Phylogenetic and structural diversity of aromatically dense pili from environmental metagenomes

M. S. Bray[1,2], J. Wu[1], C.C. Padilla[1#], F. J. Stewart[1], D. A. Fowle[3], C. Henny[4], R. L. Simister[5], K. J. Thompson[5], S. A. Crowe[2,5], J. B. Glass[1, 2*]

[1]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA;

[2]School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA, USA;

[3]Department of Geology, University of Kansas, Lawrence, KS, USA;

[4]Research Center for Limnology, Indonesian Institute of Sciences, Cibinong, Indonesia;

[5]Department of Earth, Ocean, & Atmospheric Sciences; Department of Microbiology & Immunology, University of British Columbia, Vancouver, BC, Canada

*Correspondence to: Jennifer.Glass@eas.gatech.edu

Running Title: Predicting e-pili

[#]Now at: Dovetail Genomics, LLC, Santa Cruz, CA, USA

Originality and Significance. Electroactive pili (e-pili) are used by microorganisms to respire solid metals in their environment through extracellular electron transfer. Thus, e-pili enable microbes to occupy specific environmental niches. Additionally, the conductive nature of e-pili underscores their potential for biotechnological applications. Currently the repertoire of known e-pili is small, and their environmental distribution is largely unknown. Using sequence analysis, we identified numerous genes encoding putative e-pili from diverse anoxic, metal-rich ecosystems. Our results expand the diversity of putative e-pili in environments where metal oxides may be important electron acceptors for microbial respiration.

Summary. Electroactive type IV pili, or e-pili, are used by some microbial species for extracellular electron transport. Recent studies suggest that e-pili may be more phylogenetically and structurally diverse than previously assumed. Here, we investigated putative e-pilin genes in metagenomes from long term enrichments of Lake Matano sediment and diverse ecosystems including those in which metal reduction is likely important. We found that laboratory enrichments selected for aromatically dense pilins that were phylogenetically and structurally different than those in Lake Matano sediments. The putative e-pilins recovered from environmental metagenomes were diverse in phylogeny and protein length. We found that the majority of e-pilins in the environment may be “full-length” pilins and could be used by diverse bacterial taxa. Our results expand upon the structural and phylogenetic diversity of e-pilins.

Introduction

Electroactive bacteria transport electrons through cell membranes into the extracellular environment (Sydow et al., 2014). These organisms play important roles in biogeochemical cycles in soils and sediments, bioremediation of toxic metals, and energy generation in microbial fuel cells (Lovley, 1991; Lovley and Coates, 1997; Logan, 2009; Lovley, 2011). For example, electroactive *Deltaproteobacteria* in the genus *Geobacter* transport electrons through conductive pili, also known as electroactive pili (e-pili), composed of e-pilin structural subunits (Ntarlagiannis et al., 2007). *Geobacter* use e-pili for Fe(III) respiration, direct interspecies electron transfer (DIET), and growth on anodes (Reguera et al., 2005; Reguera et al., 2006; Rotaru et al., 2014).

E-pili are members of the type IV pilin proteins, which exhibit broad distribution among the prokaryotic domains (Giltner et al., 2012). Type IV pili have evolved to perform diverse cellular functions, including twitching motility, attachment, and genetic transformation. Additionally, most bacteria encode pseudopilins proteins which are structurally similar to but phylogenetically distinct from type IV pilins and assemble into type II secretion systems instead of pili (Hobbs and Mattick, 1993; Reguera et al., 2005; Cianciotto, 2009; Ayers et al., 2010). Characterized *Geobacter* e-pilins are truncated versions of canonical type IV pilins.

Aromatic amino acid density seems to be essential for efficient electron transport in e-pili (Vargas et al., 2013). It is thought that the close packing of aromatic residues within the pilus facilitates electron transport (Feliciano et al., 2015). The most conductive e-pilus measured to date is that of *Geobacter metallireducens*, which contains pilins with 15.3% aromatics, 59 aa mature length (after signal peptide sequence removal at the prepilin cleavage site), and no aromatic-free gaps >22 aa. The *G. metallireducens* e-pilus is 5000 times more conductive than

63 the *Geobacter sulfurreducens* e-pilus, which has pilins with 9.8% aromatics, 61 aa mature
64 length, and no aromatic-free gaps >22 aa (Tan et al., 2017). The *G. sulfurreducens* e-pilus is 100
65 times more conductive than the *Geobacter uraniireducens* pilus, which is not electroactive, and
66 contains much longer pilins (193 aa), 9.1% aromatics, and a maximum 53 aa aromatic-free gap
67 (Tan et al., 2016). Longer, non-electroactive type IV pilins (120-200 aa), such as those from *G.*
68 *uraniireducens* and *Neisseria gonorrhoeae*, are thought to be incapable of electroactivity due to
69 insufficient aromatic residue packing (Feliciano et al., 2015; Malvankar et al., 2015; Kolappan et
70 al., 2016).

