Phylogeny of the class Actinobacteria revisited in the light of complete genomes. The orders ‘Frankiales’ and Micrococcales should be split into coherent entities: proposal of Frankiales ord. nov., Geodermatophilales ord. nov., Acidothermales ord. nov. and Nakamurellales ord. nov.

Arnab Sen,1 Vincent Daubin,2 Danis Abrouk,3 Isaac Gifford,4 Alison M. Berry4 and Philippe Normand3

1NBU Bioinformatics Facility, Department of Botany, University of North Bengal, Siliguri, 734013, India
2Biométrie et Biologie Evolutive, Centre National de la Recherche Scientifique UMR 5558, Université Lyon I, Université Lyon, Villeurbanne, France
3Ecologie Microbienne, Centre National de la Recherche Scientifique UMR 5557, Université Lyon I, Université Lyon, Villeurbanne, France
4Department of Plant Sciences, University of California, One Shields Avenue, Davis, CA 95616, USA

The phylogeny of the class Actinobacteria remains controversial, essentially because it is very sensitive to the choice of dataset and phylogenetic methods. We used a test proposed recently, based on complete genome data, which chooses among candidate species phylogenies based on the number of lateral gene transfers (LGT) needed to explain the diversity of histories among gene trees for a set of genomes. We used 100 completely sequenced genomes representing 35 families and 17 orders of the class Actinobacteria and evaluated eight different hypotheses for their phylogeny, including one based on a concatenate of 54 conserved proteins present in single copy in all these genomes, trees based on 16S and 23S rRNA gene sequences or their concatenation, and a tree based on the concatenation of MLSA genes (encoding AtpI, GyrA, FtsZ, SecA and DnaK). We used Prunier to infer the number of LGT in 579 proteins (different from those used to build the concatenated tree) present in at least 70 species, using the different hypothetical species trees as references. The best tree, with the lowest number of lateral transfers, was the one based on the concatenate of 54 proteins. In that tree, the orders Bifidobacteriales, Coriobacteriales, ‘Corynebacteriales’, ‘Micromonosporales’, ‘Propionibacteriales’, ‘Pseudonocardiales’, Streptomycetales and ‘Streptosporangiatales’ were recovered while the orders ‘Frankiales’ and Micrococcales were not. It is thus proposed that the order ‘Frankiales’, which has an effectively but not validly published name, be split into Frankiales ord. nov. (type family Frankiaceae), Geodermatophilales ord. nov. (Geodermatophilaceae), Acidothermales ord. nov. (Acidothermaceae) and Nakamurellales ord. nov. (Nakamurellaceae). The order Micrococcales should also be split into Micrococcales (genera Kocuria, Rothia, Micrococcus, Arthrobacter, Tropheryma, Microbacterium, Leifsonia and Clavibacter), Cellulomonales (Beutenbergia, Cellulomonas, Xylanimonas, Jonesia and Sanguibacter) and Brachybacteriales (Brachybacterium) but the formal proposal for this will have to wait until more genomes become available for a significant proportion of strains in this order.

Abbreviations: DPG, diphosphatidylglycerol; LGT, lateral gene transfer; LL-DAP, LL-diaminopimelic acid; MLSA, multi-locus sequence analysis; MN, multicopy non-universal; MU, multicopy universal; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; PI, phosphatidylinositol; PIM, phosphatidyl-inositol mannoside; UN, unicyclic non-universal; UU, universal unicyclic.

Three supplementary tables and five supplementary figures are available with the online version of this paper.
INTRODUCTION

Ray fungi, called Strahlenpilze by Lieske (1921), or actinobacteria as they are now designated, were historically considered as 'intermediary' between fungi and bacteria with a difficult taxonomy (Krassilnikov, 1941). The name itself comes from their colony morphology on agar that exhibits radial growth, a characteristic shared with fungi. They also share with fungi their mycelial shape and musty smell, and with bacteria physiological traits such as a cell-wall structure based on peptidoglycan, in common with members of the phylum Firmicutes as revealed by the Gram stain. The first taxonomical treatment of actinobacteria was that of Buchanan (1917) who had proposed to name an order Actinomycetales, although with features of debatable validity ("Mold like, not typically water forms, nor with the sheath impregnated with iron, true branching often evident..."). A few taxonomical works have been based on a perceived evolutionary trend from simple forms to more complex features such as hyphae and sporangia (Kluyver & van Niel, 1936). However many authors have argued that the search for a reliable taxonomy was vain given the numerous instances of loss of a function or of a morphological feature that would result in faulty positioning, especially in classifications essentially based on a dichotomous scheme.

