International Journal of Systematic and Evolutionary Microbiology

Evaluation of the genus Thiothrix Winogradsky 1888 (Approved Lists 1980) emend. Aruga et al. 2002: reclassification of Thiothrix disciformis to Thiolinea disciformis gen. nov., comb. nov., and of Thiothrix flexilis to Thiofilum flexile gen. nov., comb nov., with emended description of Thiothrix.

--Manuscript Draft--

Manuscript Number:	IJSEM-D-18-00034R3
Full Title:	Evaluation of the genus Thiothrix Winogradsky 1888 (Approved Lists 1980) emend. Aruga et al. 2002: reclassification of Thiothrix disciformis to Thiolinea disciformis gen. nov., comb. nov., and of Thiothrix flexilis to Thiofilum flexile gen. nov., comb nov., with emended description of Thiothrix.
Article Type:	Taxonomic Description
Section/Category:	New taxa - Proteobacteria
Keywords:	Thiothrix; Thiotrichales; Thiolinea; Thiofilum; filamentous Bacteria; Gammaproteobacteria
Corresponding Author:	Rich Boden, Ph.D B.Sc (Lond.) PGCert University of Plymouth Plymouth, Devon UNITED KINGDOM
First Author:	Rich Boden, Ph.D B.Sc (Lond.) PGCert
Order of Authors:	Rich Boden, Ph.D B.Sc (Lond.) PGCert
	Kathleen M Scott
Manuscript Region of Origin:	UNITED KINGDOM
Abstract:	Thiothrix is the type genus of the Thiotrichaceae in the Thiotrichales of the Gammaproteobacteria, comprising nine species of sulfur-oxidising filamentous Bacteria, which are variously autotrophic, heterotrophic or have mixed metabolic modes. Within the genus, 4 species show 16S rRNA gene identities lower the Yarza threshold for the rank of genus (94.5 %) - Thiothrix disciformis, Thiothrix flexilis, Thiothrix defluvii and Thiothrix eikelboomii - as they show no affiliation to extant genera, a polyphasic study was undertaken including biochemical, physiological and genomic properties and phylogeny based on the 16S rRNA gene (rrs), recombination protein A (RecA), polynucleotide nucleotidyltransferase (Pnp), translation initiation factor IF-2 (InfB), glyceraldehyde-3-phosphate dehydrogenase (GapA), glutaminyl-tRNA synthetase (GlnS), elongation factor EF-G (FusA) and concatamers of 53 ribosomal proteins encoded by rps, rpl and rpm operons, all of which support the reclassification of these species. We thus propose Thiolinea gen. nov. and Thiofilum gen. nov. for which the type species are Thiolinea disciformis gen. nov., comb. nov. and Thiofilum flexile gen. nov., comb. nov. We also propose that these genera are each circumscribed into novel families Thiolinaceae fam. nov. and Thiofilaceae fam. nov., and that Leucothrix and Cocleimonas are circumscribed into Leucotrichaceae fam. nov. and provide emended descriptions of Thiothrix and Thiotrichaceae.

Evaluation of the genus *Thiothrix* Winogradsky 1888 (Approved Lists 1980) emend. Aruga *et al.* 2002: reclassification of *Thiothrix disciformis* to *Thiolinea disciformis* gen. nov., comb. nov., and of *Thiothrix flexilis* to *Thiofilum flexile* gen. nov., comb nov., with emended description of *Thiothrix*.

Rich Boden[1,2*] and Kathleen M. Scott[3]

1. School of Biological and Marine Sciences, University of Plymouth, Plymouth, UK.

2. Sustainable Earth Institute, University of Plymouth, Plymouth, UK.

3. Department of Integrative Biology, University of South Florida, Tampa, Florida, USA.

* Corresponding author: rich.boden@plymouth.ac.uk

Running title: Reclassification of *Thiothrix* spp. to *Thiolinea* gen. nov. and *Thiofilum* gen. nov.

KEYWORDS: *Thiothrix, Thiotrichales, Thiolinea, Thiofilum*, filamentous *Bacteria*, *Gammaproteobacteria*

NON-STANDARD ABBREVIATIONS: *Tla.*, *Thiolinea*; *Tfm.*, *Thiofilum*; Δp, proton motive force; rMLST, ribosomal mutli-locus sequence typing; RuBisCO, ribulose 1,5-bisphosphate carboxylase/oxygenase.

Abstract

Thiothrix is the type genus of the *Thiotrichaceae* in the *Thiotrichales* of the *Gammaproteobacteria*, comprising nine species of sulfur-oxidising filamentous *Bacteria*, which are variously autotrophic, heterotrophic or have mixed metabolic modes. Within the genus, 4 species show 16S rRNA gene identities lower the Yarza threshold for the rank of genus (94.5 %) – *Thiothrix disciformis*, *Thiothrix flexilis*, *Thiothrix defluvii* and *Thiothrix eikelboomii* – as they show no affiliation to extant genera, a polyphasic study was undertaken including biochemical, physiological and genomic properties and phylogeny based on the 16S rRNA gene (*rrs*), recombination protein A (RecA), polynucleotide nucleotidyltransferase (Pnp), translation initiation factor IF-2 (InfB), glyceraldehyde-3-phosphate dehydrogenase (GapA), glutaminyl-tRNA synthetase (GlnS), elongation factor EF-G (FusA) and concatamers of 53 ribosomal proteins encoded by *rps*, *rpl* and *rpm* operons, all of which support the reclassification of these species. We thus propose *Thiolinea* gen. nov. and *Thiofilum* gen. nov. for which the type species are *Thiolinea disciformis* gen. nov., comb. nov. and *Thiofilum flexile* gen. nov., comb. nov. We also propose that these genera are each circumscribed into novel families *Thiolinaceae* fam. nov. and *Thiofilaceae* fam. nov., and that *Leucothrix* and *Cocleimonas* are circumscribed into *Leucotrichaceae* fam. nov. and provide emended descriptions of *Thiothrix* and *Thiotrichaceae*.

The genus *Thiothrix* Winogradsky 1888 (Approved Lists 1980) emend. Aruga *et al*. 2002 [1-3]) is the type genus of the family *Thiotrichaceae*, in turn the type family of the order *Thiothrichales* of the *Gammaproteobacteria*. It comprises filamentous sulfur-oxidising *Bacteria* that form numerous structures including rosettes and holdfasts and which grow chemolithoautotrophically, heterotrophically and/or mixotrophically or possibly chemolithoheterotrophically, with considerable debate regarding their exact metabolic mode dating back to Winogradsky's 1888 study. In addition to the 9 species with validly published names, a large number of names have been mentioned without valid or effective publication. For example, a further 10 species names appear in the *Index Bergeyana* [4] and a further name appears in the *Supplement to the Index Bergeyana* [5]; however, the strains of all 11 of these 'species' have been lost, and as they do not appear on the *Approved Lists* in any case, their names have no standing. Whilst some common features are shared across the 9 species with validly published names, there are a number of considerable differences, which has led to previous studies investigating the true nature of the genus, often concluding that in spite of an enormous phylogenetic and physiological diversity, that the genus should remain as-is [6]. Here we present a polyphasic analysis including phylogenetic and genomic evidence, and thus the case for the reclassification of members of *Thiothrix* into two novel genera, *Thiolinea* gen. nov. and *Thiofilum* gen. nov.

For the sake of clarity and to avoid confusion, we must define some metabolic modes used herein that are sometimes misinterpreted or misused in the literature. We herein use "mixotrophy" to refer to autotrophy and heterotrophy simultaneously occurring in an organism, typified by assimilation of carbon dioxide into biomass at the same time as the assimilation of *e.g.* a sugar or carboxylate, and is accompanied by the oxidation of an inorganic electron donor in order to generate proton motive force (Δp) which is used to yield ATP and NAD(P)H which are used to fuel the assimilation of carbon dioxide. We use

63 "chemolithoheterotrophy" to refer to the oxidation of an inorganic electron donor as a source
64 of auxiliary energy (ATP and/or NAD(P)H) during heterotrophic growth, which is typified by
65 a complete lack of carbon dioxide assimilation (*pace* anapleurotic reactions). The umbrella
66 term for general mixed metabolisms (*i.e.* chemolithoheterotrophy, mixotrophy *etc*) that we
67 use herein is "mixed metabolic modes" – some studies refer to general mixed metabolism as
68 "mixotrophy", though this is really incorrect and breeds significant confusion.

69 The 9 species with validly published names are (type species first) *Thiothrix nivea*
70 (Rabenhorst 1865) Winogradsky 1888 (Approved Lists 1980), emend. Larkin and
71 Shinabarger 1983 [1-2, 7]); *Thiothrix caldifontis* (Chernousova *et al*., 2009 [8]), *Thiothrix*
72 *defluvii* (Howarth *et al*., 1999 [9]), *Thiothrix disciformis* (Aruga *et al*., 2002 [3]), *Thiothrix*
73 *eikelboomii* (Howarth *et al*., 1999 [9]), *Thiothrix flexilis* (Aruga *et al*., 2002 [3]), *Thiothrix*
74 *fructosivorans* (Howarth *et al*., 1999 [9]), *Thiothrix lacustris* (Chernousova *et al*., 2009 [8])
75 and *Thiothrix unzii* (Howarth *et al*., 1999 [9]). They have been isolated principally from
76 activated sludge and sulfidic lakes or groundwaters (Table 1), and other strains have been
77 isolated from diverse locations including as symbionts of the cave amphipod *Niphargus ictus*
78 G. Karaman [10] that inhabits sulfidic waters, and the marine amphipod *Urothoe poseidonis*
79 Reibish [11]. It is worth noting that *Thiothrix* spp. have also been observed attached to other
80 members of the *Eukarya* such as *Cladophora* spp. Kütz, *Vaucheria* spp. A. P. de Candolle
81 [12], and *Drunella grandis* Eaton [13], and to members of the "*Cyanobacteria*" [12], without
82 determination of their status as symbionts or commensals. It is also worth noting that
83 intracellular parasitic *Bacteria* have been observed within *Thiothrix* spp. [13] grown *in situ*,
84 and which could give rise to apparent phenotypic or chemotaxonomic variation of cultures *in*
85 *vitro*.

86 Of the 9 species with validly published names, many are not possible to reclassify owing to
87 their absence from public culture collections, the current deposit *statūs* being: *T. caldifontis*

(DSMZ; lost by VKM); *T. defluvii* (lost, as far as we can tell – never deposited); *T. disciformis* (DSMZ, JCM); *T. eikelboomii* (ATCC); *T. flexilis* (DSMZ, JCM); *T. fructosivorans* (ATCC); *T. lacustris* (DSMZ; lost by VKM); *T. nivea* (ATCC, DSMZ), and *T. unzii* (ATCC). Thus, this study can only focus on *T. nivea*, *T. flexilis* and *T. disciformis* by way of making any formal reclassification, since type strains of each new combination must be available in two public service collections per Rules 27 and 30 of the *International Code of Nomenclature of Prokaryotes* (hereafter "the *Code*"). As such, our focus is on these taxa but we will note the positions of those species than cannot be formally reclassified at this time to facilitate further work, whilst not formally declaring new combinations, merely as "incidental mentions" per Rule 28*b* of the *Code*.

Of the other members of the *Thiotrichaceae*, the genera *Thiomargarita*, *Achromatium*, *Thiobacterium*, *Thioploca* and *Thiospira* do not have cultures available, thus are not included in our analyses. Although the genus *Beggiatoa* was included in this family in recent editions of *Bergey's Manual*, there is very low 16S rRNA (*rrs*) gene identity of members of this genus to the rest of the *Thiotrichaceae* – it was recently found to cluster entirely separately from the rest of this family [49] and the 16S rRNA gene identity of *Beggiatoa alba* B18LD^{T} (type species) to that of *Thiothrix nivea* JP2^{T} (type species of type genus of *Thiotrichaceae*) is 85.4 %, which falls below the Yarza median for members of the same class (86.35 %, the Yarza parameters being explained in the next two paragraphs), let alone the same family. As such, we have not considered *Beggiatoa* spp. herein.

