Proposal for a new class within the phylum Proteobacteria, Acidithiobacillia classis nov., with the type order Acidithiobacillales, and emended description of the class Gammaproteobacteria

Kelly P. Williams¹ and Donovan P. Kelly²

¹Systems Biology Department, Sandia National Laboratories, Livermore, CA 94551, USA
²School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK

The phylogeny Proteobacteria has its taxonomic origin as the ‘purple bacteria’, defined as four bacterial groups (alpha, beta, gamma and delta), which were classified by their 16S rRNA gene sequence structures (Woese, 1987). The phylogeny was formally established, also using phylogenetic analysis of 16S rRNA gene sequences, by Garrity et al. (2005a), with five constituent classes containing all known Gram-negative bacteria, Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, Deltaproteobacteria and Epsilonproteobacteria. Subsequently a sixth class was proposed, the ‘Zetaproteobacteria’, although this name is not yet validly published (Emerson et al., 2011). Recent molecular analyses using complete multigene/multiprotein alignment studies have made possible more in-depth probing of these class delineations (Williams et al., 2007, 2010; Yutin et al., 2012). These have confirmed the internal coherence of the classes Alphaproteobacteria and Gammaproteobacteria, enabled further taxonomic dissection within these classes, and support the exclusion of the ‘Zetaproteobacteria’ from Gammaproteobacteria. We now confirm the exclusion also of the order Acidithiobacillales from the class Gammaproteobacteria, as a new class of proteobacteria.

The order Acidithiobacillales currently comprises the two families of Order II of the class Gammaproteobacteria, as defined by Garrity et al. (2005b, c). The type genus of the type order is the genus Acidithiobacillus Kelly and Wood 2000 emend. Hallberg et al. 2010, with Acidithiobacillus thiioxidans as the type species (Waksman & Joffe 1922; Kelly & Wood 2000, 2005a). Four other species of the family Acidithiobacillaceae are currently recognized: Acidithiobacillus ferrooxidans, Acidithiobacillus caldus, Acidithiobacillus albertensis and Acidithiobacillus ferrivorans. The second family is the family Thermithiobacillaceae, currently with a single genus and species, Thermithiobacillus tepidarius (Garrity et al., 2005b, d; Kelly & Wood, 2005b).

The phylogenetic relationships of T. tepidarius and the species of the genus Acidithiobacillus have been conjectural since the proteobacterial classes were defined on the basis of rRNA signatures (Lane et al., 1992; Woese et al., 1984a, b, 1985). Lane et al. (1992) placed T. tepidarius, Acidithiobacillus thiioxidans and Acidithiobacillus ferrooxidans close to the beta–gamma root in the class Betaproteobacteria, but placed another putative Acidithiobacillus ferrooxidans strain (m-1) in the class Gammaproteobacteria. When the taxonomy of the thiobacilli was rationalized, T. tepidarius and the species of the genus Acidithiobacillus were reassigned to the class Gammaproteobacteria (Kelly & Wood, 2000), with the exception of strain m-1 which was reclassified as Acidiferrobacter thiioxidans in the gammaproteobacterial family Ectothiorhodospiraceae (Hallberg et al., 2011). Woese et al. (1984b, 1985) identified 8–12 nt signature sequences in the 16S rRNA gene that were unique to the species of the classes Betaproteobacteria and Gammaproteobacteria considered by them. One of these (AAAAACCUAAC, found in 100 % of members of the class Betaproteobacteria, but absent from members of the class Gammaproteobacteria) was found (as AACCUAAC) in all species of the genus Acidithiobacillus and in T. tepidarius; and another (AAACUCAAUG, unique to the class Gammaproteobacteria) was also found (as AAACUCAAAA) in species of the genera Acidithiobacillus and..
Thermithiobacillus (our sequence search data). Other diagnostic sequences were absent from the 16S rRNA genes of members of the order Acidithiobacillales, supporting the anomalous relationship of these genera to both the class Betaproteobacteria and the class Gammaproteobacteria. The availability of the complete genomes of some species of the genus Acidithiobacillus has now enabled a reassessment of the phylogeny of the order Acidithiobacillales and resolves this long-standing issue.