Chemical criteria were initially too few to provide a solid taxonomy, yet they have accumulated, providing eventually a solid set of data that could compensate for a few lost functions. Nowadays, to describe a species, it is required to provide a measure of the DNA G+C content that is a characteristic of the genome; an analysis of the phospholipids that are characteristic of the membrane; the diaminopimelic acid that is a characteristic of the cell wall; or the quinones that are a characteristic of the respiratory chain. None of these elements taken individually is sufficient to identify a microbe; however, taken as a whole they can yield a solid taxonomic basis in combination with morphology and growth characteristics. Sokal & Sneath (1963) used a combination of such characters as the basis of numerical taxonomical treatments. These approaches have permitted the classification of many actinobacterial lineages starting in 1967 (Goodfellow, 1967), in particular the genera Streptomyces (Williams et al., 1983) and Mycobacterium (Tsukamura et al., 1969).

The advent of molecular phylogenetic tools, 16S rRNA gene cataloguing (Stackebrandt & Woese, 1981) and later 16S rRNA gene sequencing crowned the search for a molecular clock that needed to be present in all lineages, retained the same function and ticked at comparable speed in all these lineages. These techniques resulted in the actinobacteria emerging as a coherent subdivision in the phylum Gram-Positive bacteria, with the genera Bacillus and Clostridium (Woese, 1987). This confirmation buttressed the class Actinobacteria as coherent, with short genetic distances, a generally aerobic metabolism and branched filaments with the exception of what are called the deeper branches, namely the members of the order Bifidobacteriales (Woese, 1987). This treatment was retained in many successive editions of Bergey's Manual and the class name was published in 1997 (Stackebrandt et al., 1997) although invalidly because no type order had been proposed at the time (Euzéby & Tindall, 2001).

However, 16S rRNA genes and the resulting phylogenetic reconstructions cannot be considered the golden standard of bacterial taxonomy for several reasons. First, there are instances of species whose genomes contain more than one copy of the 16S rRNA gene differing by as much as 6%, as in Thermomonospora chromogena (Yap et al., 1999), which was interpreted as evidence of lateral transfer. Secondly, the 16S rRNA gene has also been shown to be plasmid-borne in Bacillus megaterium (Kunnimalaiyan et al., 2001), further supporting the idea it could be transferred laterally. Finally, the 16S rRNA gene is a single marker that thus contravenes one basic tenet of biology, that any analysis should be assessed for reproducibility. This limitation led to multi-locus sequence analysis (MLSA) developed initially for the genus Neisseria (Maiden et al., 1998), where typically five conserved genes or more are sequenced. This approach is now frequently used, mostly to characterize species within a genus. It has led to the re-characterization of the genera Nocardia (McTaggart et al., 2010) and Streptomyces (Doroghazi & Buckley, 2010), with significant deviations from the topology yielded by 16S rRNA gene sequences, indicative of widespread recombination.

The most recent taxonomical treatment of the phylum Actinobacteria (Ludwig et al., 2012a, b), based essentially on the 16S rRNA gene phylogeny, has considerably modified all levels of their taxonomy. There are now six classes comprising five basal ones (Acidimicrobia, Coriobacteria, Nitriliruptorida, Rubrobacteria and Thermoleophilia) each having only one or two orders, and the main class Actinobacteria that comprises 15 orders (Actinomycetales with 1 family-5 genera, ‘Actinopolysporales’ 1-1, Bifidobacteriales 1-7, ‘Catenulisporales’ 2-2, ‘Corynebacteriales’ 6-13, ‘Frankiales’ 6-11, ‘Glycomycetales’ 1-3, ‘Jiangellales’ 1-2, ‘Kineosporales’ 1-3, Micrococcaceae 15-92, ‘Micromonosporales’ 1-23, ‘Propionibacteriales’ 2-18, ‘Pseudonocardiales’ 1-22, Streptomycetales 1-3 and ‘Streptosporangiales’ 3-22).

