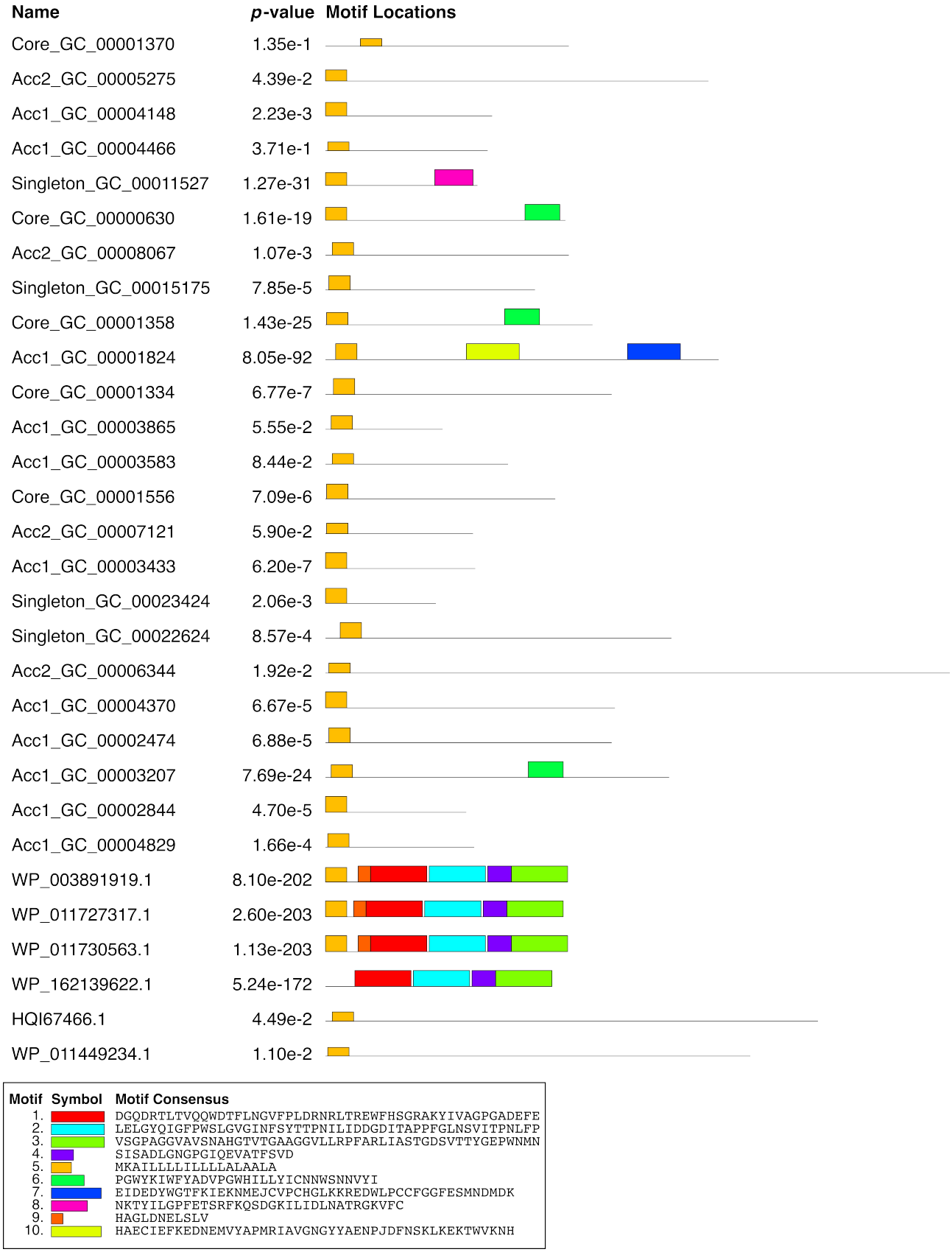


Figure 6.4: Predicted amyloidogenic regions in transcribed mspA-like genes from *Methanothrix* MAGs.

Group	Gene Cluster	inha_p3_afbr_gac_117_A	inha_p3_afbr_gac_29_B	inha_p3_afbr_gac_7_B	inha_p3_afmbr_gac_261_A	inha_p3_afmbr_gac_26_B	inha_p3_afmbr_gac_54_B	sf_fp_safmbr_gac_124_A	sf_fp_safmbr_gac_208_B	% of MAG
Acc1	GC_00001824	0	0	0	0	0	0	0	0	
Acc1	GC_00002474	0	0	0	0	0	0	1	0	1/9
Acc1	GC_00002682	0	0	0	0	0	0	0	1	1/9
Acc1	GC_00002844	0	0	0	0	0	0	0	0	
Acc1	GC_00003207	0	0	0	0	0	0	0	1	1/9
Acc1	GC_00003433	0	0	0	0	0	0	0	1	1/9
Acc1	GC_00003583	0	0	1	0	0	0	0	0	1/9
Acc1	GC_00003865	0	0	0	0	0	0	0	0	
Acc1	GC_00004148	0	0	1	0	0	0	0	0	1/9
Acc1	GC_00004370	0	0	0	0	0	0	0	1	1/9
Acc1	GC_00004466	0	0	1	0	0	0	0	0	1/9
Acc1	GC_00004829	0	0	0	0	0	0	0	0	
Acc2	GC_00005275	0	0	0	0	0	0	0	1	1/9
Acc2	GC_00006344	0	0	0	0	0	0	0	0	
Acc2	GC_00007121	0	0	0	0	0	0	0	1	1/9
Acc2	GC_00008067	0	0	0	0	0	0	0	0	
Acc2	GC_00008218	0	0	0	0	0	0	0	0	
Acc2	GC_00009103	0	0	0	0	0	0	0	0	
Core	GC_00000231	1	0	1	0	1	1	1	0	5/9
Core	GC_00000630	1	0	0	0	0	1	1	0	1/3
Core	GC_00001177	1	1	0	1	0	0	1	1	5/9
Core	GC_00001249	1	1	1	1	1	0	0	0	5/9
Core	GC_00001334	0	0	1	0	0	1	1	0	1/3
Core	GC_00001358	1	1	1	1	0	1	1	0	2/3
Core	GC_00001370	0	0	1	0	0	0	0	1	2/9
Core	GC_00001556	1	1	0	1	1	1	0	0	5/9
Core	GC_00001586	0	1	1	1	0	0	1	0	4/9
Core	GC_00001744	0	0	1	0	0	1	1	0	1/3
Singleton	GC_00011527	0	0	0	0	0	0	0	0	
Singleton	GC_00015175	0	0	0	0	0	0	0	0	
Singleton	GC_00022624	0	0	0	0	0	0	0	0	
Singleton	GC_00023424	0	0	0	0	0	0	0	0	
Singleton	GC_00023959	0	0	0	0	0	0	0	0	

Figure 6.5: Heatmap showing the presence (1) or absence (0) of gene clusters across *Methanothrix* MAGs reconstructed from this study's metagenome.

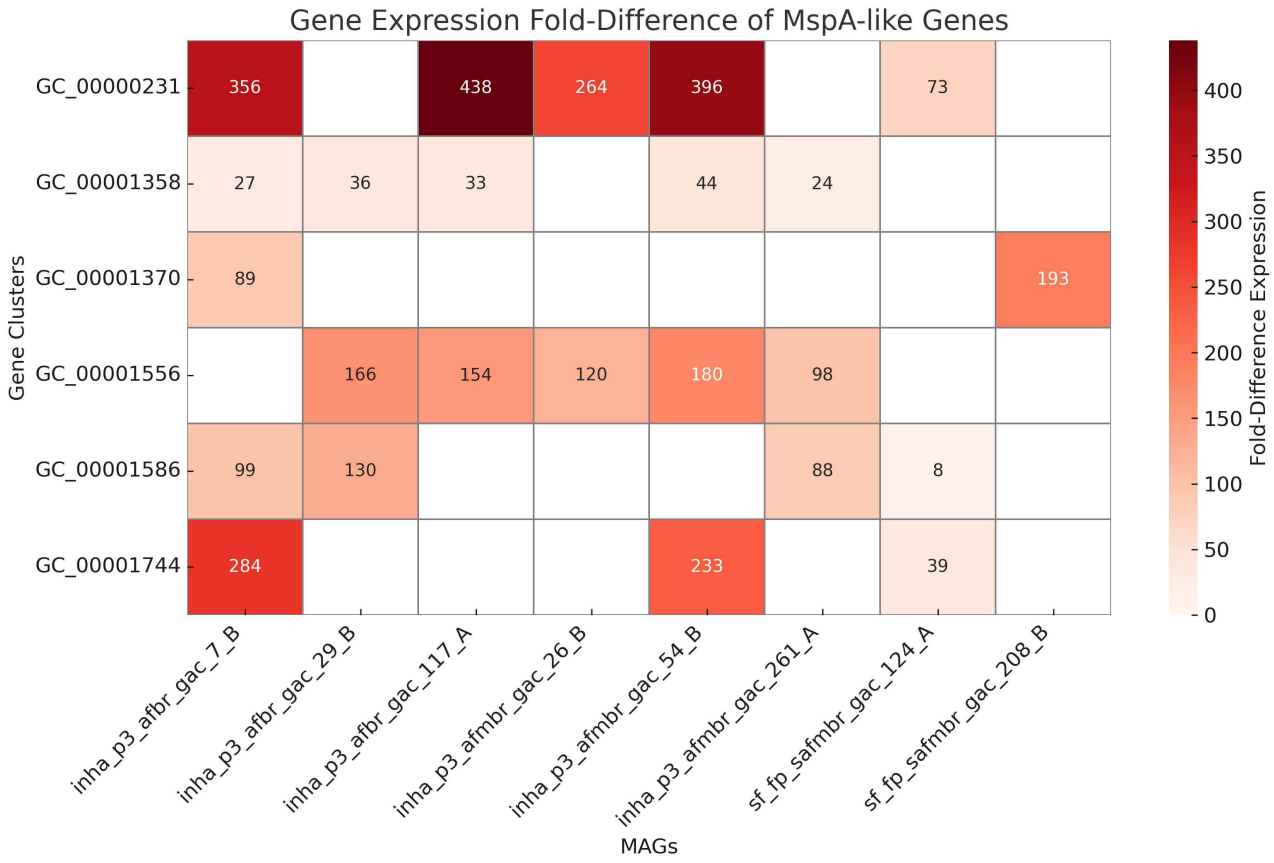


Figure 6.6: Heatmap of fold-difference in MspA-like gene expression relative to the median RPKM across *Methanothrix* MAGs from this study's metagenomes. Fold-differences were calculated within the same sample by comparing metatranscriptome data to its corresponding MAG.

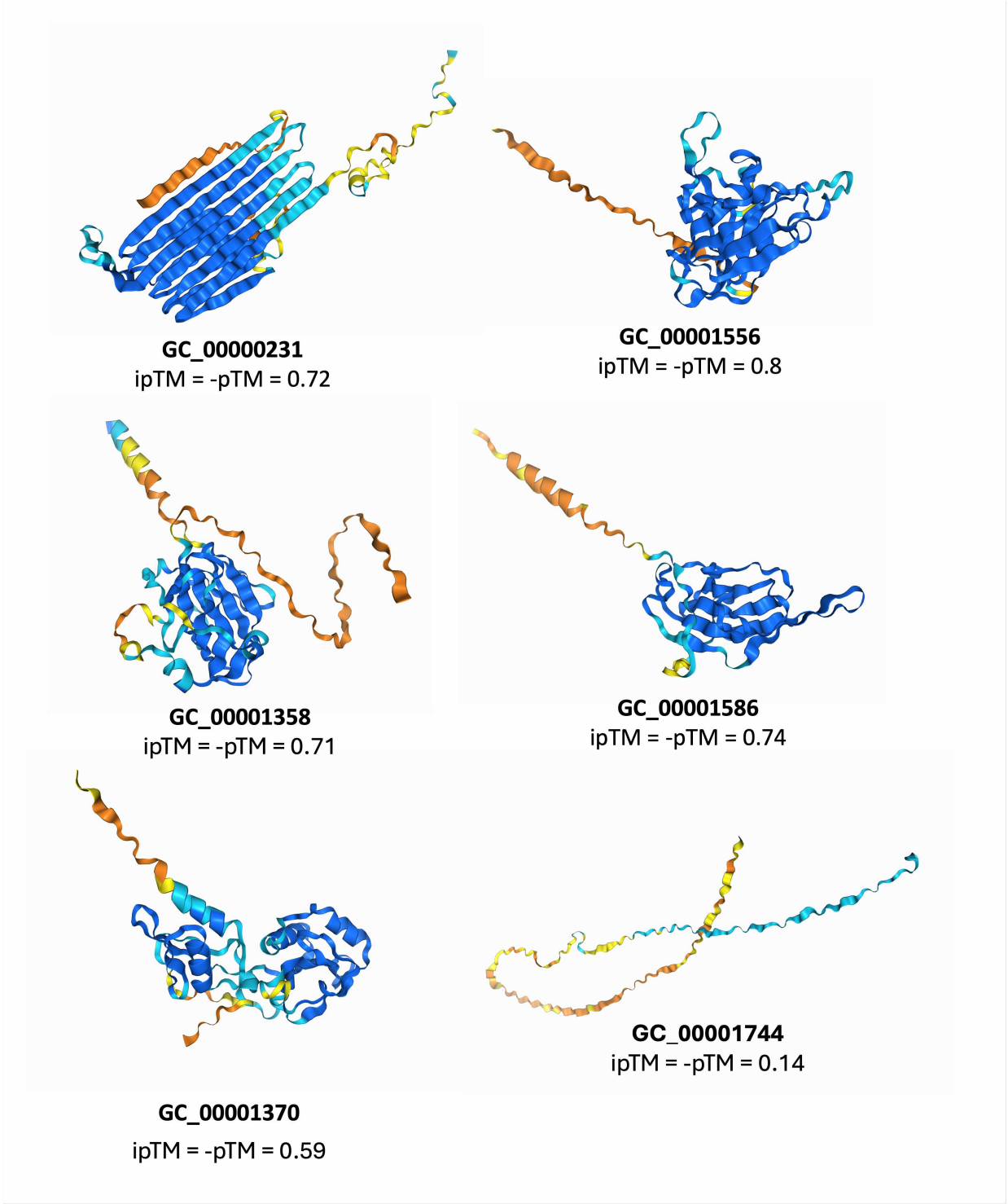


Figure 6.7: 3D protein structure prediction on MspA-like genes recovered from the MAGs with significant expression in the metatranscriptomes. Interface accuracy was assessed using the interface predicted template modeling score (ipTM). Prediction was done by using AlphaFold 3.

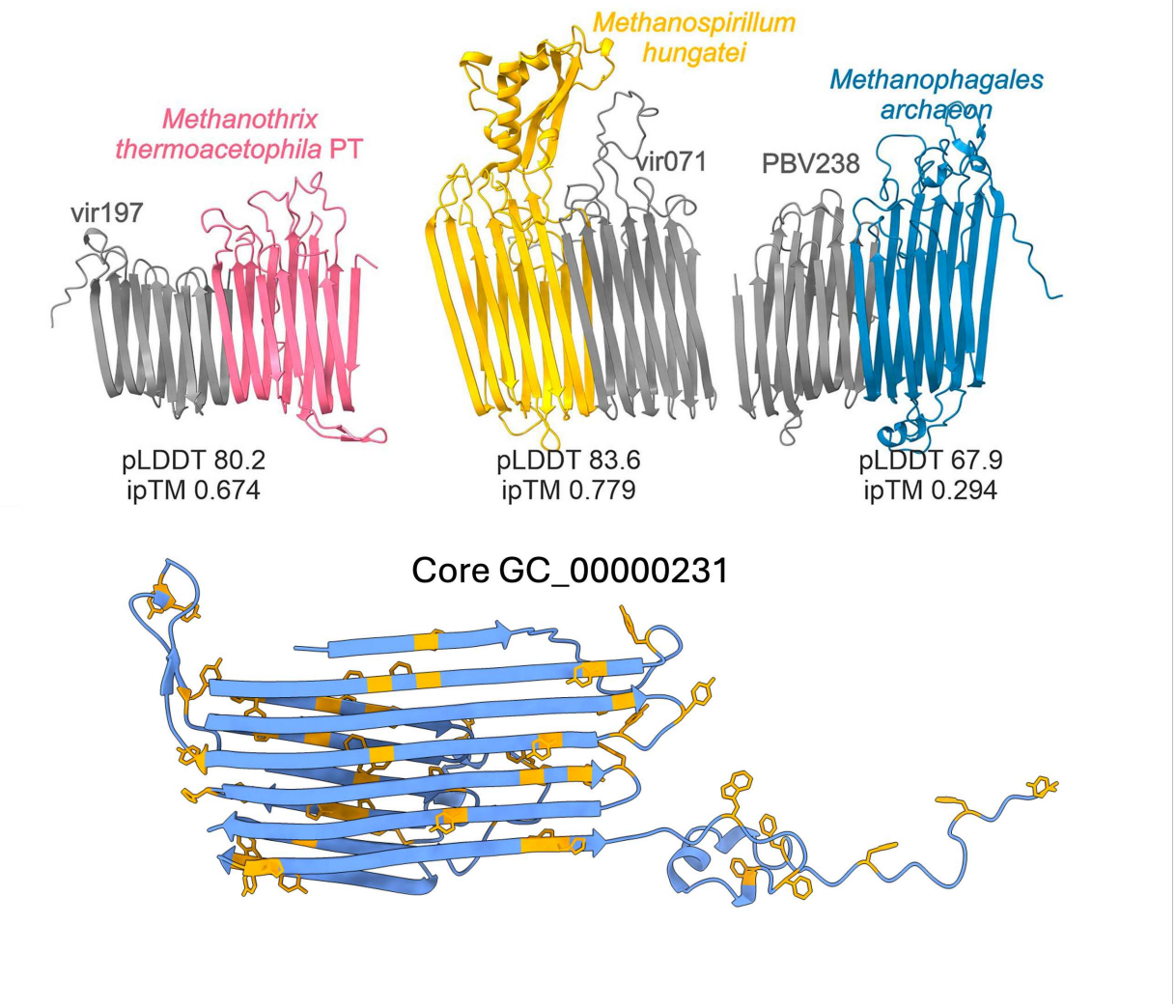


Figure 6.8: Structural comparison of MspA-like proteins. The top row displays previously annotated MspA proteins [41], predicted using AlphaFold 2. These models represent interactions between viral SH-like proteins (Vir197, Vir071, and PBV238) and their corresponding host SH proteins from *Methanotrichales*, *Methanomicrobiales*, and *Syntropharchaeia*. Model confidence is indicated by the predicted local distance difference test (pLDDT), and interface accuracy is assessed using the ipTM. The bottom row shows the GC_00000231, which is identified from the core gene cluster in this study, predicted using AlphaFold 3. This MspA-like protein structurally resembled the Aromatic residues, which are highlighted in orange, with only residues having a nearest neighbour within 20 Ådisplayed.

MspA-like sheath protein, PQQ-binding β-propeller proteins, Sqp, and the membrane-associated MP shuttle (Fig. 6.9).

The MspA-like major sheath protein forms a tubular, amyloid-based matrix that encapsulates the filamentous cells. Structural modelling and amino acid composition indicate high β-sheet content and periodic aromatic residues capable of π–π stacking, supporting a mechanism of long-range electron hopping along the sheath [171]. This conductive property allows electrons derived from extracellular partners or redox-active materials to migrate through the sheath to the cell envelope. The repeating arrangement of aromatic side chains and possible coordination with metal ions such as Fe^{2+}/Fe^{3+} may further facilitate charge delocalisation and redox coupling within the sheath network [375].

At the interface between the sheath and the cell surface, electrons are proposed to enter a PQQ-dependent redox complex comprising the surface quinoprotein (Sqp) and associated PQQ-binding β-propeller proteins. These proteins, predicted to be located in the outer cell envelope, contain highly conserved residues for PQQ binding and form an electron gateway between the sheath and membrane-bound carriers. The PQQ cofactor undergoes reversible two-electron cycling (PQQ $\leftrightarrow$ $PQQH_2$) [128, 376], thereby accepting electrons from the sheath and transferring them to a lipid-soluble electron carrier such as methanophenazine [116] [128, 376].

Methanophenazine (MP) is embedded within the archaeal cytoplasmic membrane and functions analogously to quinones in bacterial electron transport chains [112]. Reduced methanophenazine (MPH_2) donates electrons to the heterodisulfide reductase (Hdr) complex, which catalyses the reduction of the CoM-S-S-CoB heterodisulfide to coenzyme M (CoM-SH) and coenzyme B (CoB-SH). This redox step is essential for the final reduction of methyl-coenzyme M to CH_4 catalysed by methyl-coenzyme M reductase (Mcr). The EET pathway thus couples extracellular electron uptake to the methanogenic core, allowing *Methanothrix* to conserve energy from external electron donors while maintaining redox balance through membrane-associated complexes such as $F_{420}H_2$:methanophenazine dehydrogenase (Fpo) and Hdr [112, 377].

Pangenomic analysis revealed that the MspA-like proteins were conserved within the core genome of *Methanothrix* species, suggesting its fundamental role in cellular physiology. Its consistently high transcriptional levels in the SAF-MBR systems further indicate that MspA is actively involved in EET processes associated with CH_4 production. By forming conductive networks within biofilms, MspA likely mediated DIET, enabling efficient electron flow between syntrophic partners and the methanogenic cells.

6.4.2 Comparing to other methanogens

Unlike *Methanosarcina*, which can utilise both acetate and CO_2 as carbon sources, *Methanothrix* is obligately aceticlastic under normal conditions, primarily utilising acetate. However, the detection of genes associated with CO_2 reduction to CH_4 suggests that Methanothrix may also be capable of reducing CO_2 via alternative electron donors other than hydrogen, indicative of a potential EET-linked methanogenic pathway.

Comparing *Methanothrix* with other methanogens, such as *Methanosarcina* and *Methanospir-*

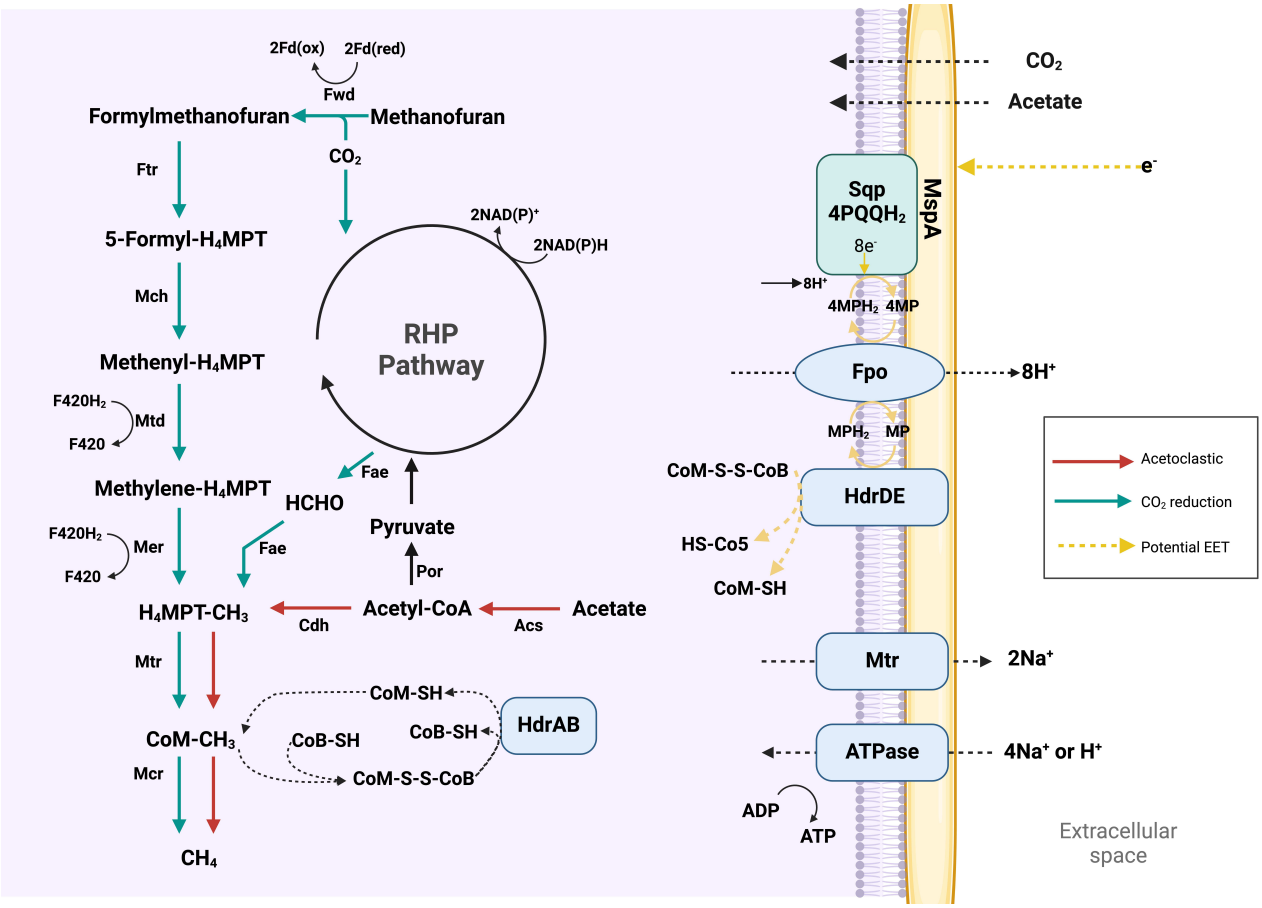


Figure 6.9: Proposed *Methanothrix* EET methanogenesis pathway. Arrows are coloured according to each methanogenesis pathway. Acs, acetyl-CoA synthase; CoB-SH, coenzyme B; Cdh, Acetyl-CoA decarbonylase/synthase complex; CoMSH, coenzyme M; Fae, formaldehyde activating enzyme; Fd, ferredoxin; Fpo, $F_{420}H_2$ dehydrogenase; Ftr, formylmethanofuran–H_4MPT formyltransferase; Fwd, formylmethanofuran dehydrogenase; H4MPT, tetrahydromethanopterin; HCHO, formaldehyde; Hdr, heterodisulfide reductase; mch, methenyl-H_4MPT cyclohydrolase; Mcr, methyl-coenzyme M reductase; Mer, F_{420}-dependent methylene-H_4MPT reductase; MP, methanophenazine; MspA, major sheath protein; Mtr, Na^+-translocating methyl-H_4MPT-coenzyme-M methyltransferase; PQQ, pyrroloquinoline quinone; RHP, reductive hexulose phosphate; Sqp, surface quinoprotein. The image was created using BioRender.

illum, reveals both similarities and distinct differences in EET mechanisms. While *Methanosarcina* employs multiheme c-type cytochromes like MmcA for electron transfer [47], *Methanothrix* relies predominantly on amyloid sheath proteins and PQQ-binding proteins, lacking homologs for cytochromes found in *Methanosarcina*. This divergence suggests evolutionary adaptations tailored to their specific ecological niches and metabolic requirements.

6.4.3 Role of structural proteins in EET

The tubular proteinaceous sheath in *Methanothrix* is crucial for maintaining cell chain architecture and facilitating efficient EET. The amyloid nature of MspA, characterised by stacked β-sheet structures, supports a hole-hopping mechanism where electrons transfer via closely stacked tyrosine residues [174]. This mechanism enables long-distance electron transport without the need for redox-active metal cofactors.

Proton-coupled electron transfer (PCET) is enhanced by proton acceptors, such as glutamine residues, and structural stabilisers like zinc ions, which maintain tyrosine stacking and hydrogen-bonding networks [121, 174]. These features ensure efficient electron transfer from the sheath to intracellular shuttles like MP, sustaining methanogenesis under varying environmental conditions.

6.4.4 Metabolic requirements and flexibility

Despite engaging in EET-mediated methanogenesis, acetate remains essential for *Methanothrix* growth. Metatranscriptomic studies indicate that while *Methanothrix* can reduce CO_2 during DIET, acetate is indispensable, particularly when utilising external electron donors such as Fe(0) or stainless steel [378]. The inability of *Methanothrix soehngenii* GT6 to respond to electrical stimulation with CO_2 as the sole carbon source emphasises its obligate aceticlastic nature[1], contrasting with *Methanosarcina*, which can utilise CO_2 directly [111].

This specialisation limits *Methanothrix*' metabolic flexibility but ensures efficient utilisation and high affinity for acetate in wastewater treatment systems with low acetate concentration effluent [32]. Future experiments employing ^{13}C-labelled acetate and CO_2 could elucidate the precise metabolic fluxes and confirm the extent of CO_2 incorporation into biomass and CH_4 under EET conditions.

6.4.5 Implications for biotechnological applications

The identification of MspA and other EET-related proteins in *Methanothrix* offers potential enhancements for bioelectrochemical systems[2] and wastewater treatment processes. By studying the conductive properties of amyloid sheath proteins, *Methanothrix* strains with superior EET ca-

[1] During the early stage of this PhD study, *Methanothrix soehngenii* GT6 did not grow using CO_2 as the sole carbon source.

[2] A bioelectrochemical system (BES) is an engineered reactor that couples microbial metabolism with electrochemical processes, enabling electron transfer between microorganisms and electrodes to achieve energy conversion or wastewater treatment.

pacity, identified through selective enrichment or developed via genetic engineering, could exhibit enhanced CH_4 production efficiency through improved electron transfer dynamics. Additionally, monitoring MspA gene expression using quantitative PCR (qPCR) enables real-time tracking of transcriptional responses during experiments or bioreactor operations, facilitating better control and optimisation of the methanogenic process.

Insights into the hole-hopping mechanism and PCET in amyloid proteins can inspire the design of bio-inspired conductive materials. These materials may find applications in bioelectronics, where efficient, metal-free electron transporters are desirable, based on the structural and functional properties of MspA and related proteins. Moreover, considering the role of Methanothrix in metal oxidation, including stainless steel, the potential implications of EET on bio-corrosion should also be factored into reactor design. Given the effectiveness of GAC in facilitating EET/DIET, exploring alternative materials that incorporate GAC, such as a hypothetical GAC-embedded concrete wall, could be an innovative approach to promoting treatment efficiency via EET in a more sustainable manner while also addressing corrosion-related challenges.

6.4.6 Critical evaluation of current models

Current models of EET in methanogens, primarily based on *Methanosarcina* and *Methanospirillum*, emphasise the role of cytochromes and multiheme proteins (Section 2.2.2. The findings in this study in *Methanothrix* suggest that amyloid sheath proteins and PQQ-binding proteins may also contribute to EET in the absence of cytochromes. This indicates that methanogens may utilise multiple strategies for EET, with mechanisms adapted to their specific metabolic pathways and environmental conditions.

While these observations do not fundamentally alter existing EET models, they provide additional insights into the diversity of electron transfer mechanisms among methanogenic archaea. This study contributes to the current understanding by presenting evidence for alternative EET pathways, highlighting the possible involvement of structural and redox-active proteins beyond cytochromes. Further research is necessary to fully integrate these findings into broader EET frameworks, which could enhance the applicability of EET models across a more diverse array of methanogenic species.

Chapter 7

Horizontal Gene Transfer of Antibiotic Resistance Gene in SAF-MBRs

7.1 Introduction

This study originally aimed to investigate HGT of EET genes, such as MHC and e-pili, within SAF-MBRs. These bioreactors rely on biofilm-based microbial communities that grow on GAC, thereby creating high-density, metabolically active environments. Given these characteristics, SAFMBRs were hypothesised to facilitate HGT via mechanisms such as conjugation or phage-mediated transduction [379]. The anticipated mobility of EET genes was of particular interest due to their roles in interspecies electron transfer, syntrophic metabolism, and biogeochemical cycling under anaerobic conditions.

However, genome-resolved metagenomic analysis using `BlastP` and `HGTector 2` did not yield strong evidence for the horizontal transfer of EET-related genes [280, 285]. Unexpectedly, the analysis revealed mobile genetic elements, phages and plasmids, carrying ARGs, particularly *sul4*, a gene conferring resistance to sulfamethoxazole (SMX). Given the established mobility of *sul* genes and their association with anthropogenic pollution, the research scope was expanded to characterise the prevalence and potential transfer mechanisms of ARGs alongside EET-related genes.

Antibiotic resistance is a globally recognised threat, particularly within wastewater treatment systems where diverse microbial populations coexist and frequently exchange genetic material [379, 380]. Among the resistance determinants identified in wastewater, *sul* genes (e.g., *sul1–4*) are of particular concern due to their widespread occurrence and role in conferring resistance to SMX, a commonly used sulfonamide antibiotic [379]. The potential for ARG dissemination via HGT in wastewater treatment bioreactors raises significant public health concerns, especially when antibiotics or their metabolites persist at low concentrations and exert selective pressure on microbial communities [280, 285].

Anaerobic bioreactors such as SAFMBRs offer promising solutions for micropollutant removal, including effective degradation of pharmaceuticals like SMX, atenolol, carbamazepine, and NDMA [381]. In these systems, GAC functions not only as a physical adsorbent but also as a scaffold

for microbial biofilms capable of pharmaceutical biodegradation [382, 383]. Previous batch experiments using GAC sourced from full-scale SAFMBRs demonstrated rapid ($>$99%) removal of SMX, attributed primarily to biodegradation rather than sorption [383]. These findings raise important questions about how such systems influence microbial community dynamics and gene mobility, particularly in the context of ARGs.

Compared to conventional wastewater treatment systems (e.g., activated sludge), SAFMBRs exhibited lower biomass concentrations but greater phage and plasmid diversity [379, 384]. This unique ecological configuration may support distinct co-evolutionary interactions between phages, plasmids, and their microbial hosts. Mobile elements in these environments often carry AMGs in addition to ARGs, potentially influencing both community structure and functional capacity. The discovery of *sul4* within mobile genetic elements underscores the need for a systematic exploration of how microbial ecology and reactor design contribute to ARG dissemination.