71 Recently, the phylogenetic and structural diversity of e-pili has expanded beyond
72 *Geobacter* spp. with the discovery of strongly conductive pili made of much longer e-pilins. E-
73 pilins from *Desulfurivibrio alkaliphilus* (*Firmicutes*) and *Calditerrivibrio nitroreducens*
74 (*Deferribacteres*) have significantly longer pilins (119 and 182 aa, respectively) than *Geobacter*
75 e-pilins (Walker et al., 2017). Yet, like *Geobacter* e-pilins, they have high aromaticity (11 and
76 13%, respectively) and small aromatic-free gaps (22 aa and 27 aa, respectively). Recently, the
77 flagellum of the archaeon *Methanospirillum hungatei*, comprised of archaellin subunits
78 containing 11% aromatics and 35 aa maximum aromatic free gaps, was also found to be
79 conductive (Walker et al., 2019). Non-conductive pili from *Desulfofervidus auxilii*, *Shewanella*
80 *oneidensis*, and *Pseudomonas aeruginosa* have pilins with 5.6-6.8% aromatics, 126-206 aa, and
81 42-52 aa aromatic-free gaps (Reguera et al., 2005; Liu et al., 2014; Walker et al., 2017).
82 Therefore, it seems that aromatic density, defined here as percentage of aromatic amino acids
83 and spacing of aromatic residues in the pilin sequence, is the key factor for identifying putative
84 e-pilins based on sequence similarity.

Outside of the organisms referenced above, the phylogenetic and structural diversity of e-pilins are largely unknown. Characterization of phylogenetically diverse e-pili is hindered by the lack of cultured representatives for the vast majority of microbial diversity (Torsvik et al., 2002). Recently, pili were discovered in metagenomes of uncultivated bacteria belonging to the candidate phyla radiation (CPR) from deep subsurface samples (Luef et al., 2015). CPR bacteria have unusually small genomes and comprise a significant fraction of the bacterial domain (Brown et al., 2015; Castelle et al., 2018). Many lineages have type IV pilin genes and form pili-like structures, the function of which remains unknown (Albertsen et al., 2013; Kantor et al., 2013; Castelle et al., 2017).

In this study, we used sequence information to search for and predict e-pilins based on aromatic density in metagenomes from anoxic and metal-rich environments. Our results suggest that organisms from diverse phyla may use e-pili, and longer e-pilins may be more prevalent than truncated e-pilins in nature. We also found that Fe(III)-reducing sediment enrichments may select for shorter *Deltaproteobacteria* e-pilins compared to those in native sediment.

Results

Aromatic density distinguishes e-pilins from other type IV pilins. First, we obtained published sequences for biochemically confirmed e-pilins from seven organisms (**Table S1**; Walker et al., 2017), predicted e-pilins from ten organisms (**Table S2**; Holmes et al., 2016), and functionally verified attachment/motility/competence pilins from 36 organisms (26 type IV pilins and 10 pseudopilins; **Table S3**). All biochemically confirmed and predicted e-pilins had >9.8% aromatics and an aromatic-free gap of ≤35 aa, although most had a gap ≤22 aa. All but two of the 36 pilins implicated in functions other than electroactivity had <9.8% aromatic percentage and

108 >22-aa aromatic-free gaps. The two aromatically-dense pilins were OxpG, an atypical
109 pseudopilin involved in metal reduction in *G. sulfurreducens* (Mehta et al., 2006), and PilA from
110 *Dichelobacter nodosus* (Han et al., 2007). The low aromatic density of 34/36 of the pilins
111 implicated in functions other than electroactivity supports the hypothesis that high aromatic
112 density is indicative of e-pilins. Therefore, we used ≥9.8% aromatics and ≤22-aa aromatic-free
113 gap as a conservative threshold for predicting putative e-pilins from metagenomes.
114
115 *Putative e-pilins can be a significant portion of type IV pilins in natural environments and are*
116 *longer and more aromatically dense than characterized e-pilins.* We used the *G. sulfurreducens*
117 e-pilin protein sequences to query either metagenomic contigs or metagenome-assembled
118 genomes (MAGs) from 16 environments chosen based on where we predicted anaerobic
119 respiration of metals to be an important microbial metabolism **(Table 1)**. Environments included
120 ferruginous sediments and lake water, lake and river sediments, lake water, deep anoxic
121 groundwater, temperate forest soils, and crustal aquifer fluids. We then screened the retrieved
122 amino acid sequences for type IV pilins using Pilfind (Imam et al., 2011).
123 Our search recovered 2,433 type IV pilins, which were then screened using the aromatic
124 density threshold established above (≥9.8% aromatic amino acids and ≤22-aa aromatic gaps).
125 After partial sequences were removed, our search recovered 183 putative e-pilins of varying
126 length (58-219 aa) and aromatic density (9.8-17.8%). Putative e-pilins comprised up to one-third
127 of the total type IV pilins recovered from a single environmental metagenome (**Table 1**). Nearly
128 three-quarters (133/183) of putative e-pilins were >100 aa; only 15% (28/183) were truncated
129 (<70 aa). Over three-quarters (144/183) had 10-15% aromatics. Eight had ≥15% aromatics.
130 Twenty-one had 15-21 aa maximum aromatic-free gaps. The most aromatically dense pilin