Complete genomes were initially obtained to gain a comprehensive understanding of the functioning of cells. The first bacterial genome was published in 1995 (Fleischmann et al., 1995) and the rapid increase since then of sequencing capabilities has resulted in the present (December 2013) tally of 4536 published finished bacterial genomes (http://www.ncbi.nlm.nih.gov/Genome?term=bacteria), of which 16% or about 731 are classified in the class Actinobacteria. The first study that aimed to use these complete genomes to revisit bacterial phylogeny resulted in marked changes relative to the 16S rRNA gene-based phylogeny, especially in the class Actinobacteria where many orders were shown to be unsupported by a concatenate of 31
universal proteins. Gao and Gupta (2012) used a similar approach with a concatenate of 35 proteins to re-examine actinobacterial phylogeny, adding a study of indels to confirm the main lines of that topology (Wu et al., 2009), an approach also followed specifically for the class Actinobacteria that showed the orders ‘Frankiales’ and Micrococcaceae were in need of reassessment (Verma et al., 2013). Beyond subsampling such as the bootstrap approach (Felsenstein, 1985), it is nevertheless what the Prunier approach is designed to do, using a large number of proteins to assess the reliability of nodes of topologies obtained with other proteins (Abby et al., 2010). It was decided to use this approach on 100 actinobacterial genomes representing the main actinobacterial lineages.

**METHODS**

**Selection of genomes.** Amino acid sequences of all protein-coding genes as well as DNA sequences of 16S and 23S rRNA genes were downloaded from the JGI-IMG (http://img.jgi.doe.gov/cgi-bin/w/main.cgi) and Genoscope (http://www.genoscope.cns.fr/spip) databases. One hundred actinobacterial genomes were chosen for the study and care was taken to include most of the orders of the class Actinobacteria. A list of all genomes used in the present study and the resulting statistics are presented in Table S1 (available in the online Supplementary Material).

**Species biotopes.** Actinobacteria can thrive under various environmental conditions and ecological niches. Based on the predominant lifestyle for a given taxon, the selected actinobacterial genera were assigned to seven different biotopes. These are water, plant, thermal, mammal, soil, arthropods and extremophiles (other than thermal). A letter code was assigned for each biotope that was later used in the phylogenetic trees.

**Reconstruction of phylogenetic trees.** Eight different phylogenetic trees were used as reference to infer lateral gene transfer (LGT): 1) based on the 16S RNA genes, 2) based on the 23S rRNA genes, 3) based on concatenation of 16S and 23S rRNA genes, 4) based on concatenation of five MLSA genes (encoding AtpI, GyrA, FtsZ, SecA and DnaK) and 5) an artificial tree based on the present classification of Bergey’s Manual. All these trees were reconstructed using RAxML (Randomized Axelerated Maximum-likelihood; see below). Finally, three trees based on concatenation of 54 proteins universal in members of the class Actinobacteria were reconstructed, 6) with maximum-likelihood in MEGA software with a distance-based initial tree in MEGA, 7) with MrBayes, and 8) with RAxML.

Maximum-likelihood phylogenetic analyses were performed with RAxML 7.2.8 to reconstruct 16S and 23S rRNA gene trees (Figs S1 and S2, respectively) keeping the number of bootstraps [N] at 1000 and using the substitution model [m] GTRGAMMAI [General Time Reversible model of nucleotide substitution with the Γ model of rate heterogeneity (Yang, 1993) and estimate of proportion of invariable sites].

The 16S and 23S rRNA genes were concatenated using Seaview version 4.3.1. (Gouy et al., 2010). The concatenated sequences were then aligned with software package MUSCLE (Edgar, 2004a, b). RAxML 7.2.8 was then used to reconstruct the tree (henceforth to be designated ‘16S–23S’ tree) (Fig. S3) with the above-mentioned parameters.