Historically, many taxonomic studies have divided *Thiothrix* spp. into the '*T. nivea* group' and the 'Eikelboom type 021N group' and so on - we have not followed this convention as it somewhat clouds the phylogenetic picture. Phylogenetic trees on the basis of the 16S rRNA (*rrs*) gene from *Thiothrix* spp. and the only other canonical *Thiotrichaceae* that have 16S rRNA gene data available - *Leucothrix* spp. and *Cocleimonas flava* - are given in Figure 1,

with full details of construction given in the figure legend. It can be seen from Figure 1 that *Thiothrix* spp. cluster into 3 distinct lineages on the basis of gross tree topology, which we have termed Clade X (*T. nivea*, *T. fructosivorans*, *T. caldifontis*, *T. lacustris* and *T. unzii*), Clade Y (*T. disciformis* and *T. eikelboomii*) and Clade Z (*T. defluvii* and *T. flexilis*). Within Clade X, the 16S rRNA gene identities to *T. nivea* are 94.4 – 95.2 %. A threshold 16S gene identity value of 94.5 % for genus level relationships was documented by Yarza *et al.* [14], which we have termed 'the Yarza threshold' [15] and have applied to clarifying the systematics of other *Thiotrichales* [16-17] and other *Gammaproteobacteria* [18], as well as the *Betaproteobacteria* and the *Hydrogenophilalia* [15]. On the basis of the Yarza threshold, the species in Clade X mostly constitute a single genus, within which *T. nivea* takes priority as the type species, however *T. fructosivorans* stands at 94.4 % 16S rRNA gene identity to *T. nivea* and thus *could* constitute a separate genus on the basis of the Yarza threshold alone. Since it does not form a separate line of descent from the rest of Clade X, we have retained it as a member of the genus *Thiothrix*, though it should be noted that it exhibits growth on fructose and catalase enzyme activity, both of which are not found in any other Clade X organisms. Clade Y members have 16S rRNA gene identities of 90.9 to 92.3 % to *T. nivea*, indicating that they, on the basis of the Yarza threshold, are not in the same genus as *T. nivea*, but since they have 16S rRNA gene identities to one another of 94.8 %, Clade Y species are both within one genus. Clade Z organisms have identities to *T. nivea* of 89.2 – 89.5 %, indicating that they are not members of the same genus either, but since they have identities to one another of 96.9 %, both Clade Z species would form one genus. All clades are sufficiently distant from *L. mucor* and *C. flava* to not belong to those genera (shown in Table 1) – though it is worth noting that all Clade X to Z organisms have the helix 18 deletion of the 16S rRNA gene, but *L. mucor* and *C. flava* do not, confirming common ancestry of these Clades as distinct from *Leucothrix* spp. and *Cocleimonas* spp.

In terms of higher taxa systematics and 16S rRNA gene identities, we have previously defined a series of 'Yarza medians' as a system of cut-off values for family (92.25%), order (89.20%), class (86.35%) and phylum (83.68%) based on values determined by Yarza *et al*. [14], and which we have, again, applied across the *Gammaproteobacteria* [15-16, 18], *Betaproteobacteria* and *Hydrogenophilalia* [15]. Consideration of these parameters with gene identities given in the previous paragraph and in Table 1 would indicate that Clade X and Clade Y all belong to the same order, but not to the same family, with Clade Y forming a distinct family from that in which Clade X falls (since the latter is *Thiothrix sensu stricto*, *Thiotrichaceae* is retained as the name for this family); as such, Clade Y will form the type genus of a novel family within the *Thiotrichales*. Clade Z also constitutes a novel family, but falls into a separate order of the *Gammaproteobacteria*, distinct from the *Thiotrichales* but we will not consider this further within this study since a much larger issue must be considered at the same time, *viz*. the systematics of the higher taxa of the *Gammaproteobacteria*. Since both *Leucothrix* and *Cocleimonas* are sufficiently distant from *T. nivea* to potentially constitute a separate class, and further work must be undertaken to delineate this. We will, however, define a new family for Clade Z, with position *incertae sedis* pending further work – given the size of the *Gammaproteobacteria* and the number of species therein and the necessity for rMLST-type approaches with long concatamers, it may not be possible to ascertain this accurately based on the current algorithm, software and hardware limitations, even within high-performance computing. We can, however, determine that *Leucothrix* and *Colceimonas* do not fall within the *Thiothrichaceae* or our two novel families, and so we circumscribe them on the basis of their 16S rRNA gene identities as a novel family, *Leucotrichaceae* fam. nov., but we place this *incertae sedis* pending further work, whilst rejecting it from the *Thiothrichales sensu stricto*.

For those taxa with sequenced genomes, phylogenetic studies were performed on the basis of a range of 'housekeeping' genes other than the 16S rRNA gene (above) commonly used in taxonomic studies (Figure 2), selected on the basis of their involvement in a diversity of pathways and systems, namely those encoding polyribonucleotide nucleotidyltransferase (EC 2.7.7.8, *pnp*); translation initiation factor IF-2 (EC 3.6.5.3, *infB*); glyceraldehyde-3-phosphate dehydrogenase (NAD^+, EC 1.2.1.12, *gapA*); glutaminyl-tRNA synthetase (EC 6.3.5.7, *glnS*) and elongation factor EF-G (EC 3.6.5.3, *fusA*) and recombination protein A (*recA*). Gene sequences were obtained from public databases of genes from *Thiothrix* and *Leucothrix* spp., and *Thiomicrospira pelophila* DSM 1534^T (also from the *Thiotrichales* of the *Gammaproteobacteria*), genes from which were used as the outgroup. We employed *Thiothrix* sp. AAV1 as a proxy for *T. fructosivorans* (16S rRNA gene identity 99.9% and 16S rRNA gene tree position shown in Supplementary Figure S1). We have provided full details of tree construction in the figure legends for these trees. A further analysis was performed using amino acyl sequences derived from genes encoding ribosomal proteins in the form of 53-gene concatamers, obtained from databases of the ribosomal multilocus sequence typing (rMLST) online platform [19], again, with full details provided in the legend (Figure 3).

Across almost all phylogenetic analyses of amino acyl sequences derived from housekeeping genes performed, *Thiothrix* spp. fell into the same three clades per the 16S rRNA gene analyses, branching from highly supported nodes (Figure 2), *pace* RecA, the tree for which did not have well-supported nodes. The GapA and GlnS trees did not show full resolution of Clade Y and Clade Z, but this was clear across all of the other trees. In spite of these minor variations in single gene trees, the concatamer analysis of 53 ribosomal protein-coding genes gave the same overall topography as the 16S rRNA gene tree (Figure 3), with three well-divided clades originating from well-supported nodes with bootstrap values of 99-100 % over

5,000 replicate reconstructions. These data support the division of *Thiothrix* into 3 genera, in line with the 16S rRNA gene identities across the genus.

Clade X (*Thiothrix* spp. *sensu stricto*) organisms are united by chemolithoautotrophic or seemingly obligately mixotrophic growth and have cells 0.7-2.5 μm wide and 0.7-6.5 μm long. They are catalase negative (*pace* a weak reaction in *T. fructosivorans*) but oxidase positive. They form filaments which do not contain knots or branches and which almost all (*pace T. unzii*) have polysaccharide sheaths. Salt is not required for growth and temperature optima are 24-30 °C, with pH optima of 7.0-8.0 and maxima of 8.2-8.6. Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO, EC 4.1.1.39) from the Calvin-Benson-Bassham cycle is found in two key forms in the '*Proteobacteria*' – form I (composed of 8 CbbL and 8 CbbS subunits) and form II (a dimer of CbbM, which is evolutionarily related to CbbL, [20-21]). The form I RuBisCOs present in the organisms considered here can be further divided into forms IAq (cytoplasmic) and IAc (carboxysomal). In Clade X, genes encoding form IAq and form II RuBisCO were found in all strains for which a genome sequence exists, with *T. nivea* and *T. caldifontis* also having form IAc. The form IAc genes in *T. nivea* and *T. caldifontis* are collocated with genes encoding carboxysome shell proteins, carboxysomal carbonic anhydrase and the RuBisCO assembly factor [22]. For both species, carboxysome genes are followed by genes homologous to those encoding inorganic carbon transporters, similar to observations in other chemolithoautotrophs (*e.g.* [16] and [47]). Based on the presence of genes encoding carboxysome components, and the (likely) inorganic carbon transporters, these two species could express mechanisms for the concentration of carbon dioxide, allowing growth under low carbon dioxide partial pressures.

All *Thiothrix sensu lato* oxidise thiosulfate to elementary sulfur and [then] sulfate, possessing genes encoding elements of the canonical Kelly-Friedrich pathway ('sox pathway') of sulfur oxidation (found complete in *Paracoccus* spp. from the *Alphaproteobateria* and in both

complete and partial forms across the *Gammaproteobacteria*, *Betaproteobacteria* and *Acidithiobacillia*, [23]), but in different degrees of fragmentation around the genome (Table 1), which we have also observed in other members of the *Thiotrichales* (Scott *et al.*, under review). Sulfide dehydrogenase (flavocytochrome *c*, EC 1.8.2.3, Fcc – used in the oxidation of sulfide to elementary sulfur) genes and sulfide:quinone reductase (EC 1.8.5.4, Sqr – used in the oxidation of sulfide to polysulfide, in the formation of elementary sulfur granules) genes were also present in all *Thiothrix sensu lato*, correlating with these species accumulating elementary sulfur granules during growth. Additionally, all Clade X organisms have adenosine 5'-phosphosulfate reductase (EC 1.8.4.8, AprAB), which is found in sulfate-reducing anaerobes but also in aerobic sulfur oxidisers, operating to oxidise sulfite to sulfate. Organisms across all three clades also have the reverse form of dissimilatory (bi)sulfite reductase (EC 1.8.99.5, DsrAB), which oxidises sulfide to sulfite in aerobes. The Clade X enzyme data are similar to the findings in "*Thiothrix ramosa*", which affiliates to this clade [40].

All members of all clades have a full complement of genes encoding enzymes of Krebs' cycle, including the E3 subunit of the 2-oxoglutarate dehydrogenase complex – dihydrolipoamide dehydrogenase (*lpdA*, EC 1.8.1.4) – which is either missing at gene level or present but does not result in *in vivo* enzyme activity of the complex in most obligate autotrophs, cutting Krebs' cycle into Smith's biosynthetic horseshoe [16, 24-27] and preventing heterotrophic growth since the cycle cannot be for energetic purposes, only for biosynthesis. It is not known if this complex is expressed in *Thiothrix* spp., but the potential for heterotrophic growth is at least present. Genes encoding the enzymes of the glyoxylate shunt were found in all *Thiothrix sensu lato*, permitting growth on acetate and other C_2 compounds, as sole carbon sources, in principle [28], but genes encoding enzymes of C_1 compound metabolism (*viz*. methanol dehydrogenases (*mxaF* and *xoxF* EC 1.1.2.7 [29]),

Quayle pathway (ribulose monophosphate pathway [30]) and serine cycle [31]) were not found. In terms of glycolytic (or gluconeogenic) pathways, the full complement of genes encoding enzymes of the Entner-Doudoroff [32], Embden-Meyerhoff-Parnas and pentose phosphate [33] was found in all clades examined.