131 (17.8% aromatics and a 16 aa gap) based on our predictions belonged to a *Falkowbacteria* (CPR)
132 MAG from groundwater.
133
134 *Sediment incubations enrich for different e-pilins than those present in the environment.* Next,
135 we identified putative e-pilin sequences in metagenomes of native ferruginous sediment from
136 Lake Matano, Indonesia where we expected e-pili due to high abundances of Fe(III), a potential
137 extracellular electron acceptor. Lake Matano is a permanently stratified tropical lake that hosts
138 one of the largest ferruginous environments on Earth, which we have previously studied for
139 Fe(III) reduction and coupled iron-methane cycling (Crowe et al., 2007; Crowe et al., 2011; Bray
140 et al., 2017). We also compared these to putative e-pilin sequences from year-long laboratory
141 incubations inoculated with Lake Matano sediment and enriched with iron and manganese oxides
142 (see **Experimental Procedures**). Putative e-pilins in the native sediment had closest similarity to
143 *Deltaproteobacteria*, *Aminicenantes* (OP8), *Planctomycetes*, and *Zixibacteria* (47-86% identity;
144 **Table S4**).
145 After anoxic incubation for over one year in the presence of iron or manganese oxides, 13
146 putative e-pilin sequences were identified in metagenomes (**Table S5**). Twelve had closest
147 BLAST hits (61-86% identity) to *Deltaproteobacteria* (including nine to *Geobacteraceae*), and
148 one to *Actinobacteria* (53% identity). Five truncated (59-72 aa) e-pilins were identified in long-
149 term enrichments, whereas no e-pilins <98 aa were identified in the native sediment. Although
150 the range of aromatic densities was similar for putative e-pilins in native and incubated
151 sediments (10.0-15.3%), aromatically dense pilins comprised a higher fraction of total type IV
152 pilins in incubations (38-100%) compared to native sediment (8%).
153

154 *Putative e-pilins are phylogenetically diverse*. To determine the phylogenetic diversity of pilins,
155 we aligned all of the pilin amino acid sequences described above and constructed a maximum
156 likelihood tree (**Fig. 1**). The phylogeny shows that the majority of recovered environmental
157 putative e-pilins are distinct from pseudopilins and other type IV pilins. A third of environmental
158 putative e-pilins, including 79% of truncated e-pilins, clustered with *Deltaproteobacteria*.
159 Twelve sequences were identified in MAGs of CPR bacteria from deep subsurface samples, and
160 nine sequences from ferruginous water from Kabuno Bay (Democratic Republic of Congo)
161 clustered with these CPR pilins. Putative e-pilins were also found in members of the
162 *Flavobacteriaceae*, *Dependentiae* (formerly TM6), *Nitrospinae*, *Acidobacteria*, *Elusimicrobia*,
163 *Alteromonas, Ectothiorhodospiraceae,* and *Betaproteobacteria* from groundwater MAGs.

164 Nine out of the ten pseudopilins from publicly available databases, as well as two
165 putative e-pilins from Kabuno Bay, formed a distinct cluster from other type IV pilins. Most of
166 the non-electroactive type IV pilins formed a distinct cluster; others were dispersed throughout
167 the tree. The *G. sulfurreducens* OxpG protein formed a distinct cluster, with eleven putative e-
168 pilins from anoxic groundwater, and anoxic and/or ferruginous sediment. The two competence
169 pilins from *Firmicutes* were divergent from most other sequences in the analysis, except for the
170 *Actinobacteria* sequence identified in Lake Matano sediment enrichments.

171

172 **Discussion**

173 *Expanding the phylogenetic diversity of e-pili*. Our identification of putative e-pilins in many
174 bacterial taxa outside of *Deltaproteobacteria* suggests that a wide diversity of organisms may
175 use e-pili for electroactivity. Perhaps the most notable among these are the CPR given that not
176 much is known about the biogeochemical role of these organisms despite them comprising >15%

of bacterial diversity (Brown et al., 2015; Castelle et al., 2018). While pilin genes have been identified in CPR genomes (Luef et al., 2015), their function has not been investigated. Although multiheme cytochromes involved in metal reduction have not been identified in CPR, our result suggests that some CPR may use pili for electroactivity. Due to their small genome size, many CPR organisms lack the genes necessary for full respiratory pathways, and images of CPR cells using pili to attach to other larger cells suggesting that they may be capable of DIET (Luef et al., 2015).

We identified putative e-pilins in taxa known or predicted to perform metal reduction, including *Alteromonas, Acidobacteria,* and *Zixibacteria*. Not only do cultured isolates of *Alteromonas* reduce Fe(III), but members of this genus can form electroactive biofilms (Vandecandelaere et al., 2008). The presence of e-pilins is also consistent with the predicted capability of *Zixibacteria* for Fe(III) reduction (Hernsdorf et al., 2017). We also identified putative e-pilins in many clades that have not been implicated in metal reduction or extracellular electron transport including *Planctomycetes*, *Elusimicrobia*, *Nitrospinae*, *Aminicenantes*, *Ectothiorhodospiraceae*, *Dependentiae* (formerly TM6), and *Flavobacteriaceae*. Multiheme cytochromes and MtrH homologs have been identified in *Planctomycetes* and *Elusimicrobia*, suggesting that the capacity for metal reduction may exist in these taxa (Sharma et al., 2010; Zhong and Shi, 2018). Detection of putative e-pilins in *Ectothiorhodospiraceae* also makes some sense given that members of this clade are capable of Fe(II) oxidation (Leguijt et al., 1993; Hallberg et al., 2011). In addition to putative e-pilins, several environmental sequences clustered with OxpG from *G. sulfurreducens*, suggesting that some sequences may not be type IV pilins, but instead atypical pseudopilins in metal-reducing type II secretion systems.

The diversity of pilin structure and characteristics is larger than previously thought. Our results suggest that truncation may be relatively rare among environmental e-pilins, and primarily associated with *Deltaproteobacteria*. We found pilins of many different lengths to be aromatically dense, suggesting that length may not be a determinant for e-pili. Three putative environmental e-pilins were more aromatically dense than the densest characterized e-pilin from *G. metallireducens* (>15.5%). Not only did some pilins have a high percentage of aromatic residues, they also had aromatic residues that were evenly spaced, leading to small aromatic-free gaps. Such features could allow these pilins to form remarkably conductive e-pili; however, this remains to be tested.