Accordingly, this chapter aims to elucidate the prevalence and distribution of ARGs, with a particular emphasis on *sul4*, within SAFMBR systems. In parallel, it investigates whether horizontal gene transfer events involving ARGs occur under the distinct physicochemical and ecological conditions created by granular activated carbon-supported biofilms in these reactors.

7.2 Materials and Methods

7.2.1 Global horizontal gene transfer event identification

To detect HGT events associated with EET genes, protein sequences from the MAGs were analysed using HGTector 2 [280]. These protein sequences were queried against the NCBI nr protein database via BLASTp with an E-value threshold of 1×10^{-5}, a minimum sequence identity of 30%, and a minimum alignment coverage of 50%.

Viral contigs (length $\geq$ 3,000 bp) were identified with VirSorter2 [385] and subsequently annotated using DRAM-v [245] to identify auxiliary metabolic gene. The annotated sequences were aligned against the core dataset from the Virulence Factor Database (VFDB 2022), retaining only those hits with identity $>$35% and query coverage $>$70% [386].

7.2.2 Plasmid and phage detection

Eleven DNA samples were collected from SAF-MBRs for plasmid and phage detection. Metagenomic sequences were pre-processed and assembled as described in Section 3.2.1. Contigs longer than 3,000 bp were used for phage detection, while those longer than 1,000 bp were evaluated for plasmid content.

For plasmid identification, assembled contigs were first aligned to the PLSDB reference database [387]. Contigs that did not map to PLSDB were further analysed with PLASMe [189], employing order-specific Transformer models.

For phage identification, PhaBOX was utilised to detect phage contigs in the genome collection, integrating multiple established methods: PhaMer for identification [190], PhaTYP for lifestyle

prediction [281], PhaGCN for taxonomy [282], and CHERRY for host prediction [283]. This combination enables efficient phage detection, lifestyle classification (lysogenic/lytic), taxonomic assignment, and host prediction. The analysis applied a 1,000 bp minimum contig length, with identity and coverage thresholds of 0.9 for accuracy.

7.2.3 Antibiotic resistance gene (ARG) detection

Antibiotic resistance genes present in MAGs, plasmids, and phages were identified using the Resistance Gene Identifier (RGI) against the Comprehensive Antibiotic Resistance Database (CARD) [388]. Alignments were performed with Diamond [233], and ARGs were categorised as "Perfect", "Strict", or "Discovery".

BLASTp searches were conducted with a strict identity cutoff of 99% [285] to investigate the potential HGT of ARGs. Candidate conjugative HGT (microbe-microbe) was inferred by comparing ARG-containing MAGs in an all-vs.-all manner. Vertical gene transfer was considered likely where ARGs were conserved within MAGs sharing $\geq 95\%$ ANI; such ARGs were excluded from further HGT analyses.

A total of 38 references *sul* genes were downloaded from the NCBI. Multiple sequence alignments were performed with T-Coffee [238], trimmed using trimAl (`automated1` option) [235], and phylogenetic analysis carried out with RAxML-NG [240] under the VT+G4 model, which achieved the best fit (weight score: 0.8625).

7.2.4 Quantifying ARGs

The overall abundance of ARGs within the metagenomes was assessed using ARG-OAP v3 [389]. This workflow normalises ARG counts to 16S rRNA gene copies (n16S) and total cell estimates (nCell). The n16S normalisation accounts for differences in microbial load while incorporating nCell, which provides an additional layer of accuracy by reflecting the actual bacterial density. As a result, ARG-OAP v3 offers a robust means of comparing ARG abundance across complex samples and evaluating the extent of resistance gene prevalence.

7.2.5 Plasmid and phage abundance

Abundances of MAGs and phages were quantified using CoverM [307] in RPKM mode, thereby normalising read counts by gene length and library size. For each sample, RPKM values for all contigs assigned to a given MAG or phage were summed to yield a single abundance estimate per taxon.

7.3 Results and Discussion

7.3.1 Overview

HGTector 2 analysis revealed no significant HGT signals in the EET-related genes. Only a limited number of carbon metabolism-related AMG were detected using the BlastP approach (21 genes in total), comprising 6 for glycolysis, 5 for the pentose phosphate pathway, 5 for the TCA cycle, and 5 for the acetyl-CoA pathway (see the full table in the GitHub repository). However, no HGT events were found in these genes. Although previous studies report AMGs in wastewater treatment plants, the essential verification step for HGT is often omitted [390, 391]. As a result, although such AMGs might be historically linked to transfer events, there is no direct evidence of HGT during bioreactor operation.

Figure 7.1 presents the quantification of ARGs normalised to both 16S rRNA gene copy number and cell counts. The mean value of ARG/16S was found to be 0.28 ± 0.083, while ARG/Cell was 0.49 ± 0.273. One advantage of using the ARG-OAP pipeline for ARG abundance is its consistent internal normalisation, facilitating cross-study comparisons. Indeed, analyses of large-scale wastewater treatment plants suggest ARG abundances can vary widely, from 0.41 to 2.82 ARG copies per 16S rRNA gene [391, 392]. By contrast, the ARG abundance in the SAF-MBR system in this study is at the lower end of this spectrum, indicating relatively reduced ARG prevalence across both bulk and GAC samples.

7.3.2 ARG distribution across samples

The distribution of ARGs in this study aligns with previous investigations of reclaimed water [384] and anaerobic sludge [393], where aminoglycoside, beta-lactam, tetracycline, sulfonamide, and streptogramin resistance genes typically dominate (Figure 7.2). Their prevalence is consistent with a human waste origin [394], a primary component in domestic wastewater streams. When considered compositionally, ARG profiles differed between bulk and GAC-associated communities. Bulk samples exhibited a more even distribution across resistance classes, with a higher proportional contribution from acriflavine resistance, whereas GAC-associated communities were dominated by erythromycin and rifampin resistance. These compositional patterns were conserved across operational phases, indicating that physical niche rather than operational regime was the primary determinant of ARG composition.

7.3.3 HGT of sulfonamide resistance genes

A total of 41 *sul* genes were identified (Figure 7.3). Of these, 14 belonged to the *sul1*, *sul2*, and *sul3* group; all observed HGT events in this group were plasmid-mediated. In contrast, 28 *sul4* genes were detected, with phage-mediated transfer observed in 22 of these instances. Notably, *sul4* showed a higher relative abundance under conditions that stimulate EET: the EET+/EET− ratio reached 0.867 for *sul4*, compared to 0.4 for *sul1-3*. Likewise, the GAC/Bulk ratio was 3.67 for

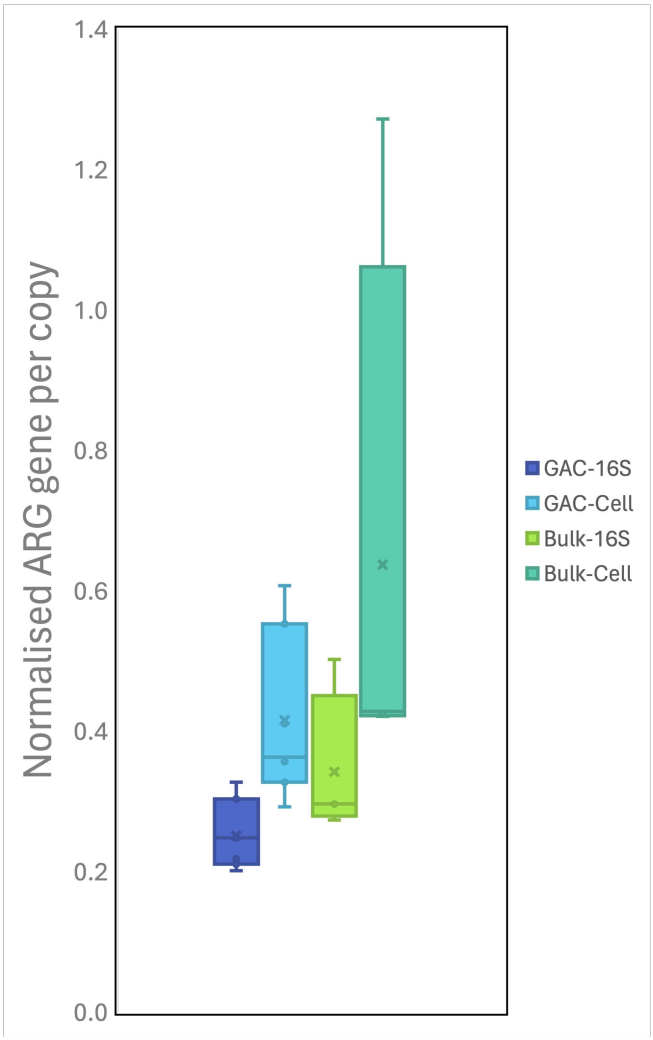


Figure 7.1: ARG quantification grouped by the sample type.

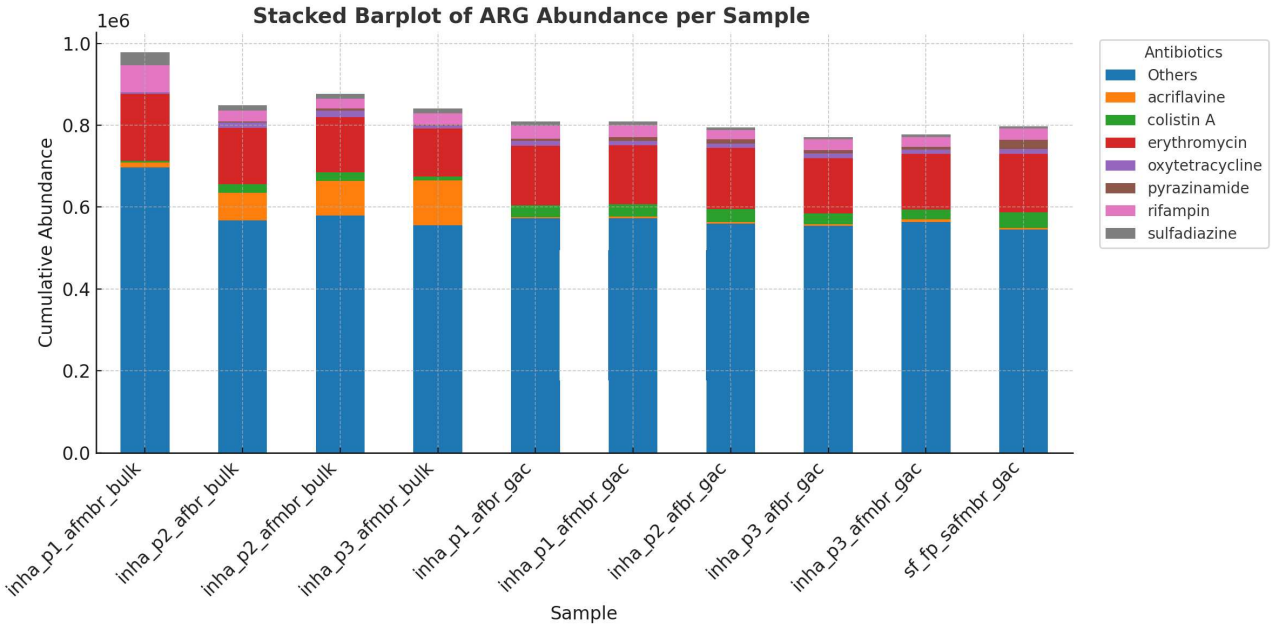


Figure 7.2: A stacked barplot of drug class abundance across different samples.

sul4 versus 1.33 for *sul1-3*.

These findings suggest that *sul1*, *sul2*, and *sul3* are primarily disseminated via plasmid-mediated conjugative HGT, often linked to mobile genetic elements such as class 1 integrons, transposons, and insertion sequences [395–397]. In contrast, *sul4* is associated with bacteriophages, indicating transduction as its main route of transfer [398]. Its enrichment under EET-stimulating conditions may reflect increased phage activity or a selective advantage for *sul4*-carrying hosts. Unlike *sul1–3*, which are tied to resistance integrons, *sul4* is linked to folate metabolism genes, suggesting additional selective pressures. A higher GAC/Bulk ratio for *sul4* further supports the influence of environmental factors such as GAC on phage-mediated gene transfer.

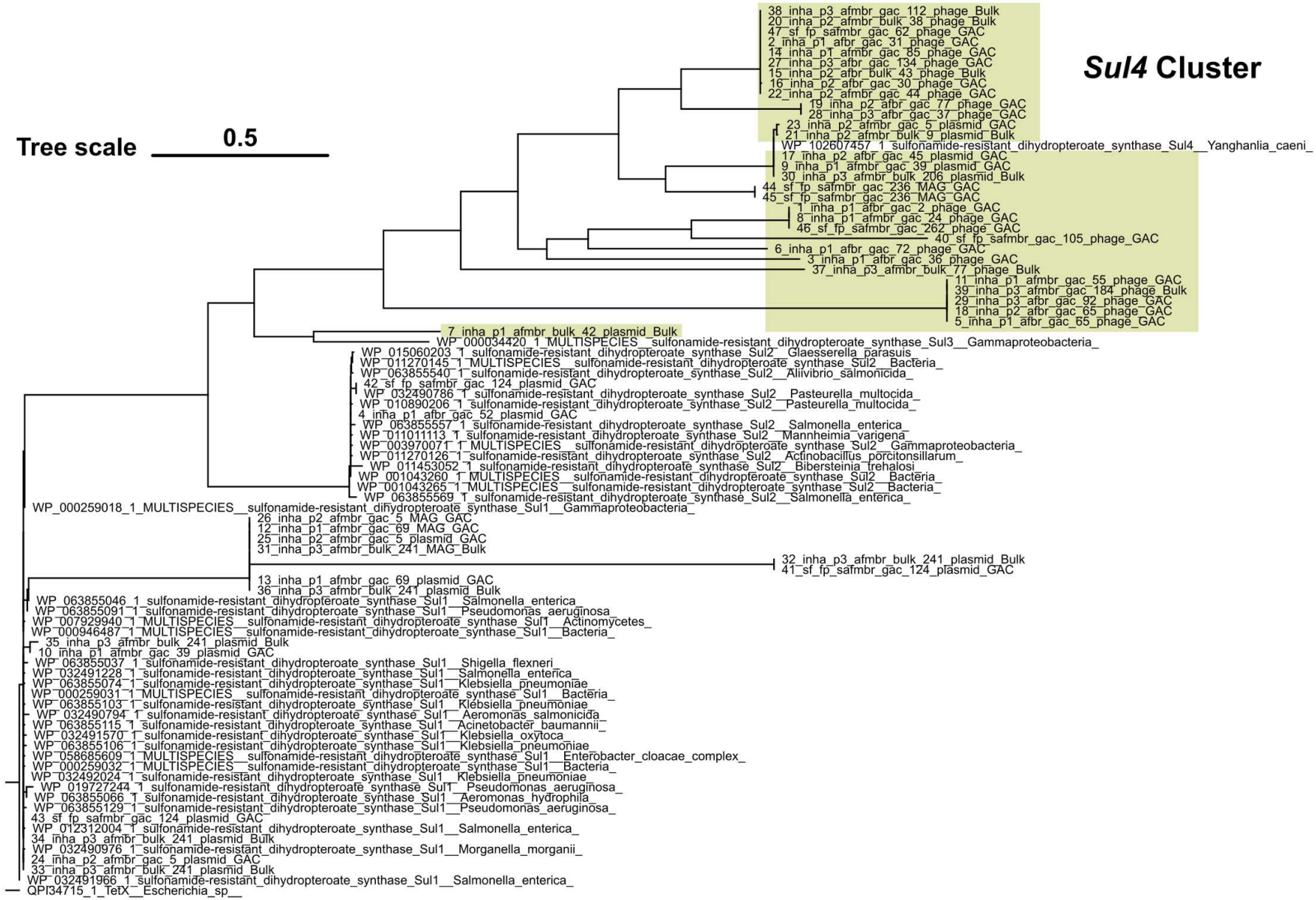
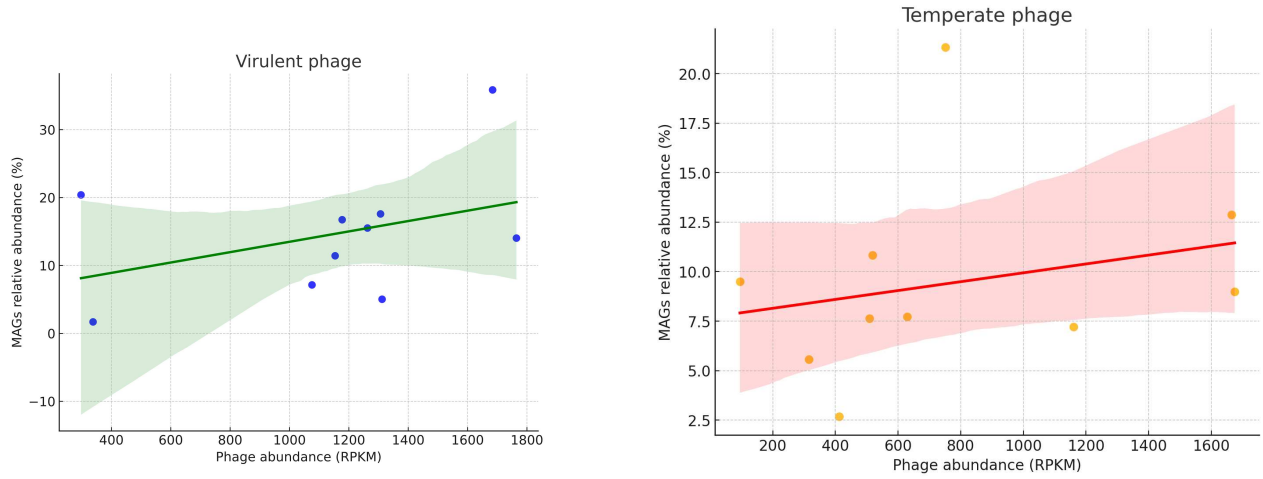


Figure 7.3: Phylogenetic distribution and putative transfer mechanisms of *sul* genes. *sul4* clusters are highlighted in yellow.

7.3.4 Phage-host interactions in the SAF-MBR

Previous research suggests that temperate phages often correlate positively with their host MAGs, potentially driving ARG transfer, whereas virulent phages typically exhibit negative correlations due to their lytic lifestyle [399]. However, in the SAF-MBR system, no pronounced difference was detected between the abundance patterns of temperate and virulent phages relative to their host MAGs (Figure 7.4). This uniformity is surprising given that temperate phages typically play a more prominent role in ARG dissemination, while virulent phages frequently induce episodic, host-killing outbreaks [399, 400].



(a) Virulent phage plot

(b) Temperate phage trend

Figure 7.4: Correlation between phage and MAG abundance in the SAF-MBR.

Despite global studies indicating that plasmids are the dominant mediators of ARG transfer (often surpassing phage-mediated transfer) [401], the SAF-MBR presented a contrasting trend. Here, bacteriophages contributed to HGT events approximately three times more frequently than plasmids (Figure 7.5). Other regional investigations have similarly shown that phage-mediated transfer is less common in conventional wastewater treatment plants [402–404], suggesting that the SAF-MBR might provide environmental conditions that favour phage-related HGT, potentially due to the structured and densely populated nature of the biofilm environment.

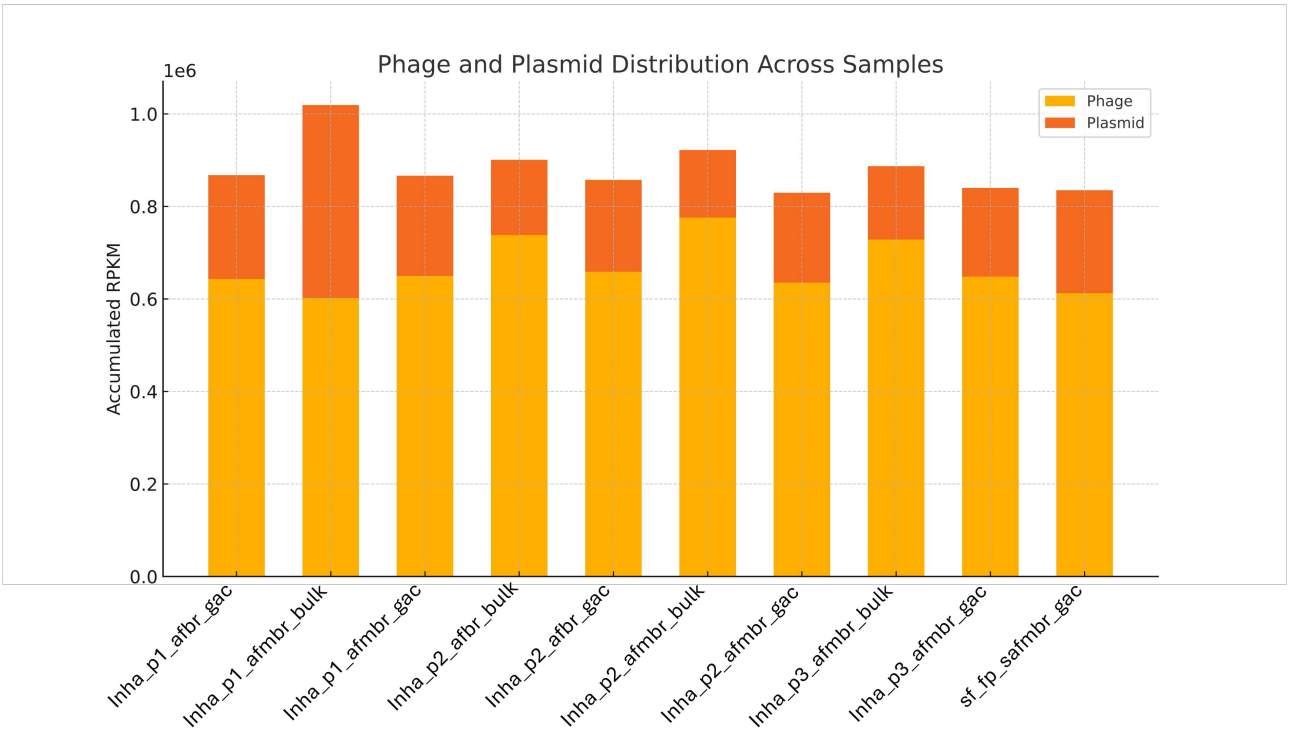


Figure 7.5: Stacked bar plot illustrating the distribution of phage- and plasmid-mediated HGT events across different samples.

Chapter 8

General Discussion

8.1 Genomic evidence for extracellular electron transfer in engineered anaerobic biofilms

Genomic analyses across SAF-MBR metagenomes demonstrate that genetic capacity for EET is widespread in engineered anaerobic biofilms, extending beyond a small number of canonical ”electrogenic” taxa. Instead, EET-associated genes were detected across multiple functional guilds, including fermentative bacteria, syntrophic oxidisers, and methanogens, indicating that extracellular redox interactions may represent a distributed metabolic strategy within the anaerobic food web.

Previous studies of EET and DIET in anaerobic digestion have primarily focused on enrichment of well-characterised model organisms, particularly *Geobacter* species and acetoclastic methanogens, under laboratory conditions or synthetic feedstocks [14, 22, 69]. In such systems, EET is often inferred from targeted enrichment experiments, conductive material amendments, or electrochemical measurements. By contrast, the present study directly infers EET-associated gene inventories from complex pilot-scale communities treating low-strength municipal wastewater, without reliance on enrichment or pure-culture assumptions.

At the level of dereplicated MAGs, genes encoding MHC, aromatic-rich e-pili candidates, and other redox-associated extracellular proteins were detected across a broad taxonomic spectrum. This pattern is consistent with emerging metagenomic surveys of anaerobic digesters and sedimentary environments, which report that EET-related genes are not restricted to classical exoelectrogens but occur in diverse lineages with distinct metabolic roles [44, 161]. However, most prior studies report EET potential at coarse taxonomic resolution or as gene presence/absence at the community level, limiting mechanistic interpretation in spatially structured systems.

A key distinction of the present work is the integration of EET gene inventories with microhabitat-resolved genomic context. Rather than treating the reactor as a homogeneous community, EET-associated genes were analysed in relation to carrier-associated and bulk biomass fractions. This revealed systematic enrichment of EET proxy genes within biofilm-associated MAGs, supporting the interpretation that EET capacity is selectively retained in attached-growth niches where

spatial proximity and prolonged residence times favour extracellular electron exchange. Similar enrichment of redox-active traits in biofilm-associated populations has been reported in laboratory-scale bioelectrochemical systems and conductive-particle-amended digesters [16, 70], but has rarely been demonstrated at pilot scale using genome-resolved approaches.

Importantly, the presence of EET-associated genes in this study does not imply that a single electron transfer mechanism dominates across all taxa or conditions. Multihaem cytochromes and pilin-like proteins are multifunctional and have been implicated in mineral respiration, biofilm formation, redox homeostasis, and surface adhesion in addition to interspecies electron transfer [135]. The genomic evidence, therefore, constrains the space of plausible electron transfer strategies rather than providing direct proof of DIET flux. Nevertheless, the consistent co-occurrence of EET proxies across fermentative, syntrophic, and methanogenic populations supports a network-level interpretation in which extracellular redox interfaces may contribute to stabilising electron disposal across multiple steps of anaerobic carbon conversion.

8.2 Microhabitat-structured organisation of EET gene inventories

Genome-resolved analyses revealed that EET-associated gene inventories were not evenly distributed across the SAF-MBR microbiome, but instead showed clear structuring between bulk liquid and carrier-associated biomass. EET proxy genes, including MHC and aromatic-rich e-pili, were consistently enriched within MAGs associated with biofilm fractions relative to those dominant in the suspended community. This pattern indicates that genetic capacity for extracellular redox interactions is selectively retained in attached-growth niches.

Most studies investigating EET or DIET in anaerobic digestion infer functional organisation at the reactor scale, implicitly assuming a well-mixed microbial community. Under this assumption, enrichment of conductive materials or DIET-capable taxa is interpreted as a bulk community response [14, 22]. However, biofilm-based systems inherently generate spatially distinct microhabitats, characterised by differences in residence time, substrate availability, redox gradients, and physical proximity between interacting organisms. The present results demonstrate that these microhabitats exert strong selective pressure on the distribution of EET-associated genes, a pressure that cannot be resolved by bulk community profiling alone.

Previous work has reported preferential colonisation of conductive particles by syntrophic bacteria and methanogens, often inferred from taxonomic enrichment or microscopy [16, 70]. While these studies support the conceptual link between conductive surfaces and DIET, they typically lack genome-resolved evidence connecting spatial organisation to specific functional gene inventories. By resolving EET proxy genes at the MAG level and explicitly distinguishing between biofilm-associated and suspended populations, this study provides direct genomic support for microhabitat-driven organisation of extracellular redox capacity in an engineered anaerobic system.

The enrichment of EET-associated genes within carrier-associated MAGs is mechanistically consistent with the physicochemical constraints of interspecies electron transfer. Direct or surface-

mediated electron exchange requires short intercellular distances and sustained physical association [44], both of which are favoured within biofilms but are transient in the bulk liquid phase. In addition, attached-growth niches allow slow-growing syntrophs and methanogens to persist under short hydraulic retention times, increasing the likelihood that energetically marginal pathways, including DIET-compatible interactions, are maintained within the community.

Importantly, this microhabitat structuring complicates simplistic interpretations of EET based on either taxonomy or community-averaged gene abundance. The presence of EET-associated genes at low relative abundance in bulk samples does not necessarily indicate limited functional relevance. Still, it may instead reflect spatial segregation of redox-active populations within biofilms. Conversely, bulk enrichment of taxa previously linked to DIET does not guarantee that those organisms participate in EET if they are not retained within appropriate physical niches. These observations reinforce the need for spatially resolved genomic interpretation when evaluating EET in complex engineered microbiomes.

8.3 Limits of taxonomic inference and the necessity of gene-centric approaches

A recurring outcome of this study is that taxonomic identity alone provides an incomplete and, in some cases, misleading basis for inferring EET capability in complex anaerobic biofilms. Across SAF-MBR metagenomes, a substantial proportion of EET-positive MAGs were affiliated with uncultured, poorly resolved, or placeholder taxa under current genome taxonomies. Even among classified genera, EET-associated gene inventories varied markedly between closely related MAGs, indicating that functional capacity cannot be reliably inferred from taxonomic assignment alone.

Much of the existing EET and DIET literature relies on taxonomic enrichment as a proxy for functional interpretation. In laboratory-scale digesters and enrichment cultures, increases in the relative abundance of taxa such as *Geobacter*, *Syntrophomonas*, or *Methanothrix* are frequently interpreted as evidence for DIET or enhanced extracellular redox coupling [22, 372, 405]. While this approach is defensible in simplified systems dominated by a small number of well-characterised lineages, it becomes increasingly problematic in engineered biofilms where community composition is diverse, spatially structured, and dominated by uncultured taxa.

The present genome-resolved analysis demonstrates that EET-associated genes are distributed heterogeneously within and across taxonomic groups. In several cases, MAGs assigned to the same genus differed in both the presence and composition of EET proxy genes, including MHC and e-pili candidates. Conversely, EET-associated gene inventories were detected in lineages not previously highlighted in DIET-focused studies, particularly among uncultured bacterial groups enriched in carrier-associated biofilms. These findings are consistent with recent metagenomic surveys of anaerobic environments, which suggest that EET-related functions are evolutionarily widespread and not confined to a small set of model organisms [161].

These observations underscore a fundamental limitation of taxonomy-led inference: genus-

or family-level labels obscure functional heterogeneity that is critical for interpreting metabolic interactions such as EET. In addition, reliance on taxonomic proxies risks circular reasoning, whereby organisms are assumed to participate in DIET because they resemble taxa previously implicated in DIET, rather than because they encode relevant molecular machinery. This limitation is amplified in pilot-scale systems, where operational heterogeneity and long-term selection favour functional convergence across phylogenetically distinct lineages.

In contrast, gene-centric approaches provide a more direct and falsifiable basis for evaluating EET potential. By explicitly defining rule-based criteria for EET+ classification based on curated inventories of redox-active extracellular proteins, this study avoids conflating taxonomic presence with functional capability. Importantly, this framework does not assume that any single gene or gene family is diagnostic of DIET. Instead, it constrains interpretation by identifying trait combinations consistent with extracellular redox interactions, while remaining agnostic about the specific electron transfer mechanism operating in situ.

The necessity of gene-centric inference is particularly evident in spatially structured systems such as SAF-MBRs, where functional roles are shaped by microhabitat-specific selection rather than by phylogenetic relatedness alone. As shown in the previous section, EET-associated genes are preferentially retained within carrier-associated biofilms, even when the same taxonomic groups are present in the bulk phase. Without genome-resolved, gene-level analysis, such patterns would remain obscured, leading to underestimation of the ecological significance of EET in engineered anaerobic microbiomes.

8.4 EET-associated functions beyond canonical methanogenesis

EET in anaerobic systems is frequently framed in the literature as a mechanism that primarily enhances methanogenesis through DIET between syntrophic bacteria and methanogens. While this interpretation is well supported by enrichment studies and pure-culture experiments [22, 44], genome-resolved evidence from complex engineered biofilms suggests that the functional roles of EET-associated genes extend beyond the terminal step of CH_4 formation.

In the SAF-MBR microbiomes analysed here, EET proxy genes were not confined to methanogens but were also prevalent among fermentative and syntrophic bacterial populations. This distribution contrasts with a methanogen-centric view of EET. It is consistent with observations from sedimentary and wastewater-associated metagenomes, where redox-active extracellular proteins have been detected in taxa spanning multiple trophic levels [161]. Such patterns indicate that EET-associated functions may support upstream metabolic processes, including fermentation and syntrophic oxidation, by facilitating electron disposal under energetically constrained conditions.

Several studies have reported that conductive materials can accelerate degradation of complex substrates and volatile fatty acids even in the absence of marked increases in methanogen abundance [16, 70]. These observations have often been interpreted indirectly as evidence of DIET-driven methanogenesis. However, the genomic patterns observed in this study suggest an

alternative or complementary explanation: extracellular redox interactions may stabilise upstream carbon transformation by alleviating redox bottlenecks in fermentative and syntrophic steps, thereby indirectly supporting methanogenesis without requiring a direct increase in methanogenic activity.

The presence of MHC and e-pili in fermentative bacteria suggests that these organisms engage in surface-associated redox interactions that facilitate the efficient release of electrons during polymer breakdown. Similarly, syntrophic bacteria encoding EET-associated genes may rely on EET to maintain low intracellular redox potential during oxidation of fatty acids such as propionate and butyrate, processes that are thermodynamically marginal under standard anaerobic conditions [8, 44]. These interpretations align with theoretical and experimental work suggesting that alternative electron disposal pathways can expand the feasible operating window for syntrophic metabolism [33].

Notably, the multifunctionality of EET-associated genes cautions against over-attribution of specific pathways. MHC and e-pili have been implicated in a range of processes, including mineral respiration, biofilm formation, and redox homeostasis [135]. The genomic evidence presented here, therefore, does not imply that all detected EET proxies mediate DIET. Instead, it supports a broader conceptual model in which extracellular redox interfaces help maintain balanced electron flow across multiple metabolic steps in anaerobic biofilms.

By extending interpretation beyond canonical methanogenesis, this study aligns with emerging views that treat EET as a system-level property of anaerobic communities rather than a pathway restricted to specific organism pairs. In engineered biofilms treating low-strength wastewater, such distributed EET-associated functions may be particularly important for stabilising carbon conversion under short hydraulic retention times and fluctuating redox conditions, where classical diffusional electron carriers alone may be insufficient to sustain efficient metabolic coupling.