To make the multi-protein-coding-gene concatenated tree (henceforth called ‘concatenated-RAxML’ tree), all the genes of all studied genomes were clustered into one file on which reciprocal BLAST was performed with BLASTp of blastall version 2.2.25, with Block Substitution Matrix (BLOSUM62) and where cost and gap extension values [G & E] were set to 11 and 1, respectively with an e-value of 1e-64. The number of database sequences to show one-line descriptions [-v] was kept at 10000 and the number of database sequences to show alignments [-b] was set at 10000.

The output of BLAST was subjected to clustering of homologous sequences using the SiLiX software (http://libe.univ-lyon1.fr/SiLiX; Miele et al., 2011) with the minimum per cent identity to accept BLAST hits for building families [-ident] set at 35, the minimum per cent overlap to accept BLAST hits for building families [-overlap] set at 80, the minimum length to accept partial sequences in families [-l] set at 100 and the minimum per cent overlap to accept partial sequences in families [-m] set at 50.

In this process several gene families were generated, which were classified into four groups: 1) UU families (unicony universal families, containing genes present in one copy in each genome and present in all the 100 genomes); 2) UN families (unicony non-universal families, same as above but not present in all genomes); 3) MU families (multicycopy universal families, present in all genomes but as more than one copy), and 4) MN families (multicycopy non-universal families, more than one copy and not present in all genomes). An in-house python (this study) program was developed which short-listed the gene families into various groups.

There were ultimately 54 UU families and around 5000 UN families recovered. Since the number of UN families was fairly large, the number of families was determined by taking only those families that were present in more than 70 genomes but not in all 100 studied genomes. This permitted the selection of 579 UN families. The 54 gene UU families were concatenated with Seaview version 4.3.1 and aligned with software CLUSTAL Omega version 1.1.0 (Sievers et al., 2011) with default options. A tree was reconstructed with RAxML using this aligned file with the parameters described above but with a substitution model [m] of PROTACAT_GAMMAIWAG -f a (Fig. 1). We also reconstructed trees using another widely used program, MEGA with default parameters (Tamura et al., 2013), except the WAG substitution model and Gamma Distributed as in the RAxML tree (Fig. S4); and MrBayes (Bayesian Analysis of Phylogeny) with the GTR model and gamma-distributed rate variation across sites and a proportion of invariable sites (Ronquist et al., 2012) (Fig. S5).

To make an MLSA amino acids tree (Margos et al., 2009), housekeeping genes in different gene families were searched: one gene (encoding membrane-bound ATP synthase, F1 sector, Atpl) was found in UU genomes and four families in MU families (encoding DNA gyrase, GyrA, tubulin-like GTP-binding protein, FtsZ, ATPase secretory preprotein translocase, SecA and chaperone Hsp70 in DNA biosynthesis, DnaK). To choose a gene from each strain from MU families (genes of these type of families present in all genomes but as more than one copy), the gene of Atopobium parvulum (DSM 20469) was taken as reference and blasted with the rest of the genes of the same family. Ultimately, five MLSA gene families were used to make a tree in a manner similar to that described for the ‘concatenated’ tree (Fig. 2).

The last tree, which was called ‘artificial’, was reconstructed on the basis of 16S rRNA genes and conventional knowledge that corresponds to the latest taxonomical treatment in Bergey’s Manual of Systematic Bacteriology (2nd edition). There were actually some deviations in the 16S rRNA gene tree (Fig. S1) generated in the present study from Bergey’s taxonomy and those changes were artificially corrected in the ‘artificial’ tree.

**Prunier.** To identify and quantify putative LGTs among the strains and their evolutionary role, and to determine the robustness of various trees, a program called ‘Prunier’ (Abby et al., 2010) was used.
Prunier essentially reconciles a gene tree with a reference (species) tree and searches for a maximum statistical agreement forest (MSAF) between them. The 'fast' Prunier used in the present study uses LR-ELW edge support values (Strimmer & Rambaut, 2002) to identify discrepancies between gene and species trees. Prunier infers LGT events when it finds significant topological conflicts between the two trees. The following settings were used for Prunier: boot.thresh.conflict=90 (i.e. support value threshold for topological conflict); fwd.depth=2 (i.e. maximal depth at which Prunier looks forward to find a significant LGT when the current LGT is not significant). The depth value was set at 2, which means that if significantly supported disagreement cannot be removed with one LGT event, Prunier will go two steps (one at a time) 'forward' to determine whether the next LGT in the list will remove a significant conflict. If Prunier finds a better solution with a depth value of 2 or 3, it provides that solution (this indicates Prunier should compute an LGT scenario for each possible root and was used because the real roots of the trees were not certain).

