ASM | mSystems

| Environmental Microbiology | Research Article

Distinguishing *Leptothrix* and *Sphaerotilus* genera by an integrated genomic-phenotypic analysis supported by new *Leptothrix* genomes

Gracee K. Tothero,[1,2,3] Jessica L. Keffer,[1,2] David Emerson,[4] Emily J. Fleming,[5] Clara S. Chan[1,2,3]

AUTHOR AFFILIATIONS See affiliation list on p. 20.

ABSTRACT The *Sphaerotilus-Leptothrix* group of bacteria includes one of the first described microorganisms, *Leptothrix ochracea*, an uncultured type strain, plus isolates of *Leptothrix* and *Sphaerotilus*. This group is unified by the ability to form sheaths and oxidize metals, although *L. ochracea* exhibits obvious ecological, morphological, and functional differences from the rest of *Sphaerotilus-Leptothrix*. Recently, there have been calls to combine the group into one genus, *Sphaerotilus*; however, these studies lacked adequate genomic representation of *L. ochracea*. Here, we present a comprehensive comparative genomic analysis of the *Sphaerotilus-Leptothrix* group, including expanded representation of *L. ochracea*, a closely related novel species, *Leptothrix toolikensis*, and two new isolates (*Leptothrix mechoopdaensis*). Analysis of 38 genomes resolves three phylogenetic and functional groups: the *ochracea*-type *Leptothrix* (Group 1), the *mobilis*-type *Leptothrix* (Group 2), and *Sphaerotilus* (Group 3). Group 1 genomes form a separate genus based on average nucleotide identity and alignment fraction. The genomes clearly diverge from the rest of *Sphaerotilus-Leptothrix* in phylogeny, size, and metabolic potential. Group 1 genomes are much smaller (2.59–3.04 Mb) than those of Groups 2 (4.55–6.06 Mb) and 3 (3.94–5.07 Mb), while encoding more metal oxidases and fewer carbohydrate-active enzymes. Group 2 clusters with Group 3 phylogenetically and is similar in organic carbon metabolisms but maintains more metal oxidation genes. Group 2 members lack homogeneity in phenotype and genotype, suggesting that additional isolates and genomes are needed for confident classification. However, Group 1 genomes (*L. ochracea* and *L. toolikensis*) show clear divergence, precluding their inclusion in *Sphaerotilus* and supporting the retention of the genus *Leptothrix*.

IMPORTANCE Researchers have long noted differences in metal oxidation, morphology, and ecology among *Sphaerotilus-Leptothrix*, but longstanding confusion over phylogeny and genus boundaries led to inconsistent taxonomic classification between the two genera. This confusion stems from previous work that used isolates that are unavailable or lost distinguishing traits in culture, and from limited genomic data. Furthermore, the *Leptothrix* type strain *L. ochracea* has never been isolated. This study provides molecular evidence that substantiates calls to reassign some *Leptothrix* members to the genus *Sphaerotilus* but adds to an emerging body of evidence that Group 1 *L. ochracea* and now *L. toolikensis* represent a functionally distinct lineage. While genomic similarity metrics left taxonomic divisions unclear, integrating metabolic potential with phylogeny resolved genus boundaries based on clear functional groupings. This polyphasic approach for delineating genera clarifies longstanding taxonomic confusion and refines our understanding of functional diversity both across and within *Sphaerotilus-Leptothrix* lineages.

KEYWORDS iron oxidation, *Leptothrix*, *Sphaerotilus*, pangenome, metagenomics

Editor Ryan J. Newton, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin, USA

Address correspondence to Clara S. Chan, cschan@udel.edu.

The authors declare no conflict of interest.

See the funding table on p. 20.

See the related article at https://doi.org/10.1128/mra.01458-25.

Received 19 December 2025
Accepted 5 May 2026
Published 12 June 2026

Copyright © 2026 Tothero et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

The bacteria of the *Sphaerotilus-Leptothrix* group (SLG) are among the earliest described microbes due to their formation of clearly identifiable filaments that form biofilms, flocs, and mats. Studies on SLG date back to 1797 (1–5) and have historically considered *Leptothrix* and *Sphaerotilus* close relatives based on morphological and physiological traits (6–10). Both genera are metal-oxidizing bacteria that exhibit filamentous growth; chains of cells are enclosed in tubular sheaths, which can become encrusted in iron and/or manganese oxides (4–7, 11–18). The close association of *Leptothrix* and *Sphaerotilus* proved justified by phylogenetic analysis, initially of 16S rRNA and *mofA* genes, and most recently whole genome comparisons (6, 19–26). In fact, by some metrics, genomic analysis suggests that the group should be a single genus (26, 27). However, *L. ochracea* has always stood out as an outlier in its preference for low organic carbon environments and its prodigious production of iron-mineralized sheaths.

Leptothrix ochracea was the first described member of SLG (1) and remains the type species of *Leptothrix*; however, it has never been isolated, making it difficult to study its physiology and niche. Instead, isolates of *Leptothrix* and *Sphaerotilus* have served as models for describing this group. All SLG isolates grow organoheterotrophically (7, 17, 18, 28, 29); *Sphaerotilus* is found in environments with higher organic carbon (wastewater, organic-polluted streams) and its growth responds more strongly to organics, while most *Leptothrix* respond minimally, with the exception of *L. cholodnii* (6). Both *Leptothrix* and *Sphaerotilus* can oxidize iron, but *Sphaerotilus* do not tend to accumulate metal oxides in the environment like *Leptothrix* (6, 30). Manganese oxidation is generally considered unique to *Leptothrix* (6, 10), and while Schmidt et al. (24) reported that a *Sphaerotilus* isolate oxidized manganese, the ability to oxidize manganese is still considered a primary method for distinguishing *Leptothrix* from *Sphaerotilus*. Our understanding of the SLG is based on a limited number of isolates; thus, a more comprehensive functional analysis across a greater diversity of the group is still required to understand true distinctions in metal oxidation and organic utilization.

Phenotypic and ecological traits clearly distinguish the original, type species *L. ochracea* from other members of *Leptothrix*. Its sheaths become heavily encrusted in iron oxides and are mostly empty, containing few cells, unlike the sheaths of isolate species, which typically are densely populated with cells (6, 7, 12, 21, 31, 32). This growth pattern suggests that *L. ochracea* must oxidize extensive amounts of iron to meet its energetic needs, consistent with the low energy yield from iron oxidation (2, 12, 33–36). Xenic cultures can grow in Fe(II)-amended natural waters without any added organics and, in fact, require Fe(II) for growth (33). Genomic analyses give further support for iron oxidation as the major energy source and also for a mixotrophic metabolism with limited organic utilization abilities, in contrast to the broader substrate utilization capabilities of its organoheterotrophic relatives (21). These contrasts in function and genomic content suggest that *L. ochracea* differs substantially from other SLG in its impact on ecology and biogeochemistry, and the combined differences in phenotype, genotype, and ecotype could merit a separate genus.

Proper genus classification is significant because genera assignments are most commonly used for interpreting physiology, metabolism, and niche. For most of their history, *Leptothrix* and *Sphaerotilus* have been distinguished from each other and from close relatives based on phenotypes: cell size, sheath morphology, iron and manganese oxidation capabilities, organic carbon substrate utilization, and response to organics (6, 7, 10, 12, 24, 37). However, there have been numerous classification mistakes due to the loss of phenotypes in cultures post-isolation (e.g., sheaths and Mn oxidation) or phenotypic variations between culture conditions (6, 8, 9, 12, 24, 38–40). This led to a convoluted history of renaming between the two genera (7, 12, 34, 40). The confusion between genera was not completely alleviated once phylogenetic options became available, since *Leptothrix* and *Sphaerotilus* are closely related to each other and to other members of the Sphaerotilaceae (19, 36, 41–45). Consequently, it is often unclear whether an organism belongs to *Leptothrix* or *Sphaerotilus*, even when molecular data are available. These difficulties are perpetuated by incorrect and incomplete

classifications of SLG in reference databases (e.g., Silva; see Tothero et al. [21]). Stable phylogenies depend on sufficient numbers of sequences with relatively even distributions across each clade. Prior genome-based reclassification efforts relied on a sparse subset of reference genomes that either missed *L. ochracea* or included only partial representation of *Leptothrix* (26, 27, 44). A fair, consistent evaluation of the *Leptothrix* and *Sphaerotilus* genera requires the use of a robust molecular data set with adequate reference sequences, including representation from more than one *L. ochracea* genome.

To better resolve phylogenetic relationships and characterize functional diversity within the group, we performed a comprehensive comparative analysis of 38 high-quality genomes of *Sphaerotilus-Leptothrix* members. These include new genomes that fill in missing diversity in *Leptothrix* with two novel species represented by four metagenome-assembled genomes (*Leptothrix toolikensis*) and two isolate genomes (*Leptothrix mechoopdaensis*). We used the full set of sequences to delineate three distinct subgroups within the SLG: the *ochracea*-type *Leptothrix* (Group 1), the *mobilis*-type *Leptothrix* (Group 2), and *Sphaerotilus* (Group 3). We searched for genes indicating key metabolisms known to distinguish *Leptothrix* from *Sphaerotilus*, including metal oxidation, carbon fixation, and organic carbon utilization. This work establishes a framework and a comprehensive reference data set with which to identify additional SLG by sequence-based methods.

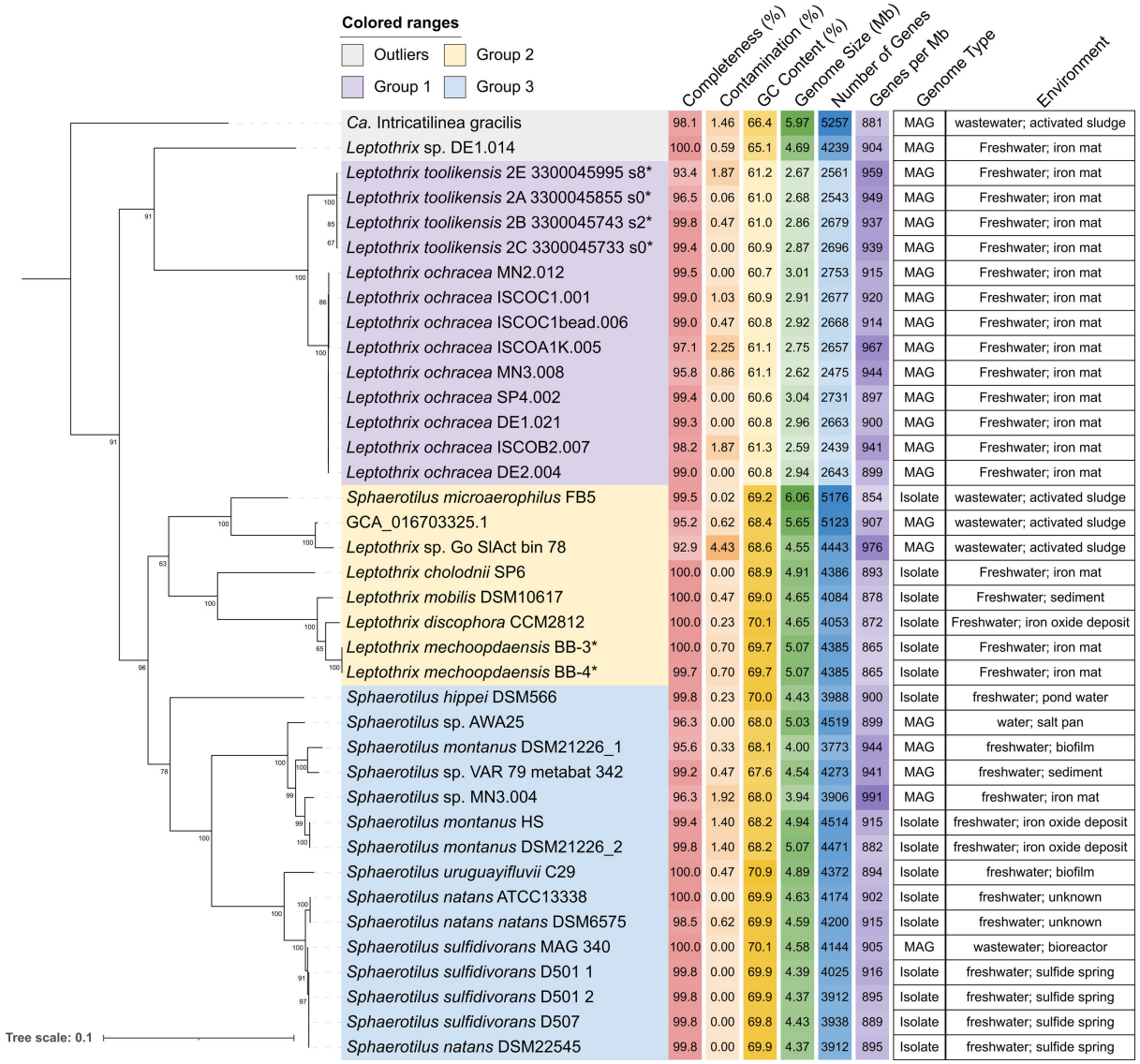
RESULTS

Phylogeny and genome features

We collected 38 high-quality SLG genomes, which include 16 isolates and 22 metagenome-assembled genomes (MAGs) from our own work and public databases (Table S1; Supplemental text), employing cutoffs of 90% completeness and 5% contamination, in accordance with MIMAG standards for high-quality draft genomes (46). This includes four new MAGs from Arctic iron microbial mats (see Michaud et al. [47] for site information), and two newly sequenced genomes of isolates (*Leptothrix mechoopdaensis*) from temperate stream iron mats (Fleming et al., [48]). Genomes were classified as SLG based on clustering in a monophyletic clade with known SLG in a concatenated ribosomal protein tree (Fig. 1).

Concatenated trees revealed three major subgroups within SLG; these groups cluster with *Leptothrix ochracea*, *Leptothrix mobilis*, and *Sphaerotilus natans* (Fig. 1; Fig. S1). We designate these three subgroups as SLG Group 1 (the *ochracea*-type *Leptothrix*), SLG Group 2 (the *mobilis*-type *Leptothrix*), and SLG Group 3 (*Sphaerotilus*). Despite their historical placement in different genera, the *L. mobilis* clade clusters more closely to the *S. natans* clade than it does to *L. ochracea* with high bootstrap support (91%–96%). Two genomes clustered with SLG but emerged as outliers (Fig. 1; Fig. S1); these include *Leptothrix* DE1.014 and one genome that was previously reclassified from *Leptothrix* to "*Candidatus* Intricatilinea gracilis" (44). However, "*Ca*. I. gracilis" exhibits the filamentous morphology characteristic of *Leptothrix* (44), supporting the inclusion of "*Ca*. I. gracilis" in this analysis. Amino acid identity (AAI) analysis of the two outliers showed they are sufficiently close (>70% identity) to *L. cholodnii* SP-6 to be considered part of the same genus (Fig. 2; [49]). Consequently, they are included but were not assigned to a specific subgroup.