8.5 *Methanothrix* as a focal taxon: pangenomic evidence for putative surface-associated EET modules

Among methanogenic taxa recovered from the SAF-MBR microbiomes, *Methanothrix* warrants particular attention because of its recurring association with DIET-capable consortia and its dominance in low-acetate, attached-growth anaerobic systems. Several coculture and conductive-material studies have demonstrated that *Methanothrix* can participate in DIET-based syntrophic interactions, most notably with *Geobacter* species, resulting in enhanced CH_4 production and altered electron flow relative to classical H_2/formate-mediated syntrophy [14, 22]. However, despite this consistent ecological association, the molecular basis of extracellular electron uptake in *Methanothrix* remains poorly resolved.

To date, only a small number of studies have proposed specific protein candidates for EET in *Methanothrix*. In *Methanothrix thermophila*, a cell-surface protein termed MspA was identified as a potential electron uptake component based on its surface localisation and association with DIET-capable phenotypes [19, 116]. Subsequent analyses suggested that MspA may play a role in extracellular electron uptake, and one study further proposed that MspA could exhibit conductive

properties consistent with long-range electron transfer [174]. Importantly, MspA lacks close homologues in existing protein databases [171, 172], complicating both comparative genomics and functional inference.

This scarcity of characterised EET-related proteins in *Methanothrix* provided a key motivation for the analytical strategy adopted in this thesis. Rather than relying on homology-based annotation, which is ineffective for lineage-specific or rapidly diverged proteins, a structure-informed amyloid protein prediction approach was developed to identify candidate extracellular conductive components without assuming sequence similarity to known bacterial e-pili or cytochromes. This approach is particularly relevant for methanogens, whose cell envelopes, surface architectures, and redox biology differ fundamentally from those of well-studied bacterial exoelectrogens.

Applying this framework to *Methanothrix* MAGs recovered from the SAF-MBR revealed a conserved set of MspA-like sheath proteins across multiple genomes. These proteins exhibited sequence and predicted structural features consistent with extracellular amyloid assemblies, including regularly spaced aromatic residues compatible with electron hopping mechanisms reported for conductive amyloids in other systems. In addition, conserved secretion signals and extracellular localisation motifs supported their classification as surface-associated structures rather than intracellular or periplasmic proteins.

Beyond the sheath proteins, *Methanothrix* MAGs also encoded additional putative surface-associated redox components, including PQQ-dependent β-propeller proteins and Sqp proteins, suggesting that extracellular electron uptake may involve a multi-component module rather than a single conductive element. The repeated co-occurrence of these features across independently reconstructed *Methanothrix* genomes supports the interpretation that surface-associated redox interaction is a conserved trait within this genus under biofilm-associated conditions.

Crucially, the evidence presented here does not constitute direct proof that the identified sheath proteins mediate EET in situ. Amyloid structures are known to be multifunctional, contributing to biofilm cohesion, cell adhesion, and stress resistance in diverse microorganisms [406]. The interpretation advanced in this thesis is therefore explicitly hypothesis-driven: existing experimental studies suggest that MspA is involved in electron uptake and may be conductive, while the genome-resolved and structure-informed analyses presented here strengthen this hypothesis by demonstrating conservation, predicted conductive features, and association with other redox-active surface proteins across *Methanothrix* populations.

Taken together, these findings extend the limited existing literature on methanogen EET by providing a genome-resolved framework for interpreting surface-associated electron transfer in *Methanothrix*. Rather than proposing a definitive mechanism, this work clarifies why amyloid-based EET represents a plausible and testable model for methanogen electron uptake, and why homology-independent approaches are necessary to investigate EET in archaeal lineages that lack close bacterial analogues.

8.6 Distribution and transfer of EET-related and accessory genes in SAF-MBR microbiomes

Biofilm-dominated anaerobic systems are frequently hypothesised to promote HGT due to high cell density, prolonged physical contact, and elevated abundance of mobile genetic elements [285, 379]. In the context of EET, it has been suggested that conductive environments may facilitate the dissemination of redox-active traits across phylogenetically diverse lineages, particularly under strong selective pressure for syntrophic efficiency [44].

Genome-resolved analyses in this study provide limited support for recent HGT of canonical EET-associated genes. Across MAGs, EET proxy genes such as MHC-encoding loci and e-pili candidates were largely conserved within taxonomic lineages. They did not exhibit consistent signatures of recent horizontal acquisition based on sequence similarity, phylogenetic incongruence, or genomic context. This contrasts with expectations derived from some laboratory-scale enrichment studies, in which rapid functional shifts have been attributed to gene transfer or to plasmid-mediated dissemination of metabolic traits under strong selection [196].

Instead, the observed distribution of EET-associated genes is more parsimoniously explained by ecological selection acting on pre-existing functional diversity. In spatially structured biofilms, lineages already equipped with extracellular redox capabilities are preferentially retained. At the same time, taxa lacking such traits are less competitive under constraints imposed by short hydraulic retention times and syntrophic coupling. Similar conclusions have been drawn from metagenomic surveys of anaerobic digesters and sedimentary environments, which report that EET-related genes are evolutionarily widespread but are selectively enriched in specific environmental contexts rather than being frequently transferred [44, 161].

In contrast to canonical EET genes, clearer signatures of HGT were detected among accessory genes, particularly ARGs. Several ARGs were identified on contigs associated with mobile genetic elements, and enrichment of specific ARGs, including *sul4*, was observed within carrier-associated biomass. Biofilm-associated enrichment of ARGs has been widely reported in wastewater treatment systems and is often attributed to elevated phage activity and localised selection pressures [407, 408].

Notably, phage-associated transfer events were more prevalent than plasmid-associated events in this system. This pattern is consistent with growing evidence that bacteriophages play a dominant role in gene exchange within engineered microbial communities, particularly under stable, high-retention conditions [409]. Such phage-mediated transfer has been shown to disproportionately affect accessory functions, including antibiotic resistance, rather than core metabolic pathways.

Importantly, ARG distribution showed no systematic association with EET proxy gene inventories or with taxa inferred to participate in extracellular redox interactions. This decoupling indicates that while SAF-MBR biofilms support active mobile gene dynamics, these processes do not appear to underpin the distribution of EET-associated traits. Instead, EET-related functions and ARG dissemination represent parallel ecological processes shaped by the same physical environment

but governed by distinct selective pressures.

Taken together, these observations caution against assuming that biofilm-rich or conductive anaerobic systems necessarily promote horizontal dissemination of functionally critical metabolic traits such as EET. In the SAF-MBR microbiomes analysed here, extracellular redox capacity appears to be maintained primarily through lineage-specific retention and ecological selection. In contrast, accessory genes are more readily mobilised through phage-mediated transfer.

8.7 Implications for interpreting EET in engineered anaerobic systems

The results of this study highlight several critical considerations for interpreting EET in complex engineered anaerobic microbiomes. Much of the existing literature infers EET or DIET from system-level responses, such as improved CH_4 yield following conductive material addition or enrichment of taxa previously associated with electroactive phenotypes [14, 22, 70]. While such observations are valuable, they provide limited mechanistic resolution in spatially structured systems where functional roles are distributed across multiple guilds and microhabitats.

A key implication of the genome-resolved analyses presented here is that EET should be interpreted as a community-level redox capability rather than a discrete pathway restricted to specific organism pairs. The widespread yet non-random distribution of EET proxy genes across fermentative bacteria, syntrophs, and methanogens is consistent with an emerging view that extracellular redox interactions support multiple steps of anaerobic carbon transformation, including upstream processes that indirectly stabilise methanogenesis [69]. This contrasts with a strictly methanogen-centric framing of DIET and suggests that EET-related functions may contribute to system stability even when changes in methanogen abundance or CH_4 yield are modest.

The microhabitat structuring observed in this study further indicates that bulk community profiles can obscure functionally important patterns. In biofilm-dominated reactors, enrichment of EET-associated genes within carrier-associated biomass reflects selection for traits that benefit from prolonged physical association and short electron transfer distances. Similar spatial partitioning has been inferred in laboratory-scale systems using microscopy or taxonomic enrichment [16], but genome-resolved evidence linking spatial niches to specific functional gene inventories remains scarce. These results demonstrate that spatial context is a prerequisite for meaningful genomic interpretation of EET in engineered biofilms.

The limitations of taxonomy-led inference are particularly evident in this context. Many studies continue to infer EET capability based on the presence or enrichment of genera historically associated with DIET, such as *Geobacter* or *Methanothrix*. However, the high prevalence of uncultured taxa and the observed heterogeneity of EET gene inventories within taxonomic groups indicate that such inferences are increasingly unreliable in complex systems. Gene-centric approaches, particularly those that do not rely exclusively on homology to bacterial model organisms, provide a more robust framework for evaluating EET potential in archaeal-dominated and biofilm-associated communities.

The pangenomic evidence for putative surface-associated EET modules in *Methanothrix* illustrates the importance of this shift in perspective. Rather than confirming a specific mechanism, the analysis demonstrates how genome-resolved and structure-informed approaches can strengthen hypotheses where experimental evidence is limited and homologous reference sequences are scarce. This approach aligns with recent calls to expand EET research beyond a narrow set of bacterial exoelectrogens and to consider alternative molecular architectures that may mediate extracellular redox interactions in anaerobic archaea [69].

Finally, the contrasting patterns observed for EET-associated genes and accessory genes underscore the need to distinguish between metabolic function and mobile gene ecology. While biofilm-rich anaerobic systems can support active horizontal gene transfer, particularly via phage-mediated pathways [285, 379], the genomic evidence here suggests that core EET-related traits are maintained primarily through ecological selection rather than rapid horizontal dissemination. This distinction is important for interpreting both functional stability and ecological risk in engineered anaerobic systems.

Collectively, these findings support an interpretive framework in which EET in engineered anaerobic biofilms is best evaluated using genome-resolved, gene-centric analyses that explicitly account for spatial organisation, functional heterogeneity, and the multifunctional nature of redox-associated proteins. Such an approach provides a more nuanced and mechanistically grounded understanding of extracellular redox interactions than inference based solely on reactor performance metrics or taxonomic enrichment.

Chapter 9

Conclusions

The aim of this research was to determine how EET contributes to the efficiency and resilience of anaerobic digestion, including its roles beyond methanogenesis, within engineered methanogenic systems. This aim was addressed using staged anaerobic fluidised-bed membrane bioreactors operated at pilot scale, in combination with integrated metagenomic, metatranscriptomic, and pangenomic analyses.

Specifically, the thesis examined how conductive media influence carbon and electron flow across multiple functional guilds, how EET-mediated interactions shape syntrophic metabolism under low-sulphate conditions, and how methanogens, particularly *Methanothrix*, facilitate extracellular electron uptake at the molecular level. Together, these approaches enabled a gene-centric and system-level evaluation of EET as an organising principle in anaerobic wastewater treatment.

9.1 Key findings and contributions

9.1.1 EET gene diversity and system-level carbon conversion

This thesis demonstrates that EET-capable microbial communities are widespread and functionally significant in SAF-MBRs. By applying a refined rule-based gene-identification framework, this work substantially improved the specificity of MHC and electrically e-pili detection relative to alignment-based approaches, reducing false-positive annotations and establishing a more reliable basis for functional inference of EET capacity in complex anaerobic microbiomes.

EET-associated genomes were strongly enriched within GAC, associated biofilms and were closely linked to enhanced hydrolysis and acidogenesis kinetics. EET genes were distributed across multiple phyla, with particular enrichment in *Desulfobacterota* and *Bacteroidota*. A high prevalence of unannotated MHC genes highlighted substantial unexplored diversity in EET mechanisms in engineered anaerobic systems. Collectively, these findings establish EET as a contributor to the upstream, rate-limiting stages of anaerobic digestion, extending its functional relevance beyond methanogenesis.

9.1.2 DIET-driven syntrophy between sulfate-reducing bacteria and methanogens

This research provides evidence that, under low-sulfate conditions, SRB in pilot-scale SAF-MBRs can engage in DIET with methanogens rather than functioning solely as competitors for shared substrates. Integrated metagenomic and metatranscriptomic analyses revealed active expression of EET-associated genes in SRB, including putative e-pili, supporting their involvement in DIET-mediated syntrophic metabolism with *Methanothrix*.

The presence of GAC was shown to facilitate this interaction by promoting conductive biofilm formation and maintaining close spatial organisation between syntrophic partners. These results challenge conventional assumptions that SRB are inherently detrimental to methanogenesis and demonstrate that, under appropriate physicochemical conditions, SRB can contribute positively to CH_4 production and process stability through EET-enabled syntrophy.

9.1.3 Molecular basis of extracellular electron uptake in *Methanothrix*

This thesis identifies a highly expressed MspA-like sheath protein as a key structural component potentially underpinning extracellular electron uptake in *Methanothrix*. Pangenomic and structural analyses indicate that this protein possesses features consistent with long-range electron hopping, providing a plausible mechanism for EET in methanogens that lack canonical MHC.

These findings expand current understanding of methanogenic EET by demonstrating that proteinaceous sheaths may serve as electron transfer conduits. Comparative analyses across methanogenic lineages further revealed both conserved and lineage-specific EET strategies, underscoring the evolutionary diversity of electron transfer mechanisms within methanogenic archaea.

9.1.4 Constraints on EET gene transfer and accessory gene dynamics

Despite initial hypotheses, no conclusive evidence was found for recent horizontal gene transfer of canonical EET genes within the SAF-MBR microbiome, suggesting that EET capacity is shaped primarily by long-term evolutionary adaptation rather than rapid horizontal dissemination. In contrast, phage-mediated transfer of antibiotic resistance genes, particularly the sulfonamide resistance gene *sul4*, was clearly detected.

The predominance of phage-associated gene transfer, coupled with low overall ARG abundance, indicates that SAF-MBR operating conditions and GAC-associated biofilms create microbial and viral dynamics distinct from those of conventional wastewater treatment systems. These results highlight a decoupling between EET evolution and mobile genetic elements, while emphasising the broader ecological implications of specialised anaerobic reactor environments.

9.2 Implications and potential applications

The findings of this thesis have significant implications for the future development of anaerobic wastewater treatment for municipal sewage, particularly for low-strength wastewaters typical of temperate regions, where conventional anaerobic processes are often limited by weak thermodynamic driving forces and poor process stability.

This research demonstrates that EET provides a mechanistic basis for overcoming several of these long-standing constraints. By enabling efficient extracellular electron exchange within conductive biofilms, EET-capable microbial consortia enhance not only methanogenesis but also upstream, rate-limiting stages of anaerobic digestion, including hydrolysis and acidogenesis. Through stabilisation of intracellular redox balance and improved coupling between fermentative, syntrophic, and methanogenic metabolisms, EET alleviates thermodynamic bottlenecks that commonly restrict anaerobic treatment of low-strength wastewater.

At the microbial interaction level, the identification of DIET between sulfate-reducing bacteria and methanogens under low-sulfate conditions further revises conventional assumptions regarding competitive dynamics in anaerobic reactors. Rather than acting solely as competitors for shared substrates, sulfate-reducing bacteria were shown to participate in cooperative electron exchange, contributing positively to CH_4 production and overall process stability. This finding underscores the importance of community structure and electron-sharing interactions in determining functional robustness in anaerobic wastewater treatment systems.

The stabilising role of EET is further supported by quantitative thermodynamic–kinetic analysis [32]. Modelling of substrate half-saturation coefficients (K) indicated that the apparent K values for acetoclastic methanogenesis and propionate degradation in the SAF-MBR were one to two orders of magnitude lower than those reported for conventional anaerobic digestion. Specifically, K values of approximately 3 mgCOD L^{-1} for acetoclastic *Methanothrix* and 2.7 mgCOD L^{-1} for propionate-degrading pathways were derived for the SAF-MBR, compared with reported values of 221 mgCOD L^{-1} for acetoclastic *Methanosarcina* and 34 mgCOD L^{-1} for propionate degradation under conventional anaerobic digestion conditions. These markedly lower K values indicate enhanced substrate affinity and reduced thermodynamic constraints, providing a quantitative explanation for the observed suppression of acetate and propionate accumulation under low-strength operating conditions.

Taken together, these results indicate that EET-enabled anaerobic wastewater treatment expands the operational performance envelope of anaerobic treatment for municipal wastewater. By improving biological stability, accelerating conversion of soluble intermediates, and reducing sensitivity to low influent substrate concentrations, EET alters the feasibility of anaerobic systems for mainstream wastewater application, rather than limiting their use to high-strength waste streams.

Beyond improvements in conventional performance metrics, this work highlights a more fundamental implication of EET for membrane-based anaerobic wastewater treatment. Membrane systems decouple solids retention from hydraulic retention, enabling anaerobic operation without conventional sludge processing and producing very low effluent COD through physical retention of

solids [410]. However, low effluent COD does not necessarily indicate biological robustness, as sustained operation still depends on effective biodegradation of soluble intermediates to prevent their accumulation and associated membrane fouling. In low-strength municipal wastewater, where thermodynamic driving forces for anaerobic metabolism are inherently weak [411], maintaining this biological stability represents a central limitation of membrane-based anaerobic treatment. The results of this thesis suggest that EET-enabled microbial consortia directly address this limitation by stabilising interspecies electron flow [44], thereby enabling stable membrane-based anaerobic treatment without reliance on upstream sludge concentration or intensive operational intervention.

From a process design perspective, these findings highlight conductive materials and biofilm architecture as active determinants of reactor function rather than passive components. GAC promotes spatial organisation, electrical connectivity, and functional integration of microbial communities, enabling EET and DIET to operate effectively at pilot scale. This suggests that future anaerobic wastewater treatment systems can be deliberately engineered to select for resilient, electron-sharing consortia through informed material choice and reactor configuration.

9.3 Limitations of the study

This study was designed to prioritise system-level understanding of EET in pilot-scale anaerobic wastewater treatment, and several limitations should be considered when interpreting the findings.

First, functional inferences regarding EET mechanisms were primarily derived from metagenomic, metatranscriptomic, and pangenomic analyses rather than direct electrochemical measurements. While gene-centric and expression-based evidence provide strong support for EET and DIET activity, the absence of in situ electrochemical characterisation limits direct quantification of electron fluxes and conductive properties within biofilms.

Second, although the SAF-MBR provided a representative and operationally relevant platform for investigating EET in low-strength municipal wastewater, the findings are necessarily constrained to the specific reactor configuration, operating conditions, and influent characteristics examined in this study. Variations in temperature, hydraulic regime, carrier material, or wastewater composition may influence the extent to which EET-mediated processes manifest in other anaerobic treatment systems.

Third, the identification of key molecular components involved in electron transfer, including the MspA-like sheath protein in *Methanothrix*, was based on comparative genomics and structural inference rather than targeted biochemical or biophysical validation. As such, while the proposed mechanisms are consistent with observed gene expression patterns and structural features, direct experimental confirmation remains outstanding.

Finally, this study focused on community-level and genome-resolved analyses and did not include controlled pure-culture or defined co-culture experiments. While this approach ensured ecological relevance at the pilot scale, it limits the ability to isolate individual metabolic interactions and to attribute causality to specific organisms or pathways with absolute certainty.

9.4 Future works

This study has provided insights into the mechanisms of EET in anaerobic digestion, yet several areas warrant further investigation. The following sections propose directions for deepening understanding and enhancing application.

9.4.1 Physical examination of the surface proteins using state-of-the-art Bio-SECM

While scanning electrochemical microscopy (SECM) was planned, it could not be carried out within the timeframe of this PhD. A new collaboration at the City University of Hong Kong is expected to realise these experiments, where Prof. Tony Feng and Dr Giin-Yu Amy Tan are establishing a facility for bio-SECM. A Zeiss Inverted Fluorescence Microscope (Axio Vert. A1 FL) has been acquired, and discussions are underway with Sensolytics to integrate an SECM system compatible with the Axio Vert. A1 FL. A1 FL.

Once operational, bio-SECM could elucidate the spatial arrangement and functional roles of conductive sheaths, pili, and other surface proteins in *Methanothrix*. By employing nanoelectrodes and real-time fluorescence imaging, researchers can pinpoint electroactive hotspots, reveal cell-surface interactions with conductive materials (such as GAC), and refine models of electron flow under anaerobic conditions.

9.4.2 Transposon sequencing for *De Novo* functional annotation

Significant portions of methanogen genomes remain functionally unannotated, including genes likely related to EET. Transposon mutagenesis—especially when combined with high-throughput sequencing—offers a systematic method to disrupt genes and observe phenotypic consequences. Although this approach was initially part of the PhD proposal, it was postponed due to the COVID-19 pandemic. Plans are underway to collaborate with Dr Buz Barstow at Cornell University to resume or expand these efforts.

A key advantage of newer transposon-sequencing methods (e.g. Knockout Sudoku) is their relative cost-effectiveness compared to traditional mutant-library approaches. Previous estimates suggest that generating and screening a transposon mutant library could cost in the range of £7,500, which is notably lower than earlier methods that might exceed £200,000. This makes large-scale functional profiling more feasible, even for slow-growing organisms. By screening mutant libraries under electron-transfer conditions, researchers can pinpoint essential EET genes and clarify their roles in methanogenesis and interspecies interactions.

9.4.3 Pure and co-culture experiments

Finally, additional insight into EET could be gained from smaller-scale pure and co-culture studies involving key microbes such as methanogens and SRB. This research, envisioned in collaboration

with Dr Andreia F. Salvador at the University of Minho, would reduce community complexity, isolating the most critical interactions and metabolic pathways for DIET. Such controlled experiments could help discern whether specific genes are universally required or context-dependent, guiding the strategic engineering of microbial consortia for enhanced performance in full-scale anaerobic digestion systems.

In summary, this work highlights the potential of EET pathways to enhance anaerobic digestion. Our investigation into microbial diversity, material-based mediation, and genetic mechanisms offers preliminary insights that may guide improvements in reactor design and bioelectrochemical approaches. However, many mechanistic details remain unclear, and further interdisciplinary research is needed to build on these findings and better understand microbial electron flow in environmental engineering and biotechnology.

Chapter 10

References

1. Nations U. 2021. The Sustainable Development Goals Report 2021. United Nations, 2021. Available from: https://unstats.un.org/sdgs/report/2021/
2. Vasco-Correa J, Khanal S, Manandhar A, and Shah A. 2018. Anaerobic digestion for bioenergy production: Global status, environmental and techno-economic implications, and government policies. Bioresource Technology; 247:1015–26
3. Piadeh F, Offie I, Behzadian K, Rizzuto JP, Bywater A, Córdoba-Pachón JR, and Walker M. 2024. A critical review for the impact of anaerobic digestion on the sustainable development goals. Journal of Environmental Management; 349:119458
4. International Energy Agency. 2020. Outlook for biogas and biomethane: Prospects for organic growth. https://www.iea.org/reports/outlook-for-biogas-and-biomethane-prospects-for-organic-growth. 2020
5. Department for Business E and Strategy I. 2020. Anaerobic Digestion Statistics: United Kingdom, 2020. Accessed: 2024-02-05. 2020. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1002826/Anaerobic_Digestion_Statistics_UK_2020.pdf
6. Dagnall S. 1995. UK strategy for centralised anaerobic digestion. Bioresource Technology; 52:275–80
7. De Mes T, Stams A, Reith J, and Zeeman G. 2003. Methane production by anaerobic digestion of wastewater and solid wastes. Bio-methane & Bio-hydrogen; 2003:58–102
8. Sieber JR, McInerney MJ, and Gunsalus RP. 2012. Genomic insights into syntrophy: the paradigm for anaerobic metabolic cooperation. Annual Review of Microbiology; 66:429–52
9. Boone DR, Johnson RL, and Liu Y. 1989. Diffusion of the interspecies electron carriers H2 and formate in methanogenic ecosystems and its implications in the measurement of K m for H2 or formate uptake. Applied and Environmental Microbiology; 55:1735–41
10. Chen Y, Cheng JJ, and Creamer KS. 2008. Inhibition of anaerobic digestion process: a review. Bioresource Technology; 99:4044–64

11. Zhao Z, Li Y, Zhang Y, and Lovley DR. 2020. Sparking anaerobic digestion: promoting direct interspecies electron transfer to enhance methane production. iScience; 23

12. Li L, Xu Y, Dai X, and Dai L. 2021. Principles and advancements in improving anaerobic digestion of organic waste via direct interspecies electron transfer. Renewable and Sustainable Energy Reviews; 148:111367

13. Jung H, Baek G, and Lee C. 2020. Magnetite-assisted in situ microbial oxidation of H2S to S0 during anaerobic digestion: A new potential for sulfide control. Chemical Engineering Journal; 397:124982

14. Morita M, Malvankar NS, Franks AE, rs ZM r, Giloteaux L, Rotaru AE, Rotaru C, and Lovley DR. 2011. Potential for direct interspecies electron transfer in methanogenic wastewater digester aggregates. mBio; 2:10–1128

15. Barua S and Dhar BR. 2017. Advances towards Understanding and Engineering Direct Interspecies Electron Transfer in Anaerobic Digestion. Bioresource Technology; 244:698–707. Available from: https://doi.org/10.1016/j.biortech.2017.08.023

16. Park JH, Kang HJ, Park KH, and Park HD. 2018. Direct Interspecies Electron Transfer via Conductive Materials: A Perspective for Anaerobic Digestion Applications. Bioresource Technology; 254:300–11. Available from: https://doi.org/10.1016/j.biortech.2018.01.095

17. Liu F, Rotaru AE, Shrestha PM, Malvankar NS, Nevin KP, and Lovley DR. 2012. Promoting direct interspecies electron transfer with activated carbon. Energy & Environmental Science; 5:8982–9

18. Lei Y, Sun D, Dang Y, Chen H, Zhao Z, Zhang Y, and Holmes DE. 2016. Stimulation of methanogenesis in anaerobic digesters treating leachate from a municipal solid waste incineration plant with carbon cloth. Bioresource Technology; 222:270–6

19. Liu C, Sun D, Zhao Z, Dang Y, and Holmes DE. 2019. Methanothrix enhances biogas upgrading in microbial electrolysis cell via direct electron transfer. Bioresource Technology; 291:121877

20. Tian T, Qiao S, Li X, Zhang M, and Zhou J. 2017. Nano-graphene induced positive effects on methanogenesis in anaerobic digestion. Bioresource Technology; 224:41–7

21. Chen Q, Liu C, Liu X, Sun D, Li P, Qiu B, Dang Y, Karpinski NA, Smith JA, and Holmes DE. 2020. Magnetite enhances anaerobic digestion of high salinity organic wastewater. Environmental Research; 189:109884

22. Rotaru AE, Shrestha PM, Liu F, Markovaite B, Chen S, Nevin KP, and Lovley DR. 2014. Direct Interspecies Electron Transfer between Geobacter Metallireducens and Methanosarcina Barkeri. Applied and Environmental Microbiology; 80. Available from: https://doi.org/10.1128/AEM.00895-14

23. Rotaru AE, Shrestha PM, Liu F, Shrestha M, Shrestha D, Embree M, Zengler K, Wardman C, Nevin KP, and Lovley DR. 2014. A new model for electron flow during anaerobic digestion: direct interspecies electron transfer to Methanosaeta for the reduction of carbon dioxide to methane. Energy & Environmental Science; 7:408–15

24. Yang P, Tan GYA, Aslam M, Kim J, and Lee PH. 2019. Metatranscriptomic Evidence for Classical and RuBisCO-Mediated CO2 Reduction to Methane Facilitated by Direct Interspecies Electron Transfer in a Methanogenic System. Scientific Reports; 9:4116. Available from: https://doi.org/10.1038/s41598-019-40830-0

25. Storck T, Virdis B, and Batstone DJ. 2015. Modelling Extracellular Limitations for Mediated versus Direct Interspecies Electron Transfer. The ISME Journal; 10:621–31. Available from: https://doi.org/10.1038/ismej.2015.139

26. Gu M, Yin Q, Liu Y, Du J, and Wu G. 2019. New insights into the effect of direct interspecies electron transfer on syntrophic methanogenesis through thermodynamic analysis. Bioresource Technology Reports; 7:100225

27. Guo B, Zhang Y, Yu N, and Liu Y. 2021. Impacts of conductive materials on microbial community during syntrophic propionate oxidization for biomethane recovery. Water Environment Research; 93:84–93

28. Liu Y, Gu M, Yin Q, Du J, and Wu G. 2020. Thermodynamic analysis of direct interspecies electron transfer in syntrophic methanogenesis based on the optimized energy distribution. Bioresource Technology; 297:122345

29. Zhang L, Zhang J, and Loh KC. 2018. Activated carbon enhanced anaerobic digestion of food waste–Laboratory-scale and Pilot-scale operation. Waste Management; 75:270–9

30. Shin C and Bae J. 2018. Current Status of the Pilot-Scale Anaerobic Membrane Bioreactor Treatments of Domestic Wastewaters: A Critical Review. Bioresource Technology. Available from: https://doi.org/10.1016/j.biortech.2017.09.002

31. Shin C, McCarty PL, Kim J, and Bae J. 2014. Pilot-Scale Temperate-Climate Treatment of Domestic Wastewater with a Staged Anaerobic Fluidized Membrane Bioreactor (SAF-MBR). Bioresource Technology; 159. Available from: https://doi.org/10.1016/j.biortech.2014.02.060

32. Shin C, Tilmans SH, Chen F, and Criddle CS. 2021. Anaerobic membrane bioreactor model for design and prediction of domestic wastewater treatment process performance. Chemical Engineering Journal; 426:131912

33. McCarty PL and Bae J. 2011. Model to couple anaerobic process kinetics with biological growth equilibrium thermodynamics. Environmental Science & Technology; 45:6838–44

34. Yoo R, Kim J, McCarty PL, and Bae J. 2012. Anaerobic treatment of municipal wastewater with a staged anaerobic fluidized membrane bioreactor (SAF-MBR) system. Bioresource Technology; 120:133–9

35. Gouveia J, Plaza F, Garralon G, Fdz-Polanco F, and Pena M. 2015. Long-term operation of a pilot scale anaerobic membrane bioreactor (AnMBR) for the treatment of municipal wastewater under psychrophilic conditions. Bioresource Technology; 185:225–33

36. Evans P, Doody A, Harclerode M, Vila P, Parameswaran P, Lim K, Bae J, Shin C, Lee P, Tan A, et al. 2018. Anaerobic Membrane Bioreactor (AnMBR) for sustainable wastewater treatment. Project ER-201434. Department of Defense Environmental Security Technology Program Final Report

37. Tejedor-Sanz S, Stevens ET, Li S, Finnegan P, Nelson J, Knoesen A, Light SH, Ajo-Franklin CM, and Marco ML. 2022. Extracellular electron transfer increases fermentation in lactic acid bacteria via a hybrid metabolism. eLife; 11:e70684

38. Walker DJ, Adhikari RY, Holmes DE, Ward JE, Woodard TL, Nevin KP, and Lovley DR. 2018. Electrically Conductive Pili from Pilin Genes of Phylogenetically Diverse Microorganisms. ISME Journal; 12. Available from: https://doi.org/10.1038/ismej.2017.141

39. Walker DJ, Nevin KP, Holmes DE, Rotaru AE, Ward JE, Woodard TL, Zhu J, Ueki T, Nonnenmann SS, McInerney MJ, and Lovley DR. 2020. Syntrophus Conductive Pili Demonstrate That Common Hydrogen-Donating Syntrophs Can Have a Direct Electron Transfer Option. ISME Journal; 14. Available from: https://doi.org/10.1038/s41396-019-0575-9

40. Zheng S, Li M, Liu Y, and Liu F. 2021. Desulfovibrio Feeding Methanobacterium with Electrons in Conductive Methanogenic Aggregates from Coastal Zones. Water Research; 202. Available from: https://doi.org/10.1016/j.watres.2021.117490

41. Medvedeva S, Borrel G, and Gribaldo S. 2024. Sheaths are diverse and abundant cell surface layers in archaea. The ISME Journal; 18:wrae225

42. Toczyłowska-Mamińska R, Szymona K, Madej H, Wong WZ, Bala A, Brutkowski W, Krajewski K, San H'ng P, and Mamiński M. 2015. Cellulolytic and electrogenic activity of Enterobacter cloacae in mediatorless microbial fuel cell. Applied Energy; 160:88–93

43. Van Steendam C, Smets I, Skerlos S, and Raskin L. 2019. Improving anaerobic digestion via direct interspecies electron transfer requires development of suitable characterization methods. Current Opinion in Biotechnology; 57:183–90

44. Lovley DR. 2017. Syntrophy Goes Electric: Direct Interspecies Electron Transfer. Annual Review of Microbiology. DOI: 10.1146/annurev-micro-030117-020420

45. Ishii S, Suzuki S, Norden-Krichmar TM, Tenney A, Chain PS, Scholz MB, Nealson KH, and Bretschger O. 2013. A novel metatranscriptomic approach to identify gene expression dynamics during extracellular electron transfer. Nature Communications; 4:1601

46. Holmes DE, Shrestha PM, Walker DJF, Dang Y, Nevin KP, Woodard TL, and Lovley DR. 2017. Metatranscriptomic Evidence for Direct Interspecies Electron Transfer between Geobacter and Methanothrix Species in Methanogenic Rice Paddy Soils. Applied and Environmental Microbiology; 83. Available from: https://doi.org/10.1128/AEM.00223-17

47. Holmes DE, Zhou J, Ueki T, Woodard T, and Lovley DR. 2021. Mechanisms for electron uptake by Methanosarcina acetivorans during direct interspecies electron transfer. mBio; 12:10–1128

48. Schoch CL, Ciufo S, Domrachev M, Hotton CL, Kannan S, Khovanskaya R, Leipe D, Mcveigh R, O'Neill K, Robbertse B, et al. 2020. NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database; 2020:baaa062

49. Kanehisa M, Furumichi M, Tanabe M, Sato Y, and Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research; 45:D353–D361

50. Blum M, Chang HY, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, et al. 2021. The InterPro protein families and domains database: 20 years on. Nucleic Acids Research; 49:D344–D354

51. Bernardes JS, Dávila AM, Costa VS, and Zaverucha G. 2007. Improving model construction of profile HMMs for remote homology detection through structural alignment. BMC Bioinformatics; 8:1–12

52. Zhang Z and Wood WI. 2003. A profile hidden Markov model for signal peptides generated by HMMER. Bioinformatics; 19:307–8