The 'boot.thresh.conflict' denotes the lowest LR-ELW edge support value for a given node in the gene tree used for reconciling its topological conflict species tree, whereas the 'fwd.depth' stands for the maximal depth at which the Prunier (fast) looks forward to find a significant LGT when the current event is not significant. The depth value was set at 2, which means that if significantly supported disagreement cannot be removed with one LGT event, Prunier will go two steps (one at a time) 'forward' to determine whether the next LGT in the list will remove a significant conflict. If Prunier finds a better solution with a depth value of 2 or 3, it provides that solution in the output. For further details about Prunier please refer to http://phl.univ-lyon1.fr/software/prunier.

RESULTS

Eight trees were generated using different genes or gene concatenates with contrasted topologies, two of which will be discussed in detail.

Topology of the concatenated-RAXML tree

All the trees generated were rooted following Gao & Gupta (2012) and Embley & Stackebrandt (1994) with the order Coriobacteriales including Atopobium parvulum (DSM 20469) that has an obligately anaerobic metabolism, as do all members of the clade (Fig. 1). The branch leading to Conexibacter woesei and Rubrobacter xylanophilus, both of which are aerobic, emerges next.

The next branch to emerge is that of the genus Frankia with cluster 2 at the root, followed by cluster 4, cluster 3 and then clusters 1a and 1b as most derived. The other genera traditionally grouped with the genus Frankia within the 'Frankiales' - Geodermatophilus, Modestobacter, Blastococcus, Acidothermus and Nakamurella were not anywhere close to the genus Frankia. Bootstraps were high that positioned these families in groups away from the genus Frankia (Figs 1 & 2).

Then emerges a dichotomy with the orders 'Glycycycteales', 'Micromonosporales', 'Pseudonocardiales', 'Corynebacteriales' and the families Nakamurellaceae and Geodermatophilaceae formerly in the 'Frankiales' on the one hand and on the other the orders Microcccales, Bifidobacteriales, Actinomycetales, 'Propionibacteriales', 'Catenulisporales', Streptomyceales, 'Streptosporangiales' and families Dermabacteraceae, Kineosporiaceae, Dermacoccaceae, Intrasporangiaceae and Acidothermaceae. The genus Acidothermus was found to group closely to thermophiles in the order 'Streptosporangiales', supported by a good (91 %) bootstrap result (Fig. 1). All orders described in the latest version of Bergey's Manual (Ludwig et al., 2012b), were recovered with the exceptions of the orders 'Frankiales' and Microcccales.

In the order 'Corynebacteriales', the genera Corynebacterium, Rhodococcus and Mycobacterium were all recovered. In the order Bifidobacteriales, the genus Gardnerella was not outside the genus Bifidobacterium but within it. All other genera with more than one genome studied were recovered as coherent.

Topology of the MLSA tree

The topology of the MLSA tree (Fig. 2) is similar to that of the concatenated tree with the order Coriobacteriales clustered with the genera Conexibacter, Rubrobacter and Acidimicrobium, then the family Frankiaceae at the root. The position of the genus Acidothermus is different, being here close to the root.

The two sets seen with the concatenated tree (I: 'Glycycycteales', 'Micromonosporales', 'Pseudonocardiales', 'Corynebacteriales' Nakamurellaceae, Geodermatophilaceae, II: Microcccales, Bifidobacteriales, Actinomycetales, 'Propionibacteriales', 'Catenulisporales', Streptomyceales, 'Streptosporangiales', Dermabacteraceae, Kineosporiaceae, Dermacoccaceae, Intrasporangiaceae) were recovered in the MLSA tree except for the family Actinothermaceae that was at the root immediately after the Frankiaceae branch.

All genera with more than one genome analysed (Arthrobacter, Bifidobacterium, Corynebacterium, Frankia, Mycobacterium, Nocardiosis, Propionibacterium, Rhodococcus, Rothia, Streptomyces) were recovered as coherent with the exception of the genus Bifidobacterium that contained the genus Gardnerella.