Sediment incubations select for aromatically dense pilins from Deltaproteobacteria. Long-term incubations of ferruginous lake sediments amended with abundant metal oxide as an electron acceptor enriched for aromatically dense pilin proteins. Nearly all type IV pilins identified from sediment enrichments were putative e-pilins, whereas a smaller percentage of pilins in the native sediment were aromatically dense. Pilins identified in native sediments were longer and more phylogenetically diverse, while laboratory incubations selected for shorter, aromatically dense pilins belonging primarily to *Deltaproteobacteria*. Proteomics of these enrichments reveal that some of these pilins are highly expressed (Szeinbaum et al., in prep). Our results suggest that sediment incubations can be a useful method for studying certain aromatically dense pilins from the environment. The discrepancy between incubated and native sediment is likely do the enrichment conditions employed in this study. Different enrichment conditions may allow the enrichment of different putative e-pilins. However, many environment organisms do not grow well under any known laboratory conditions (Harwani, 2013; Lok, 2015). Furthermore, e-pilins

in natural environments may be used for functions other than metal reduction, such as DIET or some other unknown function, and so will not be selected for under metal-reducing enrichments.

Possible functions for aromatically dense pilins besides electroactivity. Out of the 36 non-electroactive pilin and pseudopilin proteins analyzed, only two pilins were sufficiently aromatically dense to form e-pili based on the thresholds used in this study. One, the OxpG pseudopilin from *G. sulfurreducens*, is involved in metal reduction, which may explain its high aromatic density. In general, pilins dedicated to functions other than extracellular electron transport have low aromatic density, potentially because high amounts of aromatics could make proteins more susceptible to oxidative damage (Stadtman and Levine, 2003). Hence, it is unlikely that pilin proteins would contain a high percentage of evenly spaced aromatic residues if this configuration did not confer a selective advantage.

While we infer that the aromatically dense pilins in this study form e-pili, we cannot be certain that they are not being used for some other function. It is unlikely that many of the sequences could be pseudopilins, given that functionally verified pseudopilin proteins clustered separately. Other uses for aromatic amino acids include protein stability and DNA binding (Burley and Petsko, 1985; Baker and Grant, 2007). Aromatic residues may help to hold pilin structure together under harsh environmental conditions such as high temperature and low pH. However, competence pilins contained very low aromatic density, suggesting either that aromatic residues are not the method of DNA binding or that only a few specific residues are involved, and none of the environments used in this study were particularly extreme.

Conclusions. This study identified putative e-pilins in the environment using aromatic density and gaps as the predictive tool. The sequences we recovered suggest that e-pilins are both phylogenetically and structurally diverse. We conclude that e-pili may be composed of pilin monomers of a variety of lengths and aromatic densities, and that diverse bacteria may use e-pili for electron transfer or possibly other unknown functions. The lack of putative e-pilins belonging to Archaea may be due to our use of a bacterial e-pilin as the query sequence for metagenomic e-pilin identification, as well as our conservative aromatic density cutoff.

While alternate uses for aromatic amino acid residues exist (e.g. protein stability, nucleic acid binding), these seem unlikely explanations given the environments from which e-pilins were recovered and the low aromatic density of known competence pilins. Still, without biochemical verification of conductivity and/or identification of other genes involved in pili biogenesis and Fe(III)-reduction, we cannot definitively assign function to extracellular electron transport. Overall, this study suggests that electroactive bacteria may be more phylogenetically diverse and environmentally distributed than previously thought.

Acknowledgements. This research was funded by NASA Exobiology grant NNX14AJ87G with support from the NASA Astrobiology Institute (NNA15BB03A). SAC, RLS and KTJ were supported through NSERC Discovery grant 0487, CFI projects 229652 and 36071, and the Canada Research Chairs program. We thank Bianca Costa, Miles Mobley, Benjamin Reed, and Johnny Striepen for assistance with laboratory incubations, and Sean Elliott for helpful discussions.

Experimental Procedures

Sampling and enrichment of Lake Matano Sediment. Two sediment cores were obtained from 590 water depth in Lake Matano, Sulawesi Island, Indonesia in May 2010 (2°28′S, 121°20′E, *in situ* sediment temperature ~27°C) and stored under anoxic conditions. The sediments were mixed with anoxic freshwater media in a 1:5 ratio in an anoxic chamber and dispensed in stoppered serum bottles, as in Bray et al. (2017). Cultures were amended first with goethite and later with ferrihydrite. They were incubated for 490 days at 30°C, with multiple transfers, each time diluting the original sediment with freshwater media. Sediment had been diluted over 1000-fold by the time DNA was extracted for sequencing. Details on metagenomes from 395-day anoxic enrichments of Lake Matano sediment incubated with Mn(III) pyrophosphate are reported in a separate publication (Szeinbaum et al., in prep).

DNA extraction and metagenome sequencing, assembly, binning and annotation. Community DNA from Lake Matano sediment enrichments was extracted from 2 g samples and purified using a PowerSoil Isolation Kit and UltraClean® 15 Purification Kit (formerly MO BIO Laboratories, now Qiagen, Carlsbad, CA, USA) following the manufacturer's protocol. Indexed libraries were created from purified community DNA using the NexteraXT DNA Sample Prep kit (Illumina, San Diego, CA, USA) following manufacturer instructions. Libraries were pooled and sequenced on two runs of an Illumina MiSeq using a 500 cycle (paired end 250 x 250 bp) kit. Illumina reads were quality trimmed using Trim Galore! (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with a quality score and minimum length cutoff of Q25 and 100 bp, respectively, and merged with FLASH with the shortest overlap of 25 bp. Barcoded sequences were de-multiplexed, trimmed (length cutoff 100 bp), and filtered to remove low quality reads (average Phred score <25) using Trim Galore!. Forward and reverse reads were assembled using SPAdes (Nurk et al., 2013) with the 'meta' option. The number of contigs, contig length, GC content, N50, and L50 assembly statistics were calculated with metaQUAST (Mikheenko et al., 2015). Raw sequence reads, and all genomic bins were deposited in NCBI under the accession number PRJNA505658.