53. Gao K and Lu Y. 2021. Putative extracellular electron transfer in methanogenic archaea. Frontiers in Microbiology; 12:611739

54. Appels L, Lauwers J, Degrève J, Helsen L, Lievens B, Willems K, Van Impe J, and Dewil R. 2011. Anaerobic digestion in global bio-energy production: potential and research challenges. Renewable and Sustainable Energy Reviews; 15:4295–301

55. Holm-Nielsen JB, Al Seadi T, and Oleskowicz-Popiel P. 2009. The future of anaerobic digestion and biogas utilization. Bioresource Technology; 100:5478–84

56. Meynell PJ et al. 1982. Methane: planning a digester. Ed. 2. Prism Press, 1982

57. De Clercq D, Wen Z, Fan F, and Caicedo L. 2016. Biomethane production potential from restaurant food waste in megacities and project level-bottlenecks: a case study in Beijing. Renewable and Sustainable Energy Reviews; 59:1676–85

58. Jain S, Jain S, Wolf IT, Lee J, and Tong YW. 2015. A comprehensive review on operating parameters and different pretreatment methodologies for anaerobic digestion of municipal solid waste. Renewable and Sustainable Energy Reviews; 52:142–54

59. Meegoda JN, Li B, Patel K, and Wang LB. 2018. A review of the processes, parameters, and optimization of anaerobic digestion. International Journal of Environmental Research and Public Health; 15:2224

60. Neumann P, Pesante S, Venegas M, and Vidal G. 2016. Developments in pre-treatment methods to improve anaerobic digestion of sewage sludge. Reviews in Environmental Science and Bio/Technology; 15:173–211

61. Nguyen VK, Chaudhary DK, Dahal RH, Trinh NH, Kim J, Chang SW, Hong Y, La DD, Nguyen XC, Ngo HH, et al. 2021. Review on pretreatment techniques to improve anaerobic digestion of sewage sludge. Fuel; 285:119105

62. Bryant M, Wolin E, Wolin M, and Wolfe R. 1967. Methanobacillus omelianskii, a symbiotic association of two species of bacteria. Archiv für Mikrobiologie; 59:20–31

63. Kato S and Watanabe K. 2010. Ecological and evolutionary interactions in syntrophic methanogenic consortia. Microbes and Environments; 25:145–51

64. Hartshorne RS, Reardon CL, Ross D, Nuester J, Clarke TA, Gates AJ, Mills PC, Fredrickson JK, Zachara JM, Shi L, et al. 2009. Characterization of an electron conduit between bacteria and the extracellular environment. Proceedings of the National Academy of Sciences; 106:22169–74

65. Lohner ST, Deutzmann JS, Logan BE, Leigh J, and Spormann AM. 2014. Hydrogenase-independent uptake and metabolism of electrons by the archaeon Methanococcus maripaludis. The ISME Journal; 8:1673–81

66. Rowe AR, Xu S, Gardel E, Bose A, Girguis P, Amend JP, and El-Naggar MY. 2019. Methane-linked mechanisms of electron uptake from cathodes by Methanosarcina barkeri. mBio; 10:10–1128

67. Cheng Q and Call DF. 2016. Hardwiring microbes via direct interspecies electron transfer: mechanisms and applications. Environmental Science: Processes & Impacts; 18:968–80

68. Zacharoff LA, Morrone DJ, and Bond DR. 2017. Geobacter sulfurreducens extracellular multiheme cytochrome PgcA facilitates respiration to Fe (III) oxides but not electrodes. Frontiers in Microbiology; 8:2481

69. Lovley DR and Holmes DE. 2022. Electromicrobiology: the ecophysiology of phylogenetically diverse electroactive microorganisms. Nature Reviews Microbiology; 20:5–19

70. Wang W and Lee DJ. 2021. Direct interspecies electron transfer mechanism in enhanced methanogenesis: a mini-review. Bioresource Technology; 330:124980

71. Martins G, Salvador AF, Pereira L, and Alves MM. 2018. Methane production and conductive materials: a critical review. Environmental Science & Technology; 52:10241–53

72. Van Lier JB, Mahmoud N, and Zeeman G. 2008. Anaerobic wastewater treatment. Biological wastewater treatment: principles, modelling and design :415–56

73. Ren S, Usman M, Tsang DC, O-Thong S, Angelidaki I, Zhu X, Zhang S, and Luo G. 2020. Hydrochar-facilitated anaerobic digestion: evidence for direct interspecies electron transfer mediated through surface oxygen-containing functional groups. Environmental Science & Technology; 54:5755–66

74. Lei Y, Sun D, Dang Y, Feng X, Huo D, Liu C, Zheng K, and Holmes DE. 2019. Metagenomic analysis reveals that activated carbon aids anaerobic digestion of raw incineration leachate by promoting direct interspecies electron transfer. Water Research; 161:570–80

75. Wang D, Ai J, Shen F, Yang G, Zhang Y, Deng S, Zhang J, Zeng Y, and Song C. 2017. Improving anaerobic digestion of easy-acidification substrates by promoting buffering capacity using biochar derived from vermicompost. Bioresource Technology; 227:286–96

76. Wang C, Wang C, Liu J, Xu Q, Han Z, Xu X, and Zhu L. 2020. Tolerance of aceticlastic methanogenesis enhanced by magnetite under the condition of ammonia stress. ACS Sustainable Chemistry & Engineering; 8:1417–26

77. Yan W, Mukherjee M, and Zhou Y. 2020. Direct interspecies electron transfer (DIET) can be suppressed under ammonia-stressed condition–reevaluate the role of conductive materials. Water Research; 183:116094

78. Lü F, Luo C, Shao L, and He P. 2016. Biochar alleviates combined stress of ammonium and acids by firstly enriching Methanosaeta and then Methanosarcina. Water research; 90:34–43

79. Zhuang L, Ma J, Yu Z, Wang Y, and Tang J. 2018. Magnetite accelerates syntrophic acetate oxidation in methanogenic systems with high ammonia concentrations. Microbial Biotechnology; 11:710–20

80. Zhang Y, Zhang L, Guo B, Zhou Y, Gao M, Sharaf A, and Liu Y. 2020. Granular activated carbon stimulated microbial physiological changes for enhanced anaerobic digestion of municipal sewage. Chemical Engineering Journal; 400:125838

81. Yang L, Si B, Zhang Y, Watson J, Stablein M, Chen J, Zhang Y, Zhou X, and Chu H. 2020. Continuous treatment of hydrothermal liquefaction wastewater in an anaerobic biofilm reactor: Potential role of granular activated carbon. Journal of Cleaner Production; 276:122836

82. Dang Y, Holmes DE, Zhao Z, Woodard TL, Zhang Y, Sun D, Wang LY, Nevin KP, and Lovley DR. 2016. Enhancing anaerobic digestion of complex organic waste with carbon-based conductive materials. Bioresource Technology; 220:516–22

83. Weng X, Mao Z, Fu HM, Chen YP, Guo JS, Fang F, Xu XW, and Yan P. 2022. Biofilm formation during wastewater treatment: Motility and physiological response of aerobic denitrifying bacteria under ammonia stress based on surface plasmon resonance imaging. Bioresource Technology; 361:127712

84. Lovley DR. 2011. Reach out and touch someone: potential impact of DIET (direct interspecies energy transfer) on anaerobic biogeochemistry, bioremediation, and bioenergy. Reviews in Environmental Science and Bio/Technology; 10:101–5

85. Lovley DR. 2017. Happy together: microbial communities that hook up to swap electrons. The ISME Journal; 11:327–36

86. Cruz Viggi C, Rossetti S, Fazi S, Paiano P, Majone M, and Aulenta F. 2014. Magnetite particles triggering a faster and more robust syntrophic pathway of methanogenic propionate degradation. Environmental Science & Technology; 48:7536–43

87. Xu H, Wang C, Yan K, Wu J, Zuo J, and Wang K. 2016. Anaerobic granule-based biofilms formation reduces propionate accumulation under high H2 partial pressure using conductive carbon felt particles. Bioresource Technology; 216:677–83

88. Zhao Z, Zhang Y, Li Y, Dang Y, Zhu T, and Quan X. 2017. Potentially shifting from interspecies hydrogen transfer to direct interspecies electron transfer for syntrophic metabolism to resist acidic impact with conductive carbon cloth. Chemical Engineering Journal; 313:10–18

89. Li L, Gao Q, Liu X, Zhao Q, Wang W, Wang K, Zhou H, and Jiang J. 2022. Insights into high-solids anaerobic digestion of food waste enhanced by activated carbon via promoting direct interspecies electron transfer. Bioresource Technology; 351:127008

90. Bielen AA, Verhaart MR, VanFossen AL, Blumer-Schuette SE, Stams AJ, Oost J van der, Kelly RM, and Kengen SW. 2013. A thermophile under pressure: Transcriptional analysis of the response of Caldicellulosiruptor saccharolyticus to different H2 partial pressures. International Journal of Hydrogen Energy; 38:1837–49

91. Ahring BK, Westermann P, and Mah RA. 1991. Hydrogen inhibition of acetate metabolism and kinetics of hydrogen consumption by Methanosarcina thermophila TM-1. Archives of Microbiology; 157:38–42

92. Jetten MS, Stams AJ, and Zehnder AJ. 1992. Methanogenesis from acetate: a comparison of the acetate metabolism in Methanothrix soehngenii and Methanosarcina spp. FEMS Microbiology Reviews; 8:181–97

93. Yamada C, Kato S, Ueno Y, Ishii M, and Igarashi Y. 2015. Conductive iron oxides accelerate thermophilic methanogenesis from acetate and propionate. Journal of bioscience and bioengineering; 119:678–82

94. Li H, Chang J, Liu P, Fu L, Ding D, and Lu Y. 2015. Direct interspecies electron transfer accelerates syntrophic oxidation of butyrate in paddy soil enrichments. Environmental Microbiology; 17:1533–47

95. Barua S, Zakaria BS, and Dhar BR. 2018. Enhanced methanogenic co-degradation of propionate and butyrate by anaerobic microbiome enriched on conductive carbon fibers. Bioresource Technology; 266:259–66

96. Dang Y, Sun D, Woodard TL, Wang LY, Nevin KP, and Holmes DE. 2017. Stimulation of the anaerobic digestion of the dry organic fraction of municipal solid waste (OFMSW) with carbon-based conductive materials. Bioresource Technology; 238:30–8

97. Bueno-López JI, Rangel-Mendez JR, Alatriste-Mondragón F, Pérez-Rodrıguez F, Hernández-Montoya V, and Cervantes FJ. 2018. Graphene oxide triggers mass transfer limitations on the methanogenic activity of an anaerobic consortium with a particulate substrate. Chemosphere; 211:709–16

98. Yin Q, Yang S, Wang Z, Xing L, and Wu G. 2018. Clarifying electron transfer and metagenomic analysis of microbial community in the methane production process with the addition of ferroferric oxide. Chemical Engineering Journal; 333:216–25

99. Xu S, He C, Luo L, Lü F, He P, and Cui L. 2015. Comparing activated carbon of different particle sizes on enhancing methane generation in upflow anaerobic digester. Bioresource Technology; 196:606–12

100. Kim M, Lam TYC, Tan GYA, Lee PH, and Kim J. 2020. Use of polymeric scouring agent as fluidized media in anaerobic fluidized bed membrane bioreactor for wastewater treatment: System performance and microbial community. J Memb Sci; 606

101. Kwon D and al. et. 2022. Polyaniline-coated conductive media promotes direct interspecies electrons transfer (DIET) and kinetics enhancement of low-strength wastewater treatment in anaerobic fluidized bed membrane bioreactor (AFMBR). Chemical Engineering Journal; 446

102. Kwon D et al. 2021. Combined effect of activated carbon particles and non-adsorptive spherical beads as fluidized media on fouling, organic removal and microbial communities in anaerobic membrane bioreactor. Membranes (Basel); 11

103. Babauta JT and Beyenal H. 2023. Biofilm Electrochemistry. *Biofilms in Bioelectrochemical Systems: From Laboratory Practice to Data Interpretation*. Ed. by Beyenal H and Babauta JT. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2023 :121–76. DOI: 10.1002/9781118414517.ch5

104. Summers ZM, Fogarty HE, Leang C, Franks AE, Malvankar NS, and Lovley DR. 2010. Direct exchange of electrons within aggregates of an evolved syntrophic coculture of anaerobic bacteria. Science; 330:1413–5

105. Burdige DJ and Nealson KH. 1985. Microbial manganese reduction by enrichment cultures from coastal marine sediments. Applied and Environmental Microbiology; 50:491–7

106. Lovley DR and Phillips EJ. 1987. Rapid assay for microbially reducible ferric iron in aquatic sediments. Applied and Environmental Microbiology; 53:1536–40

107. Lovley DR, Ueki T, Zhang T, Malvankar NS, Shrestha PM, Flanagan KA, Aklujkar M, Butler JE, Giloteaux L, Rotaru AE, et al. 2011. Geobacter: the microbe electric's physiology, ecology, and practical applications. Advances in Microbial Physiology; 59:1–100

108. Kumar R, Singh L, Wahid ZA, and Din MFM. 2015. Exoelectrogens in microbial fuel cells toward bioelectricity generation: a review. International Journal of Energy Research; 39:1048–67

109. Biffinger JC, Ray R, Little BJ, Fitzgerald LA, Ribbens M, Finkel SE, and Ringeisen BR. 2009. Simultaneous analysis of physiological and electrical output changes in an operating microbial fuel cell with Shewanella oneidensis. Biotechnology and Bioengineering; 103:524–31

110. Firer-Sherwood M, Pulcu GS, and Elliott SJ. 2008. Electrochemical interrogations of the Mtr cytochromes from Shewanella: opening a potential window. JBIC Journal of Biological Inorganic Chemistry; 13:849–54

111. Yee MO and Rotaru AE. 2020. Extracellular electron uptake in Methanosarcinales is independent of multiheme c-type cytochromes. Scientific Reports; 10:372

112. Welte C and Deppenmeier U. 2014. Bioenergetics and anaerobic respiratory chains of aceticlastic methanogens. Biochimica et Biophysica Acta (BBA)-Bioenergetics; 1837:1130–47

113. Holmes DE, Rotaru AE, Ueki T, Shrestha PM, Ferry JG, and Lovley DR. 2018. Electron and proton flux for carbon dioxide reduction in Methanosarcina barkeri during direct interspecies electron transfer. Frontiers in Microbiology; 9:3109

114. Cheng S, Xing D, Call DF, and Logan BE. 2009. Direct biological conversion of electrical current into methane by electromethanogenesis. Environmental Science & Technology; 43:3953–8

115. Walker DJ, Martz E, Holmes DE, Zhou Z, Nonnenmann SS, and Lovley DR. 2019. The archaellum of Methanospirillum hungatei is electrically conductive. mBio; 10:10–1128

116. Zhou J, Smith JA, Li M, and Holmes DE. 2023. Methane production by Methanothrix thermoacetophila via direct interspecies electron transfer with Geobacter metallireducens. mBio :e00360–23

117. Bond DR and Lovley DR. 2002. Reduction of Fe (III) oxide by methanogens in the presence and absence of extracellular quinones. Environmental Microbiology; 4:115–24

118. Ye J, Yu J, Zhang Y, Chen M, Liu X, Zhou S, and He Z. 2019. Light-driven carbon dioxide reduction to methane by Methanosarcina barkeri-CdS biohybrid. Applied Catalysis B: Environmental; 257:117916

119. Zhang J, Dong H, Zhao L, McCarrick R, and Agrawal A. 2014. Microbial reduction and precipitation of vanadium by mesophilic and thermophilic methanogens. Chemical Geology; 370:29–39

120. Flemming HC, Wingender J, Szewzyk U, Steinberg P, Rice SA, and Kjelleberg S. 2016. Biofilms: an emergent form of bacterial life. Nature Reviews Microbiology; 14:563–75

121. Uali AS, Lam TY, Huang X, Wu Z, Shih HJ, Tan GY, and Lee PH. 2023. Role and potential of the semi-classical/-quantum mechanism of the extracellular environment and cell envelope in Direct Interspecies Electron Transfer (DIET)-driven biomethanation. Critical Reviews in Environmental Science and Technology :1–21

122. Lovley DR, Coates JD, Blunt-Harris EL, Phillips EJ, and Woodward JC. 1996. Humic substances as electron acceptors for microbial respiration. Nature; 382:445–8

123. Nevin KP and Lovley DR. 2002. Mechanisms for Fe (III) oxide reduction in sedimentary environments. Geomicrobiology Journal; 19:141–59

124. Monteverde DR, Sylvan JB, Suffridge C, Baronas JJ, Fichot E, Fuhrman J, Berelson W, and Sanudo-Wilhelmy SA. 2018. Distribution of extracellular flavins in a coastal marine basin and their relationship to redox gradients and microbial community members. Environmental Science & Technology; 52:12265–74

125. Von Canstein H, Ogawa J, Shimizu S, and Lloyd JR. 2008. Secretion of flavins by Shewanella species and their role in extracellular electron transfer. Applied and Environmental Microbiology; 74:615–23

126. Mehta-Kolte MG and Bond DR. 2012. Geothrix fermentans secretes two different redox-active compounds to utilize electron acceptors across a wide range of redox potentials. Applied and Environmental Microbiology; 78:6987–95

127. Glasser NR, Saunders SH, and Newman DK. 2017. The colorful world of extracellular electron shuttles. Annual Review of Microbiology; 71:731–51

128. Kotloski NJ and Gralnick JA. 2013. Flavin electron shuttles dominate extracellular electron transfer by Shewanella oneidensis. mBio; 4:10–1128

129. Yang Y, Ding Y, Hu Y, Cao B, Rice SA, Kjelleberg S, and Song H. 2015. Enhancing bidirectional electron transfer of Shewanella oneidensis by a synthetic flavin pathway. ACS synthetic biology; 4:815–23

130. Duszenko N and Buan NR. 2017. Physiological evidence for isopotential tunneling in the electron transport chain of methane-producing archaea. Applied and Environmental Microbiology; 83:e00950–17

131. Fu L, Zhou T, Wang J, You L, Lu Y, Yu L, and Zhou S. 2019. NanoFe3O4 as solid electron shuttles to accelerate acetotrophic methanogenesis by Methanosarcina barkeri. Frontiers in Microbiology; 10:388

132. Myers CR and Nealson KH. 1988. Bacterial manganese reduction and growth with manganese oxide as the sole electron acceptor. Science; 240:1319–21

133. Newman DK and Kolter R. 2000. A role for excreted quinones in extracellular electron transfer. Nature; 405:94–7

134. Hernandez M and Newman D. 2001. Extracellular electron transfer. Cellular and Molecular Life Sciences CMLS; 58:1562–71

135. Reguera G, McCarthy KD, Mehta T, Nicoll JS, Tuominen MT, and Lovley DR. 2005. Extracellular electron transfer via microbial nanowires. Nature; 435:1098–101

136. El-Naggar MY, Wanger G, Leung KM, Yuzvinsky TD, Southam G, Yang J, Lau WM, Nealson KH, and Gorby YA. 2010. Electrical transport along bacterial nanowires from Shewanella oneidensis MR-1. Proceedings of the National Academy of Sciences; 107:18127–31

137. Kato S, Hashimoto K, and Watanabe K. 2012. Microbial interspecies electron transfer via electric currents through conductive minerals. Proceedings of the National Academy of Sciences; 109:10042–6

138. Hernsdorf AW, Amano Y, Miyakawa K, Ise K, Suzuki Y, Anantharaman K, Probst A, Burstein D, Thomas BC, and Banfield JF. 2017. Potential for microbial H2 and metal transformations associated with novel bacteria and archaea in deep terrestrial subsurface sediments. The ISME Journal; 11:1915–29

139. Arbour TJ, Gilbert B, and Banfield JF. 2020. Diverse microorganisms in sediment and groundwater are implicated in extracellular redox processes based on genomic analysis of bioanode communities. Frontiers in Microbiology; 11:1694

140. Shi L and al. et. 2016. Extracellular electron transfer mechanisms between microorganisms and minerals. Nature Reviews Microbiology; 14:651–62

141. White GF, Edwards MJ, Gomez-Perez L, Richardson DJ, Butt JN, and Clarke TA. 2016. Mechanisms of bacterial extracellular electron exchange. Advances in Microbial Physiology; 68:87–138

142. Beblawy S, Bursac T, Paquete C, Louro R, Clarke TA, and Gescher J. 2018. Extracellular reduction of solid electron acceptors by Shewanella oneidensis. Molecular microbiology; 109:571–83

143. Rowe AR, Salimijazi F, Trutschel L, Sackett J, Adesina O, Anzai I, Kugelmass LH, Baym MH, and Barstow B. 2021. Identification of a pathway for electron uptake in Shewanella oneidensis. Communications Biology; 4:957

144. Shi L, Fredrickson JK, and Zachara JM. 2014. Genomic analyses of bacterial porin-cytochrome gene clusters. Frontiers in Microbiology; 5:657

145. Manzella MP, Reguera G, and Kashefi K. 2013. Extracellular electron transfer to Fe (III) oxides by the hyperthermophilic archaeon Geoglobus ahangari via a direct contact mechanism. Applied and Environmental Microbiology; 79:4694–700

146. Smith JA, Aklujkar M, Risso C, Leang C, Giloteaux L, and Holmes DE. 2015. Mechanisms involved in Fe (III) respiration by the hyperthermophilic archaeon Ferroglobus placidus. Applied and Environmental Microbiology; 81:2735–44

147. Cai C, Leu AO, Xie GJ, Guo J, Feng Y, Zhao JX, Tyson GW, Yuan Z, and Hu S. 2018. A methanotrophic archaeon couples anaerobic oxidation of methane to Fe (III) reduction. The ISME Journal; 12:1929–39

148. Leu AO, Cai C, McIlroy SJ, Southam G, Orphan VJ, Yuan Z, Hu S, and Tyson GW. 2020. Anaerobic methane oxidation coupled to manganese reduction by members of the Methanoperedenaceae. The ISME Journal; 14:1030–41

149. Ettwig KF, Zhu B, Speth D, Keltjens JT, Jetten MS, and Kartal B. 2016. Archaea catalyze iron-dependent anaerobic oxidation of methane. Proceedings of the National Academy of Sciences; 113:12792–6

150. McGlynn SE, Chadwick GL, Kempes CP, and Orphan VJ. 2015. Single cell activity reveals direct electron transfer in methanotrophic consortia. Nature; 526:531–5

151. Leu AO, McIlroy SJ, Ye J, Parks DH, Orphan VJ, and Tyson GW. 2020. Lateral gene transfer drives metabolic flexibility in the anaerobic methane-oxidizing archaeal family methanoperedenaceae. mBio; 11:10–1128

152. Lovley DR. 1991. Dissimilatory fe (iii) and mn (iv) reduction. Microbiological reviews; 55:259–87

153. Ouboter HT, Mesman R, Sleutels T, Postma J, Wissink M, Jetten MS, Ter Heijne A, Berben T, and Welte CU. 2024. Mechanisms of extracellular electron transfer in anaerobic methanotrophic archaea. Nature Communications; 15:1477

154. Lovley DR. 2017. Electrically conductive pili: biological function and potential applications in electronics. Current Opinion in Electrochemistry; 4:190–8

155. Malvankar NS and al. et. 2011. Tunable metallic-like conductivity in microbial nanowire networks. Nature Nanotechnology; 6

156. Malvankar N, Tuominen M, and Lovley D. 2012. Lack of involvement of c-type cytochromes in long-range electron transport in microbial biofilms and nanowires. Energy Environment Science; 5:8651–9

157. Ueki T, Nevin KP, Rotaru AE, Wang LY, Ward JE, Woodard TL, and Lovley DR. 2018. Geobacter strains expressing poorly conductive pili reveal constraints on direct interspecies electron transfer mechanisms. mBio; 9:10–1128

158. Adhikari RY, Malvankar NS, Tuominen MT, and Lovley DR. 2016. Conductivity of individual Geobacter pili. RSC Advances; 6:8354–7

159. Tan Y, Adhikari RY, Malvankar NS, Ward JE, Nevin KP, Woodard TL, Smith JA, Snoeyenbos-West OL, Franks AE, Tuominen MT, et al. 2016. The low conductivity of Geobacter uraniireducens pili suggests a diversity of extracellular electron transfer mechanisms in the genus Geobacter. Frontiers in Microbiology; 7:980

160. Vargas M and al. et. 2013. Aromatic amino acids required for pili conductivity and long-range extracellular electron transport in Geobacter sulfurreducens. mBio; 4

161. Holmes DE, Dang Y, Walker DJ, and Lovley DR. 2016. The electrically conductive pili of Geobacter species are a recently evolved feature for extracellular electron transfer. Microbial Genomics; 2

162. Tan Y, Adhikari RY, Malvankar NS, Ward JE, Woodard TL, Nevin KP, and Lovley DR. 2017. Expressing the Geobacter metallireducens PilA in Geobacter sulfurreducens yields pili with exceptional conductivity. mBio; 8:10–1128

163. Lovley DR and Walker DJF. 2019. Geobacter Protein Nanowires. Frontiers in Microbiology; 10. Preprint at https://doi.org/10.3389/fmicb.2019.02078

164. Bray MS, Wu J, Padilla CC, Stewart FJ, Fowle DA, Henny C, Simister RL, Thompson KJ, Crowe SA, and Glass JB. 2020. Phylogenetic and structural diversity of aromatically dense pili from environmental metagenomes. Environmental Microbiology Reports; 12:49–57

165. Jiang X, Wonderen JH van, Butt JN, Edwards MJ, Clarke TA, and Blumberger J. 2020. Which multi-heme protein complex transfers electrons more efficiently? Comparing MtrCAB from Shewanella with OmcS from Geobacter. The Journal of Physical Chemistry Letters; 11:9421–5

166. Nevin KP and Lovley DR. 2000. Lack of production of electron-shuttling compounds or solubilization of Fe (III) during reduction of insoluble Fe (III) oxide by Geobacter metallireducens. Applied and Environmental Microbiology; 66:2248–51

167. Makarova KS, Koonin EV, and Albers SV. 2016. Diversity and evolution of type IV pili systems in archaea. Frontiers in Microbiology; 7:667

168. Poweleit N, Ge P, Nguyen HH, Loo RRO, Gunsalus RP, and Zhou ZH. 2016. CryoEM structure of the Methanospirillum hungatei archaellum reveals structural features distinct from the bacterial flagellum and type IV pilus. Nature Microbiology; 2:1–12

169. Holmes DE, Tang H, Woodard T, Liang D, Zhou J, Liu X, and Lovley DR. 2022. Cytochrome-mediated direct electron uptake from metallic iron by Methanosarcina acetivorans. mLife; 1:443–7

170. Beveridge TJ, Patel GB, Harris BJ, and Sprott GD. 1986. The ultrastructure of Methanothrix concilii, a mesophilic aceticlastic methanogen. Canadian Journal of Microbiology; 32:703–10

171. Dueholm MS, Larsen P, Finster K, Stenvang MR, Christiansen G, Vad BS, Bøggild A, Otzen DE, and Nielsen PH. 2015. The tubular sheaths encasing Methanosaeta thermophila filaments are functional amyloids. Journal of Biological Chemistry; 290:20590–600

172. Christensen LFB, Hansen LM, Finster K, Christiansen G, Nielsen PH, Otzen DE, and Dueholm MS. 2018. The sheaths of Methanospirillum are made of a new type of amyloid protein. Frontiers in Microbiology; 9:2729

173. Forman C, Wang N, Yang Z, Mowat C, Jarvis S, Durkan C, and Barker P. 2013. Probing the location of displayed cytochrome b562 on amyloid by scanning tunnelling microscopy. Nanotechnology; 24:175102

174. Shipps C, Kelly HR, Dahl PJ, Yi SM, Vu D, Boyer D, Glynn C, Sawaya MR, Eisenberg D, Batista VS, et al. 2021. Intrinsic electronic conductivity of individual atomically resolved amyloid crystals reveals micrometer-long hole hopping via tyrosines. Proceedings of the National Academy of Sciences; 118:e2014139118

175. Zheng S, Liu F, Wang B, Zhang Y, and Lovley DR. 2020. Methanobacterium capable of direct interspecies electron transfer. Environmental Science & Technology; 54:15347–54

176. Zheng S, Zhang H, Li Y, Zhang H, Wang O, Zhang J, and Liu F. 2015. Co-Occurrence of Methanosarcina Mazei and Geobacteraceae in an Iron (III)-Reducing Enrichment Culture. Frontiers in Microbiology; 0:941. Available from: https://doi.org/10.3389/FMICB.2015.00941

177. Yee MO, Snoeyenbos-West OL, Thamdrup B, Ottosen LD, and Rotaru AE. 2019. Extracellular electron uptake by two Methanosarcina species. Frontiers in Energy Research :29

178. Bryant MP and Boone DR. 1987. Isolation and characterization of Methanobacterium formicicum MF. International Journal of Systematic and Evolutionary Microbiology; 37:171–1

179. Soucy SM, Huang J, and Gogarten JP. 2015. Horizontal gene transfer: building the web of life. Nature Reviews Genetics; 16:472–82

180. Arnold BJ, Huang IT, and Hanage WP. 2022. Horizontal gene transfer and adaptive evolution in bacteria. Nature Reviews Microbiology; 20:206–18

181. Barlow M. 2009. What antimicrobial resistance has taught us about horizontal gene transfer. Horizontal Gene Transfer: Genomes in Flux :397–411

182. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer EL, Tosatto SC, Paladin L, Raj S, Richardson LJ, et al. 2021. Pfam: The protein families database in 2021. Nucleic Acids Research; 49:D412–D419

183. Fischer S, Cornils K, Speiseder T, Badbaran A, Reimer R, Indenbirken D, Grundhoff A, Brunswig-Spickenheier B, Alawi M, and Lange C. 2016. Indication of horizontal DNA gene transfer by extracellular vesicles. PLoS One; 11:e0163665

184. Wagner A, Lewis C, and Bichsel M. 2007. A survey of bacterial insertion sequences using IScan. Nucleic Acids Research; 35:5284–93

185. Roux S, Enault F, Hurwitz BL, and Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. PeerJ; 3:e985

186. Rozov R, Brown Kav A, Bogumil D, Shterzer N, Halperin E, Mizrahi I, and Shamir R. 2017. Recycler: an algorithm for detecting plasmids from de novo assembly graphs. Bioinformatics; 33:475–82

187. Krawczyk PS, Lipinski L, and Dziembowski A. 2018. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. Nucleic Acids Research; 46:e35–e35

188. Arredondo-Alonso S, Willems RJ, Van Schaik W, and Schürch AC. 2017. On the (im) possibility of reconstructing plasmids from whole-genome short-read sequencing data. Microbial genomics; 3

189. Tang X, Shang J, Ji Y, and Sun Y. 2023. PLASMe: a tool to identify PLASMid contigs from short-read assemblies using transformer. Nucleic Acids Research; 51:e83–e83

190. Shang J, Tang X, Guo R, and Sun Y. 2022. Accurate identification of bacteriophages from metagenomic data using Transformer. Briefings in Bioinformatics; 23:bbac258

191. Allen MA, Lauro FM, Williams TJ, Burg D, Siddiqui KS, De Francisci D, Chong KW, Pilak O, Chew HH, De Maere MZ, et al. 2009. The genome sequence of the psychrophilic archaeon, Methanococcoides burtonii: the role of genome evolution in cold adaptation. The ISME Journal; 3:1012–35

192. Deppenmeier U, Johann A, Hartsch T, Merkl R, Schmitz RA, Martinez-Arias R, Henne A, Wiezer A, Bäumer S, Jacobi C, et al. 2002. The genome of Methanosarcina mazei: evidence for lateral gene transfer between bacteria and archaea. Journal of Molecular Microbiology and Biotechnology; 4:453–61

193. Nelson-Sathi S, Sousa FL, Roettger M, Lozada-Chávez N, Thiergart T, Janssen A, Bryant D, Landan G, Schönheit P, Siebers B, et al. 2015. Origins of major archaeal clades correspond to gene acquisitions from bacteria. Nature; 517:77–80

194. Fournier GP and Gogarten JP. 2008. Evolution of acetoclastic methanogenesis in Methanosarcina via horizontal gene transfer from cellulolytic Clostridia. Journal of Bacteriology; 190:1124–7

195. Wagner A, Whitaker RJ, Krause DJ, Heilers JH, Van Wolferen M, Van Der Does C, and Albers SV. 2017. Mechanisms of gene flow in archaea. Nature Reviews Microbiology; 15:492–501

196. Baker IR, Conley BE, Gralnick JA, and Girguis PR. 2022. Evidence for horizontal and vertical transmission of Mtr-mediated extracellular electron transfer among the bacteria. mBio; 13:e02904–21

197. Ballesté E, Blanch AR, Muniesa M, Garcıa-Aljaro C, Rodrıguez-Rubio L, Martın-Dıaz J, Pascual-Benito M, and Jofre J. 2022. Bacteriophages in sewage: abundance, roles, and applications. FEMS Microbes; 3:xtac009

198. Pires DP, Melo LD, and Azeredo J. 2021. Understanding the complex phage-host interactions in biofilm communities. Annual Review of Virology; 8:73–94

199. Gödeke J, Paul K, Lassak J, and Thormann KM. 2011. Phage-induced lysis enhances biofilm formation in Shewanella oneidensis MR-1. The ISME Journal; 5:613–26

200. Saunders SH, Edmund C, Yates MD, Otero FJ, Trammell SA, Stemp ED, Barton JK, Tender LM, and Newman DK. 2020. Extracellular DNA promotes efficient extracellular electron transfer by pyocyanin in Pseudomonas aeruginosa biofilms. Cell; 182:919–32