AAI and average nucleotide identity (ANI) metrics can help estimate the species-level diversity among the three groups. Based on the thresholds of 95% AAI and ANI (50, 51), Group 1 contains two SLG species, Group 2 has six, and Group 3 has nine (Fig. 2; Table S2). The majority of Group 2 and some of the Group 3 genomes are single representatives of their own species. While these AAI- and ANI-based patterns reflect considerable diversity across Groups 2 and 3, they also suggest that sequencing efforts have not yet saturated the full diversity of SLG. Group 2 includes two genomes of novel *Leptothrix* isolates, *Leptothrix mechoopdaensis* strains BB-3 and BB-4, which were isolated from an iron seep in Centerville, CA (as described by Fleming et al. [48]). They are most closely related to *L. discophora* but do not meet the 95% AAI or ANI cutoff to be considered the same species (AAI = 87.7%; ANI = 89.4%). Therefore, we designate these isolates a novel *Leptothrix*



Ca. Intricatilinea gracilis	98.1	1.46	66.4	5.97	5257	881	MAG	wastewater; activated sludge
Leptothrix sp. DE1.014	100.0	0.59	65.1	4.69	4239	904	MAG	Freshwater; iron mat
Leptothrix toolikensis 2E 3300045995 s8*	93.4	1.87	61.2	2.67	2561	959	MAG	Freshwater; iron mat
Leptothrix toolikensis 2A 3300045855 s0*	96.5	0.06	61.0	2.68	2543	949	MAG	Freshwater; iron mat
Leptothrix toolikensis 2B 3300045743 s2*	99.8	0.47	61.0	2.86	2679	937	MAG	Freshwater; iron mat
Leptothrix toolikensis 2C 3300045733 s0*	99.4	0.00	60.9	2.87	2696	939	MAG	Freshwater; iron mat
Leptothrix ochracea MN2.012	99.5	0.00	60.7	3.01	2753	915	MAG	Freshwater; iron mat
Leptothrix ochracea ISCOC1.001	99.0	1.03	60.9	2.91	2677	920	MAG	Freshwater; iron mat
Leptothrix ochracea ISCOC1bead.006	99.0	0.47	60.8	2.92	2668	914	MAG	Freshwater; iron mat
Leptothrix ochracea ISCOA1K.005	97.1	2.25	61.1	2.75	2657	967	MAG	Freshwater; iron mat
Leptothrix ochracea MN3.008	95.8	0.86	61.1	2.62	2475	944	MAG	Freshwater; iron mat
Leptothrix ochracea SP4.002	99.4	0.00	60.6	3.04	2731	897	MAG	Freshwater; iron mat
Leptothrix ochracea DE1.021	99.3	0.00	60.8	2.96	2663	900	MAG	Freshwater; iron mat
Leptothrix ochracea ISCOB2.007	98.2	1.87	61.3	2.59	2439	941	MAG	Freshwater; iron mat
Leptothrix ochracea DE2.004	99.0	0.00	60.8	2.94	2643	899	MAG	Freshwater; iron mat
Sphaerotilus microaerophilus FB5	99.5	0.02	69.2	6.06	5176	854	Isolate	wastewater; activated sludge
GCA_016703325.1	95.2	0.62	68.4	5.65	5123	907	MAG	wastewater; activated sludge
Leptothrix sp. Go SlAct bin 78	92.9	4.43	68.6	4.55	4443	976	MAG	wastewater; activated sludge
Leptothrix cholodnii SP6	100.0	0.00	68.9	4.91	4386	893	Isolate	Freshwater; iron mat
Leptothrix mobilis DSM10617	100.0	0.47	69.0	4.65	4084	878	Isolate	Freshwater; sediment
Leptothrix discophora CCM2812	100.0	0.23	70.1	4.65	4053	872	Isolate	Freshwater; iron oxide deposit
Leptothrix mechoopdaensis BB-3*	100.0	0.70	69.7	5.07	4385	865	Isolate	Freshwater; iron mat
Leptothrix mechoopdaensis BB-4*	99.7	0.70	69.7	5.07	4385	865	Isolate	Freshwater; iron mat
Sphaerotilus hippei DSM566	99.8	0.23	70.0	4.43	3988	900	Isolate	freshwater; pond water
Sphaerotilus sp. AWA25	96.3	0.00	68.0	5.03	4519	899	MAG	water; salt pan
Sphaerotilus montanus DSM21226_1	95.6	0.33	68.1	4.00	3773	944	MAG	freshwater; biofilm
Sphaerotilus sp. VAR 79 metabat 342	99.2	0.47	67.6	4.54	4273	941	MAG	freshwater; sediment
Sphaerotilus sp. MN3.004	96.3	1.92	68.0	3.94	3906	991	MAG	freshwater; iron mat
Sphaerotilus montanus HS	99.4	1.40	68.2	4.94	4514	915	Isolate	freshwater; iron oxide deposit
Sphaerotilus montanus DSM21226_2	99.8	1.40	68.2	5.07	4471	882	Isolate	freshwater; iron oxide deposit
Sphaerotilus uruguayifluvii C29	100.0	0.47	70.9	4.89	4372	894	Isolate	freshwater; biofilm
Sphaerotilus natans ATCC13338	100.0	0.00	69.9	4.63	4174	902	Isolate	freshwater; unknown
Sphaerotilus natans natans DSM6575	98.5	0.62	69.9	4.59	4200	915	Isolate	freshwater; unknown
Sphaerotilus sulfidivorans MAG 340	100.0	0.00	70.1	4.58	4144	905	MAG	wastewater; bioreactor
Sphaerotilus sulfidivorans D501 1	99.8	0.00	69.9	4.39	4025	916	Isolate	freshwater; sulfide spring
Sphaerotilus sulfidivorans D501 2	99.8	0.00	69.9	4.37	3912	895	Isolate	freshwater; sulfide spring
Sphaerotilus sulfidivorans D507	99.8	0.00	69.8	4.43	3938	889	Isolate	freshwater; sulfide spring
Sphaerotilus natans DSM22545	99.8	0.00	69.9	4.37	3912	895	Isolate	freshwater; sulfide spring

FIG 1 Phylogeny of the *Sphaerotilus-Leptothrix* group. Maximum likelihood tree of concatenated ribosomal proteins from genomes in SLG, using *Rhodoferax ferrireducens* T118 as an outgroup (not shown). Bootstrap support between 50% and 100% shown. Generated using RAxML with 1,000 bootstraps from alignment of 39 ribosomal proteins. *, newly sequenced genomes.

species, *L. mechoopdaensis*. Group 1 genomes include *L. ochracea* (21), and four newly sequenced MAGs from an Arctic tundra iron seep that are most closely related to *L. ochracea*, but cluster separately in the tree and do not meet the 95% AAI or ANI cutoff to be considered the same species as *L. ochracea* (AAI = 90.0%–90.4%; ANI = 90.0%–90.1%). Therefore, we designate these MAGs a novel *Leptothrix* species, *L. toolikensis*. This species represents the first sequenced species that is closely related to, yet separate from, *L. ochracea*.

Using AAI metrics, it is not completely clear if Groups 1, 2, and 3 represent one or two genera. Genus-level AAI thresholds are not well agreed upon but generally range from 65% to 70% (27, 49). If the most inclusive genus-level AAI threshold proposed for the family Sphaerotilaceae is used (67.12% [27]), all high-quality genomes of SLG analyzed could belong to a single genus (Fig. 2). Using the strictest cutoff of 70% does not delineate clear boundaries between Groups 2 and 3; all representatives from Groups 2 and 3 could be considered the same genus. On the other hand, the Group 1 genomes

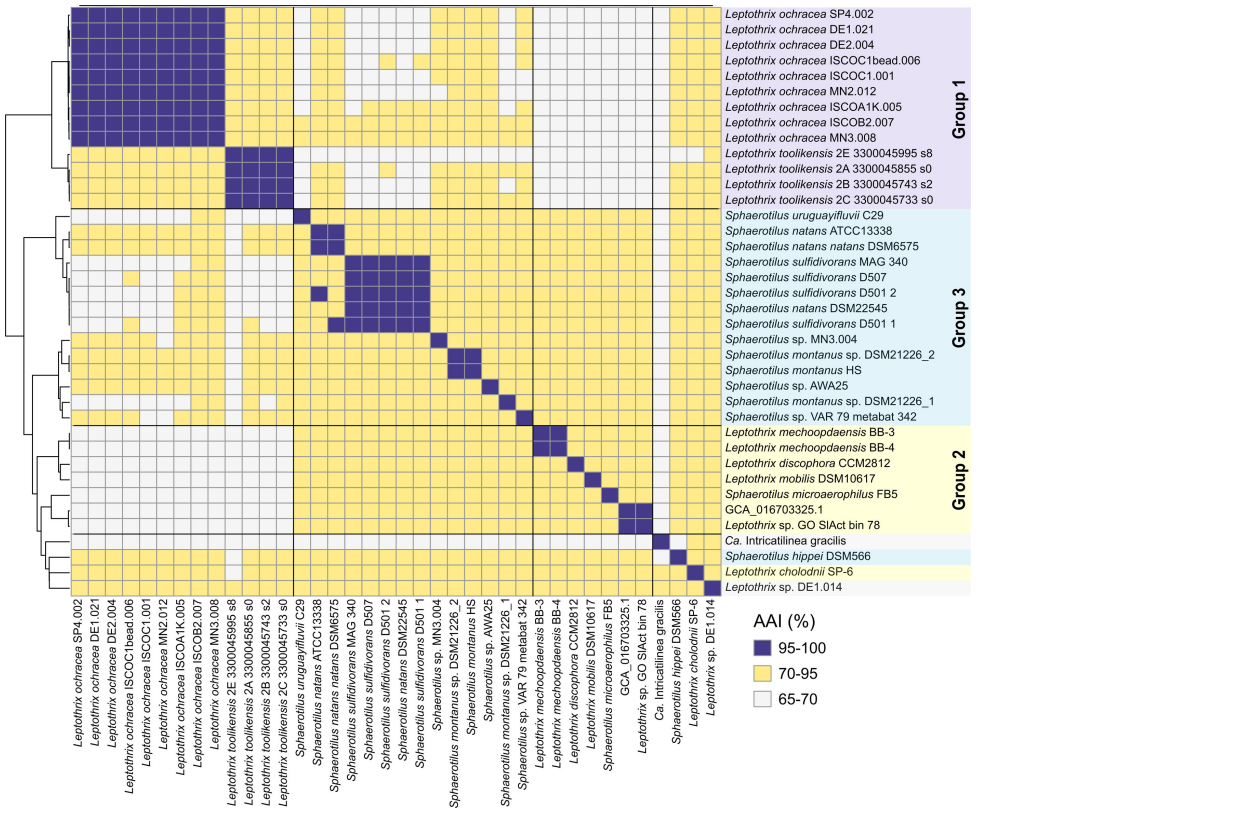


FIG 2 Amino acid identity (%) between SLG genomes. AAI values colored by species-level (95%, blue) and genus-level (70%, yellow; 65%, gray) cutoffs.

meet this threshold with a minority of genomes from Groups 2 and 3. Thus, it is not clear based on AAI alone that *L. ochracea* belongs to the same genus as other SLG.

The percentage of conserved proteins (POCP) can be used as an additional metric for genus delineation at a threshold of 50% (52, 53). Similar to AAI, the genera *Leptothrix* and *Sphaerotilus* cannot be divided cleanly using this threshold (Fig. S2). However, Group 1 genomes clearly show high POCP using this metric, and cluster separately from Groups 2 and 3, adding to evidence of the distinctiveness of Group 1.

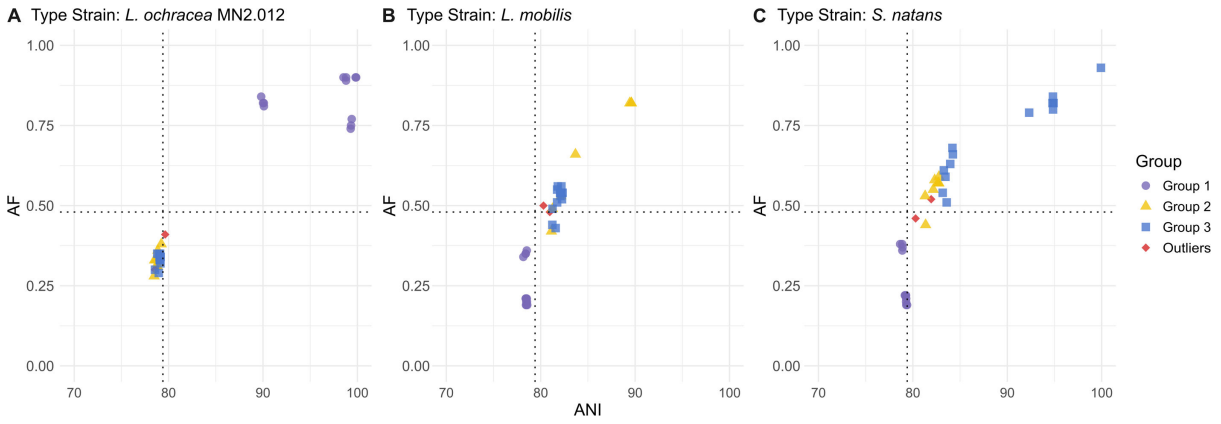


FIG 3 Average nucleotide identity (ANI) versus alignment fraction (AF) for all genomes in this study, compared to each of the three type strains. (A) *L. ochracea* MN2.012, (B) *L. mobilis* DSM10617, (C) *S. natans* CCM2812. Each panel shows pairwise comparisons to the type strain. Dashed lines indicate genus-level thresholds of 79.40% ANI and 0.480 AF, averages of the thresholds for Burkholderiaceae and Comamonadaceae defined by Barco et al. (54). Calculated using ANIcalculator 2014-127, version 1.0 (https://ani.jgi.doe.gov/html/download.php).

TABLE 1 Genome metrics across SLG subgroups

Group	Description	Avg size (Mb)	Size range (Mb)	Avg GC%	GC% range	Avg no. of genes	No. of genes (range)
1	*ochracea*-like *Leptothrix*	2.83	2.59–3.04	60.9	60.6–61.3	2,630	2,439–2,753
2	*mobilis*-like *Leptothrix*	5.08	4.55–6.06	69.2	68.4–70.1	4,504	4,053–5,176
3	*Sphaerotilus*	4.55	3.94–5.07	69.2	67.6–70.9	4,141	3,773–4519

Genomic similarity metrics can be combined in a method proposed by Barco et al. (54) to delineate genus-level boundaries based on ANI and alignment fraction (AF) to genus type strains. We applied this method using three type strains: the highest quality *L. ochracea* genome (MN2.012), *S. natans* ATCC13338, and *L. mobilis* DSM10617; we repeated this method using multiple ANI and AF calculation methods (Fig. 3; Fig. S3; Table S2). Since the thresholds have not been defined for the family Sphaerotilaceae, we used the averaged Burkholderiaceae and Comamonadaceae ratio (AF:ANI; 0.480:79.40; [54]). Based on this analysis, Group 1 is distinct from other SLG, but Groups 2 and 3 do not cluster separately from each other, consistent with the AAI and POCP analyses. Thus, various genomic metrics do not distinguish Group 2 *Leptothrix* and Group 3 *Sphaerotilus* as separate genera; however, Group 1 is clearly distinguishable.

Much like these sequence-based distinctions, genome size and GC content demonstrate that Group 1 is distinct from Groups 2 and 3 (Table 1; Table S1). While Group 1 genomes are 2.83 Mb on average, Groups 2 and 3 are 4.55 and 5.08 Mb on average, and there is no overlap between the genome size ranges of Group 1 and Groups 2 and 3. The number of genes per genome corresponds to these differences; Group 1 genomes are ~60% the size of the genomes in Groups 2 and 3 and contain ~60% the number of genes. Accordingly, the gene density in all three groups is roughly the same, with around 900 genes per Mb. In addition, the average GC content of Group 1 is 8.3% lower than that of Groups 2 and 3. These metrics support that Group 1 (*L. ochracea* and *L. toolikensis*) is distinct among SLG.

SLG pangenomic comparison

Pangenomic comparisons can help reveal the extent of commonalities and differences among individual genomes and groups. Across 38 genomes, 20,884 gene clusters were identified, including 8,948 singleton gene clusters (43% of the clusters), that is, genes that occur in only one genome. Among Groups 1, 2, and 3 (excluding the outlier genomes), 18,214 gene clusters were identified, including 6,320 singletons (35% of the clusters) (Table S1), and 865 (4.7% of the clusters) genes that were shared by 100% of the genomes (core gene set). The genomes in this study are 90%–100% complete, so SLG likely contains an even greater number of shared gene clusters. To account for genes missing due to completeness, we designated near-core gene sets based on the presence of a gene cluster in at least 85% of each group of genomes (32 of 38; Fig. 4). The near-core gene set of SLG, excluding the outliers, contained 1,219 genes (6.4% of the clusters). Considering that SLG have 2,439–5,257 genes per genome, 23%–50% of each genome is composed of near-core genes. Furthermore, Groups 2 and 3 have more singleton gene clusters than Group 1 (Table S1), reflecting broader gene content differences among Groups 2 and 3, while Group 1 showed limited variation between strains. Group 1 had the largest number of gene clusters specific to its near-core gene set, despite having the smallest genomes, and shared few gene clusters exclusively with Group 2 or Group 3 (Fig. 4).