Quantification of LGT

Revision of phylogeny of organisms, especially bacteria using complete genome sequences reveals a degree of incongruence due to biological or methodological differences (Boussau & Daubin, 2010). Therefore it is important to verify any phylogenetic tree generated. One relatively reliable method of verification is the detection of LGT based on the search for a maximum statistical agreement forest between a gene tree and a reference tree (Abby et al.,...
Fig. 2. Phylogenetic tree reconstructed with five MLSA gene families and the RAxML program for all 100 actinobacterial strains. The colour code indicates the biotopes of the sequenced genomes. Numbers at nodes are the bootstrap results for 1000 subsamplings expressed as a percentage. Bar, 0.2 substitutions per site distance.
As described above, two sets of gene trees were generated, one set of 54 gene trees with UU families and another set of 579 gene trees with UM families with LGTs. Gene trees were then simulated from various reference trees by using the subtree pruning and regrafting operations as described by Abby et al. (2010). Precisely, eight different reference trees were used, the concatenated-RAxML tree, the concatenated-MEGA tree, the concatenated-Bayesian tree, the MLSA tree, the 23S tree, the 16S–23S tree, the 16S tree and the tree based on conventional knowledge [an artificial tree that corresponds to the latest taxonomical treatment in Bergey's Manual of Systematic Bacteriology (2nd edition)].

The highest number of transfers in various UU gene family trees ranged from 69 to 375 whereas in MU gene family trees, it ranged from 19 to 39 (for details see Table 1).

It was found that the concatenated tree (RAxML) had the lowest number of transfers (69 and 1130) across all the roots in both types of tree followed by the Bayesian tree (71 and 1136), the MLSA tree (125 and 1490), 23S (123 and 1554), 16S–23S (150 and 1870), artificial (346 and 3633) and finally the 16S tree (375 and 4039). Therefore it may be inferred that among the eight trees analysed, the concatenated-sequence tree is the most reliable one.

We also sought to detect the most reliable root. Though conventionally the order Coriobacteriales including the genus Atopobium is considered as the root, the default parameters were used where Prunier proposes a scenario for every possible root of the species tree (Abby et al., 2010) and the root where least transfer took place would be the actual root. The result of transfer in various roots is shown in Tables S2 and S3 (for UU gene family trees and MU gene family trees, respectively). It was found that the concatenated-RAxML tree had the minimum transfer (69–73 and 1140–1265) across all the roots in both type of trees, followed by the Bayesian tree (72–75 and 1132–1243), the concatenated-MEGA tree (103–122 and 1261–1376), the MLSA tree (125–130 and 1501–1622), 23S (124–131 and 1583–1682), 16S–23S (151–163 and 1897–1982), artificial (347–367 and 3665–3793) and finally the 16S tree (375 and 4039). Therefore it may be inferred that among the eight trees analysed, the concatenated-RAxML tree is the most reliable one. Concerning the position of the root, statistically negligible differences of transfer were found among various roots within the trees. It was therefore decided to continue to consider the genus Atopobium in the Coriobacteriales clade as root for interpretation of the trees.

**DISCUSSION**

Actinobacteria constitute a major group of bacteria, exceeded only by the phyla Proteobacteria and Firmicutes in the number of available complete genomes (http://www.genomesonline.org/cgi-bin/GOLD/phylogenetic_distribution.cgi; Pagani et al., 2012). The taxonomic treatment of the actinobacteria has fluctuated considerably due to technical developments, the most recent being the phylogenetic analysis based on complete genomes (Wu et al., 2009), which has yielded a topology markedly different from that obtained previously with the 16S rRNA gene. There is thus a major need to assess the reliability of the different trees. The Prunier method (Abby et al., 2010) can provide such an independent measure, and we have thus applied it to actinobacterial phylogeny.