The complete transaldolase variant Calvin-Benson-Bassham cycle enzymes could be accounted for in all three clades with the exception of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO, EC 4.1.1.39), which was not found in Clade Z (or in *L. mucor*); all Clade Y species have RuBisCO form IAq (a cytoplasmic form, [20]) and some also have form IAc (carboxysomal), whereas Clade X have either form IAq or form II (also cytoplasmic), and no carboxysomal forms. These data would imply that Clade Z are obligate heterotrophs and whilst they may oxidise thiosulfate to sulfate to provide Δp and thus synthesise ATP to permit chemolithoheterotrophic growth (per some other members of the *Thiotrichales*, *e.g. Methylophaga thiooxydans* [34-35]). They thus do not fix carbon dioxide into biomass *sensu stricto*, but will no doubt assimilate it to a minor degree *via* anaplerotic fixation (*e.g.* pyruvate carboxylase, EC 6.4.1.1) and cannot be considered autotrophic. These data also imply that Clade Y species have diverse means for carbon dioxide fixation, both evolved for life at higher carbon dioxide partial pressures in which oxygen is present, without having any means to concentrate carbon dioxide or live at low carbon dioxide partial pressures. Clade X have the further ability to live at both much higher carbon dioxide partial pressures using form II RuBisCO or much lower partial pressures using the carboxysomal form, IAc, where oxygen is at lower partial pressures [20] – Clade X has been found in sulfidic wells and springs (karstic and thus high in dissolved bicarbonate), and on amphiphods living in karstic ecosystems [10], which would be in keeping with the presence of form II RuBisCO, which operates more effectively in high carbon dioxide partial pressures than form IAc, which is optimised for lower partial pressures. There is a paucity of ecological

data on Clade Y since they have only been isolated from activated sludge thus far, but given that organic matter decomposes with rapidity in activated sludge, we anticipate that the carbon dioxide partial pressures therein to be high, thus precluding the necessity of carbon dioxide concentrating mechanisms in the form of carboxysomes.

Substrate profiles for heterotrophic (or mixotrophic) growth are quite restricted in Clade X - with hexoses and disaccharides not being used (*pace* fructose and sucrose in *T. fructosivorans*), which is unusual given Krebs' cycle is complete at genome level and there are no lesions of glycolysis *pace* the pentose phosphate pathway in two species. Acetate is the only substrate used universally in Clade X. Whilst Clades Y and Z can grow on sugar alcohols (*viz*. mannitol and glycerol), Clade X organisms cannot – a distinguishing and useful diagnostic feature. It is worth noting that the growth of *T. fructosivorans* on fructose can probably be ascribed to the presence of the fructose transport system genes in the genome sequence, which are absent from all other species examined. *T. nivea* and *T. unzii* both hydrolyse gelatine and starch but the rest of the clade cannot. All species of all clades can use ammonium or nitrate as their sole nitrogen source and some species can use nitrite. *T. caldifontis* is the only *Thiothrix* sp. for which diazotrophy has been reported, which is supported by the presence of *nif* genes encoding the molybdenum-iron nitrogenase (EC 1.19.6.1), which were also found in *T. nivea*, but not in any other species examined. GC fractions in Clade X are 49.3 – 52.0 mol% *in vitro*, or 50.5 – 54.9 mol% *in silico*.

All genomes examined for Clades X, Y and Z contain genes for the cbb_3 cytochrome *c* oxidase and *bd*-I quinol oxidase, both of which are also common to *L. mucor*, which additionally has the aa_3 cytochrome *c* oxidase, but this is not found in Clades X, Y or Z. The cbb_3 oxidase has a high oxygen affinity in general and this would be in keeping with the microaerophilic nature of many organisms in these clades, growing best at low oxygen partial pressures, and the absence of the aa_3-type oxidase would be in keeping with very weak

286 growth under high oxygen partial pressures. As the *bd*-I ubiquinol oxidase has been
287 implicated in promoting survival to oxidative stress, this is also in keeping with the lifestyle
288 of these organisms [36].

289 For the sake of completeness and to benefit those reading ecological studies or literature on
290 *Thiothrix* species without validly published names, consideration was given to the various
291 "*Thiothrix*" 16S rRNA gene sequences curated into the GenBank™ and IMG databases, a
292 phylogenetic analysis of which is given in Supplementary Figure S1. Of these, two groups of
293 sequence are worthy of consideration – the first is "*Thiothrix ramosa*", a well-characterised
294 isolate from a sulfidic spring in what is now Latvia, subjected to extensive physiological and
295 phylogenetic characterisation in the 1990s [37-41] – probably the best studied *Thiothrix*
296 isolate in terms of sulfur physiology. It was identified as related to *T. nivea*, though it is now
297 lost from culture (Elena Odintsova, *personal communication*) and the name was never validly
298 or effectively published. Our analysis confirms that "*T. ramosa*" would still fall within the
299 genus *Thiothrix sensu stricto*, with the closest relative being *T. lacustris* with a 16S rRNA
300 gene identity of 97.7 % - *T. lacustris* also being obtained from a sulfidic freshwater
301 ecosystem of similar latitude, but lacking the characteristic branching observed in "*T.*
302 *ramosa*", which is absent from other *Thiothrix* spp. "*T. ramosa*" was shown to grow on a
303 range of energy sources, including substituted thiophenes (the metabolism of which is very
304 poorly understood), thus *T. lacustris* DSM 21227^{T} (genome sequence available) may prove a
305 valuable model organism for these metabolic pathways if it indeed shares this trait. The
306 second group of note are the large number of *Thiothrix*-affiliated cloned 16S rRNA and
307 rRNA gene sequences obtained from *Niphargus ictus* specimens from the Frassassi Caves in
308 Italy, which have sequences deposited in the GenBank™ database and most of these probably
309 represent close relatives of *T. fructosivorans*, the type strain of which was isolated from
310 activated sludge in the USA.

We conclude that extant *Thiothrix* spp. in culture actually represent three distinct genera, on the basis of the clades defined in this study, the properties of which are curated at genus level in Table 2, with comparison to other genera of the *Thiotrichales*. Since Clade X contains the type species (*T. nivea*), the genus that it circumscribes must retain the name *Thiothrix*, in accordance with Rule 39*a* of the *Code*. As such, *T. nivea*, *T. unzii*, *T. caldifontis*, *T. fructosivorans*, *T. lacustris* are considered *bona fide* members of the genus *Thiothrix*. We propose that Clade Y circumscribes a novel genus for which we propose the name *Thiofilum* gen. nov., for which the type species is *Thiofilum flexile* gen. nov., comb. nov. *T. defluvii* would be circumscribed into *Thiofilum* too, but the lack of available strains means we cannot make this formal reclassification and "*Thiofilum defluvii*" will have to be subject of future work, if ever *T. defluvii* cultures are found and deposited accordingly. Clade Z circumscribes another new genus, for which we propose the name *Thiolinea* gen. nov., for which the type species is *Thiolinea disciformis* gen. nov., comb. nov. Again, *T. eikelboomii* undoubtedly would be circumscribed within *Thiolinea*, the lack of strains deposited in international collections makes it impossible to formally reclassify it at this time, thus "*Thiolinea eikelboomii*" will be the subject of future work if and when the strain is made available to the public in 2 or more international service collections. Obviously our mention of "*Tfs. defluvii*" and *"Tla. eikelboomii*" are merely 'incidental mentions' under Rule 28*b* of the *Code* and do not constitute or attempt to constitute any formal prioritisation or laying any claim to future publication of these two names.

Description of *Leucotrichaceae* fam. nov.

Leucotrichaceae (Leu.co.tri.cha.ce'ae. N.L. fem. n. *Leucothrix*, type genus; *-aceae* suffix to denote family; N.L. fem. pl. n. *Leucotrichaceae*, the *Leucothrix* family).

This family is circumscribed on the basis of 16S rRNA gene sequences and comprises the genera *Leucothrix* (type genus) and *Cocleimonas*. Ubiquinone-8 (UQ-8) is the dominant respiratory quinone. Dominant fatty acids are palmitoleic ($C_{16:1}$), palmitic ($C_{16:0}$) and vaccenic ($C_{18:1}$) acids. G+C fractions are 43.4 to 49.5 mol%. Contain aa_3 and cbb_3 type cytochrome *c* oxidases and the *bd*-I ubiquinol oxidase. Found in seawater and on/in marine green algae and marine gastropods. Falls within the *Gammaproteobacteria*, with the position at order level *incertae sedis* at present.

Type genus: *Leucothrix*.

Description of *Thiolinea* gen. nov.

Thiolinea (Thi.o.li'ne.a. Gr. neut. n. *theion* sulfur, brimstone, L. transliteration *thium*, sulfur; L. fem. n. *linea*, a string, a cord; N.L. fem. n. *Thiolinea*, sulfur string)

Heterotrophic. Gram-stain-negative cells, forming filaments with holdfasts and that may exhibit branching. Rosettes are formed and filaments may contain knots, but do not have a polysaccharide sheath. Do not form endospores or exospores. Thiosulfate is oxidised during heterotrophic growth but the energetic effects have not been examined. Have RuBisCO genes. Nitrate is not reduced. The 16S rRNA genes have about 92.5 – 93.5 % sequence identity to that of *Thiothrix nivea* JP2^T. Ubiquinone-8 (UQ-8) is the dominant respiratory quinone. Major fatty acids are palmitoleic ($C_{16:1}$), palmitic ($C_{16:0}$) and vaccenic ($C_{18:1}$) acids - the dominant unsaturated fatty acid is vaccenic acid. Members of the *Thiothrichaceae* in the *Thiotrichales* of the *Gammaproteobacteria.*

Type species is *Thiolinea disciformis*, isolated from activated sludge suffering from bulking in Japan).

Description of *Thiolinea disciformis* gen. nov. comb. nov.

Thiolinea disciformis (dis.ci.for'mis. L. masc. n. *discus*, a disc (from Gr. masc. n. *diskos*); L. fem. adj. suffix *–formis* (from L. fem. n. *forma*, figure, shape or appearance) in the shape of; N.L. fem. adj. *disciformis*, disc-shaped, after the main cell morphology).

Type species of *Thiolinea.* Basonym: *Thiothrix disciformis* Aruga *et al*. 2002.

Gram-stain negative. Sugar-grown cells are oxidase-positive and violently catalase-positive. Heterotrophic, and oxidises thiosulfate or sulfide during heterotrophic growth. Forms fingerprint-like colonies on glucose-acetate agar. Cells are mostly rod shaped but morphology is variable, particularly with regard to length. Most cells are discoid or ovoid and 1.2 – 3.0 µm diameter and 0.5 – 3.0 µm in length, forming slightly bent, sheath-free filaments greater than 0.5 mm in length that can reach several millimetres. Some cells in filaments are elongate or swollen. Septa between cells of filaments are clearly defined. Volutin (polyphosphate) granules are absent. Sudanophilic (lipid) granules are present. Temperature range of growth is 14-32 °C, with optimum growth 25-30 °C and no growth at 4 °C or at 37 °C. Growth is fully inhibited by 85 mM (0.5 % *w/v*) sodium chloride. Elementary sulfur globules are deposited within the invaginated inner membrane when grown in the presence of thiosulfate or sulfide, which are oxidised to sulfate. Carbon sources for heterotrophic growth include hexose sugars (glucose, fructose, mannose), disaccharides (sucrose, maltose, trehalose), intermediates of Krebs' cycle (succinate, malate, citrate), carboxylates (pyruvate, acetate, butyrate, hydroxybutyrate), amino acids (glutamate, aspartate, alanine), alcohols (mannitol, glycerol), but not benzoate, xylose, erythritol, galactose, lactose, melibiose, raffinoise, arabinose, lactate, ethanol, propan-1-ol, sorbitol, formate, gelatine or starch. Does not require vitamins for growth. Does not reduce nitrate. Has genes encoding cbb_3 cytochrome *c* and *bd*-I ubiquinol type terminal oxidases, and complete Krebs' cycle plus the glyoxylate shunt. Has genes encoding form II RuBisCO and does not have carboxysomal genes. G+C fraction is 45.1 mol% from the genome sequence. Dominant respiratory quinone is ubiquinone-8.