201. Dakheel K, Abdul Rahim R, Al-Obaidi J, Neela V, Hun T, Mat Isa M, Razali N, and Yusoff K. 2022. Proteomic analysis revealed the biofilm-degradation abilities of the bacteriophage $UPMK_1$ and $UPMK_2$ against Methicillin-resistant Staphylococcus aureus. Biotechnology Letter; 44. DOI: 10.1007/s10529-022-03229-y

202. Kim AH and Criddle CS. 2023. Anaerobic wastewater treatment and potable reuse: energy and life cycle considerations. Environmental Science & Technology; 57:17225–36

203. Quince C, Walker AW, Simpson JT, Loman NJ, and Segata N. 2017. Shotgun metagenomics, from sampling to analysis. Nature Biotechnology; 35:833–44

204. Beghini F et al. 2021. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with biobakery 3. eLife; 10

205. Bolger AM, Lohse M, and Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics; 30:2114–20

206. Breitwieser FP, Pertea M, Zimin AV, and Salzberg SL. 2019. Human contamination in bacterial genomes has created thousands of spurious proteins. Genome Research; 29:954–60

207. Andrews S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. Accessed: Dec 13, 2023. 2010. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

208. Pevzner PA, Tang H, and Waterman MS. 2001. An Eulerian path approach to DNA fragment assembly. Proceedings of the National Academy of Sciences; 98:9748–53

209. Nurk S, Meleshko D, Korobeynikov A, and Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Research; 27:824–34

210. Prjibelski A, Antipov D, Meleshko D, Lapidus A, and Korobeynikov A. 2020. Using SPAdes de novo assembler. Current Protocols in Bioinformatics; 70:e102

211. Gurevich A, Saveliev V, Vyahhi N, and Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics; 29:1072–5

212. Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, and Nielsen PH. 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nature Biotechnology; 31:533–8

213. Garza DR and Dutilh BE. 2015. From cultured to uncultured genome sequences: metagenomics and modeling microbial ecosystems. Cellular and Molecular Life Sciences; 72:4287–308

214. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy T, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, et al. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nature Biotechnology; 35:725–31

215. Borderes M, Gasc C, Prestat E, Galvao Ferrarini M, Vinga S, Boucinha L, and Sagot MF. 2021. A comprehensive evaluation of binning methods to recover human gut microbial species from a non-redundant reference gene catalog. NAR Genomics and Bioinformatics; 3:lqab009

216. Kang DD et al. 2019. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. DOI: 10.7717/peerj.7359

217. Langmead B and Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods; 9:357–9

218. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and Subgroup 1GPDP. 2009. The sequence alignment/map format and SAMtools. Bioinformatics; 25:2078–9

219. Wu YW, Tang YH, Tringe SG, Simmons BA, and Singer SW. 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome; 2:1–18

220. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, and Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from. Cold Spring Harbor Laboratory Press Method; 1:1–31

221. Tolstoganov I, Kamenev Y, Kruglikov R, Ochkalova S, and Korobeynikov A. 2022. Binspreader: refine binning results for fuller mag reconstruction. Iscience; 25

222. Ochkalova S, Tolstoganov I, Lapidus A, and Korobeynikov A. 2023. Protocol for refining metagenomic binning with BinSPreader. STAR protocols; 4:102417

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

224. Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, Pollard KS, Sakharova E, Parks DH, Hugenholtz P, et al. 2021. A unified catalog of 204,938 reference genomes from the human gut microbiome. Nature Biotechnology; 39:105–14

225. Olm MR, Crits-Christoph A, Diamond S, Lavy A, Matheus Carnevali PB, and Banfield JF. 2020. Consistent metagenome-derived metrics verify and delineate bacterial species boundaries. Msystems; 5:10–1128

226. Menzel P, Ng KL, and Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nature Communications. DOI: 10.1038/ncomms11257

227. Chaumeil PA, Mussig AJ, Hugenholtz P, and Parks DH. 2022. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics; 38:5315–6

228. Asnicar F et al. 2020. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nature Communications

229. Chaumeil PA, Mussig AJ, Hugenholtz P, and Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. 2020

230. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, and Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics; 11:1–11

231. Matsen FA, Kodner RB, and Armbrust EV. 2010. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics; 11:1–16

232. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, Beghini F, Manghi P, Tett A, Ghensi P, et al. 2019. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. Cell; 176:649–62

233. Buchfink B, Reuter K, and Drost HG. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nature Methods; 18:366–8

234. Rozewicki J, Li S, Amada KM, Standley DM, and Katoh K. 2019. MAFFT-DASH: integrated protein sequence and structural alignment. Nucleic Acids Research; 47:W5–W10

235. Capella-Gutiérrez S, Silla-Martınez JM, and Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics; 25:1972–3

236. Price MN, Dehal PS, and Arkin AP. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PloS one; 5:e9490

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

238. Notredame C, Higgins DG, and Heringa J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. Journal of Molecular Biology; 302:205–17

239. Brown NP, Leroy C, and Sander C. 1998. MView: a web-compatible database search or multiple alignment viewer. Bioinformatics (Oxford, England); 14:380–1

240. Kozlov AM, Darriba D, Flouri T, Morel B, and Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics; 35:4453–5

241. Le SQ and Gascuel O. 2008. An improved general amino acid replacement matrix. Molecular Biology and Evolution; 25:1307–20

242. Consortium U. 2019. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research; 47:D506–D515

243. Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, Ong WK, Paley S, Subhraveti P, and Karp PD. 2020. The MetaCyc database of metabolic pathways and enzymes-a 2019 update. Nucleic Acids Research; 48:D445–D453

244. Bio-protocol - Improve Research Reproducibility. https://bio-protocol.org/. Accessed: 2024-02-18

245. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, Liu P, Narrowe AB, Rodrıguez-Ramos J, Bolduc B, et al. 2020. DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Research; 48:8883–900

246. Zhou Z, Tran PQ, Breister AM, Liu Y, Kieft K, Cowley ES, Karaoz U, and Anantharaman K. 2022. METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. Microbiome; 10:33

247. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, and Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Molecular Biology and Evolution; 38:5825–9

248. Söding J. 2005. Protein homology detection by HMM–HMM comparison. Bioinformatics; 21:951–60

249. Gough J, Karplus K, Hughey R, and Chothia C. 2001. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. Journal of Molecular Biology; 313:903–19

250. Yoon BJ. 2009. Hidden Markov models and their applications in biological sequence analysis. Current Genomics; 10:402–15

251. Eddy SR. 1998. Profile hidden Markov models. Bioinformatics (Oxford, England); 14:755–63

252. Madera M and Gough J. 2002. A comparison of profile hidden Markov model procedures for remote homology detection. Nucleic Acids Research; 30:4321–8

253. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A, and Eddy SR. 2015. HMMER web server: 2015 update. Nucleic Acids Research; 43:W30–W38

254. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, and Finn RD. 2018. HMMER web server: 2018 update. Nucleic Acids Research; 46:W200–W204

255. Finn RD, Clements J, and Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Research; 39:W29–W37

256. Eddy SR. 2011. Cryptogenomicon: hmmscan vs. hmmsearch speed: the numerology. https://cryptogenomicon.org/2011/05/27/hmmscan-vs-hmmsearch-speed-the-numerology/. Accessed on 2024-02-21. 2011

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

258. Garber AI, Nealson KH, Okamoto A, McAllister SM, Chan CS, Barco RA, and Merino N. 2020. FeGenie: a comprehensive tool for the identification of iron genes and iron gene neighborhoods in genome and metagenome assemblies. Frontiers in Microbiology; 11:37

259. Olmsted CN, Ort R, Tran PQ, McDaniel EA, Roden EE, Bond DR, He S, and McMahon KD. 2023. Environmental predictors of electroactive bacterioplankton in small boreal lakes. Environmental Microbiology; 25:705–20

260. Camacho C et al. 2009. BLAST+: Architecture and applications. BMC Bioinformatics; 10

261. Steinegger M and Söding J. 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nature Biotechnology; 35:1026–8

262. Edgar RC. 2021. MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping. BioRxiv :2021–6

263. Esquivel RN, Xu R, and Pohlschroder M. 2013. Novel archaeal adhesion pilins with a conserved N terminus. Journal of Bacteriology; 195:3808–18

264. Imam S, Chen Z, Roos DS, and Pohlschröder M. 2011. Identification of surprisingly diverse type IV pili, across a broad range of gram-positive bacteria. PloS one; 6:e28919

265. Haas BJ et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols; 8

266. Bushmanova E, Antipov D, Lapidus A, Suvorov V, and Prjibelski AD. 2016. rnaQUAST: a quality assessment tool for de novo transcriptome assemblies. Bioinformatics; 32:2210–2

267. Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner. Joint Genome Institute, Department of Energy; 13

268. Vernikos G, Medini D, Riley DR, and Tettelin H. 2015. Ten years of pan-genome analyses. Current Opinion in Microbiology; 23:148–54

269. Buan NR. 2018. Methanogens: pushing the boundaries of biology. Emerging Topics in Life Sciences; 2:629–46

270. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, et al. 2021. Community-led, integrated, reproducible multi-omics with anvi'o. Nature Microbiology; 6:3–6

271. Kans J. 2024. Entrez direct: e-utilities on the UNIX command line. *Entrez programming utilities help [Internet]*. National Center for Biotechnology Information (US), 2024

272. Delmont TO and Eren AM. 2018. Linking pangenomes and metagenomes: the Prochlorococcus metapangenome. PeerJ; 6:e4320

273. Utter DR, Borisy GG, Eren AM, Cavanaugh CM, and Mark Welch JL. 2020. Metapangenomics of the oral microbiome provides insights into habitat adaptation and cultivar diversity. Genome Biology; 21:1–25

274. Szulc N, Burdukiewicz M, Gasior-Glogowska M, Wojciechowski JW, Chilimoniuk J, Mackiewicz P, Schneideris T, Smirnovas V, and Kotulska M. 2021. Bioinformatics methods for identification of amyloidogenic peptides show robustness to misannotated training data. Scientific Reports; 11:8934

275. Teufel F, Almagro Armenteros JJ, Johansen AR, Gıslason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, Heijne G von, and Nielsen H. 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nature Biotechnology; 40:1023–5

276. Yu CS, Chen YC, Lu CH, and Hwang JK. 2006. Prediction of protein subcellular localization. Proteins: Structure, Function, and Bioinformatics; 64:643–51

277. Charoenkwan P, Ahmed S, Nantasenamat C, Quinn JM, Moni MA, Lio' P, and Shoombuatong W. 2022. AMYPred-FRL is a novel approach for accurate prediction of amyloid proteins by using feature representation learning. Scientific Reports; 12:7697

278. Walsh I, Seno F, Tosatto SC, and Trovato A. 2014. PASTA 2.0: an improved server for protein aggregation prediction. Nucleic Acids Research; 42:W301–W307

279. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Research; 47:D309–D314

280. Zhu Q, Kosoy M, and Dittmar K. 2014. HGTector: an automated method facilitating genome-wide discovery of putative horizontal gene transfers. BMC Genomics; 15:1–18

281. Shang J, Tang X, and Sun Y. 2023. PhaTYP: predicting the lifestyle for bacteriophages using BERT. Briefings in Bioinformatics; 24:bbac487

282. Shang J, Jiang J, and Sun Y. 2021. Bacteriophage classification for assembled contigs using graph convolutional network. Bioinformatics; 37:i25–i33

283. Shang J and Sun Y. 2022. CHERRY: a Computational metHod for accuratE pRediction of virus–pRokarYotic interactions using a graph encoder–decoder model. Briefings in Bioinformatics; 23:bbac182

284. Frost LS, Leplae R, Summers AO, and Toussaint A. 2005. Mobile genetic elements: the agents of open source evolution. Nature Reviews Microbiology; 3:722–32

285. Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, and Alm EJ. 2011. Ecology drives a global network of gene exchange connecting the human microbiome. Nature; 480:241–4

286. Thomas CM and Nielsen KM. 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nature Reviews Microbiology; 3:711–21

287. Jiang D, Armour CR, Hu C, Mei M, Tian C, Sharpton TJ, and Jiang Y. 2019. Microbiome multi-omics network analysis: statistical considerations, limitations, and opportunities. Frontiers in Genetics; 10:995

288. Shah N, Nute MG, Warnow T, and Pop M. 2019. Misunderstood parameter of NCBI BLAST impacts the correctness of bioinformatics workflows. Bioinformatics; 35:1613–4

289. National Center for Biotechnology Information (NCBI). 2019. BLAST+ 2.8.1 with New Databases and Better Performance. https://ncbiinsights.ncbi.nlm.nih.gov/2019/01/04/blast-2-8-1-with-new-databases-and-better-performance/. Accessed: 2024-04-20. 2019

290. National Center for Biotechnology Information (NCBI). 2020. New BLAST default parameters and search limits coming in September. https://ncbiinsights.ncbi.nlm.nih.gov/2020/06/18/new-blast-settings/. Accessed: 2024-04-20. 2020

291. McKinney W et al. 2011. pandas: a foundational Python library for data analysis and statistics. Python for high performance and scientific computing; 14:1–9

292. Harris CR, Millman KJ, Van Der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, et al. 2020. Array programming with NumPy. Nature; 585:357–62

293. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, et al. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods; 17:261–72

294. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. 2011. Scikit-learn: Machine learning in Python. The Journal of Machine Learning Research; 12:2825–30

295. Seabold S and Perktold J. 2010. Statsmodels: econometric and statistical modeling with python. SciPy; 7:1

296. Hunter JD. 2007. Matplotlib: A 2D graphics environment. Computing in Science & Engineering; 9:90–5

297. Waskom ML. 2021. Seaborn: statistical data visualization. Journal of Open Source Software; 6:3021

298. Inc. PT. 2015. Collaborative data science. Montreal, QC, 2015. Available from: https://plot.ly

299. Konopka A. 2009. What is microbial community ecology? The ISME Journal; 3:1223–30. DOI: 10.1038/ismej.2009.88

300. Bray JR and Curtis JT. 1957. An ordination of the upland forest communities of southern Wisconsin. Ecological Monographs; 27:325–49. DOI: 10.2307/1942268

301. Lozupone C and Knight R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Applied and Environmental Microbiology; 71:8228–35. DOI: 10.1128/AEM.71.12.8228-8235.2005

302. Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecology; 26:32–46. DOI: 10.1111/j.1442-9993.2001.01070.x

303. Faust K and Raes J. 2012. Microbial interactions: from networks to models. Nature Reviews Microbiology; 10:538–50. DOI: 10.1038/nrmicro2832

304. Breiman L. 2001. Random Forests. Machine Learning; 45:5–32. DOI: 10.1023/A:1010933404324

305. Zhang Y and al. et. 2022. Roles of granular activated carbon (GAC) and operational factors on active microbiome development in anaerobic reactors. Bioresource Technology; 343

306. Lovley DR. 2008. Extracellular electron transfer: wires, capacitors, iron lungs, and more. Geobiology; 6:225–31

307. Aroney STN, Newell RJP, Nissen J, Camargo AP, Tyson GW, and Woodcroft BJ. 2024 Jan. CoverM: Read coverage calculator for metagenomics. Version 0.7.0. 2024 Jan. DOI: 10.5281/zenodo.10531253

308. Peabody MA, Lau WYV, Hoad GR, Jia B, Maguire F, Gray KL, Beiko RG, and Brinkman FS. 2020. PSORTm: a bacterial and archaeal protein subcellular localization prediction tool for metagenomics data. Bioinformatics; 36:3043–8

309. Tatusov RL, Galperin MY, Natale DA, and Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Research; 28:33–6

310. Schultz J, Copley RR, Doerks T, Ponting CP, and Bork P. 2000. SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Research; 28:231–4

311. Consortium GO. 2019. The gene ontology resource: 20 years and still GOing strong. Nucleic Acids Research; 47:D330–D338

312. Xie J, Chen Y, Cai G, Cai R, Hu Z, and Wang H. 2023. Tree Visualization By One Table (tvBOT): a web application for visualizing, modifying and annotating phylogenetic trees. Nucleic Acids Research :gkad359

313. Thomas F, Hehemann JH, Rebuffet E, Czjzek M, and Michel G. 2011. Environmental and gut bacteroidetes: the food connection. Frontiers in Microbiology; 2:9588

314. Lam TY, Mei R, Wu Z, Lee PK, Liu WT, and Lee PH. 2020. Superior resolution characterisation of microbial diversity in anaerobic digesters using full-length 16S rRNA gene amplicon sequencing. Water Research; 178:115815

315. Vigneron A, Cruaud P, Guyoneaud R, and Goni-Urriza M. 2023. Into the darkness of the microbial dark matter in situ activities through expression profiles of Patescibacteria populations. Frontiers in Microbiology; 13:1073483

316. Logan BE. 2009. Exoelectrogenic bacteria that power microbial fuel cells. Nature Reviews Microbiology; 7:375–81

317. Maier B and Wong GC. 2015. How bacteria use type IV pili machinery on surfaces. Trends in Microbiology; 23:775–88

318. Whitman WB, Rainey F, Kämpfer P, Trujillo M, Chun J, DeVos P, Hedlund B, and Dedysh S. 2015. Bergey's manual of systematics of Archaea and Bacteria. Vol. 410. Wiley Online Library, 2015

319. Methe B, Nelson KE, Eisen J, Paulsen I, Nelson W, Heidelberg J, Wu D, Wu M, Ward N, Beanan M, et al. 2003. Genome of Geobacter sulfurreducens: metal reduction in subsurface environments. Science; 302:1967–9

320. Butler JE, Young ND, and Lovley DR. 2010. Evolution of electron transfer out of the cell: comparative genomics of six Geobacter genomes. BMC Genomics; 11:1–12

321. Ueki T. 2021. Cytochromes in extracellular electron transfer in Geobacter. Applied and Environmental Microbiology; 87:e03109–20

322. Dubé CD and Guiot SR. 2015. Direct Interspecies Electron Transfer in Anaerobic Digestion: A Review. Biogas Sci. Technol. :151. Available from: https://doi.org/10.1007/978-3-319-21993-6_4

323. Lam TY, Lee P, and Tan G. 2021 Jul. Novel Methanogenic Pathways from Anaerobic Digesters – A Potential Upgrade to Waste-to-Energy. *Proceedings of the 7th International Conference on Industrial and Hazardous Waste Management (CRETE 2021)*. Crete, Greece, 2021 Jul

324. Souza CL, Chernicharo CAL, and Melo GCB. 2012. Methane and hydrogen sulfide emissions in UASB reactors treating domestic wastewater. Water Science and Technology; 65

325. Zhang L, De Schryver P, De Gusseme B, De Muynck W, Boon N, and Verstraete W. 2008. Chemical and biological technologies for hydrogen sulfide emission control in sewer systems: a review. Water research; 42:1–12

326. Vu HP, Nguyen LN, Wang Q, Ngo HH, Liu Q, Zhang X, and Nghiem LD. 2022. Hydrogen Sulphide Management in Anaerobic Digestion: A Critical Review on Input Control, Process Regulation, and Post-Treatment. Bioresource Technology. Available from: https://doi.org/10.1016/j.biortech.2021.126634

327. Seemann T. 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics. DOI: 10.1093/bioinformatics/btu153

328. Baek M et al. 2021. Accurate prediction of protein structures and interactions using a three-track neural network. Science; 373

329. McCartney DM and Oleszkiewicz JA. 1993. Competition between methanogens and sulfate reducers: effect of COD:sulfate ratio and acclimation. Water Environment Research; 65

330. Omil F, Lens P, Visser A, Hulshoff Pol LW, and Lettinga G. 1998. Long-term competition between sulfate reducing and methanogenic bacteria in UASB reactors treating volatile fatty acids. Biotechnol Bioeng; 57

331. Lens P, Visser A, Janssen A, Pol LH, and Lettinga G. 1998. Biotechnological treatment of sulfate-rich wastewaters. Critical Reviews in Environmental Science and Technology; 28:41–88

332. Thorup C, Schramm A, Findlay AJ, Finster KW, and Schreiber L. 2017. Disguised as a sulfate reducer: growth of the deltaproteobacterium Desulfurivibrio alkaliphilus by sulfide oxidation with nitrate. mBio; 8:10–1128

333. Rajeev L, Chen A, Kazakov AE, Luning EG, Zane GM, Novichkov PS, Wall JD, and Mukhopadhyay A. 2015. Regulation of nitrite stress response in Desulfovibrio vulgaris Hildenborough, a model sulfate-reducing bacterium. Journal of Bacteriology; 197:3400–8

334. Hoeft SE, Kulp TR, Stolz JF, Hollibaugh JT, and Oremland RS. 2004. Dissimilatory arsenate reduction with sulfide as electron donor: experiments with Mono Lake water and isolation of strain MLMS-1, a chemoautotrophic arsenate respirer. Applied and Environmental Microbiology; 70:2741–7

335. Visser M and al. et. 2013. Genome analysis of desulfotomaculum kuznetsovii strain 17t reveals a physiological similarity with pelotomaculum thermopropionicum strain sit. Stand Genomic Sci; 8

336. Tasaki M, Kamagata Y, Nakamura K, Okamura K, and Minami K. 1993. Acetogenesis from pyruvate by Desulfotomaculum thermobenzoicum and differences in pyruvate metabolism among three sulfate-reducing bacteria in the absence of sulfate. FEMS Microbiology Letters; 106:259–63

337. Laanbroek HJ, Abee T, and Voogd IL. 1982. Alcohol conversion by Desulfobulbus propionicus Lindhorst in the presence and absence of sulfate and hydrogen. Archives of Microbiology; 133:178–84

338. Stams AJ, Kremer DR, Nicolay K, Weenk GH, and Hansen TA. 1984. Pathway of propionate formation in Desulfobulbus propionicus. Archives of Microbiology; 139:167–73

339. Tasaki M, Kamagata Y, Nakamura K, and Mikami E. 1992. Propionate formation from alcohols or aldehydes by Desulfobulbus propionicus in the absence of sulfate. Journal of Fermentation and Bioengineering; 73:329–31

340. Rittmann BE, Banaszak JE, and Reed DT. 2002. Reduction of Np(V) and precipitation of Np(IV) by an anaerobic microbial consortium. Biodegradation; 13

341. Holmes DE, Bond DR, and Lovley DR. 2004. Electron transfer by Desulfobulbus propionicus to Fe (III) and graphite electrodes. Applied and Environmental Microbiology; 70:1234–7

342. Ozuolmez D and al. et. 2015. Methanogenic archaea and sulfate reducing bacteria co-cultured on acetate: Teamwork or coexistence? Frontiers in Microbiology; 6

343. Duarte AG and al. et. 2018. An electrogenic redox loop in sulfate reduction reveals a likely widespread mechanism of energy conservation. Nature Communications; 9

344. Edwards MJ, Richardson DJ, Paquete CM, and Clarke TA. 2020. Role of multiheme cytochromes involved in extracellular anaerobic respiration in bacteria. Protein Science; 29:830–42

345. Liang D, Liu X, Woodard TL, Holmes DE, Smith JA, Nevin KP, Feng Y, and Lovley DR. 2021. Extracellular electron exchange capabilities of Desulfovibrio ferrophilus and Desulfopila corrodens. Environmental Science & Technology; 55:16195–203

346. Ueki T, Woodard TL, and Lovley DR. 2022. Genetic manipulation of Desulfovibrio ferrophilus and evaluation of Fe (III) oxide reduction mechanisms. Microbiology Spectrum; 10:e03922–22

347. Gray HB and Winkler JR. 2005. Long-range electron transfer. Proceedings of the National Academy of Sciences; 102:3534–9

348. Malvankar NS, Lovley DR, Beyenal H, and Babauta J. 2015. Electronic conductivity in living biofilms: physical meaning, mechanisms, and measurement methods. Biofilms in bioelectrochemical systems. Edited by Beyenal H, Babauta JT. Wiley :211–47

349. Dang H, Yu N, Mou A, Zhang L, Guo B, and Liu Y. 2022. Metagenomic insights into direct interspecies electron transfer and quorum sensing in blackwater anaerobic digestion reactors supplemented with granular activated carbon. Bioresource Technology; 352:127113

350. Lier J van, Mahmoud N, and Zeeman G. 2008. Biological Wastewater Treatment: Principles, Modelling and Design - Chapter 16: Anaerobic Wastewater Treatment. IWA Publishing, 2008. DOI: 10.1021/es00154a002

351. Yu N, Guo B, Zhang Y, Zhang L, Zhou Y, and Liu Y. 2021. Self-fluidized GAC-amended UASB reactor for enhanced methane production. Chemical Engineering Journal; 420:127652

352. Ekiel I, Sprott GD, and Patel GB. 1985. Acetate and CO2 assimilation by Methanothrix concilii. Journal of Bacteriology; 162:905–8

353. Smith KS and Ingram-Smith C. 2007. Methanosaeta, the forgotten methanogen? Trends in Microbiology; 15:150–5

354. Evans PN, Boyd JA, Leu AO, Woodcroft BJ, Parks DH, Hugenholtz P, and Tyson GW. 2019. An evolving view of methane metabolism in the Archaea. Nature Reviews Microbiology; 17:219–32

355. Silva SA, Salvador AF, Cavaleiro AJ, Pereira MA, Stams AJ, Alves MM, and Sousa DZ. 2016. Toxicity of long chain fatty acids towards acetate conversion by Methanosaeta concilii and Methanosarcina mazei. Microbial Biotechnology; 9:514–8

356. Lee J and Hwang S. 2019. Single and combined inhibition of Methanosaeta concilii by ammonia, sodium ion and hydrogen sulfide. Bioresource Technology; 281:401–11

357. Magdalena JA, Greses S, and González-Fernández C. 2019. Impact of organic loading rate in volatile fatty acids production and population dynamics using microalgae biomass as substrate. Scientific Reports; 9:18374

358. Song Y, Huang R, Li L, Du K, Zhu F, Song C, Yuan X, Wang M, Wang S, Ferry JG, et al. 2023. Humic acid-dependent respiratory growth of Methanosarcina acetivorans involves pyrroloquinoline quinone. The ISME Journal; 17:2103–11

359. Medini D, Donati C, Tettelin H, Masignani V, and Rappuoli R. 2005. The microbial pangenome. Current Opinion in Genetics & Development; 15:589–94

360. Simon M, Scheuner C, Meier-Kolthoff JP, Brinkhoff T, Wagner-Döbler I, Ulbrich M, Klenk HP, Schomburg D, Petersen J, and Göker M. 2017. Phylogenomics of Rhodobacteraceae reveals evolutionary adaptation to marine and non-marine habitats. The ISME Journal; 11:1483–99

361. Lee MD. 2019. GToTree: a user-friendly workflow for phylogenomics. Bioinformatics; 35:4162–4

362. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, and Ogata H. 2020. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics; 36:2251–2

363. Pritchard L, Glover RH, Humphris S, Elphinstone JG, and Toth IK. 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Analytical Methods; 8:12–24

364. Van Dongen S and Abreu-Goodger C. 2012. Using MCL to extract clusters from networks. Bacterial molecular networks: Methods and protocols :281–95

365. Shaiber A, Willis AD, Delmont TO, Roux S, Chen LX, Schmid AC, Yousef M, Watson AR, Lolans K, Esen ÖC, et al. 2020. Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome. Genome Biology; 21:1–35

366. Bailey TL, Elkan C, et al. 1994. Fitting a mixture model by expectation maximization to discover motifs in bipolymers. Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology

367. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, et al. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature :1–3

368. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, and Ferrin TE. 2018. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Science; 27:14–25. DOI: 10.1002/pro.3235

369. Kim M, Cha IT, Lee KE, Li M, and Park SJ. 2023. Pangenome analysis provides insights into the genetic diversity, metabolic versatility, and evolution of the genus Flavobacterium. Microbiology Spectrum; 11:e01003–23

370. Browne P, Tamaki H, Kyrpides N, Woyke T, Goodwin L, Imachi H, Bräuer S, Yavitt JB, Liu WT, Zinder S, et al. 2017. Genomic composition and dynamics among Methanomicrobiales predict adaptation to contrasting environments. The ISME Journal; 11:87–99

371. Kono T, Mehrotra S, Endo C, Kizu N, Matusda M, Kimura H, Mizohata E, Inoue T, Hasunuma T, Yokota A, et al. 2017. A RuBisCO-mediated carbon metabolic pathway in methanogenic archaea. Nature Communications; 8:14007

372. Chen L, Fang W, Chang J, Liang J, Zhang P, and Zhang G. 2022. Improvement of direct interspecies electron transfer via adding conductive materials in anaerobic digestion: mechanisms, performances, and challenges. Frontiers in Microbiology; 13:860749

373. Ruff KM and Pappu RV. 2021. AlphaFold and implications for intrinsically disordered proteins. Journal of Molecular Biology; 433:167208

374. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žıdek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with AlphaFold. Nature; 596:583–9

375. Patel GB, Sprott GD, Humphrey RW, and Beveridge TJ. 1986. Comparative analyses of the sheath structures of Methanothrix concilii GP6 and Methanospirillum hungatei strains GP1 and JF1. Canadian Journal of Microbiology; 32:623–31

376. Kopec KO and Lupas AN. 2013. β-Propeller blades as ancestral peptides in protein evolution. PLoS One; 8:e77074

377. Yan Z and Ferry JG. 2018. Electron bifurcation and confurcation in methanogenesis and reverse methanogenesis. Frontiers in Microbiology; 9:1322

378. Holmes DE, Woodard TL, Smith JA, Musat F, and Lovley DR. 2024. Electrobiocorrosion by microbes without outer-surface cytochromes. mLife; 3:110–8

379. Flores-Vargas G, Bergsveinson J, Lawrence JR, and Korber DR. 2021. Environmental biofilms as reservoirs for antimicrobial resistance. Frontiers in Microbiology; 12:766242

380. Berg G, Rybakova D, Fischer D, Cernava T, Vergès MCC, Charles T, Chen X, Cocolin L, Eversole K, Corral GH, et al. 2020. Microbiome definition re-visited: old concepts and new challenges. Microbiome; 8:1–22

381. McCurry DL, Bear SE, Bae J, Sedlak DL, McCarty PL, and Mitch WA. 2014. Superior removal of disinfection byproduct precursors and pharmaceuticals from wastewater in a staged anaerobic fluidized membrane bioreactor compared to activated sludge. Environmental Science & Technology Letters; 1:459–64

382. Monsalvo VM, McDonald JA, Khan SJ, and Le-Clech P. 2014. Removal of trace organics by anaerobic membrane bioreactors. Water research; 49:103–12

383. Lim M, Ahmad R, Guo J, Tibi F, Kim M, and Kim J. 2019. Removals of micropollutants in staged anaerobic fluidized bed membrane bioreactor for low-strength wastewater treatment. Process Safety and Environmental Protection; 127:162–70

384. Wang D, Shang J, Lin H, Liang J, Wang C, Sun Y, Bai Y, and Qu J. 2024. Identifying ARG-carrying bacteriophages in a lake replenished by reclaimed water using deep learning techniques. Water Research; 248:120859

385. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, Pratama AA, Gazitúa MC, Vik D, Sullivan MB, et al. 2021. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome; 9:1–13

386. Liu B, Zheng D, Zhou S, Chen L, and Yang J. 2022. VFDB 2022: a general classification scheme for bacterial virulence factors. Nucleic Acids Research; 50:D912–D917

387. Schmartz GP, Hartung A, Hirsch P, Kern F, Fehlmann T, Müller R, and Keller A. 2022. PLSDB: advancing a comprehensive database of bacterial plasmids. Nucleic Acids Research; 50:D273–D278

388. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA, Edalatmand A, Petkau A, Syed SA, Tsang KK, et al. 2023. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids Research; 51:D690–D699

389. Yin X, Zheng X, Li L, Zhang AN, Jiang XT, and Zhang T. 2022. ARGs-OAP v3. 0: Antibiotic-resistance gene database curation and analysis pipeline optimization. Engineering

390. Yuan L and Ju F. 2023. Potential auxiliary metabolic capabilities and activities reveal biochemical impacts of viruses in municipal wastewater treatment plants. Environmental Science & Technology; 57:5485–98

391. Yin X, Li L, Chen X, Liu YY, Lam TTY, Topp E, and Zhang T. 2023. Global environmental resistome: Distinction and connectivity across diverse habitats benchmarked by metagenomic analyses. Water research; 235:119875

392. Zhang L, Wang B, Su Y, Wu D, Wang Z, Li K, and Xie B. 2023. Pathogenic Bacteria Are the Primary Determinants Shaping PM2. 5-Borne Resistomes in the Municipal Food Waste Treatment System. Environmental Science & Technology; 57:19965–78

393. Wang C, Yin X, Xu X, Wang D, Liu L, Zhang X, Yang C, Zhang X, and Zhang T. 2024. Metagenomic Absolute Quantification of Antibiotic Resistance Genes and Virulence Factor Genes-Carrying Bacterial Genomes in Anaerobic Digesters. Water Research :121258

394. McInnes RS, Uz-Zaman MH, Alam IT, Ho SFS, Moran RA, Clemens JD, Islam MS, and Schaik W van. 2021. Metagenome-wide analysis of rural and urban surface waters and sediments in Bangladesh identifies human waste as a driver of antibiotic resistance. Msystems; 6:10–1128

395. Jiang H, Cheng H, Liang Y, Yu S, Yu T, Fang J, and Zhu C. 2019. Diverse mobile genetic elements and conjugal transferability of sulfonamide resistance genes (sul1, sul2, and sul3) in Escherichia coli isolates from Penaeus vannamei and pork from large markets in Zhejiang, China. Frontiers in Microbiology; 10:1787

396. Wu S, Dalsgaard A, Hammerum AM, Porsbo LJ, and Jensen LB. 2010. Prevalence and characterization of plasmids carrying sulfonamide resistance genes among Escherichia coli from pigs, pig carcasses and human. Acta Veterinaria Scandinavica; 52:1–7

397. Chen B, Liang X, Nie X, Huang X, Zou S, and Li X. 2015. The role of class I integrons in the dissemination of sulfonamide resistance genes in the Pearl River and Pearl River Estuary, South China. Journal of Hazardous Materials; 282:61–7