Genomic potential

We conducted a comparative genomic analysis to explore differences in metabolic potential across the three groups. We identified key genes for energy generation (metal oxidation, organic carbon utilization, and oxidation of other inorganic electron donors) and biomass formation (organic carbon assimilation and CO_2 fixation). In addition, we identified key genes for sheath formation. Genomic potential for oxygen reduction,

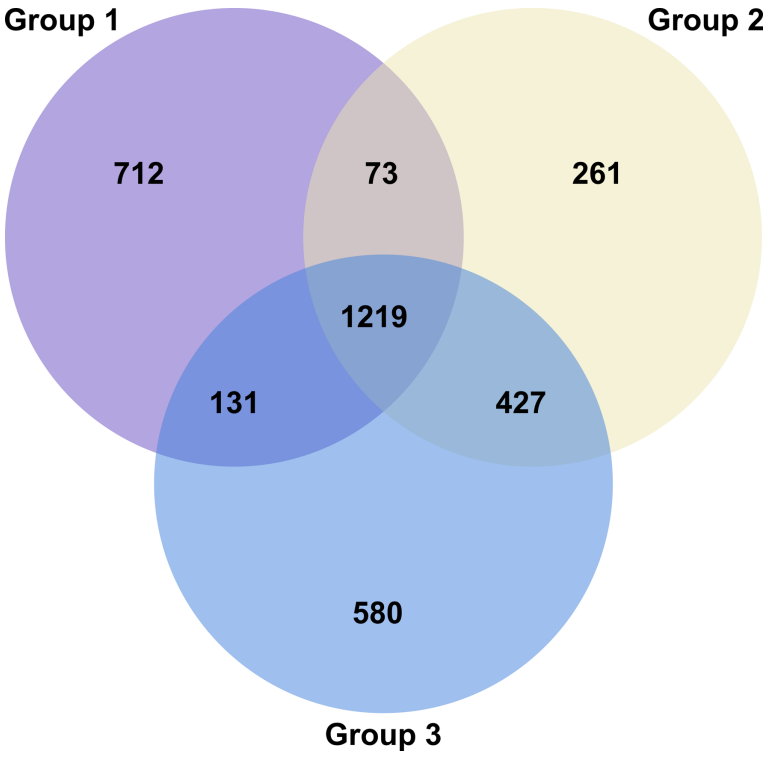


FIG 4 Venn diagram showing the distribution of near-core genes across SLG groups. Gene clusters are considered near-core if they are present in >85% of genomes in at least one SLG group.

vitamin transport, and polymer storage is described in the Supplemental text. A curated list of gene cluster frequencies and their functional annotations is provided in Table S3.

Energy generation by oxidation of metals and other inorganic electron donors

Iron oxidases

Genes for the functionally validated iron oxidase Cyc2 (55–58) were encoded in the majority of Group 1 genomes and in a limited number of genomes from Groups 2 and 3. Cluster 1 *cyc2* genes were predominantly associated with Group 1, and a single outlier (clusters defined by McAllister et al. [59]). Cluster 2 *cyc2* genes were identified in only three genomes from Groups 2 and 3, and one outlier genome (Fig. 5; Table S3).

Genes for a decaheme cytochrome-porin complex iron oxidase (*mtoAB*) (60, 61) are encoded in all Group 1 and outlier genomes (Fig. 5). The oxidative MtoA is a homolog to the iron-reducing MtrA; hence, the classification of these proteins was confirmed using a phylogeny of MtoA, MtrA, and PioA (Fig. S4). MtoA sequences from the SLG cluster with the functionally validated MtoA from *Sideroxydans lithotrophicus* ES-1; thus, their function is likely iron oxidation rather than iron reduction. These genes are absent from Group 2 and Group 3. "*Ca.* I. gracilis" is the only genome that contains the extracellular decaheme *mtrC,* and its *mtoA* gene is found in a singleton gene cluster, phylogenetically distinct from the other SLG *mtoA* genes (at the crown of Clade 2 in Fig. S4). A gene for the iron oxidase FoxE is encoded only in *L. mobilis* (Fig. 5).

Since the known iron oxidases are cytochromes, we searched the genomes for additional multiheme cytochromes, which could facilitate metal oxidation and extracellular electron uptake. All 38 genomes encode proteins with two or more heme-binding motifs ($CX_{2\text{-}4}CH$; Fig. 6). Notably, Group 3 members encode very few proteins with more than two heme-binding motifs. In contrast, all members of Groups 1 and 2 encode at least four proteins with 3–9 heme-binding motifs, and the majority also encode proteins with 10 or more motifs. In summary, genomes from Groups 1 and 2 encode

		Copies (0–8)	validated iron oxidases						multiheme cytochromes			multicopper oxidases			
			cyc1	*cyc2* cluster 1	*cyc2* cluster 2	*foxEY*	*mtoAB*	*mtrC*	PCC3	*uetJ*	unknown MHC cluster	*mcoA*	*mnxG*	*mofA*	*moxA*
Outliers	Ca. *Intricatilinea gracilis*		1	1			1	1	1	1	1	4	2	1	1
	Leptothrix sp. DE1.014		1		2		2		1					2	
Group 1	*Leptothrix toolikensis* 2E_33000045995_s8		1	1			1		1[b]					3	
	Leptothrix toolikensis 2A_33000045855_s0		1	1			1		1[b]					4	
	Leptothrix toolikensis 2B_33000045743_s2		1	1			1		1					3	
	Leptothrix toolikensis 2C_33000045733_s0		1	1			1		1[b]					3	
	Leptothrix ochracea MN2.012		1	1			1		1					7	
	Leptothrix ochracea ISCOC1.001			1			1		1[b]					5	
	Leptothrix ochracea ISCOC1bead.006		1	1			1		1					6	
	Leptothrix ochracea ISCOA1K.005		1	1			1[a]		1[b]					8	
	Leptothrix ochracea MN3.008						1		1					7	
	Leptothrix ochracea SP4.002		1	1			1[a]		1					4	
	Leptothrix ochracea DE1.021		1	1			1[a]		1[b]					4	
	Leptothrix ochracea ISCOB2.007						0[a]		1					5	
	Leptothrix ochracea DE2.004		1	1			0[a]		1					4	
Group 2	*Sphaerotilus microaerophilus* FB5										1				1
	GCA_016703325.1														
	Leptothrix sp. GO SIAct bin 78														
	Leptothrix cholodnii SP-6								1	1	1	2	3	1	1
	Leptothrix mobilis DSM10617					1			1	1	1	1	2	1	2
	Leptothrix discophora CCM2812								1	1	1	2	2	1	1
	Leptothrix mechoopdaensis BB-3				1				1	1		2	1		1
	Leptothrix mechoopdaensis BB-4				1				1	1		2	1		1
Group 3	*Sphaerotilus hippei* DSM566														2
	Sphaerotilus sp. AWA25														2
	Sphaerotilus montanus sp. DSM21226_1													1	2
	Sphaerotilus sp. VAR 79 metabat 342														1
	Sphaerotilus sp. MN3.004				1										1
	Sphaerotilus montanus HS														2
	Sphaerotilus montanus sp. DSM21226_2														2
	Sphaerotilus uruguayifluvii C29														2
	Sphaerotilus natans ATCC13338														1
	Sphaerotilus natans natans DSM6575														1
	Sphaerotilus sulfidivorans MAG 340														1
	Sphaerotilus sulfidivorans D501 1														3
	Sphaerotilus sulfidivorans D501 2														3
	Sphaerotilus sulfidivorans D507														3
	Sphaerotilus natans DSM22545														3

FIG 5 Metal oxidation genes encoded by SLG genomes. Shaded boxes indicate gene presence. Numbers indicate gene copy number. White boxes indicate gene absence. a = encodes additional truncated copies of *mtoA* and *mtoB*. b = encodes an additional extracellular cytochrome adjacent to a porin in the PCC3 gene cluster.

larger numbers of multiheme cytochromes, which also contain greater numbers of heme binding motifs per protein.

Porin cytochrome complexes (PCCs) are strong candidates for oxidizing various solid and dissolved Fe(II) substrates, transporting the electrons across the outer membrane. Iron-oxidizing bacteria frequently have functionally uncharacterized PCCs (62) that lack a demonstrated role in metal oxidation. We characterized each of the cytochromes with 10 or more heme motifs in Groups 1 and 2 with its name (if known), localization, and whether a porin was encoded within two genes on either side of the multiheme cytochrome gene (Table S4). Group 1 and one outlier encode MtoAB and a predicted porin-cytochrome complex known as PCC3 (Fig. 5) (62). Group 2 encodes PCC3 (cytochrome subunits containing 10–57 hemes, most commonly 21–29), the undecaheme UetJ (63), and its associated porin, and an unnamed decaheme cytochrome

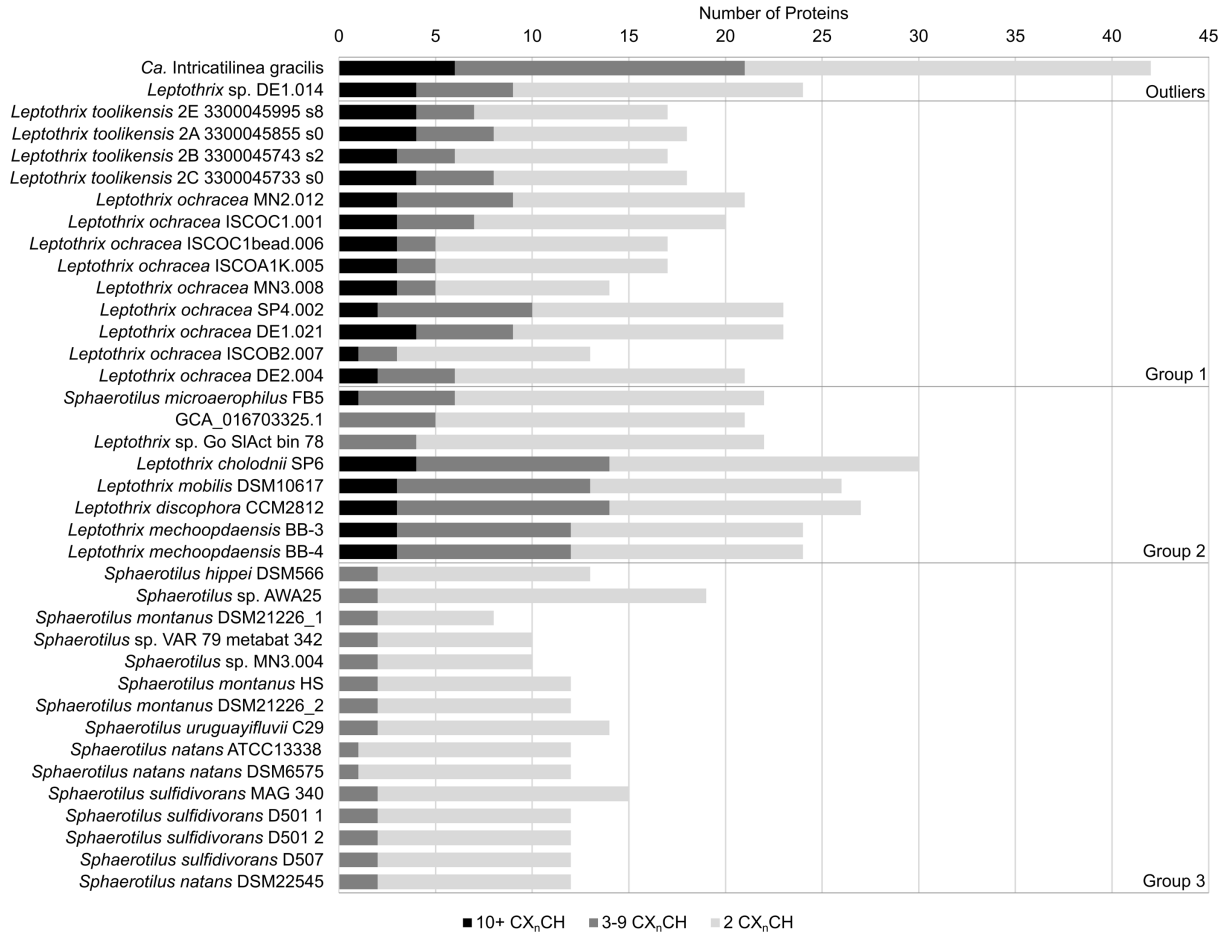


FIG 6 The number of genes encoding multiheme cytochromes. Proteins are categorized by the number of CXXCH, CXXXCH, and CXXXXCH heme-binding motifs.

for which no nearby porin was identified. The second outlier genome, "*Ca*. l. gracilis" encodes one of each type: MtoAB, PCC3, UetJ, and the unnamed decaheme cytochrome. Overall, the analysis of multiheme cytochromes suggests that Groups 1 and 2 encode additional potential metal oxidases.

Multicopper oxidases

Multicopper oxidases (MCOs) serve diverse metabolic functions; some can catalyze either iron or manganese oxidation in bacteria (64–66). Nearly all SLG genomes encode at least one of the homologs to Mn-oxidizing MCOs encoded by *L. cholodnii* SP-6 (*mnxG*, *mcoA*, *moxA*, and *mofA* (67–70), but their distribution varies among groups (Fig. 5; Table S3). The *mcoA* and *mnxG* genes are found exclusively in Group 2 and the outlier "*Ca*. l. gracilis." In contrast, *mofA* is characteristic of Group 1, in which multiple copies are common; *mofA* occurs rarely in Group 3 and is present in some Group 2 genomes as well as all outlier genomes. The *moxA* gene is present in all Group 3 genomes and one outlier. In Group 3, it is typically the sole MCO (and sole metal oxidase) encoded.

Sulfur

The *sox* genes are involved in the oxidation of thiosulfate. The *soxABXYZ* genes encode a multi-enzyme complex involved in the initial steps of thiosulfate oxidation, while complete oxidation to sulfate typically requires *soxCD*. The *soxABX* genes are found in all members of Groups 1 and 2, and most members of Group 3, while *soxYZ* is ubiquitous across all genomes (Fig. 7; Table S3). The *soxCD* genes are found in most members of Groups 1 and 2. More than half of the Group 3 genomes encode *soxABXYZ* and *soxCD*,

and the remaining six only encode *soxY* and *soxZ*. Overall, the results suggest that oxidation of thiosulfate is a viable energy-generating mechanism for the majority of SLG, although it is not a defining trait for Group 3 *Sphaerotilus*.

Sulfide dehydrogenase gene clusters (*fccAB*) are present in all 38 genomes in this study (Fig. 7). This gene is involved in dissimilatory oxidation of sulfide to elemental sulfur and represents a possible detoxification mechanism. A sulfide:quinone oxidoreductase gene (*sqrF*) was identified in all members of Group 1, and about half of the genomes from Groups 2 and 3.

The *dsrAB* genes were identified in roughly half of the genomes from Groups 1 and 2 (Fig. 7). The dissimilatory sulfite reductase DsrAB was originally shown to reduce sulfite to sulfide, although it has been shown to function in reverse to oxidize sulfide to sulfite (71, 72). In a maximum-likelihood protein tree, DsrA sequences from SLG cluster with the oxidative rDsrA sequences from S-oxidizing species (Fig. S5), indicating that the DsrAB in SLG is likely a reverse dissimilatory sulfite reductase. This suggests that some members of this group have a third mechanism for sulfur oxidation, but it is not a defining feature of SLG.