Dominant fatty acids in glucose-and-acetate grown cells are palmitoleic ($C_{16:1}$), palmitic ($C_{16:0}$) and vaccenic ($C_{18:1}$) acids.

Type strain is B3-1^T (= DSM 14473^T = JCM 11364^T), isolated from activated sludge suffering from bulking (Japan).

Description of *Thiolineaceae* fam. nov.

Thiolineaceae (Thi.o.li.ne.a.ce'ae. N.L. fem. n. *Thiolinea*, type genus; *-aceae* suffix to denote family; N.L. fem. pl. n. *Thiolineaceae*, the *Thiolinea* family)

This family is circumscribed on the basis of 16S rRNA gene sequences and comprises the genus *Thiolinea* (type genus). Ubiquinone-8 (UQ-8) is the dominant respiratory quinone. Dominant fatty acids are palmitoleic ($C_{16:1}$), palmitic ($C_{16:0}$) and vaccenic ($C_{18:1}$) acids. G+C fractions are 45.1 to 46.3 mol% on the basis of genome sequence data. Falls within the order *Thiotrichales* of the *Gammaproteobacteria*.

Type genus: *Thiolinea*

Description of *Thiofilum* gen. nov.

Thiofilum (Thi.o.fi'lum. Gr. neut. n. *theion* sulfur, brimstone, L. transliteration *thium*, sulfur; L. neut. n. *filum*, filament, thread; N.L. neut. n. *Thiofilum*, sulfur filament).

Members of the *Thiothrichaceae* in the *Thiotrichales* of the *Gammaproteobacteria.* Obligately heterotrophic. Gram-stain-negative filamentous bacteria. Filaments do not have polysaccharide sheaths but form holdfasts and rosettes and may become knotted, but are not branched. Do not form endospores or exospores. Thiosulfate can be oxidised to provide Δp, which can be used to generate ATP for chemolithoheterotrophic growth. Nitrate is reduced to nitrite. The 16S rRNA genes have about 91 – 92 % sequence identity to that of *Thiothrix*

nivea JP2^{T}. Ubiquinone-8 (UQ-8) is the dominant respiratory quinone. Major fatty acids are palmitoleic ($C_{16:1}$), palmitic ($C_{16:0}$) and vaccenic ($C_{18:1}$) acids – the dominant unsaturated fatty acid is vaccenic acid.

Type species is *Thiofilum flexile*, isolated from activated sludge suffering from bulking in Japan).

Description of *Thiofilum flexile* gen. nov. comb. nov.

Thiofilum flexile (fle'xi.le. L. neut. adj. *flexile*, flexile, pliable).

Type species of *Thiofilum.* Basonym: *Thiothrix flexilis* Aruga *et al*. 2002.

Gram-stain-negative. Sugar-grown cells are oxidase- and catalase-positive. Obligately heterotrophic. Forms fingerprint-like colonies on glucose-acetate agar. Cells are rod shaped but morphology is variable, particularly with regard to length. Most cells are discoid or ovoid and 1.0 – 4.0 µm diameter and 0.5 – 5.5 µm in length, forming slightly bent, sheath-free filaments greater than 0.5 mm in length that can reach several millimetres. Some cells in filaments are elongate or swollen. Septa between cells of filaments are clearly defined. Rosettes and holdfasts are observed in some strains. Volutin (polyphosphate) granules are usually absent. Sudanophilic (lipid) granules are present. Temperature range of growth is 14-37 °C, with optimum growth 20-30 °C and no growth at 4 °C or at 42 °C. Growth is good in 0 to 170 mM (0 – 1.0 % w/v) sodium chloride but is inhibited slightly at 340 mM (2.0 % *w/v*). A small number of elementary sulfur globules are deposited within the invaginated inner membrane when grown in the presence of thiosulfate or sulfide, which are either not oxidised or only weakly oxidised to sulfate. Trichomes collected *in situ* do not contain sulfur deposits. Carbon sources for heterotrophic growth include hexose sugars (glucose, fructose, mannose), disaccharides (sucrose, maltose, trehalose), intermediates of Krebs' cycle

(succinate, malate, citrate), carboxylates (lactate, propionate, pyruvate, acetate, hydroxybutyrate), amino acids (glutamate, aspartate, alanine), alcohols (mannitol), but not benzoate, glycerol, butanol, xylose, galactose, melibiose, erythritol, raffinoise, arabinose, rhamnose, ethanol, propan-1-ol, sorbitol, formate, gelatine or starch. Does not require vitamins for growth. Reduces nitrate to nitrite. Has genes encoding *cbb*$_3$ cytochrome *c* and *bd*-I ubiquinol type terminal oxidases, and complete Krebs' cycle plus the glyoxylate shunt. Does not have RuBisCO genes. G+C fraction is 44.3 mol% from the genome sequence. Dominant respiratory quinone is ubiquinone-8. Dominant fatty acids in glucose-and-acetate grown cells are palmitoleic ($C_{16:1}$), palmitic ($C_{16:0}$) and vaccenic ($C_{18:1}$) acids.

Type strain is EJ2M-B^T (= DSM 14609^T = JCM 11135^T), isolated from activated sludge suffering from bulking (Japan).

Description of *Thiofilaceae* fam. nov.

Thiofilaceae (Thi.o.fi.la.ce'ae. N.L. neut. n. *Thiofilum*, type genus; *-aceae* suffix to denote family; N.L. pl. n. *Thiofilaceae*, the *Thiofilum* family).

This family is circumscribed on the basis of 16S rRNA gene sequences and comprises the genus *Thiofilum* (type genus). Ubiquinone-8 (UQ-8) is the dominant respiratory quinone. Dominant fatty acids are palmitoleic ($C_{16:1}$), palmitic ($C_{16:0}$) and vaccenic ($C_{18:1}$) acids. G+C fractions are around 44.3 mol% on the basis of genome sequence data. Falls within the *Gammaproteobacteria*, with the position at order level *incertae sedis* at present.

Type genus: *Thiofilum*

Emended description of *Thiotrichaceae* Garrity *et al.* 2005

Thiotrichaceae (Thi.o.tri.cha.ce'ae. N.L. fem. n. *Thiothrix*, type genus; *-aceae*, suffix denoting family; N.L. fem. pl. n. *Thiotrichaceae*, the *Thiothrix* family).

This family is circumscribed on the basis of 16S rRNA gene sequences and includes the genus *Thiothrix* on this basis, and historically has contained *Achromatium*, *Thiobacterium*, *Thiomargarita*, *Thioploca* and *Thiospira*, but their position is impossible to determine given the paucity of sequence data. They should, however, not be regarded as rejected from this family. Type genus are filamentous organisms that deposit elementary sulfur during growth.

Type genus: *Thiothrix*.

Emended description of *Thiothrix* Winogradsky 1888 (Approved Lists 1980) emend. Aruga *et al*. 2002

Thiothrix (Thi'o.thrix. Gr. masc. n. *theion* sulfur, brimstone, L. transliteration *thium*, sulfur; Gr. fem. n. *thrix* hair; N.L. fem. n. *Thiothrix*, sulfur hair).

Members of the *Thiothrichaceae* in the *Thiotrichales* of the *Gammaproteobacteria*. Gram-stain-negative filamentous organisms. Filaments have sheaths in most species and do not form knots. Autotrophs using form IAq and form II RuBisCO. Cells are rods with rounded ends 0.7 – 2.5 µm in diameter and 0.7 – 6.5 µm in length. Form white colonies with fibrous margins. Cells have polar tufts of fimbrae. Use thiosulfate or sulfide to support chemolithoautotrophic growth, with concomitant production of sulfate. All species can assimilate carbon from acetate and most can use succinate. None grow on mannitol or glycerol. Nitrate is reduced to nitrite. Major fatty acids are palmitoleic ($C_{16:1}$), palmitic ($C_{16:0}$) and vaccenic ($C_{18:1}$) acids - the dominant unsaturated fatty acid is palmitoleic acid. Genes encoding *cbb*$_3$ and *bd*-I terminal oxidases are found in all species. G+C fractions are 51.1-54.9 mol% (from genome sequences). Can be isolated from sulfidic wells, springs and lakes, from activated sludge, and from the bodies of amphipods living in sulfidic environments.

Type species is *Thiothrix nivea*.

Acknowledgements

We thank Dr Elena V. Odintsova of the University of Birmingham (Birmingham, UK) and Dr Ann P. Wood, previously of King's College London (London, UK) for discussions on the properties of "*Thiothrix ramosa*" and confirming its loss from culture. We also acknowledge Dr Keith Jolley (University of Oxford, UK) and his team, for the rapid curation of IMG/ER sequences into the rMLST database, and Dr Jolley for providing excellent support with the platform.

Financial Information

The authors received no specific funding for this work.

Conflict of interest

The authors declare they have no competing interests.

Ethical statement

No human, animal or other work that would require ethical approval was undertaken in this study.

488 **References**

489 **[1] Rabenhorst** *Flora Europaea Algarum aquae dulcis et submarinae*, Section II. Leipzig: E.
490 Kummer; 1865. pp. 1-319.

491 **[2] Winogradsky S.** *Beiträge zur Morphologie und Physiologie der Bacterien. Heft I. Zur*
492 *Morphologie und Physiologie der Schwefelbacterien*. Leipzig*:* Arthur Felix; 1888. pp. 1-120.

493 **[3] Aruga S, Kamagata Y, Kohno T, Hanada S, Nakamura K *et al*.** Characterization of
494 filamentous Eikelboom type 021N bacteria and description of *Thiothrix disciformis* sp. nov.
495 and *Thiothrix flexilis* sp. nov. *Int J Syst Evol Microbiol* 2002;52:1309-1316.

496 **[4] Buchanon RE, Holt JG, Lessel EF.** *Index Bergeyana – an annotated alphabetic listing*
497 *of the names of the taxa of the bacteria*. Edinburgh: E & S Livingstone Ltd; 1996. p. 1154.

498 **[5] Gibbons NE, Pattee KB, Holt JG.** *Supplement to Index Bergeyana*. London: Williams &
499 Wilkins; 1981. p. 294.

500 **[6] Chernousova EYu, Belousova EV, Gavrish EYu, Dubinina G A, Tourova TP *et al*.**
501 Molecular phylogeny and taxonomy of colourless, filamentous sulfur bacteria of the genus
502 *Thiothrix*. *Microbiology (Russia)* 2012;81:332-341.

503 **[7] Larkin JM, Shinabarger DL.** Characterization of *Thiothrix nivea*. *Int J Syst Bacteriol*
504 1983;33:841-846.

505 **[8] Chernousova E, Gridneva E, Grabovich M, Dubinina G, Akimov V *et al*.** *Thiothix*
506 *caldifontis* sp. nov. and *Thiothrix lacustris* sp. nov., gammaproteobacteria isolated from
507 sulfide springs. *Int J Syst Evol Microbiol* 2009;59: 3128-3135.

508 **[9] Howarth R, Unz RF, Seviour EM, Seviour RJ, Blackall LL *et al*.** Phylogenetic
509 relationships of filamentous sulfur bacteria (*Thiothrix* spp. and Eikelboom type 021N

510 bacteria) isolated from wastewater-treatment plants and description of *Thiothrix eikelboomii*
511 sp. nov., *Thiothrix unzii* sp. nov., *Thiothrix fructosivorans* sp. nov. and *Thiothrix defluvii* sp.
512 nov. *Int J Syst Bacteriol* 1999;49:1817-1827.

513 **[10] Dattagupta S, Schaperdoth I, Montanari A, Mariani S, Kita N *et al*.** A novel
514 symbiosis between chemoautotrophic bacteria and a freshwater cave amphipod. *ISME J*
515 2009;3:935-943.