Analysis of 16S rRNA gene sequences and RecA and GyrB protein sequences of Maripondanus ferrooxidans identified a sixth class of the phylum Proteobacteria, the ‘Zetaproteobacteria’ (Emerson et al., 2007). This comprised a monophyletic group of iron-oxidizing, neutrophilic, filamentous bacteria associated with microbial mats in deep-sea habitats (McAllister et al., 2011), and is another deeply rooted branch of the phylum Proteobacteria, distinct from the classes Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, Deltaproteobacteria and Epsilonproteobacteria. Its validity was supported by a telescoping comparative multiprotein analysis of 356 protein-families among 104 gammaproteobacterial genomes, representing all 14 orders of the class Gammaproteobacteria and five outgroup taxa from the classes Alphaproteobacteria and Betaproteobacteria (Williams et al., 2010). That study showed that Acidithiobacillus ferrooxidans (representing the order Acidithiobacillales) was also excluded from the class Gammaproteobacteria, and was deduced to represent a sister group to the classes Betaproteobacteria, Gammaproteobacteria and ‘Zetaproteobacteria’, which had arisen after the divergence of the Alphaproteobacteria (Williams et al., 2010).

The displacement of Acidithiobacillus ferrooxidans from both the class Betaproteobacteria and the class Gammaproteobacteria was supported by phylogenetic trees reconstructed from a concatenated alignment of 50 ribosomal proteins from 995 bacteria (Yutin et al., 2012). These placed Acidithiobacillus ferrooxidans in a clade with 75 beta- and 251 gammaproteobacterial species, but distinct from both, and remote from 180 species of the class Alphaproteobacteria, Deltaproteobacteria and Epsilonproteobacteria (Yutin et al., 2012).

We have now widened the comparative multiprotein approach to a total of seven genomes of strains of four species of the genus Acidithiobacillus: Acidithiobacillus thiooxidans, Acidithiobacillus ferrooxidans, Acidithiobacillus caldus and Acidithiobacillus ferrivorans.

Genome sequences for seven strains of species of the genus Acidithiobacillus were available for analysis: Acidithiobacillus thiooxidans ATCC 19377T, Acidithiobacillus caldus DSM 22753T and strain SM-1, Acidithiobacillus ferrooxidans DSM 22753T, Acidithiobacillus ferrivorans ATCC 23270T and ATCC 53993 and Acidithiobacillus sp. strain GGI-221. The type species of the other genus of the order Acidithiobacillales, T. tepidarius, is targeted for genome sequencing (http://www.genomesonline.org/), but no data are yet available. Its closest phylogenetic relatives are, however, the species of the genus Acidithiobacillus, with 90.3–91.2% identity in their 16S rRNA gene sequences. The type strains of Acidithiobacillus thiooxidans, Acidithiobacillus caldus and Acidithiobacillus ferrivorans show 97.4–98.1% 16S rRNA gene sequence identity to each other and 95.2–95.6% identity to Acidithiobacillus ferrooxidans. The 16S rRNA gene sequence of strain GGI-221 shows 99.7% identity to Acidithiobacillus ferrooxidans and only 95.0–97.9% identity to the other type strains, indicating it to be a strain of Acidithiobacillus ferrooxidans.

The genomes were assessed in two phases, as described below.

The phase I phylogenetic analysis aimed for uniform coverage of the domain Bacteria (and a large archaeal outgroup), but only contained one representative of the order Acidithiobacillales, while phase II added all additional available genomes of members of the order Acidithiobacillales. Phase I began on 15 December 2009, when there were 1891 prokaryotic genomes available at NCBI. Genomes with fewer than 30 called proteins, or in more than 500 contigs, were rejected, leaving 1834 genomes. These genomes comprised 129 taxonomic groups (Table S1 available in IJSEM Online). To reduce the number of genomes to a tractable yet diverse set, one genome was selected from each of the 129 taxonomic groups. A novel scoring system was employed to select the genome most representative of the group and to avoid genomes with extremes of gene content or nucleotide bias. Extreme nucleotide bias can propagate as amino acid bias and lead to artefacts such as long-branch attraction (Williams et al., 2010). Concatenation of sequences from genes of different sizes into a supermatrix is a long-standing phylogenetic approach that strengthens overall phylogenetic signals. Certain precautions need to be taken (as our discordance filtering has done) to exclude genes with aberrant phylogenetic signals (such as those from cases of horizontal gene transfer). Longer proteins bring more phylogenetic information content and therefore rightly contribute more to the final phylogeny through their relative occupancy of the supermatrix. Thus the use of genes of different sizes is not in itself strongly biasing.