Carbon utilization

Carbon fixation

All genomes in this study, except four Group 3 *Sphaerotilus* members, have the genomic potential to perform carbon fixation. The key enzyme in the Calvin-Benson-Bassham (CBB) cycle for carbon fixation is RuBisCO, encoded by the genes *rbcL* and/or *rbcS* (Form I) or *cbbM* (Form II). RuBisCO Form II has a lower affinity for CO_2 and accepts O_2 more readily as an alternative substrate and competitive inhibitor compared to RuBisCO Form I; consequently, Form II is most effective under conditions where O_2 is low and CO_2 is relatively high, such as microaerobic environments (73). All genomes in Group 1, and four genomes in Group 2, encode Form II RuBisCO (cbbM), suggesting that these members preferentially grow in microaerobic environments. In contrast, genes for the small and large catalytic subunits of Form I RuBisCO (*rbcS/rbcL*) are found in the other four Group 2 genomes and nine Group 3 genomes (Fig. 7; Table S3). Form I is less susceptible to competitive inhibition by oxygen, and its presence in these organisms is consistent with growth under more oxic conditions. (For further discussion of oxygen preferences, see Supplemental text).

Organic carbon substrates

There are 602 gene clusters annotated as carbohydrate-active enzymes (CAZy), across six categories: auxiliary activities (AA), carbohydrate-binding modules (CBM), carbohydrate esterases (CE), glycoside hydrolases (GH), glycosyltransferases (GT), and polysaccharide lyases (PL) (Fig. 8). Group 1 genomes encode markedly fewer CAZy genes compared to Groups 2 and 3, with pronounced differences in the average numbers of glycoside hydrolase (GH) and glycosyltransferase (GT) genes. Specifically, Group 1 genomes contain an average of 15 GH and 34 GT genes, whereas Groups 2 and 3 encode averages of 45 and 43 GH genes, and 45 and 56 GT genes, respectively. These findings indicate a comparatively reduced genomic capacity for carbohydrate processing in *ochracea*-like *Leptothrix* relative to other members of the SLG.

Some genomes show potential for lactate utilization, while most genomes display acetate uptake capability. Lactate permease genes for lactate import (*lctP*) are encoded in all Group 1 genomes and in a minority of genomes from Groups 2 and 3 (Fig. 7; Table S3). Genes for metabolizing endogenous lactate (L-lactate dehydrogenase; either *lldD* or *lldEFG*) are found in all genomes, but their distribution varies: *lldD* is ubiquitous among Groups 2 and 3, and in the majority of *L. toolikensis* genomes, while *lldEFG* is ubiquitous among *L. ochracea* genomes and appears in less than half of genomes in Groups 2 and 3. Genes for a cation/acetate symporter (*actP*) are found in the full near-core gene set of the group (Fig. 7). All 38 genomes have at least one copy of this symporter, and the

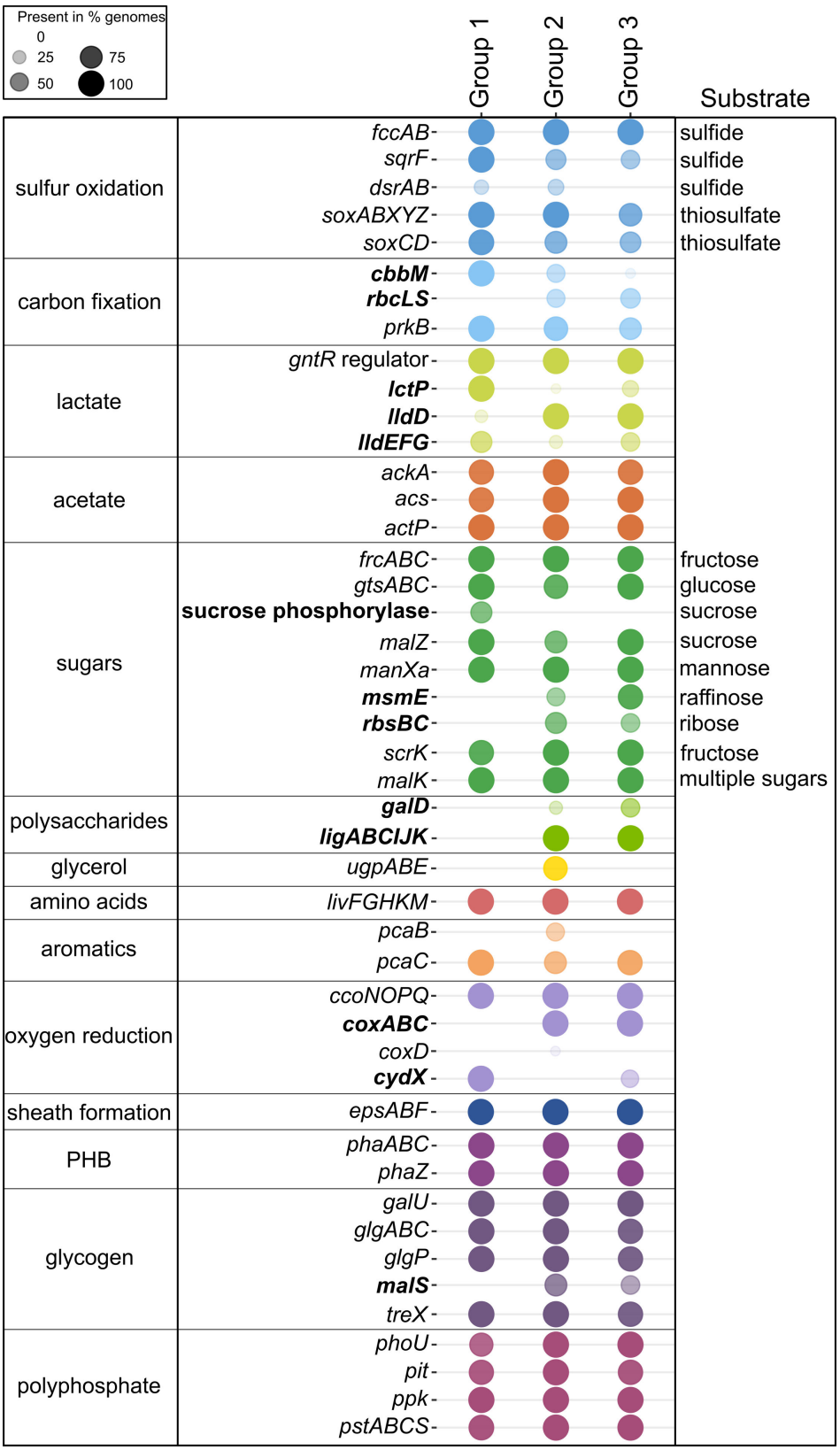
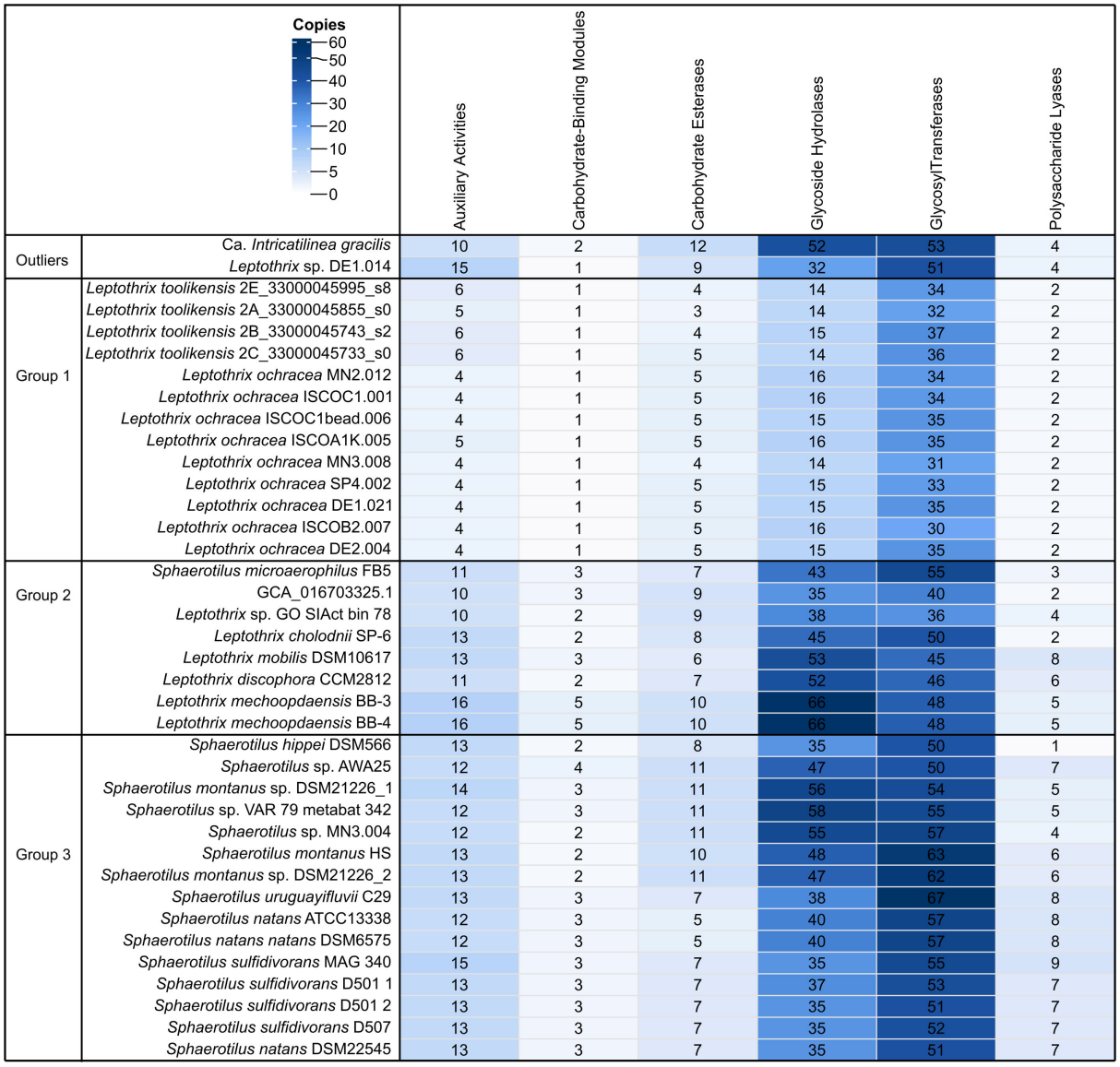


FIG 7 Genomic potential for metabolisms of interest. Based on the percentage of genomes in each SLG subgroup that encode each gene. Bolded genes are those with significant differences in presence between Group 1 and Groups 2/3.



		Auxiliary Activities	Carbohydrate-Binding Modules	Carbohydrate Esterases	Glycoside Hydrolases	GlycosylTransferases	Polysaccharide Lyases
Outliers	Ca. *Intricatilinea gracilis*	10	2	12	52	53	4
	Leptothrix sp. DE1.014	15	1	9	32	51	4
Group 1	*Leptothrix toolikensis* 2E_33000045995_s8	6	1	4	14	34	2
	Leptothrix toolikensis 2A_33000045855_s0	5	1	3	14	32	2
	Leptothrix toolikensis 2B_33000045743_s2	6	1	4	15	37	2
	Leptothrix toolikensis 2C_33000045733_s0	6	1	5	14	36	2
	Leptothrix ochracea MN2.012	4	1	5	16	34	2
	Leptothrix ochracea ISCOC1.001	4	1	5	16	34	2
	Leptothrix ochracea ISCOC1bead.006	4	1	5	15	35	2
	Leptothrix ochracea ISCOA1K.005	5	1	5	16	35	2
	Leptothrix ochracea MN3.008	4	1	4	14	31	2
	Leptothrix ochracea SP4.002	4	1	5	15	33	2
	Leptothrix ochracea DE1.021	4	1	5	15	35	2
	Leptothrix ochracea ISCOB2.007	4	1	5	16	30	2
	Leptothrix ochracea DE2.004	4	1	5	15	35	2
Group 2	*Sphaerotilus microaerophilus* FB5	11	3	7	43	55	3
	GCA_016703325.1	10	3	9	35	40	2
	Leptothrix sp. GO SIAct bin 78	10	2	9	38	36	4
	Leptothrix cholodnii SP-6	13	2	8	45	50	2
	Leptothrix mobilis DSM10617	13	3	6	53	45	8
	Leptothrix discophora CCM2812	11	2	7	52	46	6
	Leptothrix mechoopdaensis BB-3	16	5	10	66	48	5
	Leptothrix mechoopdaensis BB-4	16	5	10	66	48	5
Group 3	*Sphaerotilus hippei* DSM566	13	2	8	35	50	1
	Sphaerotilus sp. AWA25	12	4	11	47	50	7
	Sphaerotilus montanus sp. DSM21226_1	14	3	11	56	54	5
	Sphaerotilus sp. VAR 79 metabat 342	12	3	11	58	55	5
	Sphaerotilus sp. MN3.004	12	2	11	55	57	4
	Sphaerotilus montanus HS	13	2	10	48	63	6
	Sphaerotilus montanus sp. DSM21226_2	13	2	11	47	62	6
	Sphaerotilus uruguayifluvii C29	13	3	7	38	67	8
	Sphaerotilus natans ATCC13338	12	3	5	40	57	8
	Sphaerotilus natans natans DSM6575	12	3	5	40	57	8
	Sphaerotilus sulfidivorans MAG 340	15	3	7	35	55	9
	Sphaerotilus sulfidivorans D501 1	13	3	7	37	53	7
	Sphaerotilus sulfidivorans D501 2	13	3	7	35	51	7
	Sphaerotilus sulfidivorans D507	13	3	7	35	52	7
	Sphaerotilus natans DSM22545	13	3	7	35	51	7

FIG 8 Potential for carbohydrate metabolism of SLG. Genes are summarized by category according to CAZyme annotations. Shaded boxes indicate gene presence. Numbers indicate gene copy number.

majority of genomes have multiple copies. Furthermore, all but two genomes have an acetate kinase (*ackA*) gene.

Genes for the import of glucose (*gtsABC*), fructose (*frcABC*), and mannose (*manXa*) are part of the near-core gene set of SLG. The glucose transport system *gtsABC* is present in 37 genomes, while *frcABC* and *manXa* are both ubiquitous (Fig. 7; Table S3). Fructokinase genes (*scrK*) are present in all but one genome. Furthermore, a gene for the ATP-binding subunit used by many organic carbon transporters (*malK*) is ubiquitous across SLG. These patterns indicate that nearly all members of SLG have the genomic potential to import and metabolize glucose, fructose, and mannose.

Genes for sucrose phosphorylase are found in all of the *L. ochracea* genomes and in one outlier genome ("*Ca*. I. gracilis"), while this gene is absent from *L. toolikensis* and Groups 2 and 3 (Fig. 7). This confers the potential to degrade sucrose into fructose and glucose-1-phosphate. Therefore, the potential for sucrose degradation appears to be primarily limited to *L. ochracea*.

Genomes from Groups 2 and 3 encode genes for ribose (*rbsBC*) and raffinose (*msmE*) transporters, which are absent from Group 1 (Fig. 7). The raffinose transporter *msmE* is part of the Group 3 near-core and occurs in some Group 2 genomes, while *rbsBC* is found in various members of Groups 2 and 3 but not in their near-core gene sets. These results show variability in sugar transporter presence within Groups 2 and 3 compared to Group 1.

Genes for a glycerol-3-phosphate transport system (*ugpABE*) are encoded by the majority of Group 2 genomes; however, they are absent from Groups 1 and 3 (Fig. 7; Table S3). Therefore, the potential for glycerol utilization is common among Group 2 genomes and unique to this group.

Genes encoding a branched-chain amino acid transport system (*livGFHMK*) are distributed across all SLG genomes (Fig. 7; Table S3). Each of the five *liv* genes is present in all 38 genomes, typically in multiple copies. These findings suggest that all members possess the genomic potential for branched-chain amino acid transport.