516 **[11] Gillan DC, Dubilier N.** Novel epibiotic *Thiothrix* bacterium on a marine amphipod.
517 *Appl Environ Microbiol* 2004;70:3772-3775.

518 **[12] Lackey JB, Lackey WW, Morgan GB.** Taxonomy and ecology of the sulfur bacteria.
519 *Eng Prog Univ Fla Bull Ser* 1965;19:3-23.

520 **[13] Larkin JM, Henk MC, Burton, SD.** Occurrence of a *Thiothrix* sp. attached to mayfly
521 larvae and presence of parasitic bacteria in the *Thiothrix* sp. *Appl Environ Microbiol*
522 1990;56:357-361.

523 **[14] Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W *et al*.** Uniting the
524 classification of cultured and uncultured bacteria and archaea using 16S rRNA gene
525 sequences. *Nature Rev Microbiol* 2014;12:635-645.

526 **[15] Boden R, Hutt LP, Rae, AW**. Reclassification of *Thiobacillus aquaesulis* (Wood &
527 Kelly, 1996) as *Annwoodia aquaesulis* gen. nov., comb. nov., transfer of *Thiobacillus*
528 (Beijerinck, 1904) from the *Hydrogenophilales* to the *Nitrosomonadales*, proposal of
529 *Hydrogenophilalia* class. nov. within the '*Proteobacteria*', and four new families within the
530 orders *Nitrosomonadales* and *Rhodocyclales*. *Int J Syst Evol Microbiol* 2017;67:1191-1205.

531 **[16] Boden R, Scott KM, Williams J, Russel S, Antonen K *et al*.** An evaluation of
532 *Thiomicrospira*, *Hydrogenovibrio* and *Thioalkalimicrobium*: reclassification of four species

of *Thiomicrospira* to each *Thiomicrorhabdus* gen. nov. and *Hydrogenovibrio*, and reclassification of all four species of *Thioalkalimicrobium* to *Thiomicrospira*. *Int J Syst Evol Microbiol* 2017;67:1140-1151.

[17] Boden R, Scott KM, Rae AW, Hutt LP. Reclassification of *Thiomicrospira hydrogeniphila* (Watsuji *et al.* 2016) to *Thiomicrorhabdus hydrogeniphila* comb. nov., with emended description of *Thiomicrorhabdus* (Boden *et al.*, 2017). *Int J Syst Evol Microbiol* 2017;67:4205-4209.

[18] Boden R. Reclassification of *Halothiobacillus hydrothermalis* and *Halothiobacillus halophilus* to *Guyparkeria* gen. nov. in the *Thioalkalibacteraceae* fam. nov., with emended descriptions of the genus *Halothiobacillus* and family *Halothiobacillaceae*. *Int J Syst Evol Microbiol* 2017;67:3919-3928.

[19] Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C *et al*. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology* 2012;158:1005–1015.

[20] Badger MR, Bek EJ. Multiple rubisco forms in proteobacteria: their functional significance in relation to CO2 acquisition by the CBB cycle. *J Exp Bot* 2008;59:1525-1541.

[21] Tabita FR, Hanson TE, Satagopan S, Witte B, Kreel NE. Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms. *Phil Trans Roy Soc B* 2008;363:2629-2640.

[22] Axen SD, Erbilgin O, Kerfeld CA. A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Comp Biol* 2014;10:e1003898.

554 **[23] Boden R, Hutt LP, Huntemann M, Clum A, Pillay M *et al.*** Permanent draft genome
555 of *Thermithiobacillus tepidarius* DSM 3134^{T}, a moderately thermophilic, obligately
556 chemolithoautotrophic member of the *Acidithiobacillia*. *Standards Genomic Sci* 2016;**11**:74.

557 **[24] Smith AJ, London J, Stanier RY.** Biochemical basis of obligate autotrophy in blue-
558 green algae and Thiobacilli. *J Bacteriol* 1967;94:972-983.

559 **[25] Quasem I, Achille AN, Caddick BA, Carter TA, Daniels C *et al.*** A peculiar citric acid
560 cycle of hydrothermal vent chemolithoautotroph *Hydrogenovibrio crunogenus*, and insights
561 into carbon metabolism by obligate autotrophs. *FEMS Microbiol Lett* 2017;364:fxn148.

562 **[26] Wood AP, Aurikko JP, Kelly DP.** A challenge for 21st century molecular biology and
563 biochemistry: what are the causes of obligate autotrophy and methanotrophy? *FEMS*
564 *Microbiol Rev* 2004;**28**:335-352.

565 **[27] Hutt LP, Huntemann M, Clum A, Pillay M, Palaniappan K *et al.*** Permanent draft
566 genome of *Thiobacillus thioparus* DSM 505^{T}, an obligately chemolithoautotrophic member
567 of the *Betaproteobacteria*. *Standards Genomic Sci* 2017;12:10.

568 **[28] Kornberg HL, Krebs HA**. Synthesis of cell constituents from C2-units by a modified
569 tricarboxylic acid cycle. *Nature* 1957;179:988-991.

570 **[29] Anthony C**. *The biochemistry of methylotrophs*. London: Academic Press; 1982. pp.
571 152-194.

572 **[30] Quayle JR, Ferenci T.** Evolutionary aspects of autotrophy. *Microbiol Rev* 1978;42:251-
573 273.

574 **[31] Large PJ, Peel D, Quayle JR**. Microbial growth on C_1 compounds. 2. Synthesis of cell
575 constituents by methanol- and formate-growth *Pseudomonas* AM1 and methanol-grown
576 *Hyphomicrobium vulgare*. *Biochem J* 1961;81:470-480.

[32] Entner N, Doudoroff M. Glucose and gluconic acid oxidation of *Pseudomonas saccharophila*. *J Biol Chem* 1952;196:853-862.

[33] Flamholz A, Noor E, Bar-Even A, Liebermeister W, Milo R. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc Natl Acad Sci* 2013;110:10039-10044.

[34] Boden R, Kelly DP, Murrell JC, Schäfer H. Oxidation of dimethylsulfide to tetrathionate by *Methylophaga thiooxidans* sp. nov.: a new link in the sulfur cycle. *Environ Microbiol* 2010;12:2688-2699.

[35] Boden R, Ferriera S, Johnson J, Kelly DP, Murrell JC *et al*. Draft genome sequence of the chemolithoheterotrophic, halophilic methylotroph *Methylophaga thiooxydans* DMS010. *J Bacteriol* 2011;193:3154-3155.

[36] Giuffrè A, Borisov VB, Arese M, Sarti P, Forte E. Cytochrome *bd* oxidase and bacterial tolerance to oxidative and nitrosative stress. *Biochim Biophys Acta Bioenerg* 2014;1837:1178-1187.

[37] Odintsova EV, Dubinina GA. New filamentous colourless sulphur bacteria *Thiothrix ramosa* nov. sp. *Mikrobiologiia* 1990;59:637-644.

[38] Odintsova EV, Dubinina GA. The growth cycle, reproduction and ultrastructure of *Thiothrix ramosa. Mikrobiologiia* 1991;60:314-320.

[39] Odintsova EV, Dubinina GA. The role of reduced sulphur compounds on the metabolism of *Thiothrix ramosa*. *Mikrobiologiia* 1993;62:213-222.

[40] Odintsova EV, Wood AP, Kelly DP. Chemolithoautotrophic growth of *Thiothrix ramosa*. *Arch Microbiol* 1993;160:152-157.

598 **[41] Polz MF, Odintsova EV, Cavanaugh CM.** Phylogenetic relationships of the
599 filamentous sulfur bacterium *Thiothrix ramosa* based on 16S rRNA sequence analysis. *Int J*
600 *Syst Bacteriol* 1996;46:94-97.

601 **[42] Edgar RC.** MUSCLE: multiple sequence alignment with high accuracy and high
602 throughput. *Nucleic Acids Res* 2004;32,1792-1797.

603 **[43] Kumar S, Stecher G, Tamura K.** MEGA7: Molecular Evolutionary Genetic Analysis
604 version 7.0 for bigger datasets. *Mol Biol Evol* 2016;33:1870-1874.

605 **[44] Tamura K, Nei M.** Estimation of the number of nucleotide substitutions in the control
606 region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 1993;**10**: 512-526.

607 **[45] Le S, Gascuel O.** An improved general amino acid replacement matrix. *Mol Biol Evol*
608 2008;25:1307-1320.

609 **[46] Kimura M.** A simple method for estimating evolutionary rate of base substitutions
610 through comparative studies of nucleotide sequences. *Mol Biol Evol* 1980;16:111-120.

611 **[47] Scott KM, Williams J, Porter CMB, Russel S, Harmer TL *et al*.** Genomes of
612 ubiquitous marine and hypersaline *Thiomicrospira*, *Thiomicrorhabdus*, and *Hydrogenovibrio*
613 spp. encode a diversity of mechanisms to sustain chemolithoautotrophy in heterogeneous
614 environments. *Environ Microbiol* 2018, *In Press*. https://doi.org/10.1111/1462-2920.14090

615 **[49] Dubinina G, Savvichev A, Orlova M, Gavrish E, Verbarg *et al*.** *Beggiatoa*
616 *leptomitoformis* sp. nov.,the first freshwater member of the genus capable of
617 chemolithoautotrophic growth. *Int J Syst Evol Microbiol* 2017;67:197-204.

618

619

620

621 **Table 1**. Comparative properties of *Thiothrix* spp., *Leucothrix* spp. and *Cocleimonas flava*.
622 Data refer to type strains unless otherwise indicated.

623 Data are curated from [1-3] and [6-9] or obtained from genome sequence data held in the
624 Integrated Microbial Genomes (IMG) database.

625 *N.D.*, not determined/no data available; +, positive or present; -, negative or absent; ±, weakly
626 positive.

627 * Values in square brackets are *in silico* values derived from genome sequence data in the
628 IMG public database, other values are from *in vitro* determinations.

629 † Genome parameters are from *Thiothrix* sp. AAV1, which has 99.9% 16S rRNA gene
630 identity to *T. fructosivorans* Q^T.

631

632 **Table 2**. Curated properties of the genera *Thiothrix*, *Thiolinea* gen. nov. and *Thiofilum* gen.

633 nov. and their respective families. Data are from [1-3] and [6-9].

634

635

Figure 1. Phylogenetic trees on the basis of the 16S rRNA (*rrs*) gene, showing the positions of *Thiothrix* species, divided into Clade X, Clade Y and Clade Z, *versus Leucothrix* spp. and *Cocleimonas flava*. Type species of genera are shown in bold type. Nucleotide sequences were aligned using MUSCLE [42] without use of any pre-sets for speed that reduce accuracy. Aligned sequences were tested for best fit to models on the basis of the Bayesian information coefficient (BIC) in MEGA 7.0.26 [43], and the trees reconstructed accordingly, using the Tamura-Nei model [44] with a discreet gamma distribution to model rate differences across sites (maximum likelihood parameter = 0.1664; neighbour joining/minimum evolution shape parameter = 5). Trees shown are the optimal trees, with numbers at nodes indicating the percentage of 5,000 bootstrap replications in which the topology was preserved (values <70 % omitted for clarity). All positions at which there was less than 95 % coverage were omitted from the final analysis, in which 1,551 nt were used. Branch lengths are to scale and indicate the number of substitutions per site – bars represent 0.02 substitutions per site for all trees shown. The outgroup of each tree is *Thiomicrospira pelophila* DSM 1534^{T} from the *Piscirickettsiaceae*, also in the *Thiothrichales* per all members of the ingroup. Maximum likelihood tree shown had the highest log-likelihood after 5,000 replications (-4806.27). Neighbour joining and minimum evolution trees shown had the optimal sum of branch length 0.523 and 0.525, respectively. Accession numbers are given in parentheses and refer to the GenBank or Integrated Microbial Genomes (IMG) databases (the latter contain an underscore "_").