Although we aimed to avoid genomes with extremes of gene content, there was still a range of gene content among the selected genomes. The 98 selected proteins, whose functions can be considered as housekeeping, were chosen for our study on the basis of their universality. Extremely reduced genomes, such as those of less than 300 kb (as found in some insect endosymbionts), can even begin to shed the less essential of the housekeeping genes. Some of the genomes we used were small, but none reached this extreme reduction.

Within each group, genomes were scored based on their protein count (gP) and DNA G+C content (gG). First, target values were set for the group. A moderately high target protein count (tP) was set at halfway between the...
Acidithiobacillus classis nov. in the phylum Proteobacteria

mean (mP) and top count for the group. A target DNA G+C content (tG) was set at halfway between the mean (mG) and top DNA G+C content for the group if the mean was below 0.47, otherwise it was set at halfway between the mean and minimum DNA G+C content; this scheme aimed for balanced DNA G+C content to avoid compositional attraction phenomena during phylogenetic analysis. These targets were slightly incremented (by 0.001%) to systematically break ties in two-genome groups. Genomes were then scored for their approximation to the target values tP and tG. A protein score (P) was taken for each genome as (gP−tP)/(mP−tP) and a DNA G+C score (G) was taken as (gG−tG)/(mG−tG), where lower scoring genomes were closer to the targets. A combined score was taken as G+0.8×P. The lowest scoring genome with fewer than 30 contigs was chosen to represent each group, except that the classical Escherichia coli K-12 MG1655 genome replaced the lowest scoring genome with fewer than 30 contigs was chosen to represent each group, except that the classical Escherichia coli K-12 MG1655 genome replaced the lowest scoring genome in its group. The final 129 genomes comprised 124 ingroup bacterial genomes, representing each available bacterial order, and five outgroup archaeal genomes.

Phase I genomes included only one member of the order Acidithiobacillales. Phase II added the six additional genomes of members of the genus Acidithiobacillus available on 1 November 2012, taken from the Pathosystems Resource Information Center (PATRIC) website (http://www.patricbrc.org/; Gillespie et al., 2011).

In phase I with the 129-genome dataset, 340245 of all 426708 proteins were placed into 33173 non-singleton families using OrthoMCL (Li et al., 2003). Only five families were perfectly single-copy, so a relaxed single-copy criterion was employed to gain more families. Families were rejected either if they had more than seven bacterial ingroup genomes lacking that protein, or if they had more than seven ingroup genomes that were multi-copy for that protein, leaving 111 families. These families were subjected to discordance filtering to remove those with the most anomalous phylogenetic signals (Williams et al., 2010), leaving 98 families (Table S2). For the cases among these families where a genome had more than one protein in that family, all copies of the protein were rejected for that genome; this left 11949 proteins. The preferred amino acid substitution matrix was determined for each family from among the 22 available from RAxML v7.2.6 (Stamatakis and Alachiotis, 2010). The LG matrix was preferred for 83 of the families, and six additional matrices accounted for the remaining 15 families.

In phase II, protein sequences from the six additional genomes of species of the genus Acidithiobacillus were added to the 98 families as follows. A collection strategy was employed that used proteins from all phase I genomes as queries, to ensure that the collection process did not have a priori bias towards the phase I Acidithiobacillus ferrooxidans ATCC 23270 genome. The Acidithiobacillus ferrooxidans ATCC 23270 protein set, however, was in fact the best phase I protein match for all found proteins. The incomplete genome of Acidithiobacillus sp. strain GGI-221 did not have proteins annotated, so its genome was subjected to tblastn interrogation, querying with the complete 11949-protein set from phase I. Portions of each of the 98 protein genes were found, though with many apparent frameshifts. Rather than attempting to correct these frameshifts, only the highest-scoring (against any query) reading frame was taken for each family, translated and added to the protein set. In the final supermatrix (see below), 21.9% of the characters were missing for strain GGI-221, compared with 0.04% for Acidithiobacillus ferrooxidans ATCC 23270. For the last five new genome sequences of species of the genus Acidithiobacillus, their proteins were subjected to Blastp, querying with the 11949-protein set of phase I, and the highest-scoring (against any query) was taken if it scored higher than 40% of the bitscore of the perfect match to that query. Only five of the 5×98 searched proteins were not found.