Sheath formation

The genes underlying sheath formation in SLG are not well understood and are difficult to separate from other metabolic functions, including EPS formation. However, polysaccharide export genes *wza*, *wzb*, and *wzc* could function in excreting polymers for sheath formation and are encoded by all genomes analyzed here (Fig. 7; Table S3). Additional genes related to organic carbon export and processing (e.g., glycosyl transferases) likely contribute to sheath development (74–77); however, additional work is needed to validate sheath formation genes in SLG.

DISCUSSION

Leptothrix and *Sphaerotilus* have been known for over two centuries, beginning with *L. ochracea*, due to its conspicuous growth as rusty orange microbial mats and its clearly distinguishable sheaths tinted by iron oxides. *L. ochracea* lacks an isolate; hence, it no longer meets the standards for type material according to the International Code of Nomenclature of Prokaryotes (78; see also SeqCode [79]). However, extensive characterization of *L. ochracea* using xenic cultures, single-cell methods, and meta-omics has allowed for many of the same insights as would be available for an isolate, including its identity and physiology (20, 21, 33). In any case, according to the List of Prokaryotic names with Standing in Nomenclature (LPSN) (https://lpsn.dsmz.de/species/leptothrix-ochracea [80]), *Leptothrix* and *Leptothrix ochracea* are still the "correct names," validly published by Kützing in 1843 (4) and included by Sneath et al. (81) in the Approved List of Bacterial Names.

The close phenotypic and genetic relationship between *Leptothrix* and *Sphaerotilus* has led to periodic renaming of species, causing confusion (6–8, 12, 18, 40, 44, 82). Recently, there have been renewed suggestions to rename all of the *Leptothrix* species to *Sphaerotilus* (26, 27); however, such a change requires careful consideration of all available evidence. Here, we perform a comprehensive assessment of the SLG by integrating analyses of phylogenomics, genome similarity, key functional genes, and phenotypes. Our analysis differs from previous ones in our inclusion of nine high-quality *L. ochracea* genomes plus four new genomes of the closely related *L. toolikensis*. With these additions, our results justify keeping the two genera, *Leptothrix* and *Sphaerotilus*, and add valuable insights into the distinguishing genetic features.

Classification

Existing genome classification tools have previously been inconsistent in classifying SLG genomes, partly due to few reference genomes in databases and also the lack of a systematic phylogenetic analysis using the full SLG diversity. Our work provides a comprehensive data set with multiple representatives from each of the three subgroups and helps establish a framework for more rigorous genomic classification. This

framework will allow recognition of more genomes as SLG from existing or new data sets, which will add further resolution for improving classifications. Any current and future classifications of new genomes should be supported by careful application of the polyphasic approach outlined below, using concatenated conserved (e.g., ribosomal) protein trees, genome comparison metrics, environmental ecotypes, and consideration of genotypic and phenotypic lines of evidence.

When using genomes to help delineate genera, the recommendation is to use multiple genome comparison metrics alongside phylogenomics and analyses of genes that represent shared physiology and ecological niche, since thresholds for quantitative metrics like ANI or AAI are not definitive (52). These latter metrics have been used to support the merging of *Leptothrix* and *Sphaerotilus* into a single genus; Smolyakov et al. (26) applied a 65% AAI cutoff, while Liu et al. (27) used an AAI threshold of 67.12%–71.55%, specific to the Sphaerotilaceae. When developing the combined ANI/AF genome relatedness index, Barco et al. (54) found that family-specific thresholds are required due to differences in evolutionary rates between taxa. Boundary-defining thresholds often need to be adjusted to ensure coherence with phylogenomics (52, 83, 84). Therefore, as in previous SLG classification, using ANI or AAI values to delineate genera is still only an approximate estimation that needs genotypic, phenotypic, and environmental context to provide a complete taxonomic picture.

The phylogenetic analyses presented here support the existence of three subgroups within *Sphaerotilus-Leptothrix*: the *ochracea*-type *Leptothrix* (Group 1), the *mobilis*-type *Leptothrix* (Group 2), and *Sphaerotilus* (Group 3). Here, AAI results suggest that while Groups 2 and 3 could be considered the same genus, Group 1 belongs to a separate genus. Using the ANI/AF thresholds for the families that SLG have belonged to previously (Burkholderiaceae, Comamonadaceae), the genus boundary lies between Group 1 and Groups 2/3. This is supported by the phylogenetic analysis, in which the Group 2 *Leptothrix* and Group 3 *Sphaerotilus* cluster more closely with one another than either group does with Group 1, suggesting that the genus *Leptothrix* may be paraphyletic. Genome size and GC content provide further evidence of Group 1's distinctiveness. The genomes are ~60% the size of Group 2/3 genomes, indicating that Group 1 genomes have been streamlined. Group 1 has an average %GC that is consistently 8% lower than Groups 2/3, giving further evidence of a diverging evolutionary trajectory. Given this combination of phylogenomics and multiple genomic comparison metrics, Group 1 clearly stands apart.

The current standards for defining genera require a polyphasic approach that considers phylogeny, phenotypes, and environmental niches, not just genomic metrics (29, 41, 54, 85–93). Looking at the distribution of key functional genes, genotype patterns mirror phenotypes, supporting the above findings for Group 1 but causing some difficulty in characterizing Group 2. Group 1 is unique in that all members encode iron oxidases (Cyc2, MtoAB) that can yield energy, and multiple copies of *mofA*, consistent with findings that *L. ochracea* requires Fe(II) for growth (33). Group 3 almost all lack genes for iron oxidation energy metabolism, and generally only encode one gene for metal oxidation, the MCO *moxA*, suggesting that metal oxidation is not central to Group 3 energy metabolism. In contrast to Groups 1 and 3, metal oxidation gene patterns in Group 2 are less consistent, making it difficult to determine defining features. Perhaps the most notable feature of Group 2 is that many genomes have multiple metal oxidation genes, although most lack the well-characterized genes for iron oxidation, making it unclear whether Group 2 members generally oxidize iron for energy. Group 1 genomes host relatively few organic carbon utilization genes, which fits with isotopic and transcriptomic evidence of mixotrophy in *L. ochracea* (21, 33). This contrasts with Groups 2 and 3, in which all isolates grow organoheterotrophically; this is reflected in the abundance and range of organic utilization genes (e.g., CAZy) in Groups 2 and 3 genomes. Among Groups 2 and 3, the distribution of genes for metal oxidation and abundance of genes for organic carbon utilization fit with the fact that neither requires

substantial quantities of iron for energy, instead relying on organotrophy for energy metabolism.

Group 2 members lack homogeneity in genotype, phenotype, and ecotype, which makes it difficult to determine if Group 2 (or perhaps a subset) should be reassigned to *Sphaerotilus*. There is some evidence that Group 2 may not be well-defined and instead includes two subgroups that display phylogenetic separation consistent with their isolation sources. The majority of the group is formed by five freshwater *Leptothrix* isolates, all of which are metal oxidizers with multiple potential metal oxidase genes. In contrast, the three genomes derived from sludge (including one isolate) largely lack metal oxidases and instead appear to be organotrophic (45). Consequently, the freshwater Group 2 isolates differ markedly from *Sphaerotilus*, whereas the three Group 2 members derived from sludge do not. These three are in the minority, and they have a markedly lower alignment fraction to *L. mobilis* than the metal-oxidizing members of Group 2 (Fig. 3; Table S2) and form a phylogenetically distinct clade, suggesting that they could be considered as distant from *L. mobilis* as any Group 3 genome. The bootstrap of the node defining Group 2 is 63%, and it is possible that the addition of more genomes will cause this organotrophic subgroup to associate with Group 3 instead of Group 2. In order to make a definitive decision, additional isolates related to both subgroups are needed to more confidently define Group 2.

We predict that further genomic sequencing and phenotypic characterization of isolates will result in well-supported (i.e., high bootstrap value, >75%), coherent, phylogenomic, and phenotypic/genotypic groups. For now, Group 1 (*L. ochracea* and *L. toolikensis*) shows sufficient phylogenetic and functional uniqueness to merit a genus distinct from *Sphaerotilus*. Overall, genus delineation should consider combined genotype and phenotype, alongside phylogenomics and genomic metrics. This polyphasic framework gives us multiple coherent lines of evidence for retaining the *Leptothrix* designation for *L. ochracea* and *L. toolikensis,* while further isolates and genomes are required to conclusively classify Group 2; therefore, the current genus designations should be maintained until this is resolved.

Interpreting genes in the context of classic phenotypic distinctions

There are three canonical phenotypic distinctions, which have been used to identify SLG genera and species: iron and manganese oxidation, response to organics, and sheath morphology. Here, we consider how this SLG pangenome analysis gives insight into each of these phenotypic categories.

Iron and manganese oxidation

Historically, *Leptothrix* and *Sphaerotilus* were distinguished by their metal oxidation ability. While both are able to oxidize iron, only *Leptothrix* can oxidize manganese (10, 94). The genomes show a multitude of known and putative metal oxidation genes that reveal more about metal oxidation potential. However, the genotypes do not all cohere to the phenotypic distinctions, but given the ongoing research on metalloxidases, this suggests that SLG genomes include novel iron oxidases.

Known microbial manganese oxidases are multicopper oxidases, including the four encoded in the SLG pangenome: McoA, MnxG, MofA, and MoxA (67–70). While other MCOs are known to oxidize iron (65, 95, 96), these four are not known to play roles in iron oxidation. Almost all of the SLG genomes encode at least one of these four MCOs, but if they are all truly manganese oxidases, this conflicts with the understanding that *Sphaerotilus* do not oxidize manganese. *Sphaerotilus* do oxidize iron, but all but one Group 3 *Sphaerotilus* lack known or suspected iron oxidation genes, leaving the MCO MoxA homolog as the only metalloxidase, suggesting that it is a plausible iron oxidase, although this has yet to be tested.

The currently known iron oxidases are Cyc2, MtoAB, and FoxEY, while PCC3 and UetJ are predicted porin cytochrome clusters present in other iron-oxidizing bacteria

and could theoretically oxidize various Fe(II) substrates, including dissolved free Fe^{2+}, organic-chelated Fe^{2+}, and mineral-bound Fe^{2+}, which vary in redox potential. Group 1 and most of the Group 2 genomes contain these known and/or putative iron oxidases, consistent with the iron-oxidizing phenotype of *Leptothrix*. The presence of multiple iron oxidases is common among iron-oxidizing bacteria (97), and different iron oxidases may be expressed to gain energy from distinct iron substrates (61, 98).

Carbon utilization: autotrophy, mixotrophy, and heterotrophy

Another classical distinction in SLG is "response to organics." A stronger response in culture suggests organoheterotrophy, while a muted or non-response implies mixotrophy, oligotrophy, or autotrophy. Most SLG genomes possess genes for both carbon fixation and heterotrophy; so, in theory, SLG members could be autotrophic, mixotrophic, and heterotrophic, but it is an open question as to how much each group or species can exhibit autotrophic vs. heterotrophic character. There has been relatively little systematic exploration of this, as evidence for autotrophy/mixotrophy is sparse. The only isotope-labeled carbon uptake experiments have been performed in xenic cultures of *L. ochracea*, which show some uptake of inorganic carbon in the presence of organic carbon (33). This demonstrates mixotrophy, consistent with poor responses to organics in *L. ochracea* cultures, the lower number of heterotrophy genes, and metabolic modeling results (10, 21, 37). The only clear experimental evidence of autotrophy is for the isolate *S. natans* D-507, which was shown to grow autotrophically by sulfur oxidation (25). Most SLG genomes possess both sulfur oxidation and carbon fixation genes, and, thus, the potential for lithoautotrophic sulfur oxidation, but this is untested in other SLG, as is lithotrophic iron oxidation.

Mixotrophic iron oxidation is most plausible in SLG Group 1 but is also tenable in some of Group 2, while demonstration of autotrophic iron oxidation remains elusive. Iron oxidation yields relatively little energy and also requires reverse electron transport, making autotrophic iron oxidation a challenging lifestyle (36). The presence of various organic transporters and utilization genes, even in Group 1 genomes, suggests that SLG have adapted to use organics where available, tending toward mixotrophy and organoheterotrophy. The variability in organic utilization genes, particularly in Groups 2 and 3, where they occur in higher numbers, is consistent with the variation in substrates that can support SLG growth (10). Thus, the overall diversity in carbon metabolism likely accounts for the ability of SLG to colonize a range of aquatic environments, but it remains to be seen the range and flexibility of each SLG in terms of organic and inorganic carbon utilization.

How metal and carbon metabolisms influence sheath morphology

Before the advent of molecular taxonomy, sheath morphology played an important role in characterizing SLG. The starkest difference lies in the comparison of *L. ochracea* sheaths, which are mainly empty, versus Groups 2 and 3 sheaths that are filled with cells. These differences in sheath habits can be explained, given our understanding of SLG metal oxidation and carbon metabolism geno/phenotypes.

The typical SLG sheaths of the Groups 2 and 3 isolates are organic. Studies of *L. cholodnii* showed that terminal cells at either end of a filamentous chain excrete sheath fibrils that form a robust organic matrix held together by disulfide bonding (99); the *S. natans* sheath is also proposed to form similarly (100). The organic backbone of these sheaths is consistent with an organoheterotrophic metabolism, and metal oxidation only occurs once sheaths are formed (101, 102). Because the organotrophic SLG do not require metal oxidation for growth, less metal is oxidized per cell, and the sheaths are not completely encrusted; so, the cells can continue to inhabit sheaths without being cut off from nutrient supply. The resulting morphology is long filaments of cells encased in a manganese and/or iron oxyhydroxide-coated sheath.

In contrast, Group 1 appears to exhibit more autotrophic character in phenotype and genotype. Group 1 has multiple pathways to connect iron oxidation to electron

transport and energy generation (Cyc2, MtoAB, and PCC3) but relatively few organic utilization genes. This supports the likely niche of Group 1 *Leptothrix* as mixotrophs that use iron oxidation as a major energy-generating mechanism, which would result in copious oxides. The largely empty sheaths of *L. ochracea* could be considered analogous to the stalks of the autotrophic iron oxidizer *Gallionella*, in which one end is anchored as a holdfast, and cells package ferric iron waste in a mineralized stalk with a cell at the leading end (31). Vesenka et al. (74) showed that *L. ochracea* sheath formation is driven by filaments of 10–20 cells moving in aqueous media and simultaneously excreting nanofibrils that are immediately encrusted with (and largely composed of) ferric oxyhydroxide byproducts of their metabolism. In essence, as cells form the sheaths and move past iron oxides, they are performing a unique form of metal-mediated motility. The result is a trailing, empty tubular mineralized sheath that is the signature of *L. ochracea*, again reflecting its mixotrophic iron oxidation metabolism.

Conclusions and implications

This work provides a comprehensive view of the SLG, giving a robust framework for consistent identification and interpretation. The SLG is separated into three functional subgroups based on a polyphasic classification that incorporates phylogenomic and functional evidence. Combining Groups 2 and 3 as *Sphaerotilus* is justifiable based on phylogeny, genome similarity, and, to some extent, functional potential. However, there are clear distinctions between Group 2 and Group 3 in metal oxidation genes, which merit distinguishing these subgroups; the diversity in Group 2 relative to the number of genomes suggests that more genomes/isolates are needed to resolve its relationship to the other groups. For now, we recommend retaining the existing genus designations for Group 2 until further genomes improve the phylogenetic resolution. The clear differences between Group 1 and Groups 2/3 require the retention of the genus *Leptothrix* for members of Group 1, rather than fully combining the two genera as one (8, 26, 27, 40).

Genus nomenclature is important for understanding and communicating how microbes influence environmental outcomes. The metabolic differences between the more autotrophic/mixotrophic vs. organoheterotrophic SLG result in distinct biogeochemical and ecological roles. The organoheterotrophic *Leptothrix* and *Sphaerotilus* likely oxidize much less iron and thus have more influence on carbon rather than metal cycles and restrict these taxa to more organic-rich environments. In contrast, Group 1 oxidizes large quantities of iron in the environment, thereby acting as major drivers of iron cycling and resulting nutrient and metal sequestration and release. Their ability to grow in low carbon environments and produce iron-mineralized sheaths makes them keystone species that engineer mat structures which are colonized by diverse microbial communities. These iron mats are critical biogeochemical hotspots that occur where groundwater discharges into surface water, and being able to precisely identify and interpret metabolisms of Group 1 *L. ochracea* and *L. toolikensis* will improve our understanding of freshwater metal and nutrient cycles.