Figure 2. Maximum likelihood trees for the genera *Thiothrix* and *Leucothrix*, on the basis of amino acyl sequences derived from a range of housekeeping genes encoding polyribonucleotide nucleotidyltransferase (EC 2.7.7.8, *pnp*); translation initiation factor IF-2 (EC 3.6.5.3, *infB*); glyceraldehyde-3-phosphate dehydrogenase (NAD^+, EC 1.2.1.12, *gapA*); glutaminyl-tRNA synthetase (EC 6.3.5.7, *glnS*); elongation factor EF-G (EC 3.6.5.3, *fusA*), and recombination protein A (*recA*). *Thiothrix* clades X, Y and Z are indicated and type species of genera are shown in bold type. Nucleotide sequences were converted into amino acyl sequences using the bacteriological code, in frame, and were aligned using MUSCLE [42] without use of any pre-sets for speed that reduce accuracy. Aligned amino acyl sequences were tested for best fit to models on the basis of the Bayesian information coefficient (BIC) in MEGA 7.0.26 [43], and the trees reconstructed accordingly, using the Le and Gascuel model [45] with a discreet gamma distribution to model rate differences across sites, and invariant sites. Trees shown are the optimal trees, with numbers at nodes indicating the percentage of 5,000 bootstrap replications in which the topology was preserved (values <70 % omitted for clarity). All positions at which there was less than 95 % coverage were omitted from the final analysis, in which the number of amino acyl residues used was: Pnp 348, InfB 789, GapA 332, GlnS 549, FusA 689, RecA 341. Branch lengths are to scale and indicate the number of substitutions per site – bars represent 0.05 or 0.10 substitutions per site. The outgroups are sequences from *Thiomicrospira pelophila* DSM 1534^T from the *Piscirickettsiaceae*, also in the *Thiothrichales* per all members of the ingroups. Maximum likelihood trees shown had the highest log-likelihoods after 5,000 replications. Accession numbers in parenthesis relate to the IMG database.

680 **Figure 3.** Maximum likelihood tree of *Thiothrix* and *Leucothrix* spp. using amino acyl
681 concatamer sequences derived from 53 ribosomal protein genes, namely *rpsA* to *rpsU, rplA*
682 to *rplF*, *rplL* to *rplX* and *rpmA* to *rpmJ*. Sequences were extracted at gene level and
683 concatenated using the ribosomal multilocus sequence typing (rMLST) online platform [19]
684 from genome sequences publically available in the IMG database. Gene concatamers were
685 then translated in frame using the bacterial code in MEGA 7.0.26 [43] and the derived amino
686 acyl concatamers were then aligned using MUSCLE [42] without any pre-sets, and the
687 aligned dataset tested for the best model fit on the basis of the lowest BIC in MEGA. The tree
688 was thus reconstructed using the Le and Gascuel model [45] with amino acid frequencies
689 estimated from the data and with a discreet gamma distribution to model rate differences
690 across sites, and invariant sites (parameter = 0.5754). Tree shown is the optimal tree, with
691 numbers at nodes indicating the percentage of 5,000 bootstrap replications in which the
692 topology was preserved (values <70 % omitted for clarity). All positions at which there was
693 less than 95 % coverage were omitted from the final analysis, in which 6,668 aa were used.
694 Branch lengths are to scale and indicate the number of substitutions per site – bar represents
695 0.1 substitutions per site for all trees shown. The outgroup is *Thiomicrospira pelophila* DSM
696 1534^{T} from the *Piscirickettsiaceae*, also in the *Thiothrichales* per all members of the ingroup.
697 Maximum likelihood tree shown had the highest log-likelihoods after 5,000 replications (-
698 49,806.6). Values in parentheses are Genome ID numbers relating to the rMLST database.

699

700

Table 1

Click here to download Table IJSEM-D-18-00034R3 Table 1.docx

Clade	***N/A***			**Clade Z**		**Clade Y**		**Clade X**				
Order	***Incertae sedis*, pending further study**					***Thiotrichales***						
Family	***Leucotrichaceae* fam. nov.**			***Thiofilaceae* fam. nov.**		***Thiolineaceae* fam. nov.**		***Thiotrichaceae***				
Genus	***Leucothrix***		***Cocleimonas***	***Thiofilum gen. nov.***		***Thiolinea* gen. nov.**		***Thiothrix sensu stricto***				
Characteristic	***Leucothrix mucor* DSM 2157^T**	***Leucothrix pacifica* XH122^T**	***Cocleimonas flava* KMM 3898^T**	***Thiothrix defluvii* Ben57^T**	***Thiothrix flexilis* EJ2M-B^T**	***Thiothrix disciformis* B3-1^T**	***Thiothrix eikelboomii* AP3^T**	***Thiothrix nivea* JP2^T**	***Thiothrix unzii* A1^T**	***Thiothrix lacustris* DSM 21227^T**	***Thiothrix caldifontis* DSM 21228^T**	***Thiothrix fructosivorans* ATCC 49748^T**
Source of type strain	*Monostroma* sp., USA	Surface seawater, South Pacific Gyre	*Umbonium* sp., Russia	Activated sludge, Australia	Activated sludge, Japan	Activated sludge, Japan	Activated sludge, USA	Sulfidic well water, USA	Activated sludge, USA	Sulfidic lake water, Russia	Sulfidic spring water, USA	Activated sludge, USA
% 16S rRNA (*rrs*) gene sequence identity to:												
Thiothrix nivea JP2^T	85.6	88.8	87.9	89.5	89.2	90.9	92.3	100.0	94.8	95.0	95.2	94.4
Thiothrix flexilis Ben57^T	88.8	88.0	87.4	96.9	100.0	91.3	90.5	89.1	88.0	90.2	89.9	90.1
Thiothrix disciformis B3-1^T	89.2	87.6	86.4	92.2	91.3	100.0	94.8	90.8	90.5	91.5	90.9	90.9
Helix 18 deletion of 16S rRNA gene	-	-	-	+	+	+	+	+	+	+	+	+
Colonial properties:												
Colour (reflected light)	White	White	Yellow	*N.D.*	White	White	White	White	*N.D.*	White	White	*N.D.*
Margin	Entire	Entire	Entire	*N.D.*	Fingerprint-like	Fingerprint-like	Fingerprint-like	Fibrous margin	*N.D.*	Fibrous margin	Fibrous margin	*N.D.*
Cell properties:												
Morphology	Rod	Rod	Rod	Rods, cylinders, barrels	Rods, discs, ovoid	Rods, discs, ovoid	Rods, discs, cubes, barrels,	Rods	Rods	Rods	Rods	Rods
Length (µm)	1.8-2.8	2.5-3.0	1.6-1.8	5.0-10.0	0.5-5.0	0.5-3.0	1.0-8.0	3.0	0.7-3.0	4.4-6.3	3.2-6.5	1.2-2.5
Diameter (µm)	0.5-1.0	0.4-0.5	0.3-0.4	1.0-2.0	1.0-4.0	1.2-3.0	0.6-8.0	1.0-1.5	0.70-1.5	0.9-2.3	0.9-2.2	1.2-2.5
Fimbrae	*N.D.*	*N.D.*	*N.D.*	*N.D.*	*N.D.*	*N.D.*	+	+	+	*N.D.*	*N.D.*	+
Filament properties:												
Filaments formed	+	+	-	+	+	+	+	+	+	+	+	+
Sheathed	-	-	-	-	-	-	-	+	-	+	+	+
Knots found	+	-	-	+	-	*N.D.*	+	-	-	-	-	-
pH optimum	*N.D.*	8.0	8.5-9.5	*N.D.*	*N.D.*	*N.D.*	*N.D.*	*N.D.*	*N.D.*	7.0	8.0	*N.D.*
Oxidation of thiosulfate	-	-	+	*N.D.*	*N.D.*	*N.D.*	+	+	+	+	+	+
Carbon sources for heterotrophic growth:												
Sucrose	+	+	*N.D.*	*N.D.*	+	+	+	-	-	-	-	+
D-fructose	+	+	*N.D.*	*N.D.*	+	+	+	-	-	-	-	+
D-glucose	+	+	+	*N.D.*	+	+	+	-	-	-	-	-
Acetate	*N.D.*	*N.D.*	-	*N.D.*	+	+	+	+	+	+	+	+
Glycerol	+	+	-	*N.D.*	-	+	+	-	-	-	-	-
Mannitol	+	+	+	*N.D.*	+	+	+	-	-	-	-	-
Genes present encoding:												
cbb$_3$ cytochrome *c* oxidase	+	*N.D.*	*N.D.*	*N.D.*	+	+	+	+	*N.D.*	+	+	+†
aa$_3$ cytochrome *c* oxidase	+	*N.D.*	*N.D.*	*N.D.*	-	-	-	-	*N.D.*	-	-	-†
bd-I ubiquinol oxidase	+	*N.D.*	*N.D.*	*N.D.*	+	+	+	+	*N.D.*	+	+	+†
RuBisCO	-	*N.D.*	*N.D.*	*N.D.*	-	Form II	Form IAq	Form IAq Form IAc Form II	*N.D.*	Form IAq Form II	Form IAq Form IAc Form II	Form IAq† Form II
Kelly-Friedrich pathway gene cluster fragmentation	-	*N.D.*	*N.D.*	*N.D.*	*soxYZAXB*	*soxZY* *soxY* *soxXA* *soxY*	*soxBXAZY* *soxY* *soxY*	*soxZY* *soxXYZA_Y* *soxZ* *soxB* *soxY* *soxXA*	*N.D.*	*soxYZA* *soxZ* *soxB* *soxXA* *soxY*	*soxB* *soxZYAZYZ* *soxXA*	*soxY*† *soxB* *soxYZ* *soxY* *soxZAY*

Enzyme activities:												
Oxidase	+	+	+	*N.D.*	+	+	+	+	+	N.D.	+	+
Catalase	+	+	+	*N.D.*	+	+	+	-	-	-	-	+
G+C fraction (mol%) *	49.5 [47.8]	46.2	43.4	*N.D.*	44.0 [44.3]	44.0-45.0 [45.1]	44.0-45.0 [46.3]	52.0 [54.9]	49.3	51.4 [51.3]	52.0 [50.5]	N.D. [51.1]†

Table 2

	***Leucotrichaceae* fam. nov.**	***Thiofilaceae* fam. nov.**	***Thiolineaceae* fam. nov.**	***Thiotrichaceae***
Genera	*Leucothrix* *Cocleimonas*	*Thiofilum*	*Thiolinea*	*Thiothrix*
Colony colour	White, Yellow	White	White	White
G+C fraction (mol%)	43.4 – 47.8	44.0	44.1-46.3	49.3-54.9
Cell diameter × length (µm)	0.3-1.0 × 1.6-3.0	1.0-4.0 × 0.5-10.0	0.6-8.0 × 1.0-8.0	0.7-2.5 × 0.7-6.5
Catalase activity	+	+	+	-
Pigments	Carotenoids in some species	-	-	-
Gliding motility	-	+	+	+
Carbon sources				
Glycerol	±	-	+	-
Mannitol	+	+	+	-
Acetate	-	+	+	+
Butyrate	-	-	+	-
Glucose	+	+	+	-
Mannose	-	+	+	-
Trehalose	+	+	+	-
Reduction of nitrate	-	+	±	+
Temperature minima (°C)	0-4	10-14	14	4-7
pH minima	5.5	7.0	6.5-7.0	6.2-7.0
pH maxima	11.0	7.9	7.9-8.5	8.2-8.6
Metabolic modes	Heterotrophic	Heterotrophic Chemolithoheterotrophic	Autotrophic Heterotrophic Mixotrophic	Autotrophic Heterotrophic Mixotrophic
Dominant non-substituted fatty acid	Palmitoleic or vaccenic	Vaccenic	Vaccenic	Palmitoleic or palmitic