Pilin identification from microbial metagenomes. Environmental metagenomes and metagenome assembled genomes (MAGs) were downloaded from IMG JGI and NCBI (see **Table 1** for taxon object IDs). For all metagenomes, Prodigal (Hyatt et al., 2010) was used to predict genes from contig files and write them to amino acid fasta files. Amino acid sequences from Mags were downloaded directly from NCBI. Predicted protein files were then used as databases for protein BLAST, using the *G. sulfurreducens* PilA protein as query. Hits with a bit score greater than 55 were pulled from the databases. These recovered sequences were then further verified as type IV pili using Pilfind (http://signalfind.org/pilfind.html), a web tool that identifies type IV pilin signal sequences (Imam et al., 2011). Pilin amino acid sequences were then run through a python script that calculated the mature pilin length, percent aromatic amino acids, and aromatic free gaps (https://github.com/GlassLabGT/Python-scripts). Partial genes were retained if truncated on the N-terminus before the signal peptide and removed if truncated on the C-terminus.

Pilin multiple sequence alignment and phylogenetic analysis. Identified pilin amino acid sequences were aligned using ClustalW and a maximum likelihood tree was constructed using

MEGA. The alignment is provided as Supplemental Material. Duplicate sequences were removed before tree construction. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones et al., 1992). The tree with the highest log likelihood is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 182 amino acid sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were 53 positions total in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016). Relative conductivities of characterized e-pilin genes were calculated using conductance values from Walker et al. (2017) and Tan et al. (2017).