Description of *Leptothrix toolikensis* sp. nov

Leptothrix toolikensis (too.li.ken'sis. N.L. fem. adj. of Toolik, referring to the Toolik Lake Field Station; from the Iñupiaq word tutlik, meaning yellow-billed loon). The organism was identified from four metagenome-assembled genomes recovered from an iron seep adjacent to the Oksruykik River, North Slope, AK, USA. Genomes assigned to this species range from 2.67 to 2.87 Mbp in length with a G+C content of 60.9%–61.2%. Phylogenomic analysis places the organism within the genus *Leptothrix* in the family Sphaerotilaceae. Its closest relative is *L. ochracea*. A 16S rRNA gene sequence exists for the genome *L. toolikensis* 2E 3300045995 s8. Genomic analysis revealed the potential for lithotrophic iron oxidation, carbon fixation via the CBB cycle, and utilization of organic matter, indicative of a mixotrophic metabolism for cell growth. The type material is the genome sequence *L. toolikensis* 2C 3300045733 s0.

Description of *Leptothrix mechoopdaensis* sp. nov

Leptothrix mechoopdaensis (me.choop.da.en'sis. N.L. fem. adj. referring to Mechoopda, meaning "the marshy place," and honoring the Mechoopda tribe whose ancestral lands encompass the isolates' original iron seep). Cells grow as streptobacilli within a sheath, as observed from cultures grown under laboratory conditions. Cells grow on Lepto medium plates with $MnCO_3$ and $FeSO_4$ (16, 48). Strains BB-3 and BB-4 were isolated from an iron seep near Centerville, CA, USA. Genomes assigned to this species are 5.07 Mbp in length with a G+C content of 69.7%. Phylogenomic analysis places the organism within the *Sphaerotilus-Leptothrix* group in the family Sphaerotilaceae. Its closest relative is *L. discophora*. A 16S rRNA gene sequence exists for both isolates. The type material is the isolate strain *L. mechoopdaensis* BB-3.

MATERIALS AND METHODS

Genome collection

Thirty-eight high-quality genomes were included in this analysis (Table S1). Eleven MAGs were reconstructed from iron microbial mats at the Savannah River Site and from Spruce Point as described in Tothero et al. (21). Two *L. mechoopdaensis* isolates were sequenced from an iron seep in Centerville, CA, in Butte Creek Canyon as described in Fleming et al. (48).

Four MAGs (*L. toolikensis*) were reconstructed from samples collected at a unique iron seep adjacent to the Oksruykik River (68.6722, −149.1352), near the Toolik Lake field station (North Slope, AK) (47). Four separate samples, each consisting of 50 mL of aqueous mat, were collected on June 28, 2019, for metagenome analysis and MAG reconstruction. Each sample was concentrated by centrifugation and resuspended in 1 mL of deionized water. This suspension was added to a Powersoil Pro (Qiagen) extraction tube for DNA isolation. DNA was sent to the Joint Genome Institute for sequencing and metagenomic analysis. Illumina libraries were constructed and sequenced using the Illumina NovaSeq S4 platform, generating 2 × 151 paired-end reads; 132,294,094 reads totaling 19,976,408,194 bp were generated.

BBDuk v38.90 was used to remove contaminants and trim adapter sequences, homopolymers, and regions where the quality score drops to 0 (103), as well as to remove reads that contained four or more "N" bases, had an average quality score less than 3, or had a minimum length ≤51 bp. Reads were mapped with BBMap and filtered to remove reads that mapped to common eukaryotic (humans, cats, dogs, and mice) and microbial contaminants. Reads were corrected using bbcms v38.86 (103). The readset was assembled using metaSPAdes v3.14.1 using kmer lengths of 33, 55, 77, 99, and 127 (104). The input read set was mapped to the final assembly, and coverage information was generated with BBMap v38.86 (103).

The IMG/M, NCBI Genome, GenBank, and RefSeq databases were searched for existing genomes classified as *Leptothrix* and *Sphaerotilus*. In addition, the existing IMG BLAST all database was used to BLAST three ribosomal proteins (S8, L6, and L4; selected based on those identified in Olm et al. (105) for species-level delineation) from *L. cholodnii* and *L. mobilis*. Genomes with a hit >70% identity to at least one of these ribosomal proteins were pulled (n = 1,327), and hits were screened for genomes not classified to the genus level (n = 51). Genome fasta files were downloaded from IMG, and genomes were classified as *Leptothrix* and *Sphaerotilus* based on concatenated ribosomal protein trees.

All 16S rRNA gene sequences (n = 863,832) from GTDB genomes passing QC (≥200 bp, E-value ≥ 1e−06) were downloaded from the GTDB-tk release r220. Sequences were filtered to include only those in the Burkholderiaceae family (n = 21,725), and these were placed in a maximum-likelihood tree. The smallest monophyletic clade containing all verified *Leptothrix* and *Sphaerotilus* was selected from this tree, and full genomes from unclassified sequences within that clade were downloaded from GenBank and RefSeq for further classification.

Tree generation

Alignments were generated using MUSCLE v.5.1 (106) and trimmed and masked (sites with >30% gaps) in Geneious Prime. Maximum likelihood trees were generated in RAxML v1.2.1 with 1,000 bootstraps, using the LG+G model for trees of amino acid sequences and GTR+G model for trees of nucleotide sequences (107, 108). Trees were visualized in iTOL v7 (109).

Genome classification and data set curation

Genomes were annotated in Anvio-v8 (110) using Prodigal v.2.6.3 (111). Concatenated ribosomal protein trees were generated using alignments of 39 large and small ribosomal protein sequences (S12, L1, L13, L14, L16, L17, L18p, L19, L2, L20, L21p, L22, L23, L27, L27A, L28, L29, L3, L32p, L35p, L4, L5, L6, L9_C, S10, S11, S13, S15, S16, S17, S19, S2, S20p, S3_C, S6, S7, S8, S9, and L24). A phylogenetic tree was generated using the Up-to-date Bacterial Core Genes 2 pipeline using 81 core bacterial genes (112, 113), using prodigal v2.6.3 for gene calling (111), MAFFT v7.525 for alignment (114), and RAxML v8.2.13 for tree construction (107). AAI was calculated using EzAAI v.1.2.4 (115). Genome quality was determined using CheckM v1.0.18 (116). Average Nucleotide Identity (ANI) and Alignment Fraction (AF) were calculated using FastANI v0.1.3 (117, 118) and the ANIcalculator 2014-127, version 1.0 (https://ani.jgi.doe.gov/html/download.php) (119). Percentage of conserved proteins (POCP) was calculated using the Bio-py POCP-matrix calculator (53, 120).

Genomes were included in the analysis based on (i) clustering with a known *Leptothrix* or *Sphaerotilus* genome in the concatenated tree; (ii) >70% AAI to at least one known *Leptothrix* or *Sphaerotilus* genome; and (iii) greater than 90% completeness and less than 5% contamination. The final filtered data set, referred to as "SLG" contained 38 genomes. Genome groups, referred to as the "*ochracea*-type *Leptothrix*" or Group 1 (n = 13), "*mobilis*-type *Leptothrix*" or Group 2 (n = 8), "*Sphaerotilus*" or Group 3 (n = 15), or "outliers" (n = 2), were separated based on monophyletic clades in the concatenated ribosomal protein tree.

Pangenome generation

Anvio-v8 was used to generate a pangenome database of the final filtered data set (110, 121). Genes were clustered using a min-bit parameter of 0.5 and an mcl inflation parameter of 2, and using NCBI blastp for sequence similarity (122–124) and hierarchical clustering. Functional annotations were assigned to gene clusters in using FeGenie (125), NCBI COGs, Pfams, Kegg Kofams, CAZymes, a modified heme counter script to search for CXXCH, CXXXCH, and CXXXXCH motifs (126), and a custom Python script that detected conserved MCO motifs. Near-core gene sets were identified based on the presence of gene clusters in > 85% of genomes in the final filtered data set or within each genome group. Venn diagram generation was done using goodcalculators.com/venn-diagram-maker/.

Gene sequence analyses

The localization of protein sequences with 10 or more heme-binding motifs was determined using PSORTb v3.0.3 (127). If needed, signal peptides were identified with SignalP 6.0 and/or Phobius (128–130). Porins were searched for using a combination of protein localization prediction, BLAST, and HHpred (131, 132).

To predict the oxidative vs. reductive function of DsrA in SLG, a maximum-likelihood protein tree was constructed using reference sequences from Loy et al. (72). To validate the FeGenie annotations of MtoA sequences, a maximum-likelihood protein tree was constructed using reference sequences (133), including functionally validated MtoA, MtrA, and PioA reference sequences.

Data visualization was performed in R v4.3.2 using the tidyverse ecosystem. Scatter plots and bubble plots were generated using ggplot2 (134). Heatmaps were visualized using pheatmap and complexheatmap (135, 136). Dendrograms were generated using dendextend (137).

ACKNOWLEDGMENTS

We thank the CSU Chico General Microbiology class, Abigail P Mcintyre, and Joyce Zimmer for work in isolating and maintaining *Leptothrix mechoopdaensis* BB-3 and BB-4. We acknowledge that field sampling was conducted on the ancestral and contemporary lands of the Mechoopda Indian Tribe. We are grateful to the Mechoopda Indian Tribe of Chico Rancheria for guidance and consultation regarding the naming of *L. mechoopdaensis*. We are grateful to Jackson Tsuji for advisement on genome search methods. We acknowledge Nicole Nordone for assisting with database searching and Austin Chambers and Christopher Blanda for writing a custom Python script for detecting MCO motifs. We thank Rene Hoover for guidance on performing the pangenome analysis in Anvio.

This work was supported by an NSF grant EAR-2243577 to C.S.C. and J.L.K., DOE Environmental Systems Science Grant DE-SC0021010 to C.S.C., and NSF grant EAR-2243578 to E.J.F. Metagenomic sequencing and assembly were done with support from CSP proposal 503499; DOI 10.46936/10.25585/60001141, conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. Sample collection in the Arctic was supported by the NSF grant DEB-1754358 to D.E. Support from the University of Delaware Bioinformatics Data Science Core Facility (RRID:SCR_017696), including use of the BIOMIX and BioStore computational resources, was made possible through funding from Delaware INBRE (NIGMS P20GM103446), NIH Shared Instrumentation Grant (NIH S10OD028725), the State of Delaware, and the Delaware Biotechnology Institute.

AUTHOR AFFILIATIONS

[1]Delaware Biotechnology Institute, Newark, Delaware, USA
[2]Department of Earth Sciences, University of Delaware, Newark, Delaware, USA
[3]Microbiology Graduate Program, University of Delaware, Newark, Delaware, USA
[4]Bigelow Laboratory for Ocean Sciences, East Boothbay, Maine, USA
[5]Department of Biological Sciences, California State University, Chico, California, USA

AUTHOR ORCIDs

Gracee K. Tothero http://orcid.org/0000-0001-7792-3880
Jessica L. Keffer http://orcid.org/0000-0002-0302-3588
David Emerson http://orcid.org/0000-0003-1737-356X
Emily J. Fleming http://orcid.org/0009-0002-2944-5575
Clara S. Chan http://orcid.org/0000-0003-1810-4994

FUNDING

Funder	Grant(s)	Author(s)
National Science Foundation	EAR-2243577	Gracee K. Tothero Jessica L. Keffer Clara S. Chan
National Science Foundation	EAR-2243578	Emily J. Fleming
Joint Genome Institute	503499	Gracee K. Tothero Clara S. Chan

Funder	Grant(s)	Author(s)
National Science Foundation	DEB-1754358	David Emerson
Delaware IDeA Network of Biomedical Research Excellence		Gracee K. Tothero Jessica L. Keffer Clara S. Chan
National Institutes of Health	S10OD028725	Grace K. Tothero Jessica L. Keffer Clara S. Chan
State of Delaware (Delaware)		Gracee K. Tothero Jessica L. Keffer Clara S. Chan
Delaware Biotechnology Institute		Gracee K. Tothero Jessica L. Keffer Clara S. Chan

AUTHOR CONTRIBUTIONS

Gracee K. Tothero, Conceptualization, Data curation, Investigation, Visualization, Writing – original draft | Jessica L. Keffer, Investigation, Writing – review and editing | David Emerson, Resources, Writing – review and editing | Emily J. Fleming, Resources, Writing – review and editing | Clara S. Chan, Conceptualization, Funding acquisition, Project administration, Writing – review and editing

DATA AVAILABILITY

Genome sequences for *Leptothrix mechoopdaensis* BB-3 and BB-4 are deposited in NCBI under BioProject PRJNA1336152, under GenBank accessions GCA_053893855.1 (BB-3) and GCA_053893815.1 (BB-4). Genome sequences for *Leptothrix toolikensis* strains are deposited in NCBI under BioProject PRJNA1438219. The taxonomic names proposed in this study are established under SeqCode at https://registry.seqco.de/registers/r:3pq_uc4e.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Supplemental figures and text (mSystems01768-25-S0001.pdf). Supplemental text on genomes and Figures S1-S5.
Supplemental Tables (mSystems01768-25-S0002.xlsx). Tables S1-S4.

REFERENCES

1. Roth AW. 1797. Catalecta botanica: quibus plantae novae et minus cognitae descriuntur atque illustratur. I.G. Müller, Lipsiae. https://www.biodiversitylibrary.org/bibliography/124909.
2. Winogradsky S. 1922. Eisenbakterien als anorgoxydanten. Zbl Bakt II Abt 57:1–21.
3. Harder EC. 1919. Iron-depositing bacteria and their geologic relations. Geological Survey professional paper 113. https://doi.org/10.3133/pp113
4. Kützing FT. 1843. Phycologia generalis oder anatomie, physiologie und systemkunde der tange. Brockhaus.
5. Winogradsky S. 1888. Uber eisenbakterien. Bot Zeitschr 46:261–270. https://www.biodiversitylibrary.org/page/5144120.
6. Spring S. 2006. The genera *Leptothrix* and *Sphaerotilus*, p 758–777. *In* Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (ed), The prokaryotes. Springer New York, New York, NY.
7. Mulder EG, Van Veen WL. 1963. Investigations on the *Sphaerotilus-Leptothrix* group. Antonie Van Leeuwenhoek 29:121–153. https://doi.org/10.1007/BF02046045
8. Pringsheim EG. 1949. The filamentous bacteria *Sphaerotilus*, *Leptothrix*, *Cladothrix*, and their relation to iron and manganese. Phil Trans R Soc Lond B 233:453–482. https://doi.org/10.1098/rstb.1949.0002
9. Spring S, Kämpfer P. 2005. Incertae sedis III. *Leptothrix*. *In* Brenner DJ, Krieg NR, Staley JT, Garrity GM (ed), Bergey's manual of systematics of archaea and bacteria v. 2. Springer, New York.
10. Kämpfer P, Spring S. 2005. Incertae sedis VI. *Sphaerotilus*. *In* Brenner DJ, KriegNR, StaleyJT, GarrityGM (ed), Bergey's manual of systematics of archaea and bacteria v. 2. Springer, New York.
11. Emerson D, Weiss JV. 2004. Bacterial iron oxidation in circumneutral freshwater habitats: findings from the field and the laboratory. Geomicrobiol J 21:405–414. https://doi.org/10.1080/01490450490485881