398. Shindoh S, Kadoya A, Kanechi R, Watanabe K, and Suzuki S. 2023. Marine bacteria harbor the sulfonamide resistance gene sul4 without mobile genetic elements. Frontiers in Microbiology; 14:1230548

399. Wang D, Hunt KA, Candry P, Tao X, Wofford NQ, Zhou J, McInerney MJ, Stahl DA, Tanner RS, Zhou A, et al. 2023. Cross-Feedings, Competition, and Positive and Negative Synergies in a Four-Species Synthetic Community for Anaerobic Degradation of Cellulose to Methane. mBio; 14:e03189–22

400. Abby SS, Kazemzadeh K, Vragniau C, Pelosi L, and Pierrel F. 2020. Advances in bacterial pathways for the biosynthesis of ubiquinone. Biochimica et Biophysica Acta (BBA)-Bioenergetics; 1861:148259

401. Munk P, Brinch C, Møller FD, Petersen TN, Hendriksen RS, Seyfarth AM, Kjeldgaard JS, Svendsen CA, Bunnik B van, Berglund F, et al. 2022. Genomic analysis of sewage from 101 countries reveals global landscape of antimicrobial resistance. Nature Communications; 13:7251

402. Garner E, Maile-Moskowitz A, Angeles LF, Flach CF, Aga DS, Nambi I, Larsson DJ, Burgmann H, Zhang T, Vikesland PJ, et al. 2024. Metagenomic Profiling of Internationally Sourced Sewage Influents and Effluents Yields Insight into Selecting Targets for Antibiotic Resistance Monitoring. Environmental Science & Technology; 58:16547–59

403. Che Y, Xia Y, Liu L, Li AD, Yang Y, and Zhang T. 2019. Mobile antibiotic resistome in wastewater treatment plants revealed by Nanopore metagenomic sequencing. Microbiome; 7:1–13

404. Calderón-Franco D, Sarelse R, Christou S, Pronk M, Loosdrecht MC van, Abeel T, and Weissbrodt DG. 2022. Metagenomic profiling and transfer dynamics of antibiotic resistance determinants in a full-scale granular sludge wastewater treatment plant. Water Research; 219:118571

405. Zhao Z, Zhang Y, Wang L, and Quan X. 2015. Potential for direct interspecies electron transfer in an electric-anaerobic system to increase methane production from sludge digestion. Scientific Reports; 5:11094

406. Wolferen M van, Orell A, and Albers SV. 2018. Archaeal biofilm formation. Nature Reviews Microbiology; 16:699–713

407. Karkman A, Do TT, Walsh F, and Virta MPJ. 2018. Antibiotic-Resistance Genes in Waste Water. Environmental Science & Technology; 52:13988–99. DOI: 10.1021/acs.est.8b02281. Available from: https://www.sciencedirect.com/science/article/am/pii/S0966842X1730210X

408. Ju F, Beck K, Yin X, Maccagnan A, McArdell CS, Singer HP, Johnson DR, Zhang T, and Bürgmann H. 2019. Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes. The ISME Journal; 13:346–60. DOI: 10.1038/s41396-018-0277-0. Available from: https://www.researchgate.net/publication/327849276_Wastewater_treatment_plant_resistomes_are_shaped_by_bacterial_composition_genetic_exchange_and_upregulated_expression_in_the_effluent_microbiomes

409. Wang Q, Wang M, Yang Q, Feng L, Zhang H, Wang R, and Wang R. 2025. The role of bacteriophages in facilitating the horizontal transfer of antibiotic resistance genes in municipal wastewater treatment plants. Water Research; 268. phage-associated transfer often outweighs conjugative plasmid transfer in wastewater environments:122776. DOI: 10.1016/j.watres.2024.122776. Available from: https://www.sciencedirect.com/science/article/pii/S0043135424016750

410. Skouteris G, Hermosilla D, López P, Negro C, and Blanco Á. 2012. Anaerobic membrane bioreactors for wastewater treatment: A review. Chemical Engineering Journal; 198:138–48

411. McCarty PL, Bae J, and Kim J. 2011. Domestic wastewater treatment as a net energy producer–can this be achieved? 2011

412. Castelle C, Guiral M, Malarte G, Ledgham F, Leroy G, Brugna M, et al. 2008. A new iron-oxidizing/O2-reducing supercomplex spanning both inner and outer membranes, isolated from the extreme acidophile Acidithiobacillus ferrooxidans. Journal of Biological Chemistry; 283:25803–11. DOI: 10.1074/jbc.M802496200

413. Barco RA, Emerson D, Sylvan JB, Orcutt BN, Jacobson Meyers ME, Ramírez GA, et al. 2015. New insight into microbial iron oxidation as revealed by the proteomic profile of an obligate iron-oxidizing chemolithoautotroph. Applied and Environmental Microbiology; 81:5927–37. DOI: 10.1128/AEM.01374-15

414. Liu J, Wang Z, Belchik SM, Edwards MJ, Liu C, Kennedy DW, et al. 2012. Identification and characterization of MtoA: a decaheme c-type cytochrome of the neutrophilic Fe(II)-oxidizing bacterium Sideroxydans lithotrophicus ES-1. Frontiers in Microbiology; 3:37. DOI: 10.3389/fmicb.2012.00037

415. Deng X, Dohmae N, Nealson KH, Hashimoto K, and Okamoto A. 2018. Multi-heme cytochromes provide a pathway for survival in energy-limited environments. Sci. Adv. 4. DOI: 10.1126/sciadv.aao5682

416. Light SH, Su L, Rivera-Lugo R, Cornejo JA, Louie A, Iavarone AT, Ajo-Franklin CM, and Portnoy DA. 2018. A flavin-based extracellular electron transfer mechanism in diverse Gram-positive bacteria. Nature; 562:140–4

417. Bathe S and Norris PR. 2007. Ferrous iron- and sulfur-induced genes in Sulfolobus metallicus. Applied and Environmental Microbiology; 73:2491–7. DOI: 10.1128/AEM.02589-2586

418. Croal LR, Jiao Y, and Newman DK. 2007. The fox Operon from Rhodobacter strain SW2 Promotes Phototrophic Fe (II) Oxidation in Rhodobacter capsulatus SB1003. Journal of Bacteriology; 189:1774–82. DOI: 10.1128/JB.01395-1396

419. Pitts KE, Dobbin PS, Reyes-Ramirez F, Thomson AJ, Richardson DJ, and Seward HE. 2003. Characterization of the Shewanella oneidensis MR-1 decaheme cytochrome MtrA: expression in Escherichia coli confers the ability to reduce soluble Fe(III) chelates. Journal of Biological Chemistry; 278:27758–65. DOI: 10.1074/jbc.M302582200

420. Jiao Y and Newman DK. 2007. The pio operon is essential for phototrophic Fe(II) oxidation in Rhodopseudomonas palustris TIE-1. Journal of Bacteriology; 189:1765–73. DOI: 10.1128/JB.00776-776

421. Santos TC, Silva MA, Morgado L, Dantas JM, and Salgueiro CA. 2015. Diving into the redox properties of Geobacter sulfurreducens cytochromes: a model for extracellular electron transfer. Dalton. Trans. 44:9335–44. DOI: 10.1039/C5DT00556F

422. Castelle CJ, Roger M, Bauzan M, Brugna M, Lignon S, Nimtz M, et al. 2015. The aerobic respiratory chain of the acidophilic archaeon Ferroplasma acidiphilum: a membrane-bound complex oxidizing ferrous iron. Biochim. Biophys. Acta; 1847:717–28. DOI: 10.1016/j.bbabio.2015.04.006

423. Levar CE, Chan CH, Mehta-Kolte MG, and Bond DR. 2014. An inner membrane cytochrome required only for reduction of high redox potential extracellular electron acceptors. mBio; 5:10–1128

424. Pimenta AI, Paquete CM, Morgado L, Edwards MJ, Clarke TA, Salgueiro CA, Pereira IA, and Duarte AG. 2023. Characterization of the inner membrane cytochrome ImcH from Geobacter reveals its importance for extracellular electron transfer and energy conservation. Protein Science; 32:e4796

425. Joshi K, Chan CH, and Bond DR. 2021. Geobacter sulfurreducens inner membrane cytochrome CbcBA controls electron transfer and growth yield near the energetic limit of respiration. Molecular Microbiology; 116:1124–39

426. Jiménez Otero F, Chadwick GL, Yates MD, Mickol RL, Saunders SH, Glaven SM, Gralnick JA, Newman DK, Tender LM, Orphan VJ, et al. 2021. Evidence of a streamlined extracellular electron transfer pathway from biofilm structure, metabolic stratification, and long-range electron transfer parameters. Applied and Environmental Microbiology; 87:e00706–21

427. Liu X, Wang S, Xu A, Zhang L, Liu H, and Ma LZ. 2019. Biological synthesis of high-conductive pili in aerobic bacterium Pseudomonas aeruginosa. Applied Microbiology and Biotechnology; 103:1535–44

428. Ewels P, Magnusson M, Lundin S, and Käller M. 2016. MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report. Bioinformatics; 32. Accessed: Dec 13, 2023:3047–8. DOI: 10.1093/bioinformatics/btw354. Available from: https://doi.org/10.1093/bioinformatics/btw354

429. Burdukiewicz M, Sobczyk P, Rodiger S, Duda-Madej A, Mackiewicz P, and Kotulska M. 2017. Amyloidogenic motifs revealed by n-gram analysis. Scientific Reports; 7:12961

Appendix A

Supplementary Notes

A.1 Active biomass estimation in SAF-MBR

The following content is the result of discussion with Chungheon Shin. Experiments result was provided by Shin with consent.

Assessing active microbial biomass in the SAF-MBR is challenging. The bulk of the biomass resides within the AFBR's GAC biocarrier. It was impractical to directly measure active biomass in the biocarrier pores (X_a^f) using PCR-based methods. Attempts to directly extract microbial biomass from the GAC using sonication and the PowerSoil biofilm extraction kit were unreliable due to the GAC's pore size (approximately 10 μm), which is much smaller than the kit's bead size (about 100 μm). Therefore, an alternative method was adopted to estimate active biomass concentrations inside the GAC for both X_a^f and X_a^s as described below.

Experimental Setup

Separate batch tests assessed active biomass concentrations on GAC and in the liquid phase. Serum bottles for the X_a^f tests included 18.75 g of GAC from the pilot-scale SAF-MBR 2.0, while those for X_a^s contained 15 mL of bulk solution from the same system. These tests utilised a sodium acetate solution (1000 mgCOD/L) and a control with no carbon source to determine CH_4 production rates, helping to estimate active acetoclastic methanogen concentrations and necessary active biomass for MCRT calculations.

Substrate Degradation Rate

The substrate degradation rate during these tests ($\frac{dS}{dt}$) is given by:

$$\frac{dS}{dt} = -\hat{\mu}/Y \frac{S}{K+S} X_a \tag{S1}$$

For high substrate concentrations ($S \gg K$), the equation simplifies to:

$$\frac{dS}{dt} = -\hat{\mu}/Y X_a \quad (S \gg K) \tag{S2}$$

This simplification allows expressing the CH_4 production rate as:

$$\frac{dV_{CH_4}}{dt} = \hat{\mu}(1/Y - 1)X_a \tag{S3}$$

Under these conditions, the initial CH_4 production rate is directly proportional to X_a, assuming $\frac{S}{S+K}$ nears unity.

The initial CH_4 production rates ($\frac{dV_{CH_4}}{dt}$) were measured at 0.15 gCOD/d for X_a^f and 0.0022 gCOD/d for X_a^s. Using the kinetic data from Table 3 in a previous study [32], the active biomass concentrations were determined to be $X_a^f = 1878$ mgCODcell/kg GAC (wet mass) and $X_a^s = 28$ mgCODcell/L in the bulk solution. Thus the active biomass in GAC accounted for 98.5% of the total active biomass.

The pilot-scale SAF-MBR 2.0 system contained 2028 kg of wet GAC (M_{GAC}) and had an effective volume (V) of 4500 L. Based on these figures, the active biomass was primarily attached to the GAC (97%, $f_{X_a}^f = 0.97$) with a smaller fraction in the bulk solution (3%, $f_{X_a}^s = 0.03$) as dispersed biomass.

A.2 Refinement of Metagenome-Assembled Genomes

This study generated 1,796 metagenome-assembled genomes (MAGs) from twelve metagenomic datasets derived from three distinct types of samples: bulk, granular activated carbon (GAC), and polyvinylidene fluoride (PVDF). The primary aim was to determine whether the medium affects the growth patterns of microbes capable of extracellular electron transfer (EET), focusing on their predilection for bulk solution, conductive GAC, or non-conductive PVDF.

Initially, 651 MAGs satisfied the quality criteria, including a minimum of 50% completeness and a maximum redundancy of 10%. Refinement was executed using the BinSPreader software, which operated across eight modes. Ultimately, the result from method 'BinSPreader-Rcorr' without lineage data was selected for downstream analysis, as evidenced by their F1 scores. Notably, the 'Rcorr-pmax' mode introduced significant redundancy and was discarded from further analysis. This increase in redundancy could be attributed to numerous phylogenetically neighbouring MAGs, a common characteristic observed in environmental samples.

Details on the outcomes of the refinement process can be accessed at the following Github Repository. This repository provides a rank index of the various refinement strategies concerning genome completeness, redundancy (contamination), and F1 score. Post-refinement, the number of MAGs meeting the quality standards increased by 10.93%, bringing the total to 796 MAGs eligible for further analysis.

A.2.1 Four Modes of Refinement

Four operational modes of the BinSPreader tool were utilised:

1. Propagation mode

2. Correction mode

3. Correction mode with multiple assignments

4. Correction mode with sparse propagation

BinSPreader facilitates two primary approaches for refining bin assignments: correction (denoted as 'Rcorr') and propagation ('Rprop'). The correction approach employs assembly graph data to potentially amend initial bin assignments. This method may rectify misassignments, particularly in shorter contigs, but it also introduces a risk of increased redundancy in cases where initial binning is not highly pure. In contrast, the propagation approach is markedly conservative, assigning only unbinned sequences to adjacent bins, thereby preserving the integrity of initial scaffold assignments. This mode is particularly advantageous when dealing with binning that exhibits minimal contamination, aiming to integrate short unbinned scaffolds into existing bins.

Additionally, BinSPreader operates in two distinct running modes: the default mode and the sparse mode. The sparse mode, activated via the "–sparse-propagation" option, is designed for scenarios where the total length of binned contigs is significantly lower than that of the total assembly. It restricts binning propagation to a predefined graph distance threshold, making it especially pertinent when initial binning has undergone subsequent filtering or processing, or when there is a substantial presence of unbinned sequences.

Moreover, BinSPreader supports both single and multiple bin assignments. In the multiple assignment mode, a contig may be allocated to several bins simultaneously. This feature is particularly beneficial for appending highly conserved sequences, such as those encoding housekeeping genes (e.g., rRNA), to recovered MAGs.

This refinement enabled identifying key extracellular electron transfer (EET) genes. However, the process risked introducing redundancy. To maintain the integrity of the MAGs and ensure robust analysis, it was crucial to keep this redundancy below 10%, balancing enhanced gene recovery with the reliability of the genomic assemblies.

The refined MAGs achieved an average completeness of 78.31% ($\pm$14.01%), with a range from 50% to 100%. The redundancy across these genomes was minimal, averaging at 2.32% ($\pm$2.17%), with half of the MAGs exhibiting a redundancy below 1.75% (Fig. A.1).

The refinement process not only improved the completeness of the MAGs but also facilitated the identification of genes associated with EET capabilities. In this context, multiheme cytochromes (MHC) and electrically conductive pili (e-pili) were utilised as indicators of EET functionality, based on their experimentally confirmed electrical conductivity [85]. Subsequent pHMM searches and meticulous manual filtering led to the identification of 860 e-pili genes and 1,365 MHC genes, present within 237 and 256 MAGs respectively, with 72 MAGs containing both types of genes (Fig. A.2). The refinement increased the proportion of discovered EET genes by 8.51%, from 23.05% to 31.56%, enhancing our understanding of EET mechanisms within these bioreactors. Ultimately, 421 MAGs, representing 51.15% of the quality-assured MAGs (hereafter referred to as EET+ MAGs), were found to harbour EET genes, indicating a significant prevalence of these genes in the studied environments.

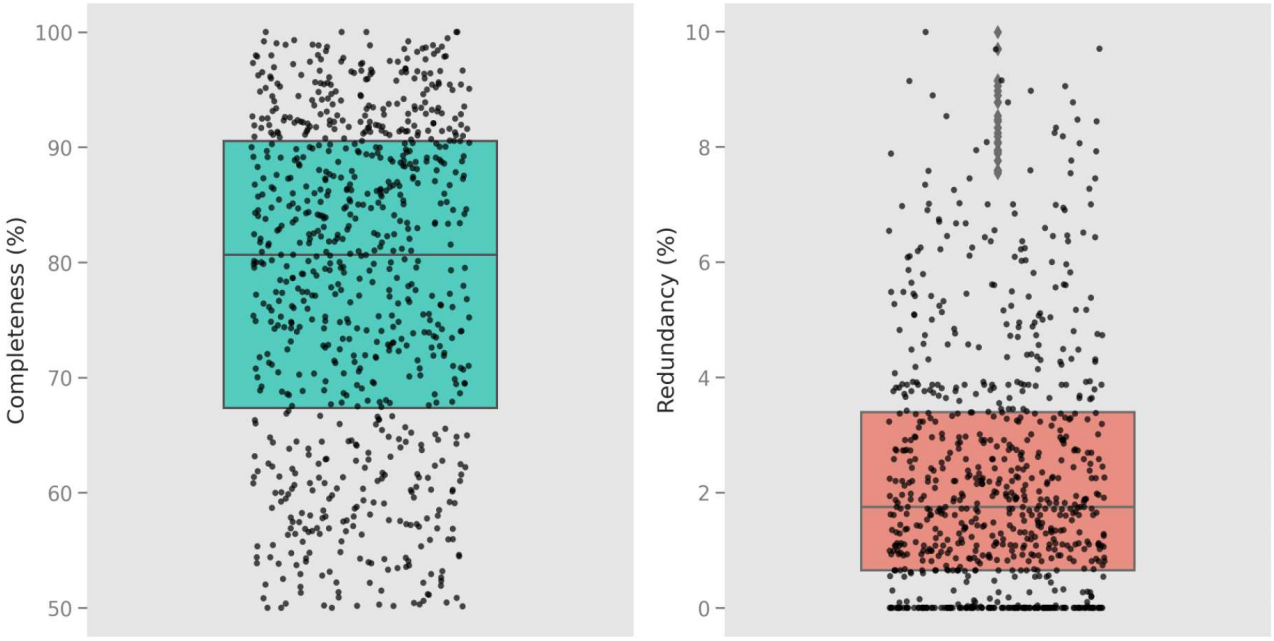


Figure A.1: Scattered boxplots of the refined MAGs' completeness and redundancy.

A.3 Phylogenetic trees oddity

This section compares two popular tools for taxonomy assignment of MAGs, PhyloPhlAn and GTDB-TK [227].

Both pipelines are designed for accurate and rapid large-scale microbial genome characterisation. However, notable differences were observed in the clustering of genomes at the phylum level between the PhyloPhlAn and GTDB-TK phylogenetic trees (Fig. A.3).

In the GTDB-TK-based tree, phylum-level clustering was more consistent, with genomes of the same phylum generally positioned closer together. In contrast, the PhyloPhlAn-based tree displayed a more scattered distribution of phyla. Additionally, GTDB-TK assigned a higher number of diverse phyla, indicating its suitability for taxonomy assignment in largely uncultured microbial groups. This difference likely arises because PhyloPhlAn relies on species-level genome bins as guides for metagenomic analysis, which may lead to genome misplacement when environmental genomes do not reach species-level identification. Consequently, GTDB-TK was selected for downstream analysis.

A.4 E-pili identification

This Python script analyses sequences of putative type IV pilin proteins to identify and characterise electrically conductive pili, often referred to as e-pili. The script processes input sequences from a FASTA file, identifies the mature pilin sequence based on a specific cleavage motif, counts aromatic amino acids in the mature sequence, calculates the percentage of aromatic amino acids, and identifies the largest gap between these aromatic residues. This script is adapted from Python-scripts/Pilinstats.py at master · GlassLabGT/Python-scripts (github.com) and changing

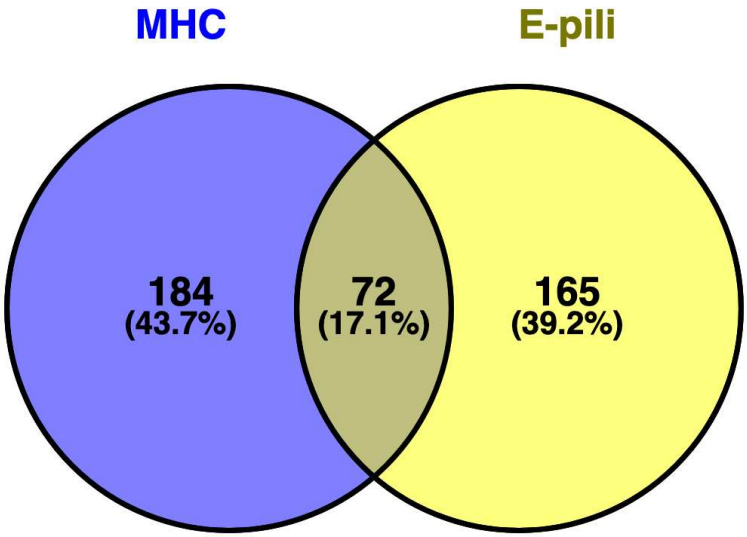


Figure A.2: Venn diagram of MHC and e-pili genes identified in the MAGs.

"Bio.Alphabet" module to better fit the current FASTA file format. The steps are as follow:

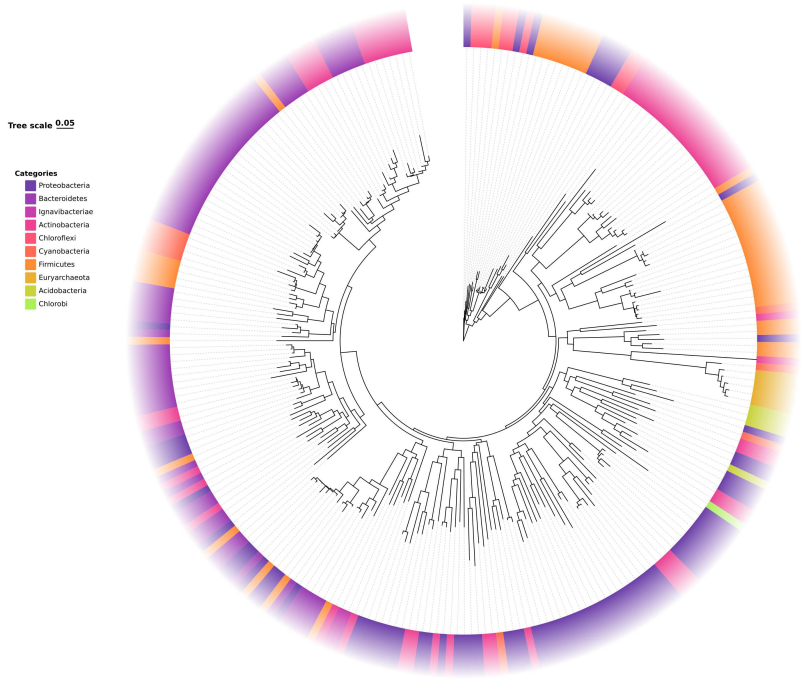
1. Mature Pilin Sequence Identification:

    - The `mature_pilin_sequence` function looks for a specific amino acid motif (`[GAS][ACFGILMNPQSTVWY]{4}[DE]`) that signifies the cleavage site for the mature pilin sequence. The `[GAS]` represents glycine, alanine, or serine, followed by any four amino acids from the specified set, ending with aspartic acid or glutamic acid.
    - When the motif is found, the function prints the position and sequence of the cleavage site and returns everything after this site as the mature pilin sequence.

2. Aromatic Amino Acid Counting: The `count_aromatics` function counts the number of tyrosines (Y), phenylalanines (F), tryptophans (W), and histidines (H) in the mature sequence. These amino acids are considered aromatic due to their aromatic side chains.

3. Calculating the Largest Aromatic-Free Gap: The `aromatic_gap` function calculates the largest gap between the aromatic amino acids in the mature sequence, which is a measure of how clustered or spread out these residues are.

4. Parsing Command-Line Arguments: The script uses `argparse` to define how it should be



(a)



(b)

Figure A.3: Phylogenetic tree constructed using MAGs that contain EET genes. The taxonomy (phylum level) was assigned based on the result done by (a) PhyloPhlAn and (b) GTDB-TK.

called from the command line, including specifying the input FASTA file of putative type IV pilin proteins and the output file for the results.

5. Processing the Input File:

    - For each sequence in the input FASTA file, the script generates the mature sequence, counts the aromatics, calculates the mature length and aromatic percentage, and finds the largest aromatic-free gap.
    - These data are written to the output file in a tab-delimited format, including sequence ID, counts of each aromatic amino acid, total aromatic count, mature pilin length, percentage of aromatics, and the largest gap.

6. Identifying e-pili Candidates: Sequences with an aromatic percentage greater than a threshold (in this case, implicitly considered with the comment about appending to `epili` if the aromatic percentage is high) are considered candidates for being e-pili. These are saved to a separate FASTA file.

```
1
2  #!/usr/bin/env python3
3
4  from Bio import SeqIO
5  import sys
6  import re
7  import argparse
8  from argparse import RawTextHelpFormatter
9
10 # Function that generates the sequence of the mature pilin protein
11 def mature_pilin_sequence(seq):
12     m = re.search(r'[GAS][ACFGILMNPQSTVWY][ACFGILMNPQSTVWY][ACFGILMNPQSTVWY
       ↪ ][ACFGILMNPQSTVWY][DE]', seq)
13     if m:
14         cleavage_motif = m.group()
15         cleavage_pos = m.start()
16         print('Found cleavage site at position %i with sequence %s' % (
       ↪ cleavage_pos, cleavage_motif))
17         mature_seq = seq[cleavage_pos + 1:]
18         return mature_seq
19
20 # Function that counts the aromatic amino acids in the mature pilin sequence
21 def count_aromatics(protein):
22     tyrosines = protein.upper().count('Y')
23     phenylalanines = protein.upper().count('F')
24     tryptophans = protein.upper().count('W')
25     histidines = protein.upper().count('H')
26     aromatics = tyrosines + phenylalanines + tryptophans + histidines
27     return aromatics, tyrosines, phenylalanines, histidines, tryptophans
28
```

```
# Function that calculates the largest aromatic free gap in the amino acid sequence
def aromatic_gap(protein):
    gaps = re.split(r'[YFWH]', protein.upper())
    lengths = [len(gap) for gap in gaps]
    return max(lengths)

# Script help and usage
parser = argparse.ArgumentParser(
    description='This script takes an amino acid fasta file of putative type IV pilin proteins and writes a tab-delimited file containing the mature pilin sequence, the mature length, and the number and percentage of aromatic amino acids within the mature length of the pilin.',
    formatter_class=RawTextHelpFormatter)
parser.add_argument('inputfile', help='Amino acid fasta file containing putative type IV pilin proteins')
parser.add_argument('outfile', help='Output file containing your pilin stats in a tab-delimited table')
args = parser.parse_args()

# Define the input and output files
Inputfile = args.inputfile
Output_name = args.outfile

# Open outfile to write
with open(Output_name + '.txt', 'w') as output:
    headers = 'SequenceID\tAromatic_count\tMature_pilin_length\tPercent_aromatics\tLargest_gap\tTyrosines\tPhenylalanines\tHistidines\tTryptophans\tFull_sequence'
    output.write(headers + '\n')

    # Open our fasta input file and iterate over fasta file to parse out sequences
    epili = []
    with open(Inputfile, 'r') as f:
        for record in SeqIO.parse(f, "fasta"):
            print('Analyzing sequence: ' + record.id)

            # Generate the full sequence
            full_sequence = str(record.seq).upper()

            # Generate the mature sequence
            mature_sequence = mature_pilin_sequence(full_sequence)
            if not mature_sequence:
                continue
            print('Mature sequence is: ' + mature_sequence)

            # Count the aromatics of the mature sequence
```

```
68              aromatics, tyrosines, phenylalanines, histidines, tryptophans =
    ↪ count_aromatics(mature_sequence)
69
70              # Determine the length of the mature sequence
71              mature_length = len(mature_sequence)
72
73              # Calculate the percent of aromatic amino acids in the mature
    ↪ sequence
74              aromatic_percent = (aromatics / mature_length) * 100 if
    ↪ mature_length > 0 else 0
75
76              # Calculate largest aromatic free gap
77              free_gap = aromatic_gap(mature_sequence)
78
79              print('Aromatic percentage is: ' + str(round(aromatic_percent,
    ↪ 2)))
80
81              # Writes output in tab delimited format
82              output_line = f"{record.id}\t{aromatics}\t{mature_length}\t{
    ↪ round(aromatic_percent, 2)}\t{free_gap}\t{tyrosines}\t{phenylalanines
    ↪ }\t{histidines}\t{tryptophans}\t{record.seq}\n"
83              output.write(output_line)
84
85              if aromatic_percent > 9.80:
86                  epili.append(record)
87
88      # Write e-pili sequences to a fasta file
89      with open(Output_name + "_epili.fasta", "w") as epili_file:
90          SeqIO.write(epili, epili_file, "fasta")
91
92  print("Processing complete. Results saved to output files.")
```

Listing A.1: PilinStat.py

Appendix B

Supplementary Tables

Table B.1: DNA quality of the biomass samples

Sample ID	Sample Volume (ml)	Sample Mass (g)	Final Volume (μl)	Nucleic Acid (ng/μl)	A_{260}/A_{280}	A_{260}/A_{230}
inha_p1_afmbr_bulk	50	–	100	110	1.87	2.02
inha_p2_afbr_bulk	50	–	100	102.9	1.89	2.11
inha_p2_afmbr_bulk	50	–	100	80	1.85	2.13
inha_p3_afmbr_bulk	50	–	100	83.3	1.87	2.2
inha_p1_afbr_gac	–	0.4	50	78.6	1.78	1.85
inha_p1_afmbr_gac	–	0.4	50	90.9	1.86	2.21
inha_p2_afbr_gac	–	0.4	50	107	2.03	1.92
inha_p2_afmbr_gac	–	0.4	50	84.6	1.89	2.17
inha_p3_afbr_gac	–	0.4	50	32.5	1.82	1.88
inha_p3_afmbr_gac	–	0.4	50	42.7	1.84	1.98
sf_fp_safmbr_gac	–	0.4	50	106.8	1.87	2.18
inha_control_pvdf	–	0.4	50	28.6	1.84	2

Table B.2: RNA quality of the biomass sample

Sample ID	Final Volume (µL)	Nucleic Acid (ng/µL)	RIN	23S/16S
inha_p3_afbr_gac	35	219	7.9	1.2
inha_p3_afmbr_gac	35	268	8.1	1.3
sf_fp_safmbr_gac	35	330	8.9	1.3

Table B.3: General statistics of the metagenomes collected in this study. % Dups is the percentage of duplicated sequences. % GC is the percentage of GC content, which refers to the percentage of nitrogenous bases in a DNA or RNA molecule that are either guanine (G) or cytosine (C). M seq is sequence in million.

Sample Name	% Dups	% GC	Average Read Length	Median Read Length	% Failed	M Seqs
inha_p0_pvdf_1	28.30%	54%	149 bp	150 bp	0%	71.6
inha_p0_pvdf_2	27.90%	54%	148 bp	150 bp	0%	71.6
sf_fp_safmbr_gac_1	29.30%	54%	149 bp	150 bp	0%	125.5
sf_fp_safmbr_gac_2	28.40%	54%	148 bp	150 bp	0%	125.5
inha_p2_afbr_bulk_1	7.50%	46%	147 bp	150 bp	0%	28.7
inha_p2_afbr_bulk_2	7.70%	46%	140 bp	150 bp	0%	28.7
inha_p2_afbr_gac_1	7.30%	51%	147 bp	150 bp	0%	31.4
inha_p2_afbr_gac_2	7.50%	50%	139 bp	150 bp	0%	31.4
inha_p2_afmbr_bulk_1	9.40%	44%	148 bp	150 bp	0%	33.3
inha_p2_afmbr_bulk_2	9.60%	44%	144 bp	150 bp	0%	33.3
inha_p2_afmbr_gac_1	7.20%	51%	147 bp	150 bp	0%	33.2
inha_p2_afmbr_gac_2	7.40%	50%	139 bp	150 bp	0%	33.2
inha_p3_afbr_gac_1	12.30%	53%	149 bp	150 bp	0%	115.1
inha_p3_afbr_gac_2	12.10%	53%	148 bp	150 bp	0%	115.1
inha_p3_afmbr_bulk_1	15.10%	47%	148 bp	150 bp	0%	115.6
inha_p3_afmbr_bulk_2	15.00%	47%	147 bp	150 bp	0%	115.6
inha_p3_afmbr_gac_1	12.20%	52%	149 bp	150 bp	0%	129.6
inha_p3_afmbr_gac_2	12.10%	52%	149 bp	150 bp	0%	129.6
inha_p1_afbr_gac_1	6.70%	52%	148 bp	150 bp	0%	31.2
inha_p1_afbr_gac_2	6.60%	52%	138 bp	148 bp	0%	31.2
inha_p1_afmbr_bulk_1	9.50%	41%	149 bp	150 bp	0%	37.5
inha_p1_afmbr_bulk_2	9.60%	41%	142 bp	150 bp	0%	37.5
inha_p1_afmbr_gac_1	6.10%	52%	148 bp	150 bp	0%	31.3
inha_p1_afmbr_gac_2	6.10%	52%	139 bp	150 bp	0%	31.3

Table B.4: List of Type C Multiheme Cytochrome Genes for Iron Reduction and Oxidation. This set of genes was used for HMM database training.