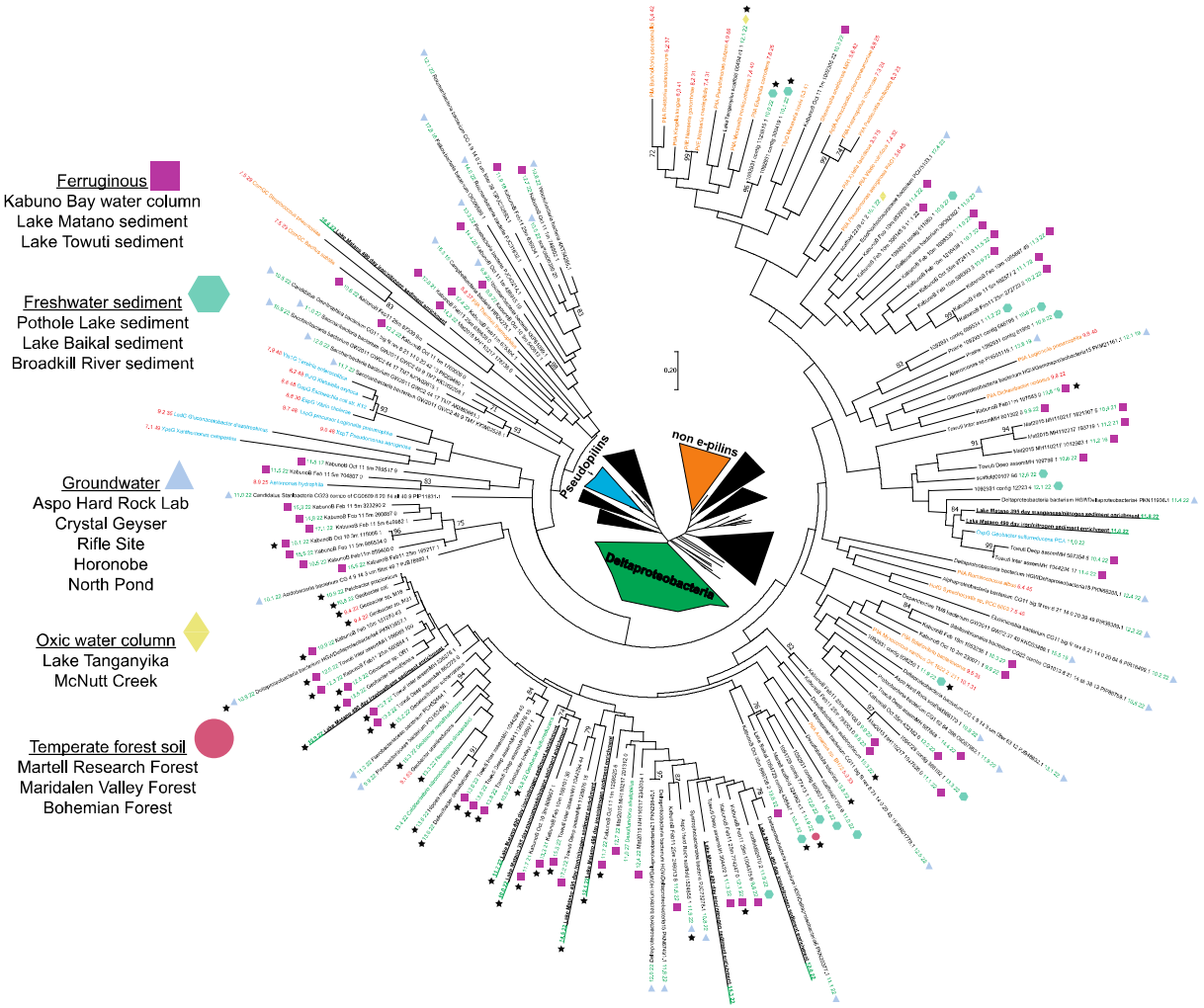

Figure 1. Maximum likelihood phylogenetic tree of pilin sequences. Functionally verified pseudopilins are shown in blue while type IV pilins implicated in functions other than electroactivity are shown in orange. Characterized e-pilins are shown in green. Putative e-pilins from Lake Matano sediment enrichments are bolded and underlined. Putative e-pilins from environmental genomes have symbols beside them denoting the type of environment from which they were recovered. Aromatic percentage and smallest aromatic free gap are listed beside each sequence. Red numbers refer to low aromatic density (<9.8% aromatics and >22 aa gap). Green numbers refer to high aromatic density (≥9.8% aromatics and ≤22 aa gap). Black starred sequences are <70 aa long. The small radiation tree in the middle highlights the major clusters on the larger tree.

339 **Table 1. Putative e-pilin genes identified from 16 environments analyzed in this study.**
340 Putative e-pilins are defined as containing ≥9.8% aromatics and ≤22 aa aromatic gap. See **Fig. 1**
341 for details about each putative e-pilin.

Environment (Data type)	Sample	IMG ID/ GenBank BioProject	Protein-coding genes	Putative e-pilins / type 4 pilins	Mature length (aa)
Lake Matano, Indonesia (contigs)	Ferruginous sediment (metagenomic contigs)	PRJNA521166	1,157,873	7/93	90-184
Lake Towuti, Indonesia (contigs)		3300010328, 3300010324	10,203,152	16/113	58-131
Kabuno Bay, Lake Kivu, Democratic Republic of Congo (contigs)	Ferruginous water column	3300014720, 3300013122, 3300013123, 3300013125, 3300013126, 3300013131, 3300013132, 3300013136, 3300013137	24,761,074	88/ 1,118	57-220
Prairie Pothole Lake, North Dakota, USA (contigs)	Sediment	3300027975	1,493,055	8/24	60-119
Lake Baikal, Russia (contigs)		3300025843	1,180,048	4/32	67-174
Broad Kill River, Delaware, USA (contigs)		3300019765	669,074	0/1	N/A
Aspo, Sweden (contigs)	Ground-water	3300014656	3,495,273	7/140	60-185
Crystal Geyser, Utah, USA (MAGs)		PRJNA297582, PRJNA362739	1,537,765	29/413	56-214
Rifle, Colorado, USA (MAGs)		PRJNA273161	633,116	11/389	106-158
Horonobe, Japan (MAGs)		PRJNA321556	557,865	7/37	64-146
Lake Tanganyika, Africa (contigs)	Water column	3300020183, 3300021092	4,124,017	1/13	58
Georgia, USA (contigs)		3300021437	464,898	0/5	129
Indiana, USA (contigs)	Forest Soil	3300024317	530,797	1/12	67
Norway (contigs)		3300028037	161,101	0/0	N/A
Czech Republic (contigs)		3300023019	80,325	0/0	N/A
North Pond, Mid-Atlantic Ridge (contigs)	Crustal aquifer	PRJNA391950	494,807	4/43	131-146
			Total	**183/ 2,433**	**56-220**

342

343 **Table 2. Putative e-pilin genes identified in long term enrichment of Lake Matano sediment in the presence of metal oxides.**
344 Putative e-pilins are defined as containing ≥9.8% aromatics and no >22-aa aromatic gaps.

Enrichment conditions	GenBank BioSample or BioProject	Protein-coding genes	Putative e-pilins / type IV pilins	Mature length (aa)
490-d Fe/CH_4 (311FMe)	SAMN10492640	66,924	2/4	64, 115
490-d Fe/N_2 (311FN)	SAMN10411764	268,143	6/16	59, 61, 64, 72, 109, 128
395-d Mn/N_2 (s9)	PRJNA489678	7,050	2/2	60, 127
395-d Mn/N_2 (s11)	PRJNA489678	45,534	2/3	60, 127
		Total:	**12/25**	**59-128**

345
346

References

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W., and Nielsen, P.H. (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnology* **31**: 533.

Ayers, M., Howell, P.L., and Burrows, L.L. (2010) Architecture of the type II secretion and type IV pilus machineries. *Future Microbiology* **5**: 1203-1218.

Baker, C.M., and Grant, G.H. (2007) Role of aromatic amino acids in protein–nucleic acid recognition. *Biopolymers: Original Research on Biomolecules* **85**: 456-470.

Bray, M.S., Wu, J., Reed, B.C., Kretz, C.B., Belli, K.M., Simister, R.L. et al. (2017) Shifting microbial communities sustain multi-year iron reduction and methanogenesis in ferruginous sediment incubations. *Geobiology* **15**: 678-689.

Brown, C.T., Hug, L.A., Thomas, B.C., Sharon, I., Castelle, C.J., Singh, A. et al. (2015) Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**: 208-211.

Burley, S., and Petsko, G.A. (1985) Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science* **229**: 23-28.

Castelle, C.J., Brown, C.T., Thomas, B.C., Williams, K.H., and Banfield, J.F. (2017) Unusual respiratory capacity and nitrogen metabolism in a Parcubacterium (OD1) of the Candidate Phyla Radiation. *Scientific reports* **7**: 40101.