Figure 1

Click here to download Figure FINAL Figure 1.pdf

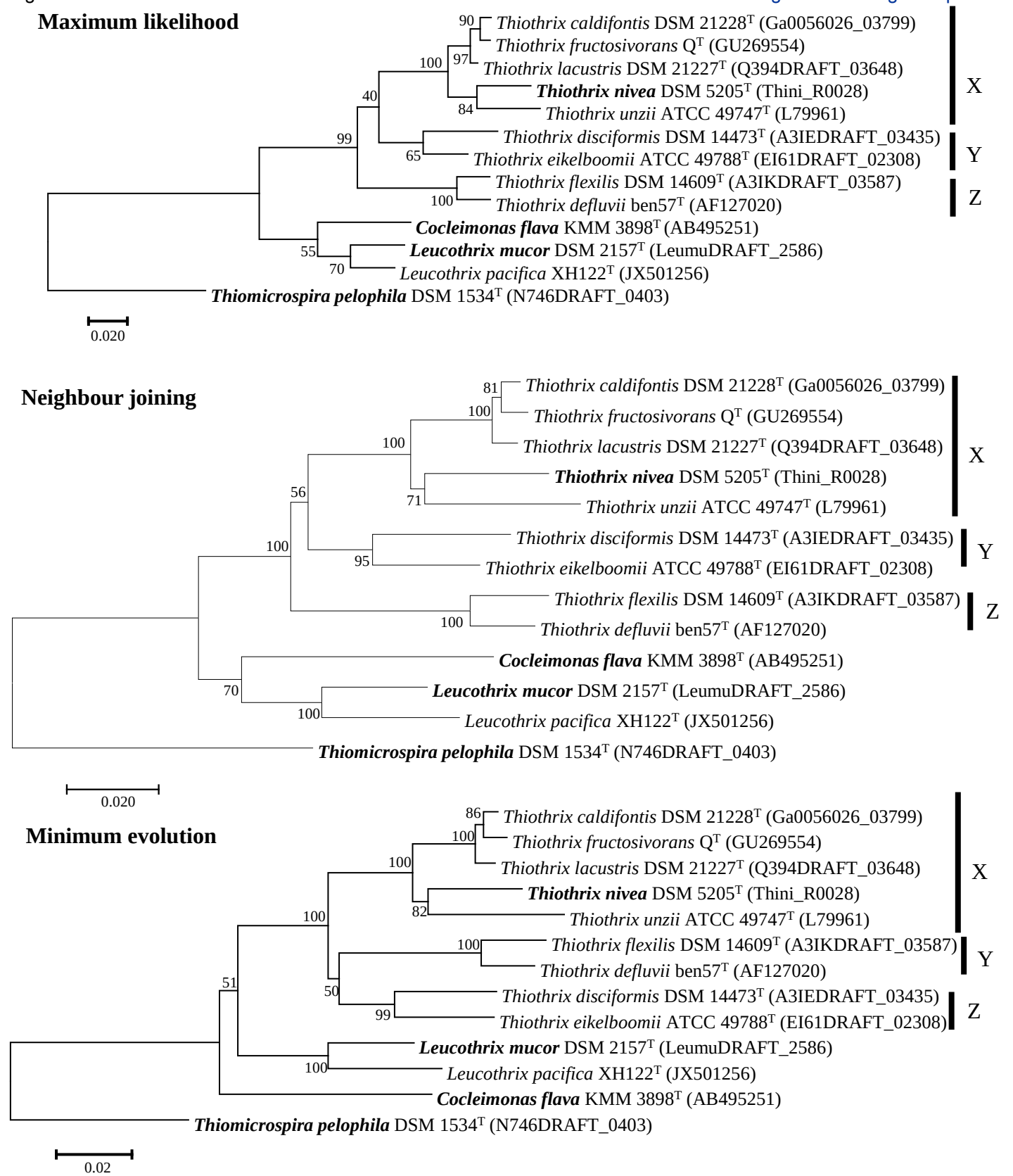
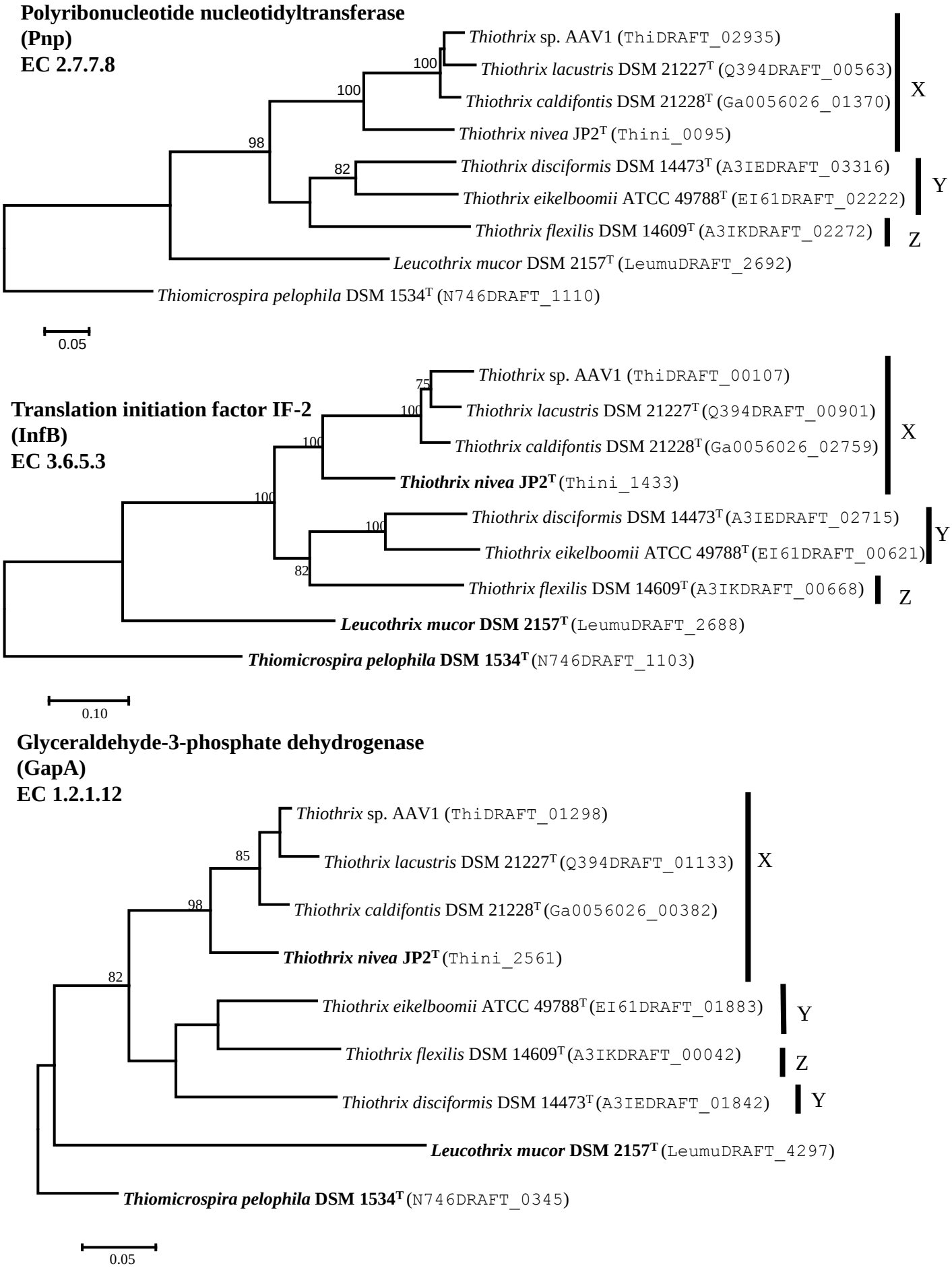
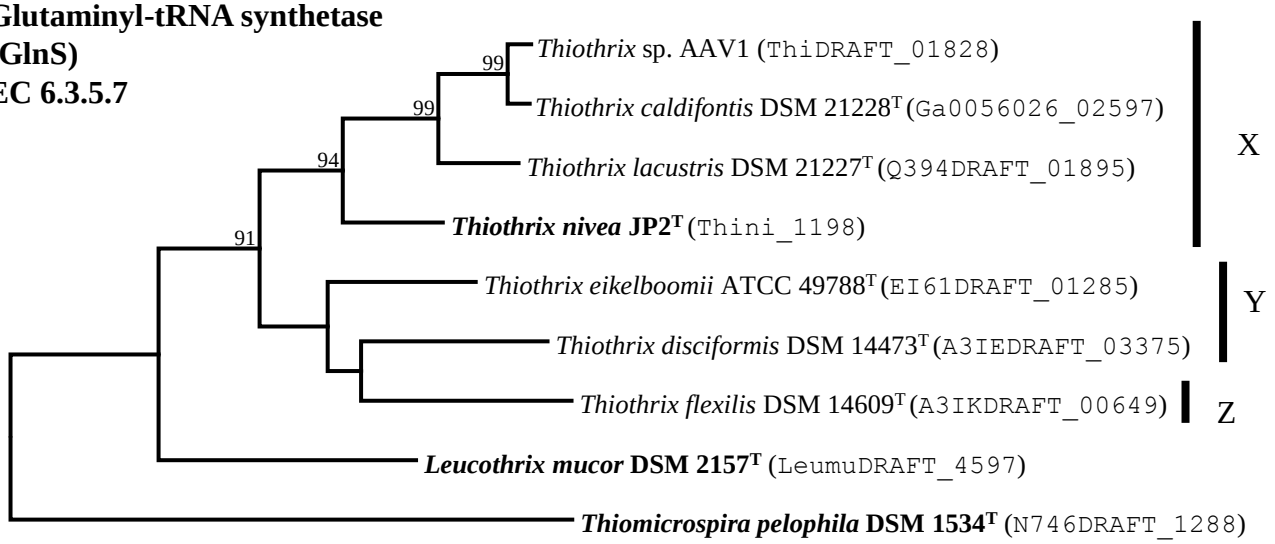