The concatenated-RAxML tree with 1130 lateral transfers was the one with the lowest number of predicted LGTs, and thus may be regarded as the most reliable. This tree represents 54 gene families with 17642 amino acids in total. The major difference that emerged compared with previous treatments in Bergey's Manual (Ludwig et al., 2012a, b) is in the topology of the order 'Frankiales'. This order was previously described as artificial, comprising families with considerable differences in morphology, physiology or genome features (Barabote et al., 2009; Gtari et al., 2012; Normand et al., 2007). The use of complete genomes thus permits resolution of that troubling aspect by splitting the paraphyletic order, yielding at least five orders, each with a single family: Frankiales ord. nov. (family Frankiaceae), Geodermatophilales ord. nov. (Geodermatophilaceae), Nakamurellales ord. nov. (Nakamurellaceae) and Acidothermales ord. nov. (Acidothermaceae) as well as the order ‘Kineosporiales’ already proposed in 2012 (Normand & Benson, 2012a).

**Table 1.** Characteristics of the eight gene groups analysed

<table>
<thead>
<tr>
<th>Characteristic*</th>
<th>16S</th>
<th>23S</th>
<th>16S-23S</th>
<th>Concatenated RAxML</th>
<th>Concatenated MEGA</th>
<th>MLSA</th>
<th>Bayes</th>
<th>Artificial</th>
</tr>
</thead>
<tbody>
<tr>
<td>LGT (max) from 579 MU</td>
<td>39</td>
<td>26</td>
<td>27</td>
<td>22</td>
<td>22</td>
<td>19</td>
<td>22</td>
<td>27</td>
</tr>
<tr>
<td>LGT (sum) from 579 MU</td>
<td>4039</td>
<td>1554</td>
<td>1870</td>
<td>1130</td>
<td>1238</td>
<td>1490</td>
<td>1136</td>
<td>3633</td>
</tr>
<tr>
<td>LGT from 54 MU</td>
<td>375</td>
<td>123</td>
<td>150</td>
<td>69</td>
<td>103</td>
<td>125</td>
<td>71</td>
<td>346</td>
</tr>
<tr>
<td>Mean no. of positions (AA/NA)</td>
<td>1525</td>
<td>3069</td>
<td>4594</td>
<td>17642</td>
<td>17642</td>
<td>2624</td>
<td>17642</td>
<td>n.a.</td>
</tr>
<tr>
<td>Length of line-ups (AA/NA)</td>
<td>3066</td>
<td>3487</td>
<td>6553</td>
<td>25589</td>
<td>25589</td>
<td>2630</td>
<td>25589</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

*LGT, lateral gene transfer; AA, amino acids; NA, nucleic acids; n.a., not applicable.
An order for the ‘Frankiales’/Frankiaceae is the easiest to defend, given that it is positioned at the root of the whole class Actinobacteria in the RAxML-based tree (Fig. 1). Additionally, the genus Frankia occupies a unique ecological niche in the symbiotic root tissues of plants and a unique physiological function, that of biological nitrogen-fixation. Members of the genus Frankia contain a cluster of *nif* genes, plasmid-borne in at least one strain (Simonet et al., 1986), but not in all other lineages (Normand et al., 2007), the phylogeny of which positions the genus Frankia away from all other lineages (Normand & Bousquet, 1989; Normand et al., 1992). All strains analysed also have a rare sugar, 2-O-methyl-D-mannose (Mort et al., 1983). Many more strains can be hypothesized to exist in nodules but are so far only known by direct characterization in various species (Nazaret et al., 1989). It is thus proposed to create the order Frankiales ord. nov. with a single family, Frankiaceae. This order may be augmented in the coming years to include additional micro-organisms identified from soil, in a manner, similar to the finding of 16S rRNA gene sequences recovered from the rhizosphere of *Alnus* (Normand and Chapelon, 1997).

The family Geodermatophilaceae forms a coherent cluster with the orders ‘Corynebacteriales’, ‘Micromonosporales’ and ‘Pseudonocardiales’, away from the family Frankiaceae. Members of the order Geodermatophilales ord. nov. also occupy unique ecological niches, such as stone walls and desert soils, and possess unusual resistance to oxidative stress (Gtari et al., 2012). They share generic genomic features such as multiple copies of the *trwC* gene (Chouaia et al., 2012; Normand et al., 2012). The family Geodermatophilaceae is evolving rapidly with several novel species proposed recently (Montero-Calasanz et al., 2013) with isolates from desert soils and plant rhizospheres (Normand & Benson, 2012b). It is thus proposed to create the order Geodermatophilales ord. nov. with a single family, Geodermatophilaceae.