Gene	Current FeGenie Category	Operon, Microbe	Using These Sequences or Pfam Family(ies)	Ref
Cyc1	Iron oxidation	*Acidithiobacillus ferrooxidans*	Sequence from IMG: 643531504	[412], [413]
Cyc2	Iron oxidation	*Acidithiobacillus ferrooxidans*, *Mariprofundus ferroxydans*, *Leptospirillum rubarum*	Sequence from IMG: 643531503	[412], [413]
CymA	Iron reduction/oxidation	cymA (*Sideroxydans lithotrophicus ES-1*)	Sequence from IMG: 646690174	[414]
DFE_0448	Iron reduction	DFE_0448-0451	Sequence from NZ_AP017378.1	[415]
DFE_0449	Iron reduction	DFE_0448-0451	Sequence from NZ_AP017378.1	[415]
DmkA	Iron reduction	FmnA-dmkA-fmnB-pplA-ndh2-eetAB-dmkB (*Listeria monocytogenes*)	Sequence from NCBI AEO07617.1	[416]
EetA	Iron reduction	FmnA-dmkA-fmnB-pplA-ndh2-eetAB-dmkB (*Listeria monocytogenes*)	Sequence from NCBI AEO07621.1	[416]
FoxA	Iron oxidation	FoxABC (*Sulfolobus tokodaii*)	Sequence from IMG: 638195097	[417]
FoxE	Iron oxidation	FoxEYZ (*Rhodobacter ferroxidans*)	Sequence from IMG: 647131574	[418]

Continued on next page

Gene	Current FeGenie Category	Operon, Microbe	Using These Sequences or Pfam Family(ies)	Ref
MtrA	Iron reduction	MtrCAB (*Shewanella oneidensis MR-1*), pioABC (*Rhodopseudomonas palustris TIE-1*)	Sequence from IMG: 637343663	[419], [420]
OmcS	Iron reduction	*Geobacter sulfurreducens*	Sequence from IMG: 637127194	[421]
PplA	Iron reduction	FmnA-dmkA-fmnB-pplA-ndh2-eetAB-dmkB (*Listeria monocytogenes*)	Sequence from NCBI AEO07619.1	[416]
Sulfocyanin	Iron oxidation	*Ferroplasma acidarmanus*, *Sulfolobus acidocaldarius*, *Ferroplasma acidiphilum*	Sequence from IMG: 638194880	[422]
ImcH	Iron reduction	*Geobacter sulfurreducens*	Sequence from NCBI: WP_119333799.1	[423], [424]
CbcA	Iron reduction	*Bdellovibrionales bacterium*	Sequence from NCBI: GWB1_55_8	[425]
ExtA_pcc_cyt	Iron reduction	*Geobacter sulfurreducens*	Sequence from NCBI: WP_119331864	[426]
ExtC_pcc_lp_cyt	Iron reduction	*Geobacter sulfurreducens*	Sequence from NCBI: WP_039643292	[426]

Table B.5: List of e-pili genes identified in literature review. This set of genes was used for HMM database training.

GenBank Accession / Locus Tag	Species and Abbreviation	Mature Length (aa)	% Aromatics	Max Aromatic-Free Gap (aa)	Conductive? (Y/N/U)	Reference
P04739	*Pseudomonas aeruginosa* (Pa)	143	5.6	45	N	[427]
WP_011070777	*Shewanella oneidensis* (So)	126	5.6	42	N	[135]
AMM39924	*Desulfofervidus auxilii* (Daux)	206	6.8	51	N	[38]
Gura_2677	*Geobacter uraniireducens* (Gu)	193	9.1	53	N	[159]
GSU1496	*Geobacter sulfurreducens* (GsPilA)	61	9.8	22	Y	[135]
Glov_2096	*Geobacter lovleyi* (Glov)	60	10	22	U	
KN63_04755	*Smithella* sp. F21 (F21)	109	10.1	22	U	
WP_028893309	*Syntrophorhabdus aromaticivorans* (Sar)	128	10.2	27	U	
WP_084710848	*Syntrophaceticus schinkii* (Synsch)	105	10.5	25	U	
WP_039645155	*Geobacter soli* (Gsoli)	65	10.8	22	U	
Ppro_1656	*Pelobacter propionicus* (Ppro)	64	10.9	22	U	
WP_027714750	*Desulfuromonas* sp. TF (TF)	64	10.9	22	U	

Continued on next page

GenBank Accession / Locus Tag	Species and Abbreviation	Mature Length (aa)	% Aromatics	Max Aromatic-Free Gap (aa)	Conductive? (Y/N/U)	Reference
SYN_00814	*Syntrophus aciditrophicus* (Sa)	110	10.9	22	Y	[39]
DaAHT2_2283	*Desulfurivibrio alkaliphilus* (Da)	182	11	27	Y	[38]
Mhun_3140	*Methanospirillum hungatei* (Mh)	164	11	35	Y	[115]
SPF50166	*Syntrophobacter* sp. SbD1 (SbD1)	145	11.7	21	U	
WP_039742239	*Geobacter pickeringii* (Gpick)	60	11.7	22	U	
OR1_03834	*Geobacter* sp. OR-1	64	12.5	22	U	
Flexsi_2288	*Flexistipes sinusarabici* (Fs)	60	13.3	22	Y	[38]
Calni_0149	*Calditerrivibrio nitroreducens* (Cn)	119	13.4	22	Y	[38]
WP_092349663	*Desulfuromusa kysingii* (Desky)	60	15	22	U	
HRM2_27700	*Desulfobacterium autotrophicum* (Deauto)	59	15.3	22	U	
Gmet_1399	*Geobacter metallireducens* (Gm)	59	15.3	22	Y	[162]

Table B.6: Generalisation of cell membrane structure according to the latest version of Bergey's Manual of Systematics of Archaea and Bacteria [318]

Phylum Name	Assignment for Subcellular Localisation Prediction	Membrane Structure Description
Euryarchaeota	Archaeal	Unique membrane structures that differ from those of bacteria. They have ether-linked lipids with isoprenoid chains and can have monolayer or bilayer membranes.
Thaumarchaeota	Archaeal	As archaea, they have unique lipid membranes similar to other archaeal groups, with ether-linked isoprenoid chains.
Proteobacteria	Gram-negative	Gram-negative, with an outer membrane containing lipopolysaccharide (LPS) and phospholipids, a peptidoglycan layer, and a periplasm.
Fusobacteria Chloroflexi Chlorobi Cyanobacteria Bacteroidetes Synergistetes Elusimicrobia Chlamydiae Chrysiogenetes Ignavibacteriae	Gram-negative	Typically Gram-negative, with an outer membrane containing LPS, a thin peptidoglycan layer, and a periplasmic space.
Actinobacteria	Gram-positive	Gram-positive, with some exceptions in the Corynebacteriales order which have a unique type of outer membrane.
Firmicutes	Gram-positive	Mostly Gram-positive, but there are diderm (double membrane) Firmicutes that challenge the traditional Gram-positive/Gram-negative divide.

Table B.7: Shannon diversity index and richness across bioreactor samples

Sample ID	Media Type	Shannon Diversity Index	Richness
inha_control_pvdf	PVDF	2.41	21
inha_p1_afbr_gac	GAC	2.20	15
inha_p1_afmbr_bulk	Bulk	1.54	7
inha_p1_afmbr_gac	GAC	1.81	9
inha_p2_afbr_bulk	Bulk	1.23	6
inha_p2_afbr_gac	GAC	1.83	10
inha_p2_afmbr_bulk	Bulk	1.90	11
inha_p2_afmbr_gac	GAC	1.80	11
inha_p3_afbr_gac	GAC	1.73	18
inha_p3_afmbr_bulk	Bulk	1.45	13
inha_p3_afmbr_gac	GAC	1.77	21
sf_fp_safmbr_gac	GAC	2.10	19

Table B.8: ANOVA Results and Tukey's HSD Test for Shannon Index

Analysis	df	sum_sq	mean_sq	F	PR(¿F)
ANOVA					
C(SampleType)	1.0	0.332514	0.332514	6.932424	0.027227
Residual	9.0	0.431686	0.047965	NaN	NaN
Tukey's HSD					
Bulk vs. GAC		meandiff = 0.3614		p-adj = 0.0272	
		lower = 0.0509, upper = 0.672		reject = True	

Table B.9: Electron Donors and EET Mechanisms of Microbes

Microbe	Phylum	Habitat	Electron Donors	EET Mechanism
Syntrophus aciditrophicus	Thermodesulfoba	Anaerobic digesters	Fatty acids, benzoate	Electrically conductive pili (e-pili)
Electrothrix and Electronema	Desulfobacterota	Diverse anaerobic soils and sediments	Sulfide	Conductive cable for intercellular electron transport
Geothrix fermentans	Acidobacteriota	Subsurface	Fatty acids	Cytochromes, electron shuttle
Geovibrio, Deferribacter, Flexistipes	Deferribacterota	Subsurface	Organic acids, amino acids, H_2	Cytochromes, e-pili

Continued on next page

Microbe	Phylum	Habitat	Electron Donors	EET Mechanism
Ferroglobus and Geoglobus	Thermoproteota	Hydrothermal systems	Organic acids, aromatics, H_2	Cytochromes
Pyrobaculum	Thermoproteota	Hydrothermal zones	Hydrogen (H_2), peptone	Unknown (no cytochromes)
Pyrodictium delaneyi	Thermoproteota	Hydrothermal zones	Hydrogen (H_2)	Cytochromes
Acidianus sulfidivorans	Thermoproteota	Solfatara	Hydrogen (H_2)	Unknown
Acidimicrobium ferrooxidans	Actinomycetota	Hot springs	Glycerol	Unknown
Prosthecochloris aestuarii	Chlorobiota	Photosynthetic mats	Light-dependent DIET, S_0, sulfide	Unknown, possibly cytochromes
Methanothrix	Methanothrichota	Anaerobic soils, sediments, digesters	DIET	Unknown (no cytochromes)
Methanosarcina type II spp.	Methanosarcinota	Anaerobic soils, sediments	C-1 compounds, DIET	Cytochrome
Desulfosporosinus acididurans	Bacillota	Moderately acidic sediments	Glycerol	Unknown, possibly cytochromes
Listeria monocytogenes	Bacillota	Soil, water, gut	Fermentable substrates	Flavoprotein, electron shuttle
Faecalibacterium prausnitzii	Bacillota	Human intestine	Fermentable compounds	Electron shuttle
Thermincola	Bacillota	Hydrothermal springs, anaerobic digesters	Acetate, H_2, amino acids	Cytochrome conduit
Shewanella oneidensis	Pseudomonadota	Aquatic sediments	Lactate, pyruvate, H_2	Porin–cytochrome conduit, electron shuttle
Rhodoferax ferrireducens	Burkholderiota	Subsurface	Organic acids, aromatics, sugars	Porin–cytochrome conduit, cytochromes
Desulfuromonas acetoxidans	Desulfobacterota	Anaerobic soils, sediments	Acetate	Cytochromes

Continued on next page

Microbe	Phylum	Habitat	Electron Donors	EET Mechanism
Pseudomonas aeruginosa	Pseudomonadota	Various environments (e.g., water, soil)	Organic acids	Pyocyanin (electron shuttle)

Table B.10: Quality control and phylogenetic classification of the MAGs.

Bin	Taxonomic Rank	Closest Assignment	Total Length	Num Contigs	N50	GC Content (%)	CheckM Completeness (%)	CheckM Redundancy (%)
bin 35	family	Desulfovibrionaceae	2 349 486	14	286 591	54.89	96.15	0.00
bin 61	family	Desulfovibrionaceae cluster FLUQ01	3 255 699	306	13 909	59.93	90.40	5.39
bin 112	family	Desulfovibrionaceae	2 613 420	112	41 373	56.27	91.97	1.58
bin 142	genus	Desulfobulbus	3 965 222	268	22 208	58.92	90.64	9.73
bin 172	genus	Desulfovibrio	2 237 464	114	31 839	49.92	92.60	0.00
bin 177	genus	Desulfovibrio	2 238 124	52	93 305	51.97	96.33	1.18
bin 195	family	Desulfovibrionaceae	2 440 641	136	29 167	58.52	92.21	1.31
bin 7	genus	Syntrophobacter	3 146 870	151	34 168	57.25	93.55	4.19
bin 94	genus	Syntrophomonas	2 351 445	247	12 093	51.16	90.94	2.75
bin 97	genus	Smithella sp F21	3 401 515	212	26 703	58.12	96.13	4.84
bin 87	genus	Geobacter	4 036 425	481	10 623	57.06	84.86	5.86
bin 28	genus	Methanothrix	1 474 962	229	7684	53.32	71.35	1.31
bin 124	genus	Methanothrix sp002505805	1 654 054	257	7559	54.40	82.21	2.61
bin 167	genus	Methanothrix	1 141 622	235	5171	61.55	64.89	0.66

Table B.11: The pili sequences' aromaticity and largest gap between neighbouring aromatic amino acids identified in bin 142. Tyrosines (Tyr), Phenylalanines (Phe), Histidines (His), and Tryptophans (Trp). The pili genes with aromaticity higher than 9.8% are bolded and considered to be putative e-pili.

Bin	SequenceID	Aromatic Count	Mature Pilin Length	Percent Aromatics	Largest Gap	Tyr	Phe	His	Trp
7	AMIOHCLB_00692	14	167	8.38%	56	4	6	2	2
7	AMIOHCLB_00693	13	137	9.49%	48	5	4	1	3
7	AMIOHCLB_00978	7	119	5.88%	28	2	4	0	1
7	AMIOHCLB_00979	8	114	7.02%	29	4	3	0	1
7	AMIOHCLB_01711	29	334	8.68%	51	16	8	1	4
7	AMIOHCLB_01906	9	121	7.44%	27	3	3	1	2
7	AMIOHCLB_02174	13	144	9.03%	33	6	5	1	1
7	**AMIOHCLB_02805**	**15**	**145**	**10.34%**	**36**	**4**	**8**	**2**	**1**
94	**PFKNKFAJ_02270**	**16**	**155**	**10.32%**	**40**	**8**	**4**	**1**	**3**
97	ELLBNBPI_01416	8	170	4.71%	56	5	2	1	0
97	ELLBNBPI_01417	13	137	9.49%	25	4	5	1	3
97	ELLBNBPI_01635	14	153	9.15%	35	4	5	3	2
97	**ELLBNBPI_01636**	**13**	**128**	**10.16%**	**30**	**4**	**7**	**1**	**1**
142	FHFJCJBA_00457	13	142	9.15%	24	6	5	0	2
142	FHFJCJBA_00458	16	248	6.45%	40	4	11	0	1
142	FHFJCJBA_00459	13	181	7.18%	42	2	7	3	1
142	**FHFJCJBA_03220**	**7**	**31**	**22.58%**	**7**	**0**	**6**	**0**	**1**

Table B.12: Overview of Antibiotic Classes, Their Mechanisms of Action, and Common Resistance Mechanisms

Antibiotic Class	**Mechanism of Action**	**Common Resistance Mechanism**
Beta-lactams (Penams, Cephalosporins, Carbapenem, Monobactam, Penem, Oxacephem)	Inhibit cell wall synthesis	Beta-lactamase production
Macrolides	Inhibit protein synthesis by acting on ribosome	Methylation of ribosome binding sites
Tetracyclines, Glycylcyclines	Inhibit protein synthesis by binding to ribosome	Efflux pumps; Ribosomal protection proteins
Aminoglycosides	Bind to ribosome and inhibit protein synthesis	Modification enzymes (acetylation, phosphorylation, adenylation)
Quinolones, Fluoroquinolones	Inhibit DNA gyrase and topoisomerase IV	Mutations in target genes; Efflux pumps
Sulfonamides	Inhibit folate synthesis	Altered enzyme targets
Glycopeptides	Inhibit cell wall synthesis	Altered target sites
Oxazolidinones	Inhibit protein synthesis by binding to ribosome	Mutations in 23S rRNA
Pleuromutilins	Bind to ribosome and inhibit protein synthesis	Mutation or methylation of ribosomal binding site
Nitroimidazoles	Damage to DNA	Reduction in enzyme activity or mutations reducing drug uptake
Phenicol antibiotics	Inhibit protein synthesis by binding to 50S ribosomal subunit	Efflux pumps; Acetyltransferases
Rifamycins	Inhibit RNA synthesis by targeting DNA-dependent RNA polymerase	Mutations in RNA polymerase

Continued on next page

Lincosamides	Inhibit protein synthesis by acting on ribosome	Methylation of ribosome binding sites
Diaminopyrimidines	Inhibit folate synthesis	Altered enzyme targets
Streptogramins (Streptogramin A, Streptogramin B)	Inhibit protein synthesis by acting on ribosome	Efflux pumps; Enzymatic inactivation
Mupirocin-like antibiotics	Inhibit isoleucyl-tRNA synthetase	Target modification
Antibacterial free fatty acids	Various	Various
Nucleoside antibiotics	Inhibit RNA or DNA synthesis	Enzymatic modification; Target alteration
Fusidane antibiotics	Inhibit protein synthesis	Target modification
Phosphonic acid antibiotics	Inhibit cell wall synthesis	Altered target pathway
Pyrazine antibiotics	Inhibit mycolic acid synthesis	Enzymatic modification
Salicylic acid antibiotic	Inhibits bacterial enzymes	Efflux pumps; Enzymatic inactivation
Nitrofuran antibiotics	Damage to DNA and proteins	Enzymatic modification; Reduced uptake
Bicyclomycin-like antibiotics	Inhibit RNA synthesis	Target modification
Polyamine antibiotics	Disrupt cell membrane or inhibit protein synthesis	Efflux pumps; Enzymatic modification
Orthosomycin antibiotics	Inhibit protein synthesis	Target modification
Peptide antibiotics, Polypeptides	Disrupt cell membrane or inhibit protein synthesis	Modification of membrane lipids; Enzymatic modification
Aminocoumarin antibiotics	Inhibit DNA gyrase	Target modification
Cephamycin	Inhibit cell wall synthesis	Beta-lactamase production
Elfamycin antibiotics	Inhibit protein synthesis	Target modification

Table B.13: Removal Efficiency of Various Substances Under Different Conditions in AFMBR

Condition	**Description**	**Removal Efficiency**			**Interpretation**
		Sulfamethoxazole	**Diclofenac**	**Ibuprofen**	
2 & 4	Bulk suspension & solid phase in AFMBR	Complete in 30 min	Complete in 30 min	Complete in 30 min	High sorption & biofilm-aided biodegradation on GAC.
3	Liquid phase in AFMBR	$\approx$20% in 24 hr	No removal in 24 hr	No removal in 24 hr	Minimal biodegradation; biofilm on GAC crucial.
5	New GAC (25% v/v)	Almost complete in 6 hr	Almost complete in 3 hr	Almost complete in 2 hr	Good adsorption by GAC; slower without biofilm.

Appendix C

Supplementary Figures

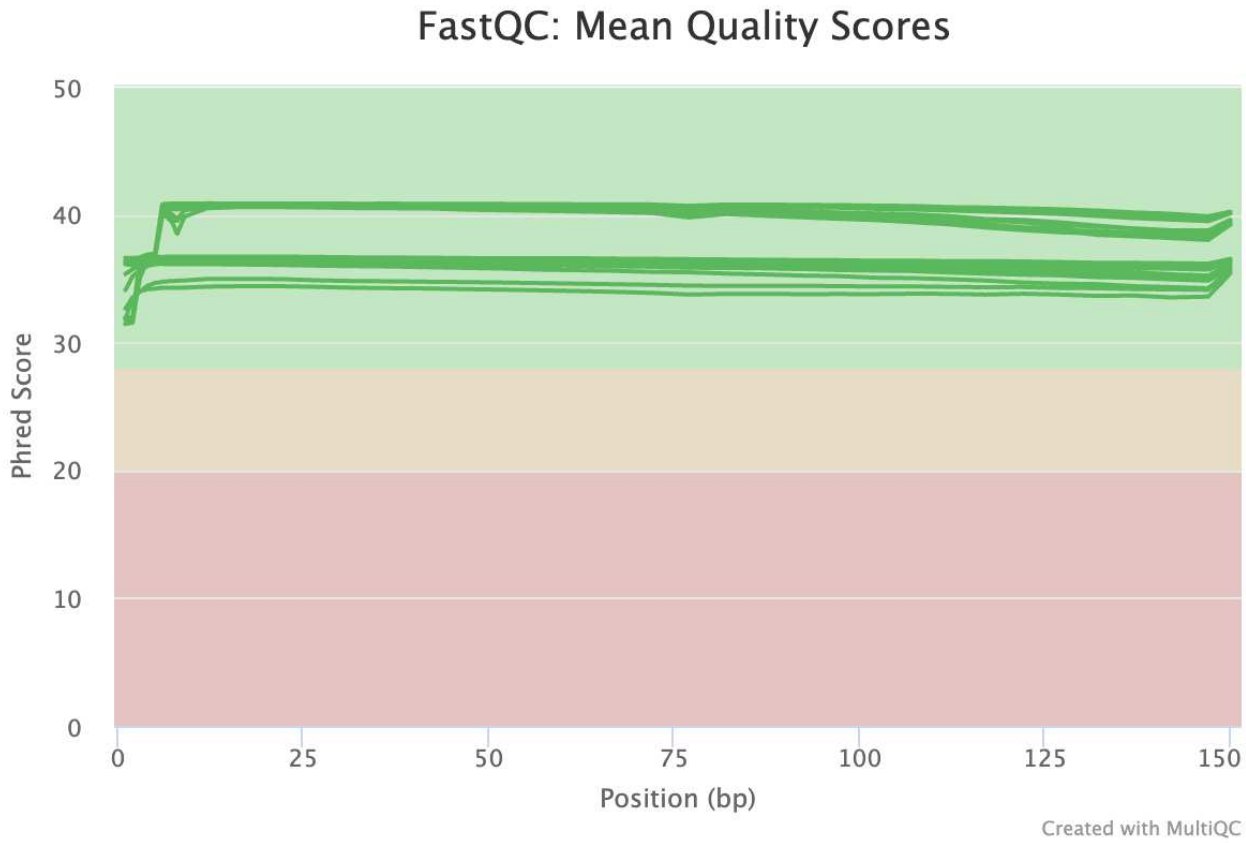


Figure C.1: Sequence quality histograms of the metagenomes collected in this study accessed using FastQC [207] and summarised using MultiQC [428]. A phred score above 30 thus all samples are high quality.

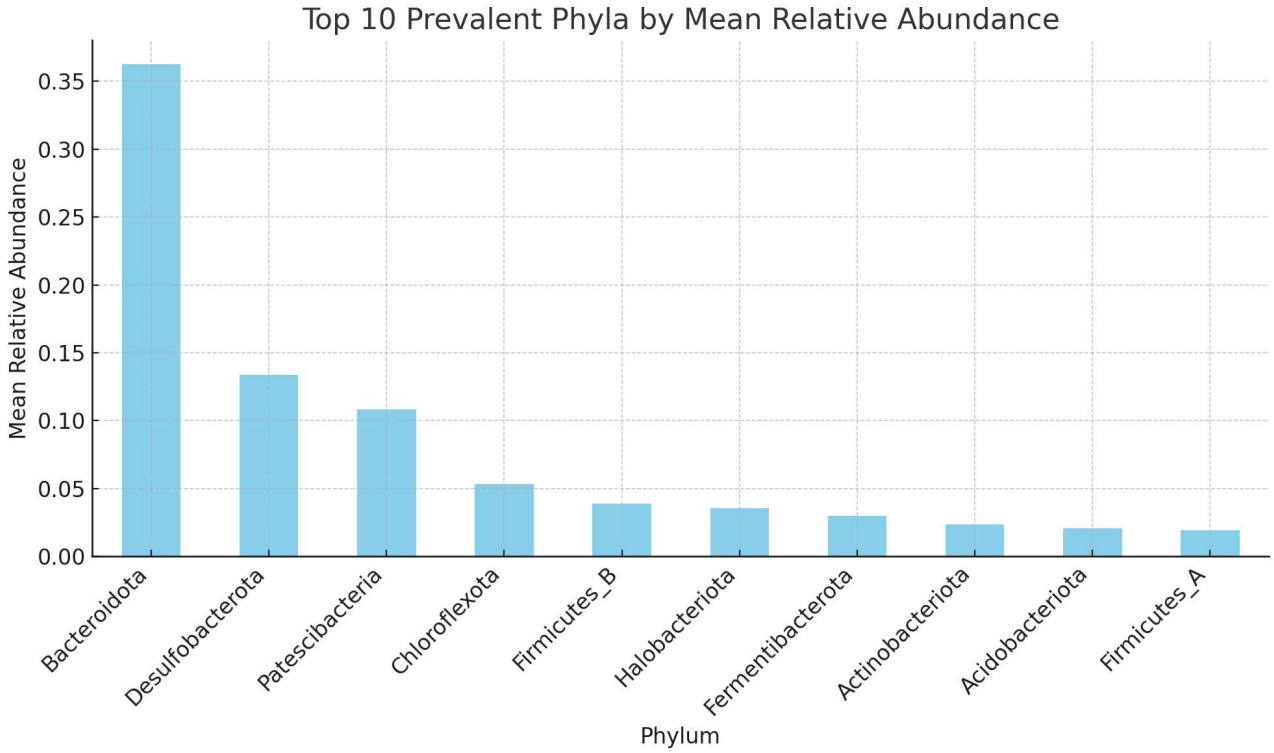


Figure C.2: Top ten prevalent phyla by mean relative abundance based on the percentage reads assigned to EET MAGs.

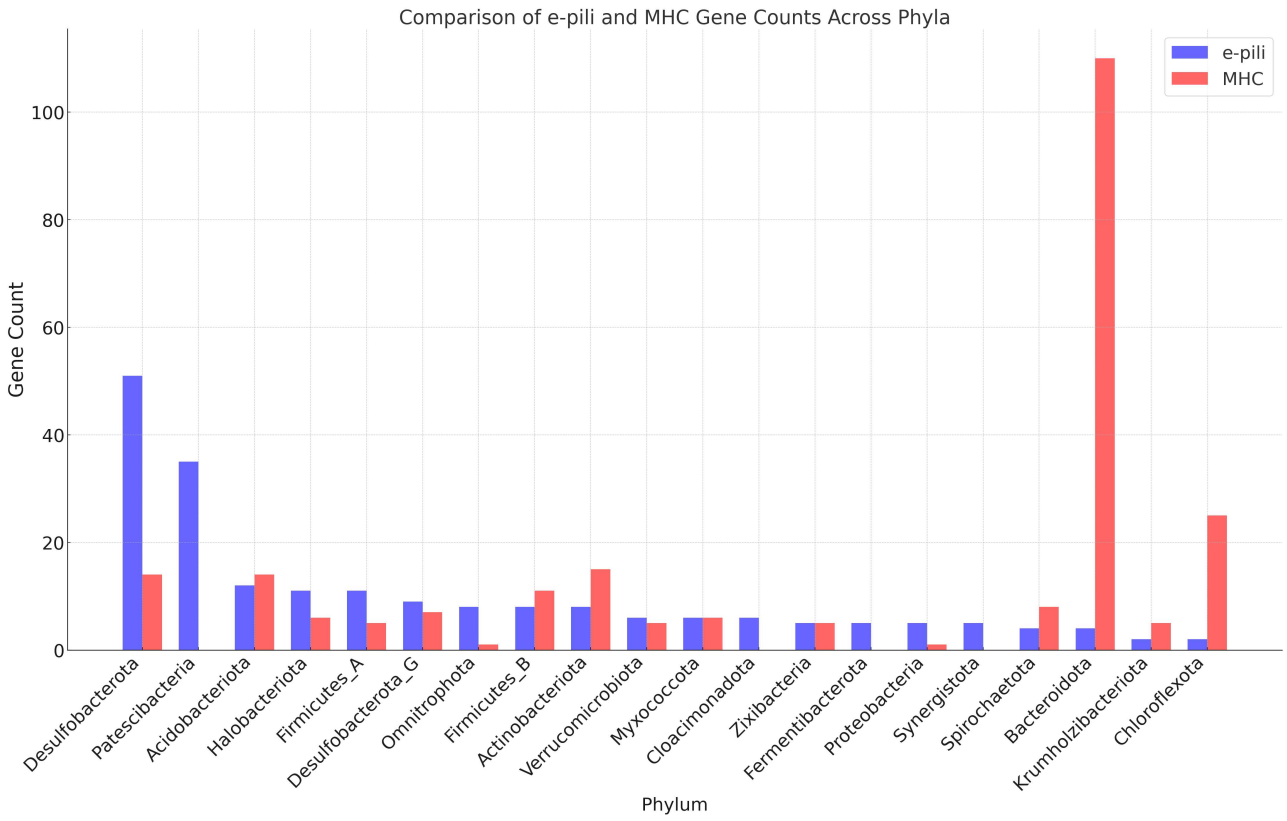


Figure C.3: e-pili and MHC gene count across various phyla. Each bar represents the count of genes identified within each phylum. Blue coloured bar represent e-pili genes and red is for MHC genes. For visualisation purpose, only phylum with gene count more than 5 was shown.

Category	Item	Pearson Correlation	Random Forest Feature Importance
Carbon metabolism	Citrate cycle (TCA cycle, Krebs cycle)	0.3808	0.0420
Carbon metabolism	Dicarboxylate-hydroxybutyrate cycle	0.1790	0.0222
Carbon metabolism	Entner-Doudoroff pathway, glucose-6P => glyceraldehyde-3P + pyruvate	0.1311	0.0230
Carbon metabolism	Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate	0.1663	0.0277
Carbon metabolism	Glyoxylate cycle	0.2876	0.0221
Carbon metabolism	Hydroxypropionate-hydroxybutylate cycle	0.1927	0.0251
Carbon metabolism	Methanogenesis, CO2 => methane	0.0184	0.0347
Carbon metabolism	Pentose phosphate pathway (Pentose phosphate cycle)	0.2033	0.0233
Carbon metabolism	Reductive acetyl-CoA pathway (Wood-Ljungdahl pathway)	0.0757	0.0225
Carbon metabolism	Reductive citrate cycle (Arnon-Buchanan cycle)	0.3540	0.0319
Carbon metabolism	Reductive pentose phosphate cycle (Calvin cycle)	0.0541	0.0269
CAZy	Amorphous Cellulose	0.2488	0.0116
CAZy	Arabinan	0.3659	0.0286
CAZy	Xyloglucan	0.3728	0.0260
Complex I	NADH:quinone oxidoreductase, prokaryotes	0.3920	0.0598
Complex II	Succinate dehydrogenase, prokaryotes	0.4195	0.0596
Complex III	Cytochrome bd ubiquinol oxidase	0.3258	0.0162
Complex IV High affinity	Cytochrome bd ubiquinol oxidase	0.3258	0.0183
Complex IV High affinity	Cytochrome c oxidase, cbb3-type	0.4023	0.0421
Complex IV Low affinity	Cytochrome c oxidase	0.3821	0.0315
Complex IV Low affinity	Cytochrome c oxidase, prokaryotes	0.4125	0.0378
Complex V	F-type ATPase, prokaryotes and chloroplasts	0.2154	0.0209
Complex V	V/A-type ATPase, prokaryotes	-0.1694	0.0166
Nitrogen metabolism	Nitrogen fixation altennative	-0.0241	0.0172
Nitrogen metabolism	nitrate => nitrite	0.0263	0.0280
SCFA and alcohol conversions	acetate, pt 1	0.0982	0.0161
Sulfur metabolism	tetrathionate => thiosulfate	-0.0118	0.0139



Figure C.4: Heatmap showing the correlation between MHC genes and metabolic capacity. Only statistically significant Pearson correlation coefficients (p-value < 0.05, determined using Student's *t*-test) with a random forest feature importance greater than 0.01 are displayed.

Category	Item	Pearson Correlation	Random Forest Feature Importance
Carbon metabolism	Acetyl-CoA pathway, CO2 => acetyl-CoA	0.0668	0.0105
Carbon metabolism	Citrate cycle (TCA cycle, Krebs cycle)	0.1529	0.0414
Carbon metabolism	Dicarboxylate-hydroxybutyrate cycle	0.1991	0.0359
Carbon metabolism	Entner-Doudoroff pathway, glucose-6P => glyceraldehyde-3P + pyruvate	0.0212	0.0179
Carbon metabolism	Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate	-0.0068	0.0407
Carbon metabolism	Glyoxylate cycle	0.1639	0.0247
Carbon metabolism	Hydroxypropionate-hydroxybutylate cycle	0.0958	0.0243
Carbon metabolism	Methanogenesis, CO2 => methane	0.0417	0.0218
Carbon metabolism	Pentose phosphate pathway (Pentose phosphate cycle)	0.0639	0.0371
Carbon metabolism	Reductive acetyl-CoA pathway (Wood-Ljungdahl pathway)	0.0259	0.0298
Carbon metabolism	Reductive citrate cycle (Arnon-Buchanan cycle)	0.1269	0.0346
Carbon metabolism	Reductive pentose phosphate cycle (Calvin cycle)	-0.0267	0.0494
CAZy	Amorphous Cellulose	-0.0695	0.0101
CAZy	Mixed-Linkage glucans	-0.0554	0.0114
Complex I	NADH:quinone oxidoreductase, prokaryotes	0.1504	0.0414
Complex II	Succinate dehydrogenase, prokaryotes	0.0200	0.0122
Complex III	Cytochrome bd ubiquinol oxidase	0.0184	0.0118
Complex IV High affinity	Cytochrome bd ubiquinol oxidase	0.0184	0.0106
Complex IV Low affinity	Cytochrome c oxidase	0.1199	0.0146
Complex V	F-type ATPase, prokaryotes and chloroplasts	0.1154	0.0363
Complex V	V/A-type ATPase, prokaryotes	-0.0403	0.0240
Methanogenesis and methanotrophy	Key functional gene	-0.0941	0.0121
Methanogenesis and methanotrophy	acetate => methane, pt 1	0.1335	0.0131
Methanogenesis and methanotrophy	acetate => methane, pt 2	-0.0240	0.0138
Nitrogen metabolism	Nitrogen fixation altennative	0.0961	0.0134
SCFA and alcohol conversions	Alcohol production	0.0867	0.0117
SCFA and alcohol conversions	acetate, pt 1	-0.0122	0.0106
Sulfur metabolism	Thiosulfate oxidation by SOX complex, thiosulfate => sulfate	-0.0338	0.0170

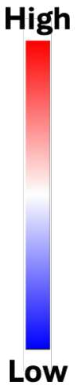


Figure C.5: Heatmap showing the correlation between e-pili genes and metabolic capacity. Only statistically significant Pearson correlation coefficients (p-value < 0.05, determined using Student's *t*-test) with a random forest feature importance greater than 0.01 are displayed.