12. Van Veen WL, Mulder EG, Deinema MH. 1978. The *Sphaerotilus-Leptothrix* group of bacteria. Microbiol Rev 42:329–356. https://doi.org/10.1128/mr.42.2.329-356.1978
13. Johnson AH, Stokes JL. 1966. Manganese oxidation by *Sphaerotilus discophorus*. J Bacteriol 91:1543–1547. https://doi.org/10.1128/jb.91.4.1543-1547.1966
14. Kunoh T, Hashimoto H, Suzuki T, Hayashi N, Tamura K, Takano M, Kunoh H, Takada J. 2016. Direct adherence of Fe(III) particles onto sheaths of *Leptothrix sp.* strain OUMS1 in culture. Minerals 6:4. https://doi.org/10.3390/min6010004
15. Corstjens PL, De Vrind JP, Westbroek P, De Vrind-de Jong EW. 1992. Enzymatic iron oxidation by *Leptothrix discophora*: identification of an iron-oxidizing protein. Appl Environ Microbiol 58:450–454. https://doi.org/10.1128/aem.58.2.450-454.1992
16. Boogerd FC, De Vrind JP. 1987. Manganese oxidation by *Leptothrix discophora*. J Bacteriol 169:489–494. https://doi.org/10.1128/jb.169.2.489-494.1987
17. Stokes JL. 1954. Studies on the filamentous sheathed iron bacterium *Sphaerotilus natans*. J Bacteriol 67:278–291. https://doi.org/10.1128/jb.67.3.278-291.1954
18. Dondero NC. 1975. The *Sphaerotilus-Leptothrix* group. Annu Rev Microbiol 29:407–428. https://doi.org/10.1146/annurev.mi.29.100175.002203
19. Siering PL, Ghiorse WC. 1996. Phylogeny of the *Sphaerotilus-Leptothrix* group inferred from morphological comparisons, genomic fingerprinting, and 16S ribosomal DNA sequence analyses. Int J Syst Bacteriol 46:173–182. https://doi.org/10.1099/00207713-46-1-173
20. Fleming EJ, Langdon AE, Martinez-Garcia M, Stepanauskas R, Poulton NJ, Masland EDP, Emerson D. 2011. What's new is old: resolving the identity of *Leptothrix ochracea* using single cell genomics, pyrosequencing and FISH. PLoS One 6:e17769. https://doi.org/10.1371/journal.pone.0017769
21. Tothero GK, Hoover RL, Farag IF, Kaplan DI, Weisenhorn P, Emerson D, Chan CS. 2024. *Leptothrix ochracea* genomes reveal potential for mixotrophic growth on Fe(II) and organic carbon. Appl Environ Microbiol 90:e00599-24. https://doi.org/10.1128/aem.00599-24
22. Willems A, De Ley J, Gillis M, Kersters K. 1991. NOTES: Comamonadaceae, a new family encompassing the acidovorans rRNA complex, including *Variovorax paradoxus* gen. nov., comb. nov., for *Alcaligenes paradoxus* (Davis 1969). Int J Syst Bacteriol 41:445–450. https://doi.org/10.1099/00207713-41-3-445
23. Corstjens P, Muyzer G. 1993. Phylogenetic analysis of the metal-oxidizing bacteria *Leptothrix discophora* and *Sphaerotilus natans* using 16S rDNA sequencing data. Syst Appl Microbiol 16:219–223. https://doi.org/10.1016/S0723-2020(11)80471-6
24. Schmidt B, Sánchez LA, Fretschner T, Kreps G, Ferrero MA, Siñeriz F, Szewzyk U. 2014. Isolation of Sphaerotilus-Leptothrix strains from iron bacteria communities in Tierra del Fuego wetlands. FEMS Microbiol Ecol 90:454–466. https://doi.org/10.1111/1574-6941.12406
25. Grabovich MY, Smolyakov DD, Beletsky AV, Mardanov AV, Gureeva MV, Markov ND, Rudenko TS, Ravin NV. 2021. Reclassification of *Sphaerotilus natans* subsp. sulfidivorans Gridneva et al. 2011 as *Sphaerotilus sulfidivorans* sp. nov. and comparative genome analysis of the genus *Sphaerotilus*. Arch Microbiol 203:1595–1599. https://doi.org/10.1007/s00203-020-02158-6
26. Smolyakov DD, Beletsky AV, Rudenko TS, Gureeva MV, Ravin NV, Grabovich MY. 2025. Comparative genome analysis of the *Sphaerotilus-Leptothrix* group supports the unification of the genera *Sphaerotilus* and *Leptothrix* into a single emended genus *Sphaerotilus*. Syst Appl Microbiol 48:126666. https://doi.org/10.1016/j.syapm.2025.126666
27. Liu Y, Du J, Pei T, Du H, Feng G-D, Zhu H. 2022. Genome-based taxonomic classification of the closest-to-Comamonadaceae group supports a new family Sphaerotilaceae fam. nov. and taxonomic revisions. Syst Appl Microbiol 45:126352. https://doi.org/10.1016/j.syapm.2022.126352
28. Kämpfer P. 1998. Some chemotaxonomic and physiological properties of the genus *Sphaerotilus*. Syst Appl Microbiol 21:245–250. https://doi.org/10.1016/S0723-2020(98)80029-5
29. Pellegrin V, Juretschko S, Wagner M, Cottenceau G. 1999. Morphological and biochemical properties of a *Sphaerotilus* sp. isolated from paper mill slimes. Appl Environ Microbiol 65:156–162. https://doi.org/10.1128/AEM.65.1.156-162.1999
30. Park S, Kim D-H, Lee J-H, Hur H-G. 2014. *Sphaerotilus natans* encrusted with nanoball-shaped Fe(III) oxide minerals formed by nitrate-reducing mixotrophic Fe(II) oxidation. FEMS Microbiol Ecol 90:68–77. https://doi.org/10.1111/1574-6941.12372
31. Chan CS, McAllister SM, Leavitt AH, Glazer BT, Krepski ST, Emerson D. 2016. The architecture of iron microbial mats reflects the adaptation of chemolithotrophic iron oxidation in freshwater and marine environments. Front Microbiol 7:796. https://doi.org/10.3389/fmicb.2016.00796
32. Mulder EG. 1964. Iron bacteria, particularly those of the *Sphaerotilus - Leptothrix* group, and industrial problems. J Appl Bacteriol 27:151–173. https://doi.org/10.1111/j.1365-2672.1964.tb04824.x
33. Fleming EJ, Woyke T, Donatello RA, Kuypers MMM, Sczyrba A, Littmann S, Emerson D. 2018. Insights into the fundamental physiology of the uncultured Fe-oxidizing bacterium *Leptothrix ochracea*. Appl Environ Microbiol 84:e02239-17. https://doi.org/10.1128/AEM.02239-17
34. Ghiorse WC. 1984. Biology of iron- and manganese-depositing bacteria. Annu Rev Microbiol 38:515–550. https://doi.org/10.1146/annurev.mi.38.100184.002503
35. Sobolev D, Roden EE. 2004. Characterization of a neutrophilic, chemolithoautotrophic Fe(II)-oxidizing *β*-proteobacterium from freshwater wetland sediments. Geomicrobiol J 21:1–10. https://doi.org/10.1080/01490450490253310
36. Emerson D, Fleming EJ, McBeth JM. 2010. Iron-oxidizing bacteria: an environmental and genomic perspective. Annu Rev Microbiol 64:561–583. https://doi.org/10.1146/annurev.micro.112408.134208
37. Fleming EJ, Cetinić I, Chan CS, Whitney King D, Emerson D. 2014. Ecological succession among iron-oxidizing bacteria. ISME J 8:804–815. https://doi.org/10.1038/ismej.2013.197
38. Adams LF, Ghiorse WC. 1985. Influence of manganese on growth of a sheathless strain of *Leptothrix discophora*. Appl Environ Microbiol 49:556–562. https://doi.org/10.1128/aem.49.3.556-562.1985
39. Emerson D, Ghiorse WC. 1992. Isolation, cultural maintenance, and taxonomy of a sheath-forming strain of *Leptothrix discophora* and characterization of manganese-oxidizing activity associated with the sheath. Appl Environ Microbiol 58:4001–4010. https://doi.org/10.1128/aem.58.12.4001-4010.1992
40. PringsheimEG. 1949. Iron Bacteria. Biological Reviews 24:200–245. https://doi.org/10.1111/j.1469-185x.1949.tb00575.x
41. Gridneva E, Chernousova E, Dubinina G, Akimov V, Kuever J, Detkova E, Grabovich M. 2011. Taxonomic investigation of representatives of the genus *Sphaerotilus*: descriptions of *Sphaerotilus montanus* sp. nov., *Sphaerotilus hippei* sp. nov., *Sphaerotilus natans* subsp. *natans* subsp. nov. and *Sphaerotilus natans* subsp. *sulfidivorans* subsp. nov., and an emended description of the genus *Sphaerotilus*. Int J Syst Evol Microbiol 61:916–925. https://doi.org/10.1099/ijs.0.023887-0
42. Lechner U, Brodkorb D, Geyer R, Hause G, Härtig C, Auling G, Fayolle-Guichard F, Piveteau P, Müller RH, Rohwerder T. 2007. *Aquincola tertiaricarbonis* gen. nov., sp. nov., a tertiary butyl moiety-degrading bacterium. Int J Syst Evol Microbiol 57:1295–1303. https://doi.org/10.1099/ijs.0.64663-0
43. Manaia CM, Nunes OC, Nogales B. 2003. *Caenibacterium thermophilum* gen. nov., sp. nov., isolated from a thermophilic aerobic digester of municipal sludge. Int J Syst Evol Microbiol 53:1375–1382. https://doi.org/10.1099/ijs.0.02622-0
44. Seguel Suazo K, Nierychlo M, Kondrotaite Z, Petriglieri F, Peces M, Singleton C, Dries J, Nielsen PH. 2025. Diversity and abundance of filamentous and non-filamentous "*Leptothrix*" in global wastewater treatment plants. Appl Environ Microbiol 91:e01485-24. https://doi.org/10.1128/aem.01485-24
45. Narihara S, Chida S, Matsunaga N, Akimoto R, Akimoto M, Hagio A, Mori T, Nittami T, Sato M, Mun S, Kang H, Back JH, Takeda M. 2024. Taxonomic characterization of Sphaerotilus microaerophilus sp. nov., a sheath-forming microaerophilic bacterium of activated sludge origin. Arch Microbiol 206:252. https://doi.org/10.1007/s00203-024-03991-9
46. BowersRM, KyrpidesNC, StepanauskasR, Harmon-SmithM, DoudD, ReddyTBK, SchulzF, JarettJ, et al, The Genome Standards Consortium. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35:725–731. https://doi.org/10.1038/nbt.3893
47. Michaud AB, Massé RO, Emerson D. 2023. Microbial iron cycling is prevalent in water-logged Alaskan Arctic tundra habitats, but sensitive to disturbance. FEMS Microbiol Ecol 99:fiad013. https://doi.org/10.1093/femsec/fiad013
48. Fleming EJ, McIntyre A, Tothero GK, Zimmer JA, Keffer JL, Pierce LN, Beilsmith KR, Wilton R, Chan CS. 2026. Complete genome sequences of metal-oxidizing bacteria *Leptothrix mechoopdaensis* BB-3 and BB-4.