Figure 2

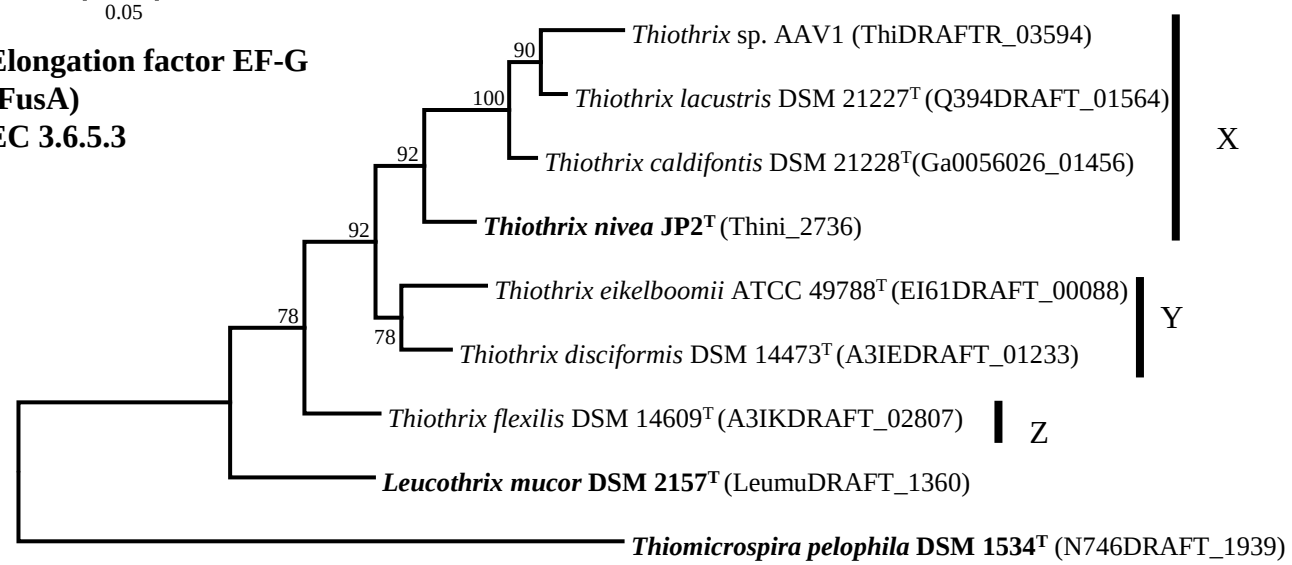


Glutaminyl-tRNA synthetase (GlnS) EC 6.3.5.7



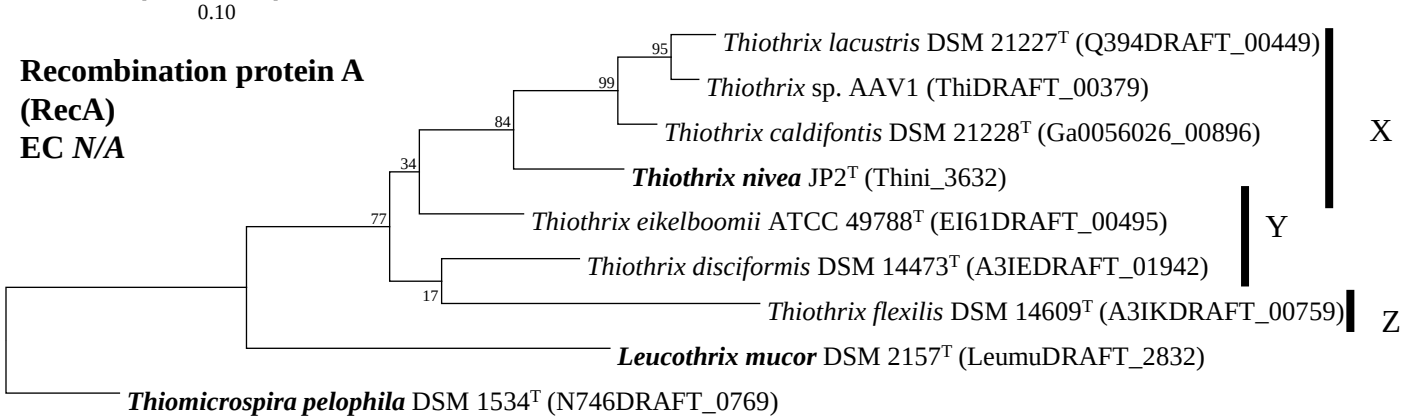
0.05

Elongation factor EF-G (FusA) EC 3.6.5.3



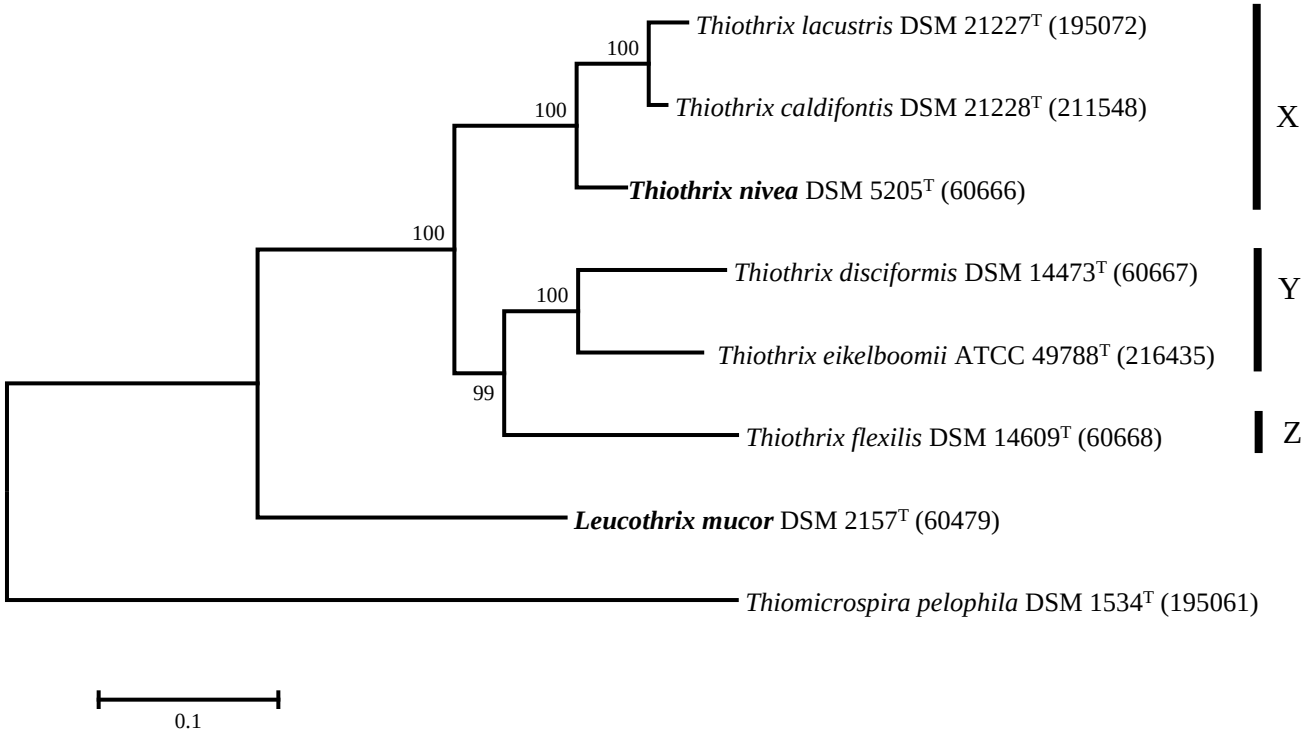
0.10

Recombination protein A (RecA) EC *N/A*



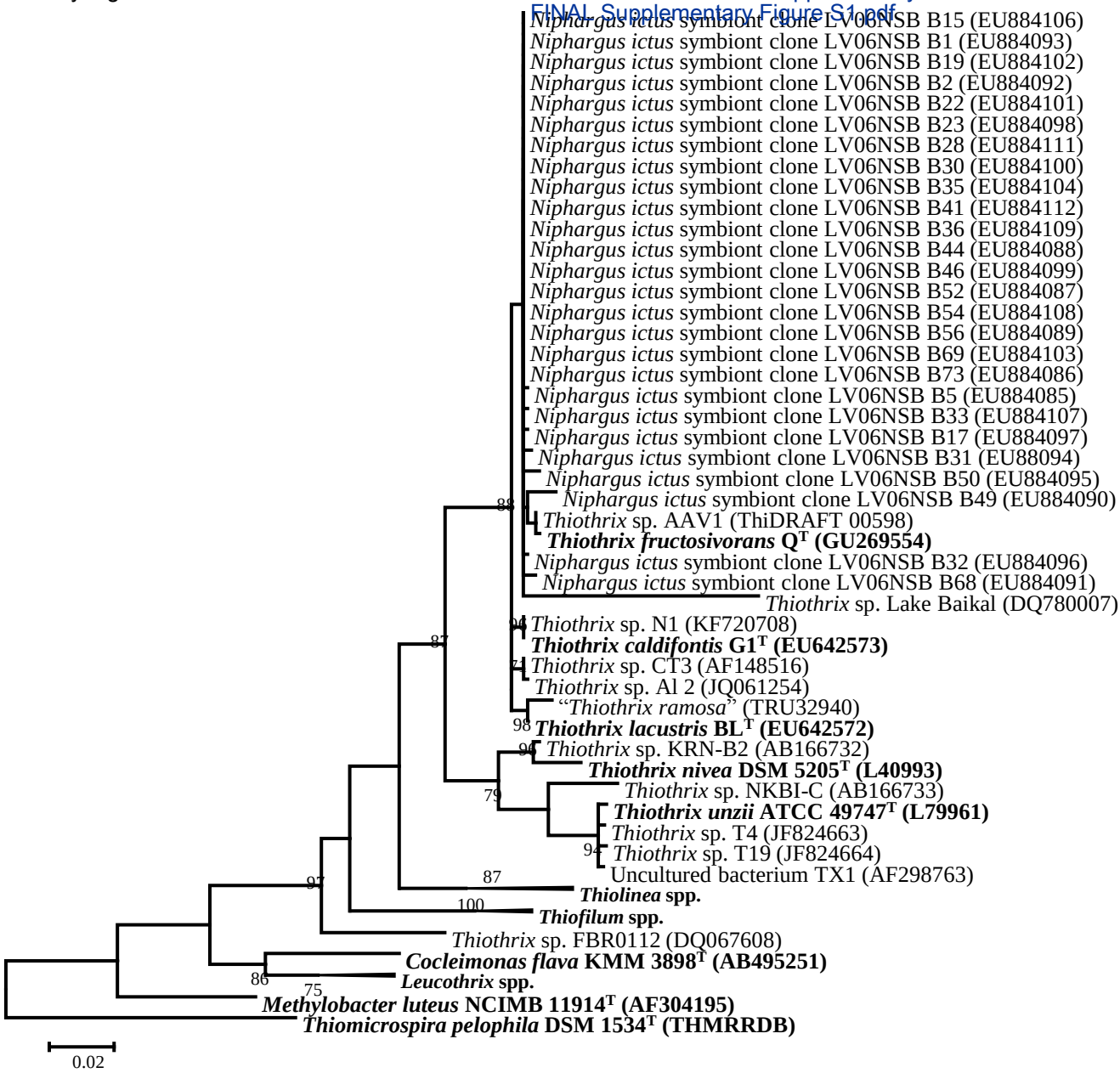
0.050

Figure 3



Supplementary Figure S1

Click here to download Supplementary Material Files
FINAL_Supplementary_Figure_S1.pdf



Supplementary Figure S1. Maximum likelihood tree on the basis of a the 16S rRNA (*rrs*) gene, showing *Thiothrix*, *Thiolinea* gen. nov. and *Thiofilum* gen. nov., with relation to sequences in the GenBank and IMG databases that purport to be "*Thiothrix*", including *Thiothrix* sp. AAV1, which we have used the genome sequence of as a proxy for *T. fructosivorans* (16S rRNA gene identity of AAV1 to Q^T is 99.9%) and "*T. ramosa*", a well-studied organism that has since been lost. Sequences were extracted at gene level from the GenBank and IMG databases and aligned using MUSCLE without any pre-sets, and the aligned dataset tested for the best model fit on the basis of the lowest BIC in MEGA. The tree was thus reconstructed using the two parameter method of Kimura [46] with a discreet gamma distribution to model rate differences across sites, and invariant sites (parameter = 0.3695). Tree shown is the optimal tree, with numbers at nodes indicating the percentage of 5,000 bootstrap replications in which the topology was preserved (values <70 % omitted for clarity). All positions at which there was less than 95 % coverage were omitted from the final analysis, in which 809 nt were used. Branch lengths are to scale and indicate the number of substitutions per site – bar represents 0.02 substitutions per site for all trees shown. The outgroup is *Thiomicrospira pelophila* DSM 1534^T from the *Piscirickettsiaceae*, also in the *Thiothrichales*. Maximum likelihood tree shown had the highest log-likelihoods after 5,000 replications (-3989.24). Values in parentheses are GenBank and IMG accession numbers – the latter have an underscore ("_").