Castelle, C.J., Brown, C.T., Anantharaman, K., Probst, A.J., Huang, R.H., and Banfield, J.F. (2018) Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. *Nature Reviews Microbiology* **16**: 629.

Cianciotto, N.P. (2009) Many substrates and functions of type II secretion: lessons learned from Legionella pneumophila. *Future Microbiology* **4**: 797-805.

Crowe, S., Katsev, S., Leslie, K., Sturm, A., Magen, C., Nomosatryo, S. et al. (2011) The methane cycle in ferruginous Lake Matano. *Geobiology* **9**: 61-78.

Crowe, S.A., O'Neill, A.H., Kulczycki, E., Weisener, C.G., Roberts, J.A., and Fowle, D.A. (2007) Reductive dissolution of trace metals from sediments. *Geomicrobiology Journal* **24**: 157-165.

Feliciano, G., Steidl, R., and Reguera, G. (2015) Structural and functional insights into the conductive pili of *Geobacter sulfurreducens* revealed in molecular dynamics simulations. *Physical Chemistry Chemical Physics* **17**: 22217-22226.

Giltner, C.L., Nguyen, Y., and Burrows, L.L. (2012) Type IV pilin proteins: Versatile molecular modules. *Microbiology and Molecular Biology Reviews* **76**: 740-772.

Hallberg, K.B., Hedrich, S., and Johnson, D.B. (2011) *Acidiferrobacter thiooxydans*, gen. nov. sp. nov.; an acidophilic, thermo-tolerant, facultatively anaerobic iron-and sulfur-oxidizer of the family *Ectothiorhodospiraceae*. *Extremophiles* **15**: 271-279.

382 Han, X., Kennan, R.M., Parker, D., Davies, J.K., and Rood, J.I. (2007) Type IV fimbrial biogenesis is
383 required for protease secretion and natural transformation in *Dichelobacter nodosus*. *Journal of*
384 *Bacteriology* **189**: 5022-5033.

385 Harwani, D. (2013) The great plate count anomaly and the unculturable bacteria. *Microbiology* **2**: 350-
386 351.

387 Hernsdorf, A.W., Amano, Y., Miyakawa, K., Ise, K., Suzuki, Y., Anantharaman, K. et al. (2017) Potential
388 for microbial H2 and metal transformations associated with novel bacteria and archaea in deep terrestrial
389 subsurface sediments. *ISME J* **11**: 1915-1929.

390 Hobbs, M., and Mattick, J.S. (1993) Common components in the assembly of type 4 fimbriae, DNA
391 transfer systems, filamentous phage and protein‐secretion apparatus: a general system for the formation
392 of surface‐associated protein complexes. *Molecular Microbiology* **10**: 233-243.

393 Holmes, D.E., Dang, Y., Walker, D.J., and Lovley, D.R. (2016) The electrically conductive pili of
394 Geobacter species are a recently evolved feature for extracellular electron transfer. *Microbial Genomics* **2**.

395 Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal:
396 prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.

397 Imam, S., Chen, Z., Roos, D.S., and Pohlschröder, M. (2011) Identification of surprisingly diverse type
398 IV pili, across a broad range of gram-positive bacteria. *PloS One* **6**: e28919.

399 Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992) The rapid generation of mutation data matrices
400 from protein sequences. *Bioinformatics* **8**: 275-282.

401 Kantor, R.S., Wrighton, K.C., Handley, K.M., Sharon, I., Hug, L.A., Castelle, C.J. et al. (2013) Small
402 genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. *MBio* **4**:
403 e00708-00713.

404 Kolappan, S., Coureuil, M., Yu, X., Nassif, X., Egelman, E.H., and Craig, L. (2016) Structure of the
405 *Neisseria meningitidis* Type IV pilus. *Nature Communications* **7**: 13015.

406 Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis
407 version 7.0 for bigger datasets. *Molecular biology and evolution* **33**: 1870-1874.

408 Leguijt, T., Engels, P., Crielaard, W., Albracht, S., and Hellingwerf, K. (1993) Abundance, subunit
409 composition, redox properties, and catalytic activity of the cytochrome bc1 complex from alkaliphilic and
410 halophilic, photosynthetic members of the family Ectothiorhodospiraceae. *Journal of bacteriology* **175**:
411 1629-1636.

412 Liu, X., Tremblay, P.-L., Malvankar, N.S., Nevin, K.P., Lovley, D.R., and Vargas, M. (2014) A
413 *Geobacter sulfurreducens* strain expressing *Pseudomonas aeruginosa* type IV pili localizes OmcS on pili
414 but is deficient in Fe (III) oxide reduction and current production. *Appl Environ Microbiol* **80**: 1219-1224.

415 Logan, B.E. (2009) Exoelectrogenic bacteria that power microbial fuel cells. *Nature Reviews*
416 *Microbiology* **7**: 375-381.

417 Lok, C. (2015) Mining the microbial dark matter. *Nature News* **522**: 270.

418 Lovley, D.R. (1991) Dissimilatory Fe(III) and Mn(IV) reduction. *Microbiological Reviews* **55**: 259-287.

419 Lovley, D.R. (2011) Live wires: direct extracellular electron exchange for bioenergy and the
420 bioremediation of energy-related contamination. *Energy & Environmental Science* **4**: 4896-4906.

421 Lovley, D.R., and Coates, J.D. (1997) Bioremediation of metal contamination. *Current Opinion in*
422 *Biotechnology* **8**: 285-289.