Microbiol Resour Announc:e01458-25. https://doi.org/10.1128/mra.01458-25
49. Rodriguez-R LM, Konstantinidis KT. 2014. Bypassing cultivation to identify bacterial species: culture-independent genomic approaches identify credibly distinct clusters, avoid cultivation bias, and provide true insights into microbial species. Microbe Magazine 9:111–118. https://doi.org/10.1128/microbe.9.111.1
50. Konstantinidis KT, Tiedje JM. 2007. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. Curr Opin Microbiol 10:504–509. https://doi.org/10.1016/j.mib.2007.08.006
51. Kim M, Oh H-S, Park S-C, Chun J. 2014. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol 64:346–351. https://doi.org/10.1099/ijs.0.059774-0
52. Riesco R, Trujillo ME. 2024. Update on the proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int J Syst Evol Microbiol 74. https://doi.org/10.1099/ijsem.0.006300
53. Qin Q-L, Xie B-B, Zhang X-Y, Chen X-L, Zhou B-C, Zhou J, Oren A, Zhang Y-Z. 2014. A proposed genus boundary for the prokaryotes based on genomic insights. J Bacteriol 196:2210–2215. https://doi.org/10.1128/JB.01688-14
54. Barco RA, Garrity GM, Scott JJ, Amend JP, Nealson KH, Emerson D. 2020. A genus definition for Bacteria and Archaea based on a standard genome relatedness index. mBio 11:e02475-19. https://doi.org/10.1128/mBio.02475-19
55. Keffer JL, McAllister SM, Garber AI, Hallahan BJ, Sutherland MC, Rozovsky S, Chan CS. 2021. Iron oxidation by a fused cytochrome-porin common to diverse iron-oxidizing bacteria. mBio 12:e01074-21. https://doi.org/10.1128/mBio.01074-21
56. Castelle C, Guiral M, Malarte G, Ledgham F, Leroy G, Brugna M, Giudici-Orticoni M-T. 2008. A new iron-oxidizing/O_2-reducing supercomplex spanning both inner and outer membranes, isolated from the extreme acidophile *Acidithiobacillus ferrooxidans*. J Biol Chem 283:25803–25811. https://doi.org/10.1074/jbc.M802496200
57. Jeans C, Singer SW, Chan CS, Verberkmoes NC, Shah M, Hettich RL, Banfield JF, Thelen MP. 2008. Cytochrome 572 is a conspicuous membrane protein with iron oxidation activity purified directly from a natural acidophilic microbial community. ISME J 2:542–550. https://doi.org/10.1038/ismej.2008.17
58. Barco RA, Emerson D, Sylvan JB, Orcutt BN, Jacobson Meyers ME, Ramírez GA, Zhong JD, Edwards KJ. 2015. New insight into microbial iron oxidation as revealed by the proteomic profile of an obligate iron-oxidizing chemolithoautotroph. Appl Environ Microbiol 81:5927–5937. https://doi.org/10.1128/AEM.01374-15
59. McAllister SM, Polson SW, Butterfield DA, Glazer BT, Sylvan JB, Chan CS. 2020. Validating the Cyc2 neutrophilic iron oxidation pathway using meta-omics of *Zetaproteobacteria* iron mats at marine hydrothermal vents. mSystems 5:e00553-19. https://doi.org/10.1128/mSystems.00553-19
60. Liu J, Wang Z, Belchik SM, Edwards MJ, Liu C, Kennedy DW, Merkley ED, Lipton MS, Butt JN, Richardson DJ, Zachara JM, Fredrickson JK, Rosso KM, Shi L. 2012. Identification and characterization of MtoA: a decaheme c-type cytochrome of the neutrophilic Fe(II)-oxidizing bacterium *Sideroxydans lithotrophicus* ES-1. Front Microbio 3. https://doi.org/10.3389/fmicb.2012.00037
61. Zhou N, Keffer JL, Polson SW, Chan CS. 2022. Unraveling Fe(II)-oxidizing mechanisms in a facultative Fe(II) oxidizer, *Sideroxydans lithotrophicus* strain ES-1, via culturing, transcriptomics, and reverse transcription-quantitative PCR. Appl Environ Microbiol 88:e01595-21. https://doi.org/10.1128/AEM.01595-21
62. He S, Barco RA, Emerson D, Roden EE. 2017. Comparative genomic analysis of neutrophilic iron(II) oxidizer genomes for candidate genes in extracellular electron transfer. Front Microbiol 8:1584. https://doi.org/10.3389/fmicb.2017.01584
63. Eddie BJ, Bird LJ, Pelikan C, Mussmann M, Martinez-Perez C, Pinamang P, Malanoski AP, Glaven SM. 2022. Conservation of energetic pathways for electroautotrophy in the uncultivated candidate order Tenderiales. mSphere 5:e00223-22. https://doi.org/10.1128/msphere.00223-22
64. Zhang Z, Zhang Z, Chen H, Liu J, Liu C, Ni H, Zhao C, Ali M, Liu F, Li L. 2015. Surface Mn(II) oxidation actuated by a multicopper oxidase in a soil bacterium leads to the formation of manganese oxide minerals. Sci Rep 5:10895. https://doi.org/10.1038/srep10895
65. Huston WM, Jennings MP, McEwan AG. 2002. The multicopper oxidase of *Pseudomonas aeruginosa* is a ferroxidase with a central role in iron acquisition. Mol Microbiol 45:1741–1750. https://doi.org/10.1046/j.1365-2958.2002.03132.x
66. Kaur K, Sharma A, Capalash N, Sharma P. 2019. Multicopper oxidases: biocatalysts in microbial pathogenesis and stress management. Microbiol Res 222:1–13. https://doi.org/10.1016/j.micres.2019.02.007
67. Corstjens PLAM, de Vrind JPM, Goosen T, Jong EWDV. 1997. Identification and molecular analysis of the *Leptothrix discophora* SS - 1 *mofA* gene, a gene putatively encoding a manganese - oxidizing protein with copper domains. Geomicrobiol J 14:91–108. https://doi.org/10.1080/01490459709378037
68. Butterfield CN, Soldatova AV, Lee S-W, Spiro TG, Tebo BM. 2013. Mn(II,III) oxidation and MnO_2 mineralization by an expressed bacterial multicopper oxidase. Proc Natl Acad Sci USA 110:11731–11735. https://doi.org/10.1073/pnas.1303677110
69. Geszvain K, McCarthy JK, Tebo BM. 2013. Elimination of manganese(II,III) oxidation in *Pseudomonas putida* GB-1 by a double knockout of two putative multicopper oxidase genes. Appl Environ Microbiol 79:357–366. https://doi.org/10.1128/AEM.01850-12
70. Ridge JP, Lin M, Larsen EI, Fegan M, McEwan AG, Sly LI. 2007. A multicopper oxidase is essential for manganese oxidation and laccase-like activity in *Pedomicrobium sp.* ACM 3067. Environ Microbiol 9:944–953. https://doi.org/10.1111/j.1462-2920.2006.01216.x
71. Schedel M, Trüper HG. 1979. Purification of *Thiobacillus denitrificans* siroheme sulfite reductase and investigation of some molecular and catalytic properties. Biochim Biophys Acta Enzymol 568:454–466. https://doi.org/10.1016/0005-2744(79)90314-0
72. Loy A, Duller S, Baranyi C, Mußmann M, Ott J, Sharon I, Béjà O, Le Paslier D, Dahl C, Wagner M. 2009. Reverse dissimilatory sulfite reductase as phylogenetic marker for a subgroup of sulfur - oxidizing prokaryotes. Environ Microbiol 11:289–299. https://doi.org/10.1111/j.1462-2920.2008.01760.x
73. Badger MR, Bek EJ. 2008. Multiple Rubisco forms in Proteobacteria: their functional significance in relation to CO_2 acquisition by the CBB cycle. J Exp Bot 59:1525–1541. https://doi.org/10.1093/jxb/erm297
74. Vesenka J, Havu J, Hruby K, Emerson D. 2018. A model for sheath formation coupled to motility in *Leptothrix ochracea*. Geomicrobiol J 35:366–374. https://doi.org/10.1080/01490451.2017.1370516
75. Reid AN, Whitfield C. 2005. Functional analysis of conserved gene products involved in assembly of *Escherichia coli* capsules and exopolysaccharides: evidence for molecular recognition between Wza and Wzc for colanic acid biosynthesis. J Bacteriol 187:5470–5481. https://doi.org/10.1128/JB.187.15.5470-5481.2005
76. Cuthbertson L, Mainprize IL, Naismith JH, Whitfield C. 2009. Pivotal roles of the outer membrane polysaccharide export and polysaccharide copolymerase protein families in export of extracellular polysaccharides in gram-negative bacteria. Microbiol Mol Biol Rev 73:155–177. https://doi.org/10.1128/MMBR.00024-08
77. Kunoh T, Yamamoto T, Ono E, Sugimoto S, Takabe K, Takeda M, Utada AS, Nomura N. 2023. Identification of lthB, a gene encoding a putative glycosyltransferase family 8 protein required for *Leptothrix* sheath formation. Appl Environ Microbiol 89:e01919-22. https://doi.org/10.1128/aem.01919-22
78. Oren A, Arahal DR, Göker M, Moore ERB, Rossello-Mora R, Sutcliffe IC. 2023. International code of nomenclature of prokaryotes. Prokaryotic code (2022 revision). Int J Syst Evol Microbiol 73. https://doi.org/10.1099/ijsem.0.005585
79. Hedlund BP, Chuvochina M, Hugenholtz P, Konstantinidis KT, Murray AE, Palmer M, Parks DH, Probst AJ, Reysenbach A-L, Rodriguez-R LM, Rossello-Mora R, Sutcliffe IC, Venter SN, Whitman WB. 2022. SeqCode: a nomenclatural code for prokaryotes described from sequence data. Nat Microbiol 7:1702–1708. https://doi.org/10.1038/s41564-022-01214-9
80. Freese HM, Meier-Kolthoff JP, Sardà Carbasse J, Afolayan AO, Göker M. 2026. TYGS and LPSN in 2025: A global core biodata resource for genome-based classification and nomenclature of prokaryotes within DSMZ digital diversity. Nucleic Acids Res 54:D884–D891. https://doi.org/10.1093/nar/gkaf1110
81. Sneath PHA, McGowan V, Skerman VBD. 1980. Approved lists of bacterial names. Int J Syst Evol Microbiol 30:225–420. https://doi.org/10.1099/00207713-30-1-225
82. Rouf MA, Stokes JL. 1964. Morphology, nutrition and physiology of *Sphaerotilus discophorus*. Archiv Mikrobiol 49:132–149. https://doi.org/10.1007/BF00422137

83. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol 36:996–1004. https://doi.org/10.1038/nbt.4229
84. Medlar AJ, Törönen P, Holm L. 2018. AAI-profiler: fast proteome-wide exploratory analysis reveals taxonomic identity, misclassification and contamination. Nucleic Acids Res 46:W479–W485. https://doi.org/10.1093/nar/gky359
85. Benedek T, Táncsics A, Szabó I, Farkas M, Szoboszlay S, Fábián K, Maróti G, Kriszt B. 2016. Polyphasic analysis of an *Azoarcus-Leptothrix*-dominated bacterial biofilm developed on stainless steel surface in a gasoline-contaminated hypoxic groundwater. Environ Sci Pollut Res 23:9019–9035. https://doi.org/10.1007/s11356-016-6128-0
86. Cleenwerck I, De Vos P. 2008. Polyphasic taxonomy of acetic acid bacteria: an overview of the currently applied methodology. Int J Food Microbiol 125:2–14. https://doi.org/10.1016/j.ijfoodmicro.2007.04.017
87. Gillis M, Vandamme P, de Vos P, Swings J, Kersters K. 2001. Polyphasic taxonomy. In Garrity GM, BooneDR, CastenholzRW (ed), Bergey's manual of systematic bacteriology v.1. Springer, New York.
88. Prakash O, Verma M, Sharma P, Kumar M, Kumari K, Singh A, Kumari H, Jit S, Gupta SK, Khanna M, Lal R. 2007. Polyphasic approach of bacterial classification — an overview of recent advances. Indian J Microbiol 47:98–108. https://doi.org/10.1007/s12088-007-0022-x
89. Ramasamy D, Mishra AK, Lagier J-C, Padhmanabhan R, Rossi M, Sentausa E, Raoult D, Fournier P-E. 2014. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol Microbiol 64:384–391. https://doi.org/10.1099/ijs.0.057091-0
90. Spring S, Kampfer P, Ludwig W, Schleifer K-H. 1996. Polyphasic characterization of the genus *Leptothrix*: new descriptions of *Leptothrix mobilis* sp. nov. and *Leptothrix discophora* sp. nov. nom. rev. and emended description of *Leptothrix cholodnii* emend. Syst Appl Microbiol 19:634–643. https://doi.org/10.1016/S0723-2020(96)80036-1
91. Uilenberg G, Goff WL. 2006. Polyphasic taxonomy. Ann N Y Acad Sci 1081:492–497. https://doi.org/10.1196/annals.1373.073
92. Vandamme P, Pot B, Gillis M, de Vos P, Kersters K, Swings J. 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. Microbiol Rev 60:407–438. https://doi.org/10.1128/mr.60.2.407-438.1996
93. Palmer M, Venter SN, Coetzee MPA, Steenkamp ET. 2019. Prokaryotic species are *sui generis* evolutionary units. Syst Appl Microbiol 42:145–158. https://doi.org/10.1016/j.syapm.2018.10.002
94. Mulder EG. 1989. Sheathed bacteria, p 1994–2003. *In* Staley JT (ed), Bergey's manual of systematic bacteriology. Vol. 3. Williams and Wilkins, Baltimore.
95. Singh SK, Grass G, Rensing C, Montfort WR. 2004. Cuprous oxidase activity of CueO from *Escherichia coli*. J Bacteriol 186:7815–7817. https://doi.org/10.1128/JB.186.22.7815-7817.2004
96. Rowland JL, Niederweis M. 2013. A multicopper oxidase is required for copper resistance in *Mycobacterium tuberculosis*. J Bacteriol 195:3724–3733. https://doi.org/10.1128/JB.00546-13
97. Hoover RL, Keffer JL, Polson SW, Chan CS. 2023. Gallionellaceae pangenomic analysis reveals insight into phylogeny, metabolic flexibility, and iron oxidation mechanisms. mSystems 8:e00038-23. https://doi.org/10.1128/msystems.00038-23
98. Zhou N, Kupper RJ, Catalano JG, Thompson A, Chan CS. 2022. Biological oxidation of Fe(II)-bearing smectite by microaerophilic iron oxidizer *Sideroxydans lithotrophicus* using dual mto and Cyc2 iron oxidation pathways. Environ Sci Technol 56:17443–17453. https://doi.org/10.1021/acs.est.2c05142
99. Takeda M, Kawasaki Y, Umezu T, Shimura S, Hasegawa M, Koizumi J. 2012. Patterns of sheath elongation, cell proliferation, and manganese(II) oxidation in *Leptothrix cholodnii*. Arch Microbiol 194:667–673. https://doi.org/10.1007/s00203-012-0801-6
100. Takeda M, Umezu T, Kawasaki Y, Shimura S, Kondo K, Koizumi J. 2012. A spatial relationship between sheath elongation and cell proliferation in *Sphaerotilus natans*. Biosci Biotechnol Biochem 76:2357–2359. https://doi.org/10.1271/bbb.120616
101. Suzuki T, Kanagawa T, Kamagata Y. 2002. Identification of a gene essential for sheathed structure formation in *Sphaerotilus natans*, a filamentous sheathed bacterium. Appl Environ Microbiol 68:365–371. https://doi.org/10.1128/AEM.68.1.365-371.2002
102. Takeda M, Nakano F, Nagase T, Iohara K, Koizumi J. 1998. Isolation and chemical composition of the sheath of *Sphaerotilus natans*. Biosci Biotechnol Biochem 62:1138–1143. https://doi.org/10.1271/bbb.62.1138
103. Bushnell B. 2014. BBTools software package. http://bbtools.jgi.doe.gov.
104. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824–834. https://doi.org/10.1101/gr.213959.116
105. Olm MR, Crits-Christoph A, Diamond S, Lavy A, Matheus Carnevali PB, Banfield JF. 2020. Consistent metagenome-derived metrics verify and delineate bacterial species boundaries. mSystems 5:e00731-19. https://doi.org/10.1128/mSystems.00731-19
106. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform 5:113. https://doi.org/10.1186/1471-2105-5-113
107. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033
108. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35:4453–4455. https://doi.org/10.1093/bioinformatics/btz305
109. Letunic I, Bork P. 2024. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res 52:W78–W82. https://doi.org/10.1093/nar/gkae268
110. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, et al. 2020. Community-led, integrated, reproducible multi-omics with anvi'o. Nat Microbiol 6:3–6. https://doi.org/10.1038/s41564-020-00834-3
111. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform 11:119. https://doi.org/10.1186/1471-2105-11-119
112. Chalita M, Kim YO, Park S, Oh H-S, Cho JH, Moon J, Baek N, Moon C, Lee K, Yang J, Nam GG, Jung Y, Na S-I, Bailey MJ, Chun J. 2024. EzBioCloud: a genome-driven database and platform for microbiome identification and discovery. Int J Syst Evol Microbiol 74:006421. https://doi.org/10.1099/ijsem.0.006421
113. Kim J, Na S-I, Kim D, Chun J. 2021. UBCG2: up-to-date bacterial core genes and pipeline for phylogenomic analysis. J Microbiol 59:609–615. https://doi.org/10.1007/s12275-021-1231-4
114. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010
115. Kim D, Park S, Chun J. 2021. Introducing EzAAI: a pipeline for high throughput calculations of prokaryotic average amino acid identity. J Microbiol 59:476–480. https://doi.org/10.1007/s12275-021-1154-0
116. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114
117. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. https://doi.org/10.1038/s41467-018-07641-9
118. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57:81–91. https://doi.org/10.1099/ijs.0.64483-0
119. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, Pati A. 2015. Microbial species delineation using whole genome sequences. Nucleic Acids Res 43:6761–6771. https://doi.org/10.1093/nar/gkv657
120. Lin H. 2021. SilentGene/Bio-py: Bio-py v1.0
121. Delmont TO, Eren AM. 2018. Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. PeerJ:6:e4320. https://doi.org/10.7717/peerj.4320
122. Van Dongen S. 2008. Graph clustering via a discrete uncoupling process. SIAM J Matrix Anal & Appl 30:121–141. https://doi.org/10.1137/040608635
123. Enright AJ. 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res 30:1575–1584. https://doi.org/10.1093/nar/30.7.1575
124. Van Dongen S, Abreu-Goodger C. 2012. Using MCL to extract clusters from networks, p 281–295. *In* Helden J, Toussaint A, Thieffry D (ed), Bacterial molecular networks. Springer, New York, NY.

125. Garber AI, Nealson KH, Okamoto A, McAllister SM, Chan CS, Barco RA, Merino N. 2020. FeGenie: a comprehensive tool for the identification of iron genes and iron gene neighborhoods in genome and metagenome assemblies. Front Microbiol 11:37. https://doi.org/10.3389/fmicb.2020.00037
126. McAllister SM. 2016. Heme counter. GitHub repository
127. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FSL. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26:1608–1615. https://doi.org/10.1093/bioinformatics/btq249
128. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, von Heijne G, Nielsen H. 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol 40:1023–1025. https://doi.org/10.1038/s41587-021-01156-3
129. Kall L, Krogh A, Sonnhammer ELL. 2007. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. Nucleic Acids Res 35:W429–32. https://doi.org/10.1093/nar/gkm256
130. Käll L, Krogh A, Sonnhammer ELL. 2004. A combined transmembrane topology and signal peptide prediction method. J Mol Biol 338:1027–1036. https://doi.org/10.1016/j.jmb.2004.03.016
131. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. 2018. A completely reimplemented mpi bioinformatics toolkit with a new hhpred server at its core. J Mol Biol 430:2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007
132. Gabler F, Nam S, Till S, Mirdita M, Steinegger M, Söding J, Lupas AN, Alva V. 2020. Protein sequence analysis using the MPI bioinformatics toolkit. Curr Protoc Bioinform 72:e108. https://doi.org/10.1002/cpbi.108
133. Baker IR, Conley BE, Gralnick JA, Girguis PR. 2022. Evidence for horizontal and vertical transmission of Mtr-mediated extracellular electron transfer among the bacteria. mBio 13:e02904-21. https://doi.org/10.1128/mbio.02904-21
134. Wickham H. 2016. ggplot2: elegant graphics for data analysis. 2nd. Springer International Publishing, New York, NY.
135. Gu Z, Eils R, Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32:2847–2849. https://doi.org/10.1093/bioinformatics/btw313
136. Kolde R. 2025. Pheatmap: pretty heatmaps. R package version 1.0.13. https://github.com/raivokolde/pheatmap.
137. Galili T. 2015. Dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics 31:3718–3720. https://doi.org/10.1093/bioinformatics/btv428