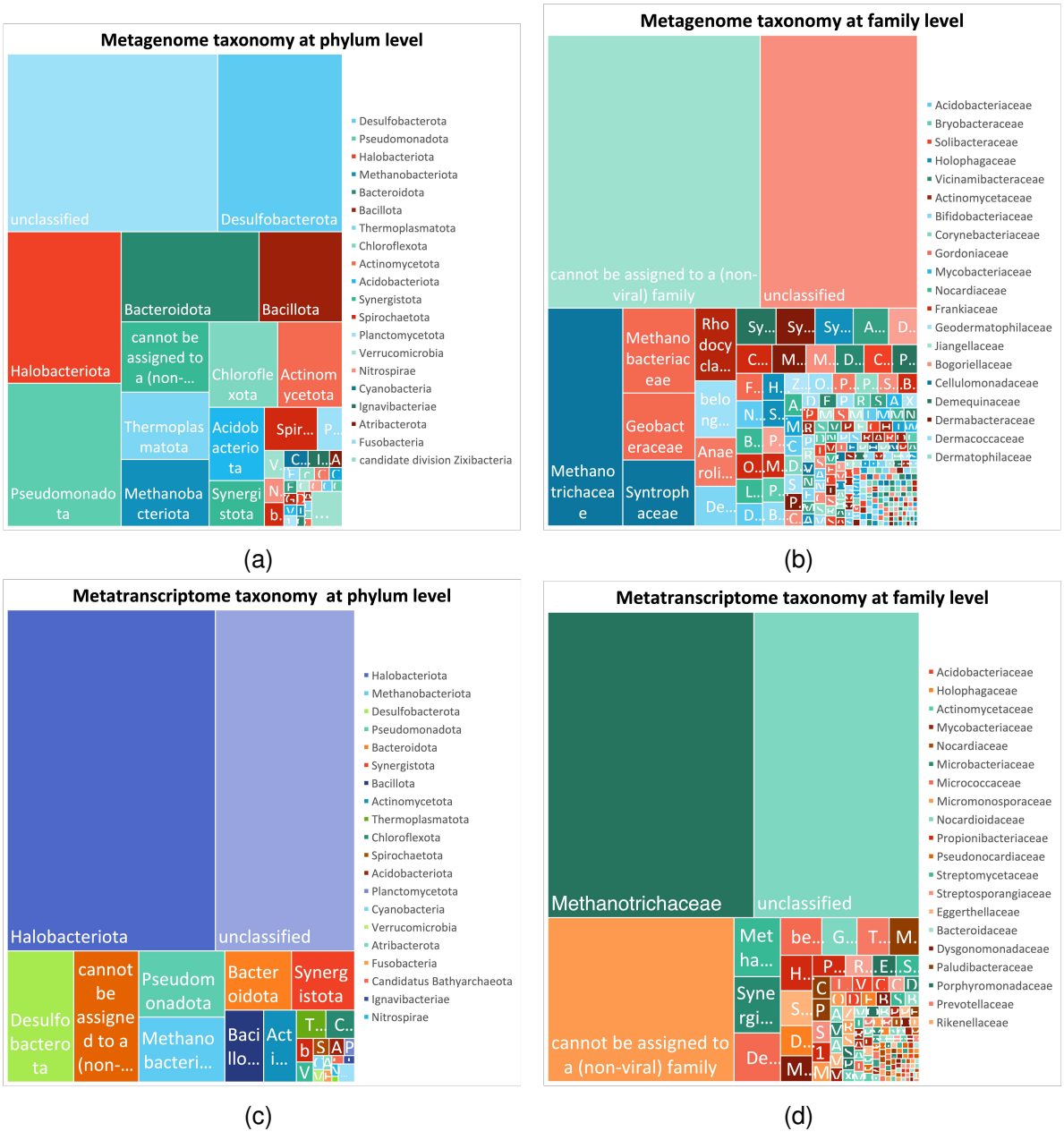


Figure C.6: Taxonomy profile classified via Kaiju. (a) search against metagenome at phylum level. (b) search against metagenome at family level. (c) search against metatranscriptome at phylum level. (d) search against metatranscriptome at family level. The cutoff is at 0.01% relative abundance.

Wood-Ljungdahl acetate oxidation I Log2 RPKM

		Syntrophs			Sulfate-reducing bacteria						
Enzyme name	**Gene**	**bin 7**	**bin 94**	**bin 97**	**bin 35**	**bin 61**	**bin 112**	**bin 142**	**bin 172**	**bin 177**	**bin 195**
Acetate kinase	*acK*	0.00	11.19	0.00	8.18	10.66	6.90	10.05	9.49	0.00	7.86
Phosphate acetyltransferase	*acyP*	0.00	10.87	0.00	9.53	9.19	9.26	8.51	9.06	6.97	9.24
Carbon monoxide dehydrogenase	*codh*	0.00	4.58	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5,10-methylenetetrahydrofolate reductase	*metF*	0.00	0.00	9.39	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5,10-methylenetetrahydrofolate reductase	*metF*	0.00	0.00	0.00	0.00	8.54	9.66	8.78	9.36	9.20	9.31
Methenyltetrahydrofolate cyclohydrolase	*mtc*	0.00	0.00	0.00	9.10	0.00	0.00	0.00	0.00	0.00	8.28
Methenyltetrahydrofolate cyclohydrolase	*mtc*	0.00	0.00	0.00	9.10	0.00	0.00	0.00	0.00	0.00	8.28
Methenyltetrahydrofolate cyclohydrolase	*mtc*	0.00	0.00	0.00	0.00	7.66	0.00	8.95	8.55	9.40	9.52
Methenyltetrahydrofolate cyclohydrolase	*mtc*	0.00	0.00	0.00	7.68	7.68	7.44	0.00	8.51	0.00	0.00
Methenyltetrahydrofolate cyclohydrolase	*mtc*	0.00	0.00	0.00	8.80	9.65	9.58	9.96	9.47	0.00	0.00
Methenyltetrahydrofolate cyclohydrolase	*mtc*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Methenyltetrahydrofolate cyclohydrolase	*mtc*	0.00	0.00	0.00	9.10	8.12	7.90	0.00	0.00	0.00	0.00
Formyltetrahydrofolate deformylase	*purU*	0.00	0.00	0.00	0.00	7.57	0.00	8.18	8.42	9.17	0.00
Formate dehydrogenase	*fdh*	12.89	11.20	12.43	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Formate dehydrogenase	*fdh*	0.00	0.00	0.00	0.00	0.00	0.00	8.53	0.00	0.00	0.00
Formate dehydrogenase	*fdh*	0.00	10.15	9.58	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Formate dehydrogenase	*fdh*	11.61	8.50	11.82	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Formate dehydrogenase	*fdh*	0.00	0.00	9.77	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Figure C.7: Transcription level of genes involved in Wood-Ljungdahl acetate oxidation pathway.

Methylmalonyl-CoA propionate degradation | log_2 RPKM

		Syntrophs			Sulfate-reducing bacteria						
Enzyme name	**Gene**	**bin 7**	**bin 94**	**bin 97**	**bin 35**	**bin 61**	**bin 112**	**bin 142**	**bin 172**	**bin 177**	**bin 195**
Glutaconate CoA-transferase	*gct*	11.88	10.21	9.60	0.00	0.00	0.00	11.05	0.00	0.00	0.00
Glutaconate CoA-transferase	*gct*	10.78	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Methylmalonyl-CoA decarboxylase	*mcd*	0.00	0.00	0.00	0.00	0.00	0.00	9.88	0.00	0.00	0.00
Propionyl-CoA carboxylase	*pcc*	8.55	0.00	0.00	0.00	0.00	0.00	10.45	0.00	0.00	0.00
Propionyl-CoA carboxylase	*pcc*	10.76	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
methylmalonyl-CoA epimerase	*mce*	10.47	0.00	0.00	0.00	0.00	0.00	8.29	0.00	0.00	0.00
Methylmalonyl-CoA mutase	*mcm*	11.26	10.61	8.17	0.00	0.00	0.00	9.22	0.00	0.00	0.00
Methylmalonyl-CoA mutase	*mcm*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Methylmalonyl-CoA mutase	*mcm*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Succinate-CoA ligase	*scs*	0.00	0.00	0.00	0.00	0.00	0.00	9.78	0.00	0.00	0.00
Succinate-CoA ligase	*scs*	0.00	0.00	0.00	0.00	0.00	0.00	10.25	0.00	0.00	0.00
Succinate dehydrogenase	*sdh*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Succinate dehydrogenase	*sdh*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Succinate dehydrogenase	*sdh*	0.00	0.00	0.00	0.00	0.00	0.00	9.62	0.00	0.00	0.00
Succinate dehydrogenase	*sdh*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fumarate reductase	*fht*	0.00	0.00	9.50	0.00	7.46	0.00	0.00	0.00	0.00	0.00
Fumarate reductase	*fht*	0.00	0.00	8.73	0.00	10.48	0.00	10.31	0.00	0.00	0.00
Fumarate reductase	*fht*	0.00	0.00	0.00	0.00	7.61	0.00	10.20	0.00	0.00	0.00
Fumarate reductase	*fht*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fumarate reductase	*fht*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fumarate reductase	*fht*	0.00	0.00	0.00	0.00	10.29	0.00	0.00	0.00	0.00	0.00
Fumarate reductase	*fht*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fumarate hydratase	*fht*	0.00	0.00	0.00	0.00	6.85	0.00	8.67	0.00	0.00	0.00
Fumarate hydratase	*fht*	0.00	0.00	0.00	10.52	0.00	0.00	10.65	0.00	0.00	0.00
Fumarate hydratase	*fht*	0.00	0.00	8.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Malate dehydrogenase	*mdh*	9.04	0.00	0.00	0.00	0.00	8.07	10.09	9.64	9.52	8.04
Malate dehydrogenase	*mdh*	0.00	0.00	0.00	0.00	8.46	0.00	9.62	0.00	0.00	0.00
Malate dehydrogenase	*mdh*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Oxaloacetate decarboxylase	*oad*	0.00	0.00	0.00	0.00	10.18	0.00	10.77	0.00	0.00	0.00
Oxaloacetate decarboxylase	*oad*	0.00	0.00	0.00	0.00	0.00	0.00	8.25	0.00	0.00	0.00
Oxaloacetate decarboxylase	*oad*	0.00	0.00	0.00	0.00	0.00	0.00	7.66	0.00	0.00	0.00
Pyruvate:ferredoxin oxidoreductase	*pfor*	11.52	11.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pyruvate:ferredoxin oxidoreductase	*pfor*	0.00	11.64	0.00	0.00	10.21	9.06	0.00	0.00	0.00	0.00
Pyruvate:ferredoxin oxidoreductase	*pfor*	12.15	0.00	0.00	0.00	10.07	0.00	0.00	0.00	0.00	0.00
Pyruvate:ferredoxin oxidoreductase	*pfor*	9.90	0.00	0.00	0.00	0.00	0.00	8.52	0.00	0.00	0.00
Pyruvate:ferredoxin oxidoreductase	*pfor*	9.38	0.00	0.00	0.00	0.00	0.00	8.24	0.00	0.00	0.00
Pyruvate:ferredoxin oxidoreductase	*pfor*	10.87	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pyruvate:ferredoxin oxidoreductase	*pfor*	0.00	0.00	0.00	10.26	11.92	10.39	0.00	9.82	9.74	11.20

Figure C.8: Transcription level of genes involved in methylmalonyl-CoA propionate degradation pathway.

Formate/H_2 MIET I log_2 RPKM

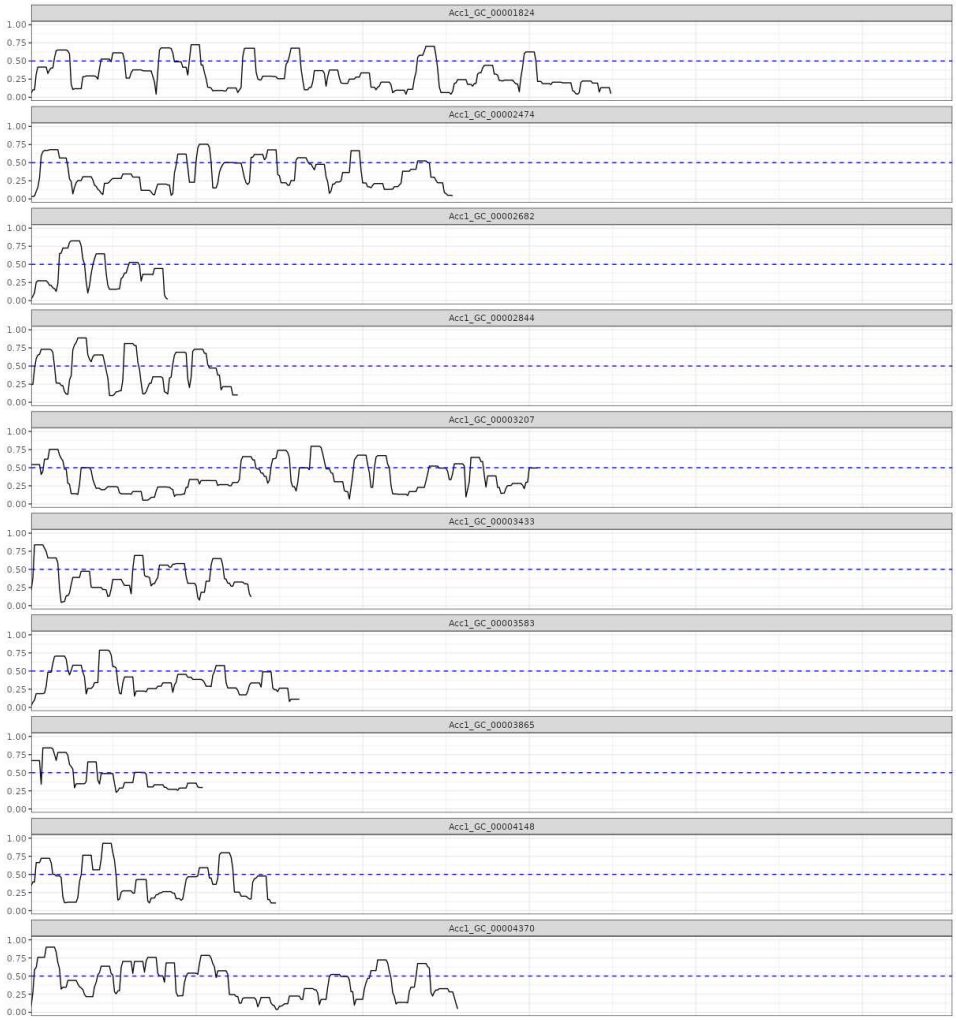
		Sulfate-reducing bacteria						
Enzyme name	**Gene**	**bin 35**	**bin 61**	**bin 112**	**bin 142**	**bin 172**	**bin 177**	**bin 195**
NADP-reducing hydrogenase	*hndA*	0.00	9.40	10.69	7.93	0.00	0.00	0.00
NADP-reducing hydrogenase	*hndC*	0.00	11.64	9.95	13.07	0.00	0.00	0.00
NADP-reducing hydrogenase	*hndD*	10.10	0.00	8.96	7.36	0.00	0.00	0.00
formate dehydrogenase	*fdhA*	0.00	0.00	0.00	8.53	0.00	0.00	0.00
cytochrome c3 family protein	*cyt c*	9.52	0.00	7.78	0.00	0.00	0.00	0.00
cytochrome c	n.a.	9.52	0.00	0.00	0.00	0.00	0.00	0.00
cytochrome c	n.a.	0.00	0.00	0.00	7.42	0.00	0.00	0.00
cytochrome c biogenesis protein	*ccsA*	0.00	0.00	0.00	10.27	0.00	0.00	0.00
cytochrome c biogenesis protein	*ccsB*	0.00	0.00	0.00	8.57	0.00	0.00	0.00
cytochrome c	n.a.	0.00	0.00	0.00	6.30	0.00	0.00	0.00
cytochrome c	n.a.	0.00	0.00	0.00	7.75	0.00	0.00	0.00
cytochrome c	n.a.	0.00	0.00	0.00	8.83	0.00	0.00	0.00
cytochrome c	n.a.	0.00	0.00	0.00	7.20	0.00	0.00	0.00
cytochrome c	n.a.	0.00	0.00	0.00	8.15	0.00	0.00	0.00
Cytochrome bd-I ubiquinol oxidase	*cydA*	0.00	8.48	0.00	6.38	0.00	0.00	0.00
Cytochrome bd-I ubiquinol oxidase	*cydB*	0.00	9.30	0.00	6.80	0.00	0.00	0.00
Heme exporter protein C	*ccmC*	0.00	7.85	0.00	0.00	0.00	0.00	0.00
cytochrome c maturation protein	*ccmE*	0.00	8.64	0.00	0.00	0.00	0.00	0.00
cytochrome c maturation protein	*ccmF*	0.00	0.00	7.81	0.00	0.00	0.00	0.00
ubiquinone/menaquinone biosynthesis methyltransferase	*ubiE*	0.00	7.88	7.55	9.07	0.00	0.00	0.00
NAD(P)H:quinone oxidoreductase	*ndh*	0.00	10.36	0.00	0.00	10.49	9.75	0.00
NADH-quinone oxidoreductase	*nuoB*	9.99	0.00	10.21	9.01	11.95	0.00	10.18
NADH-quinone oxidoreductase	*nuoC*	0.00	0.00	8.65	0.00	0.00	9.36	0.00
NADH-quinone oxidoreductase	*nuoD*	0.00	8.69	0.00	6.84	0.00	0.00	0.00
NADH-quinone oxidoreductase	*nuoF*	0.00	0.00	0.00	0.00	0.00	0.00	10.21
NADH-quinone oxidoreductase	*nuoH*	10.67	9.28	10.40	8.01	0.00	9.24	0.00
NADH-quinone oxidoreductase	*nuoJ*	0.00	9.28	0.00	0.00	0.00	0.00	0.00
NADH-quinone oxidoreductase	*nuoK*	0.00	8.98	0.00	0.00	0.00	0.00	0.00
NADH-quinone oxidoreductase	*nuoM*	0.00	9.31	0.00	8.73	0.00	0.00	0.00
NADH-quinone oxidoreductase	*nuoN*	0.00	7.84	0.00	8.09	0.00	0.00	0.00
Na(+)-translocating NADH-quinone reductase	*nqrF*	0.00	0.00	0.00	8.71	0.00	0.00	0.00
quinone oxidoreductase	*yhfP*	0.00	0.00	0.00	9.01	0.00	0.00	0.00

Figure C.9: Transcription level of genes involved in formate/H_2 MIET pathways.

Electron transfer | log_2 RPKM

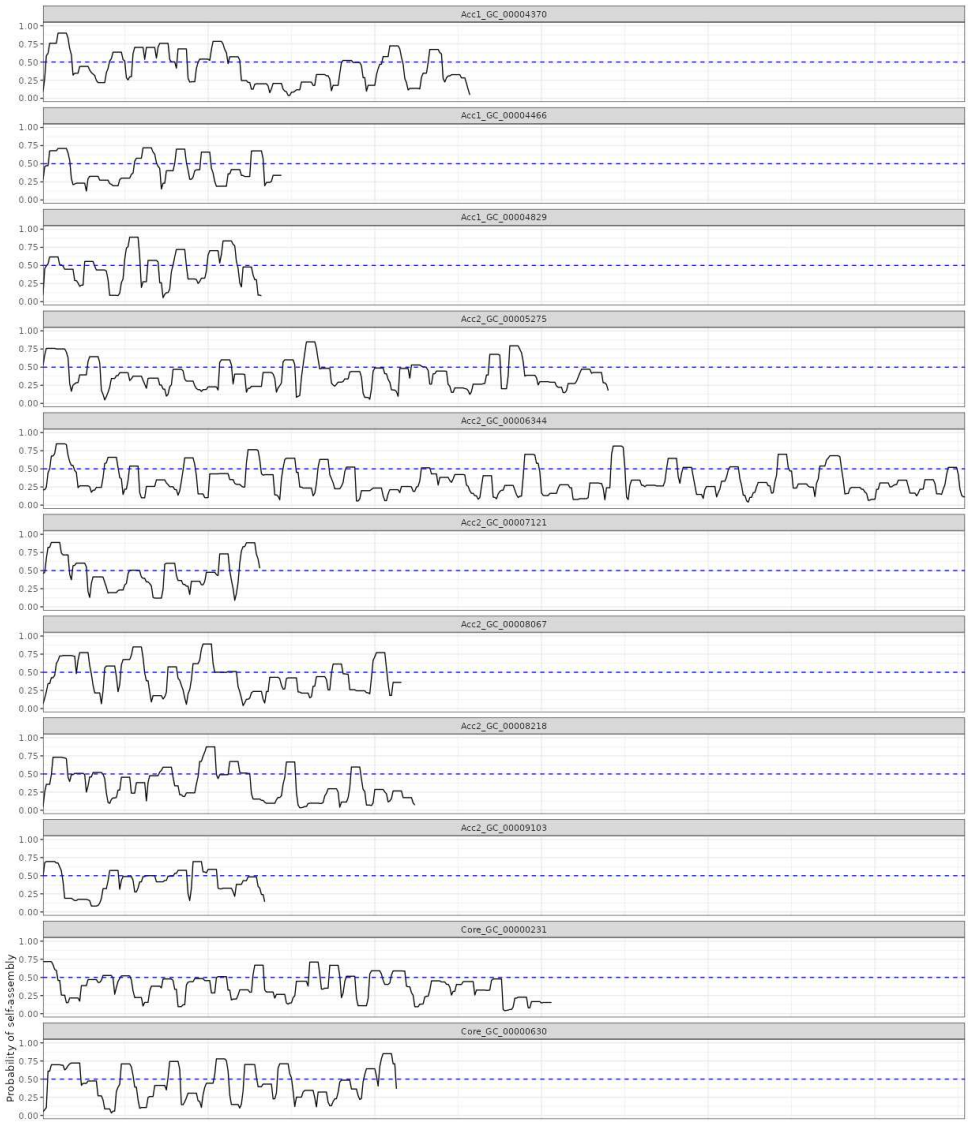
		Sulfate-reducing bacteria						
Enzyme name	**Gene**	**bin 35**	**bin 61**	**bin 112**	**bin 142**	**bin 172**	**bin 177**	**bin 195**
F0F1 ATP synthase	*atpA*	11.45	0.00	11.49	9.56	11.42	9.81	11.82
F0F1 ATP synthase	*atpB*	11.98	12.58	12.14	9.08	9.88	10.13	12.06
F0F1 ATP synthase	*atpC*	9.62	0.00	9.66	10.65	10.65	10.75	9.64
F0F1 ATP synthase	*atpD*	8.30	0.00	0.00	0.00	0.00	8.90	8.01
F0F1 ATP synthase	*atpE*	10.72	9.64	9.46	8.26	0.00	9.33	9.42
cytochrome c3 family protein	*cyt c*	9.52	0.00	7.78	0.00	0.00	0.00	0.00
cytochrome c	n.a.	9.52	0.00	0.00	0.00	0.00	0.00	0.00
cytochrome c	n.a.	0.00	0.00	0.00	7.42	0.00	0.00	0.00
cytochrome c biogenesis protein	*ccsA*	0.00	0.00	0.00	10.27	0.00	0.00	0.00
cytochrome c biogenesis protein	*ccsB*	0.00	0.00	0.00	8.57	0.00	0.00	0.00
cytochrome c	n.a.	0.00	0.00	0.00	6.30	0.00	0.00	0.00
cytochrome c	n.a.	0.00	0.00	0.00	7.75	0.00	0.00	0.00
cytochrome c	n.a.	0.00	0.00	0.00	8.83	0.00	0.00	0.00
cytochrome c	n.a.	0.00	0.00	0.00	7.20	0.00	0.00	0.00
cytochrome c	n.a.	0.00	0.00	0.00	8.15	0.00	0.00	0.00
Cytochrome bd-I ubiquinol oxidase	*cydA*	0.00	8.48	0.00	6.38	0.00	0.00	0.00
Cytochrome bd-I ubiquinol oxidase	*cydB*	0.00	9.30	0.00	6.80	0.00	0.00	0.00
Heme exporter protein C	*ccmC*	0.00	7.85	0.00	0.00	0.00	0.00	0.00
cytochrome c maturation protein	*ccmE*	0.00	8.64	0.00	0.00	0.00	0.00	0.00
cytochrome c maturation protein	*ccmF*	0.00	0.00	7.81	0.00	0.00	0.00	0.00
NADP-reducing hydrogenase	*hndA*	0.00	9.40	10.69	7.93	0.00	0.00	0.00
NADP-reducing hydrogenase	*hndC*	0.00	11.64	9.95	13.07	0.00	0.00	0.00
NADP-reducing hydrogenase	*hndD*	10.10	0.00	8.96	7.36	0.00	0.00	0.00
Formate dehydrogenase	*fdh*	0.00	0.00	0.00	8.53	0.00	0.00	0.00
quinone-interacting membrane-bound oxidoreductase	*qmoC*	0.00	0.00	0.00	8.31	0.00	0.00	0.00
menaquinone reductase	*qrcA*	0.00	0.00	7.78	0.00	0.00	0.00	0.00
Sulfate adenylyltransferase	*satA*	0.00	0.00	0.00	10.30	0.00	0.00	0.00
Sulfate adenylyltransferase	*satB*	0.00	0.00	8.90	0.00	0.00	0.00	0.00
Sulfite reductase	*dsrC*	9.09	12.90	10.83	10.96	9.05	10.69	11.60
Sulfite reductase	*dsrM*	0.00	8.32	0.00	0.00	0.00	0.00	0.00
Dissimilatory sulfite reductase	*mccA*	0.00	10.84	0.00	0.00	0.00	0.00	0.00
Dissimilatory sulfite reductase	*mccB*	0.00	10.85	0.00	0.00	0.00	0.00	0.00
Dissimilatory sulfite reductase	*mccD*	0.00	9.46	0.00	0.00	0.00	0.00	0.00
Polysulfide reductase	*psr*	8.10	8.71	7.78	8.36	7.93	8.79	7.68

Figure C.10: Transcription level of genes involved in electron transfer in SRB.



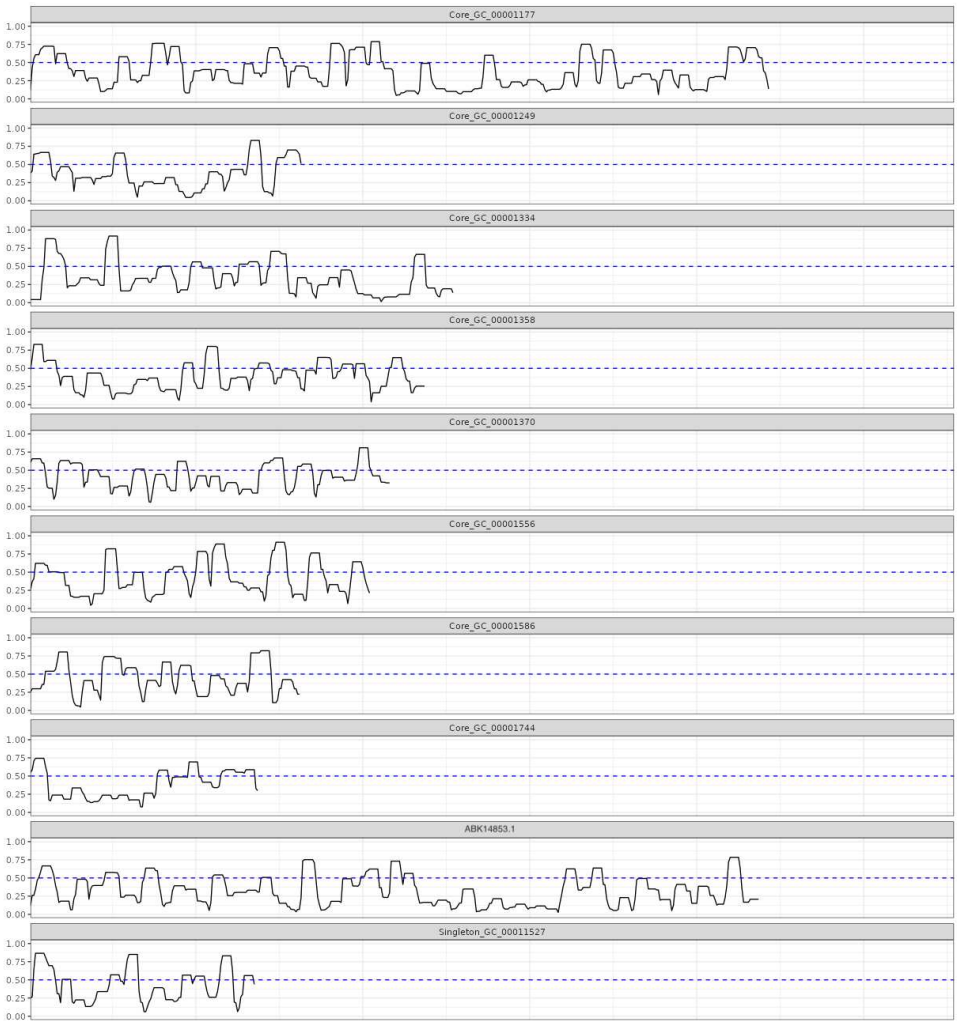
(a) First quarter of the MspA amyloidogenic region analysis.

Figure C.11: First quarter of the MspA amyloidogenic region analysis.



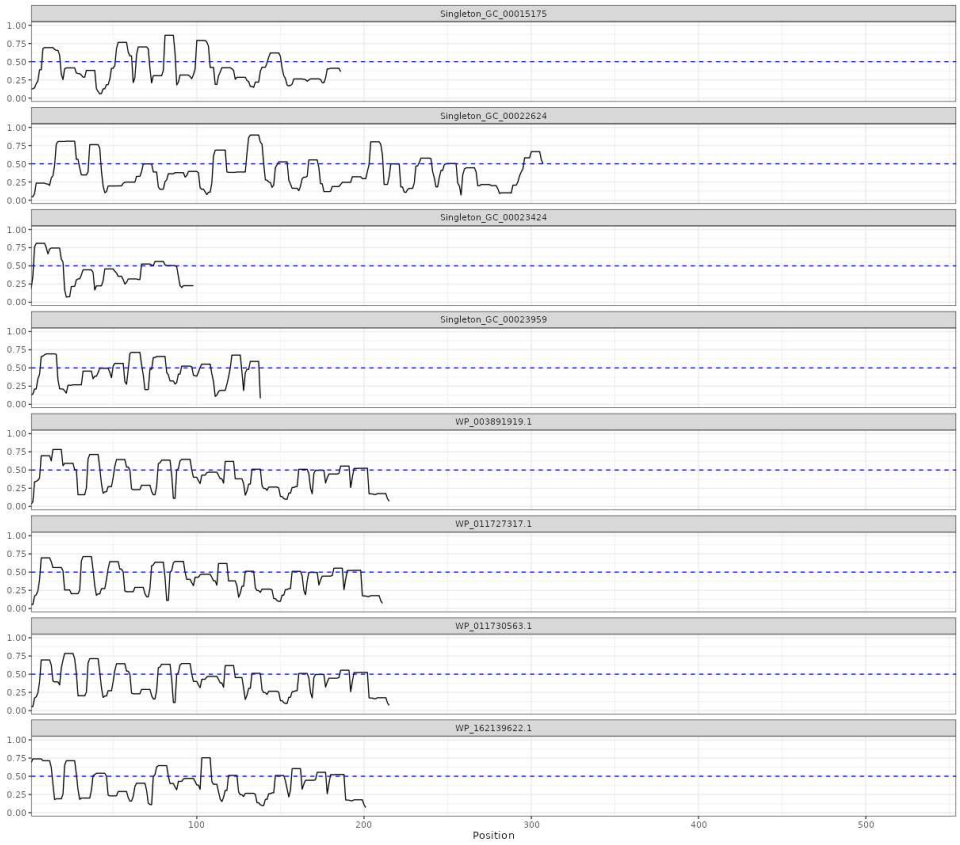
(a) Second quarter of the MspA amyloidogenic region analysis.

Figure C.12: Second quarter of the MspA amyloidogenic region analysis.



(a) Third quarter of the MspA amyloidogenic region analysis.

Figure C.13: Third quarter of the MspA amyloidogenic region analysis.



(a) Fourth quarter of the MspA amyloidogenic region analysis.

Figure C.14: Fourth quarter of the MspA amyloidogenic region analysis.

Figure C.15: MspA contains evenly distributed amyloidogenic regions. Analysis of the MspA and MspA-like sequences using AmyloGram [429] identified multiple regions with predicted high amyloidogenic potential (indicated above the horizontal lines). The sequences analysed include: (i) 33 expressed MspA-like genes predicted from the *Methanothrix* pangenome, (ii) WP_003891919, WP_011727317, WP_011730563, and WP_162139622 from *Mycobacterium smegmatis*, (iii) WP_011449234.1 from *Methanospirillum hungatei*, and (iv) ABK14853.1 from *Methanothrix thermoacetophila* PT.