423 Luef, B., Frischkorn, K.R., Wrighton, K.C., Holman, H.-Y.N., Birarda, G., Thomas, B.C. et al. (2015)
424 Diverse uncultivated ultra-small bacterial cells in groundwater. *Nature Communications* **6**.

425 Malvankar, N.S., Vargas, M., Nevin, K., Tremblay, P.-L., Evans-Lutterodt, K., Nykypanchuk, D. et al.
426 (2015) Structural basis for metallic-like conductivity in microbial nanowires. *MBio* **6**: e00084-00015.

427 Mehta, T., Childers, S.E., Glaven, R., Lovley, D.R., and Mester, T. (2006) A putative multicopper protein
428 secreted by an atypical type II secretion system involved in the reduction of insoluble electron acceptors
429 in *Geobacter sulfurreducens*. *Microbiology* **152**: 2257-2264.

430 Mikheenko, A., Saveliev, V., and Gurevich, A. (2015) MetaQUAST: evaluation of metagenome
431 assemblies. *Bioinformatics* **32**: 1088-1090.

432 Ntarlagiannis, D., Atekwana, E.A., Hill, E.A., and Gorby, Y. (2007) Microbial nanowires: Is the
433 subsurface "hardwired"? *Geophysical Research Letters* **34**.

434 Nurk, S., Bankevich, A., Antipov, D., Gurevich, A.A., Korobeynikov, A., Lapidus, A. et al. (2013)
435 Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *Journal of*
436 *Computational Biology* **20**: 714-737.

437 Reguera, G., McCarthy, K.D., Mehta, T., Nicoll, J.S., Tuominen, M.T., and Lovley, D.R. (2005)
438 Extracellular electron transfer via microbial nanowires. *Nature* **435**: 1098-1101.

439 Reguera, G., Nevin, K.P., Nicoll, J.S., Covalla, S.F., Woodard, T.L., and Lovley, D.R. (2006) Biofilm and
440 nanowire production leads to increased current in *Geobacter sulfurreducens* fuel cells. *Applied and*
441 *Environmental Microbiology* **72**: 7345-7348.

442 Rotaru, A.-E., Shrestha, P.M., Liu, F., Markovaite, B., Chen, S., Nevin, K.P., and Lovley, D.R. (2014)
443 Direct interspecies electron transfer between *Geobacter metallireducens* and *Methanosarcina barkeri*.
444 *Applied and Environmental Microbiology* **80**: 4599-4605.

445 Sharma, S., Cavallaro, G., and Rosato, A. (2010) A systematic investigation of multiheme c-type
446 cytochromes in prokaryotes. *JBIC Journal of Biological Inorganic Chemistry* **15**: 559-571.

447 Stadtman, E., and Levine, R. (2003) Free radical-mediated oxidation of free amino acids and amino acid
448 residues in proteins. *Amino acids* **25**: 207-218.

449 Sydow, A., Krieg, T., Mayer, F., Schrader, J., and Holtmann, D. (2014) Electroactive bacteria—molecular
450 mechanisms and genetic tools. *Applied microbiology and biotechnology* **98**: 8481-8495.

451 Szeinbaum, N., Nunn, B., Cavazos, A., Crowe, S., Stewart, F., DiChristina, T. et al. (in prep) Expression
452 of extracellular multiheme cytochromes in betaproteobacteria during Mn(III) reduction.

Tan, Y., Adhikari, R.Y., Malvankar, N.S., Ward, J.E., Woodard, T.L., Nevin, K.P., and Lovley, D.R. (2017) Expressing the *Geobacter metallireducens* PilA in *Geobacter sulfurreducens* yields pili with exceptional conductivity. *mBio* **8**: e02203-02216.

Tan, Y., Adhikari, R.Y., Malvankar, N.S., Ward, J.E., Nevin, K.P., Woodard, T.L. 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.

Torsvik, V., Øvreås, L., and Thingstad, T.F. (2002) Prokaryotic diversity--magnitude, dynamics, and controlling factors. *Science* **296**: 1064-1066.

Vandecandelaere, I., Nercessian, O., Segaert, E., Achouak, W., Mollica, A., Faimali, M. et al. (2008) *Alteromonas genovensis* sp. nov., isolated from a marine electroactive biofilm and emended description of *Alteromonas macleodii* Baumann et al. 1972 (Approved Lists 1980). *International journal of Systematic and Evolutionary Microbiology* **58**: 2589-2596.

Vargas, M., Malvankar, N.S., Tremblay, P.-L., Leang, C., Smith, J.A., Patel, P. et al. (2013) Aromatic amino acids required for pili conductivity and long-range extracellular electron transport in *Geobacter sulfurreducens*. *mBio* **4**.

Walker, D.J.F., Martz, E., Holmes, D.E., Zhou, Z., Nonnenmann, S.S., and Lovley, D.R. (2019) The archaellum of *Methanospirillum hungatei* is electrically conductive. *mBio* **10**: e00579-00519.

Walker, D.J.F., Adhikari, R.Y., Holmes, D.E., Ward, J.E., Woodard, T.L., Nevin, K.P., and Lovley, D.R. (2017) Electrically conductive pili from pilin genes of phylogenetically diverse microorganisms. *The ISME Journal*.

Zhong, Y., and Shi, L. (2018) Genomic Analyses of the Quinol Oxidases and/or Quinone Reductases Involved in Bacterial Extracellular Electron Transfer. *Frontiers in microbiology* **9**.