For each of the 98 protein families, sequences were aligned with Muscle v3.8.31 (http://www.drive5.com/muscle/downloads.htm; Edgar, 2004a, b), and the alignments were masked with Gblocks v0.91b (Talavera & Castresana, 2007), using intermediate gap tolerance, to remove unreliably aligned regions. The masked alignments were concatenated into a supermatrix. RAxML v7.2.8 (Stamatakis, 2006) was run in a mixed model mode where protein sets were partitioned according to their preferred substitution matrix. First the maximum-likelihood tree was generated in Protdigamma mode, then 150 bootstrap trees were produced in the rapid bootstrapping mode, according to the autoFC bootstopping criterion (Stamatakis et al., 2008).

The phase I analysis using the type strain of Acidithiobacillus ferrooxidans as the representative of the order Acidithiobacillales clearly excluded it from the classes Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, Deltaproteobacteria and Epsilonproteobacteria and from Mariifundus ferrooxidans, representing the class ‘Zetaproteobacteria’. The phase II analysis, using all seven genomes of species of the genus Acidithiobacillus, recapitulated this result and showed the genomes of the order Acidithiobacillales to represent a coherent clade (Fig. 1). On this basis, we propose that the members of the order Acidithiobacillales represent a novel, distinct class of the phylum Proteobacteria, Acidithiobacillia classis nov. Three nodes marked in Fig. 1, each with 100% support, demonstrate that the members of the class Acidithiobacillia classis nov. are: (i) united, (ii) distinct from both the class Betaproteobacteria and the class Gammaproteobacteria and (iii) distinct from the remaining proteobacterial classes. The nearby position of the class ‘Zetaproteobacteria’ shows some uncertainty (five bootstraps had the class ‘Zetaproteobacteria’ sister with the class Alphaproteobacteria and two had the class ‘Zetaproteobacteria’ basal to the Alphaproteobacteria/Betaproteobacteria/Gammaproteobacteria/Acidithiobacillia groups); the ‘Zetaproteobacteria’ clade was, however, always distinct from the Acidithiobacillia clade.
It is noteworthy that the genus *Acidithiobacillus* is phylogenetically broad. The *Acidithiobacillus caldus/Acidithiobacillus ferrooxidans* intragenus phylogenetic distance is nearly the same (90%) as that between representatives of the orders *Pasteurellales* and *Enterobacteriales*.

Our conclusion from this multiprotein comparative analysis, using genomes taken from across the whole of the domains *Bacteria* and *Archaea*, is that the members of the order *Acidithiobacillales* have to be excluded from all previously defined proteobacterial classes, and represent a new class of the phylum *Proteobacteria*, *Acidithiobacillia* class nov.

**Emended description of the class**

**Gammaproteobacteria Garrity et al. 2005**

Class III of the phylum *Proteobacteria* (Gammaproteobacteria) as designated by Garrity *et al.* (2005a) needs to be emended by deletion of the order *Acidithiobacillales* from the list of constituent orders. The order *Pseudomonadales* remains the type order of the class.

**Description of Acidithiobacillia classis nov.**

*Acidithiobacillia* (A.c.i.di.thio.ba.cil’i.a. N.L. n. Acidithiobacillus type genus of the order Acidithiobacillales, the type order of the class; -ia suffix used to designate a class; N.L. neut. pl. n. Acidithiobacillia the class of the order Acidithiobacillales).

This designation is consistent with the Rules 8 and 15 of the International Code of Nomenclature of Bacteria (the Bacteriological Code; Ezby, 2007; Lapage *et al.*, 1992), which requires a class name to be created by the addition of ‘-ia’ to the stem of the type genus of the type order of the class. The type order is *Acidithiobacillales*, which contains the two families, *Acidithiobacillaceae* and *Thermithiobacillaceae*. The type genus of the order remains *Acidithiobacillus*, and the descriptions of the two families are as given in Bergey’s Manual of Systematic Bacteriology (Garrity *et al.*, 2005b, c; Kelly and Wood, 2005a, b), with the addition of *Acidithiobacillus ferrivorans* to the genus *Acidithiobacillus* (Hallberg *et al.*, 2010). The class *Acidithiobacillia* becomes the seventh class of the phylum *Proteobacteria*, together with the Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, Deltaproteobacteria, Epsilonproteobacteria and Zetaproteobacteria.

**Acknowledgements**

K. P. W. was supported by a DOE Early Career Laboratory Directed Research and Development award. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the US Department of Energy’s National Nuclear Security Administration under contract DE-AC04-94AL85000. We are grateful to editors of the IJSEM for advice on the Bacteriological Code and the correct etymology to describe the new class.

**References**