The remaining two families, Nakamurellaceae and Acidothermales, are some distance from the family Frankiaceae in the concatenated-sequence trees with the best Prunier score, and thus could be taken out of the current order ‘Frankiales’. However, the presence of a single genome in each of the two cases, contrary to Frankiaceae ord. nov. and Geodermatophilaceae ord. nov. with several genomes published, prevents defining generic features for the moment. Nevertheless, the two genomes have long branches and no specific association with any other lineages and it is thus proposed to create new orders for them: Nakamurellales ord. nov. and Acidothermales ord. nov. in the class Actinobacteria.

The cluster comprising Acidothermales ord. nov. and the order ‘Streptosporangiiales’ is quite solid with 91% bootstrap. In this cluster there are a majority of genomes that belong to thermophiles, a fact that given the strong biochemical pressures present in a high-temperature environment may have contributed to exchange of genes in their shared biotopes.

The order ‘Glycomycetales’ is a small order comprising only one family, three genera and only one available genome at present. It is thus hard to discern general characteristics of the group. Conversely, the order ‘Micromonosporales’ is a large order with 23 genera and several available genomes, isolated mostly from soil and oceanic waters. *Stackebrandtia* *nassauensis*, which has the only characterized genome of the order ‘Glycomycetales’, is a soil microbe, non-motile and produces substrate and aerial mycelia but no spores. It is aerobic, with a fatty acid pattern dominated by saturated branched-chain acids, anteiso-C₁₇₀ (26.8%), iso-C₁₅₀ (8.7%), iso-C₁₆₀ (8.7%) and iso-C₁₇₀ (9.0%). The identified polar lipids are phosphatidylglycerol (PG) and diphosphatidylglycerol (DPG), and the predominant menaquinones are MK-10(H₄), MK-10(H₆), MK-11(H₄) and MK-11(H₆) (Munk et al., 2009). Members of the order ‘Micromonosporales’ are soil and marine microbes, with several lineages recently isolated from a variety of plant tissues (Carro et al., 2013). They are motile with substrate and scant aerial hyphae, carry spores, and have a fatty acid pattern dominated by saturated branched-chain acids, iso-C₁₅₀ (anteiso-C₁₅₀ and anteiso-C₁₇₀). The identified polar lipid is phosphatidylethanolamine (PE), and predominant menaquinones are all types of the MK-9 and MK-10 series. There is therefore little in common between the two orders. However, the apparent phylogenetic proximity between the two orders may need to wait for further chemical and genomic data before the two orders are fused and an *incertae sedis* status in the class Actinobacteria is proposed.

The order ‘Catenulisporales’, which contains only two genera (*Actinospica* with two species and *Catenulispora* with five species) is systematically associated with the order Streptomycetales (three genera, 561 species), in the 16S rRNA gene tree, the MLSA tree and the concatenated tree. At the same time, the order ‘Catenulisporales’ is characterized by the presence of LL-diaminopimelic acid (LL-DAP), the predominance of iso-C₁₆₀ and anteiso-C₁₇₀ fatty acids, MK₉-(H₆), -(H₈) and -(H₄) isoprenoids, PG, DPG, phosphatidylglycerol (PG) and phosphatidylinositol (PI) and phosphatidylinositol mannosides (PIM) as predominant phospholipids, aerial hyphae and chains of cylindrical arthrospores (Donadio et al., 2012) while the order Streptomycetales is characterized by the presence of LL-DAP and meso-DAP, MK₉-(H₆), -(H₈) and -(H₄) isoprenoids, DPG, PE, PI and PIM as predominant phospholipids, aerial hyphae and chains of arthrospores (Kämpfer, 2012a, b). On balance, the inclusion of ‘Catenulisporales’ within the order Streptomycetales is thus to be contemplated.

The order Micrococcales is not recovered intact in the concatenated tree, rather it is found with three anomalous groups: order Bifidobacteriales, order Actinomycetales and order ‘Kinesporales’. Since the physiology and ecology of the anaerobic bifidobacteria are quite distinct from those of the micrococcales beyond a preponderance of mammal-inhabiting strains, reuniting the order Micrococales
with the other three orders should be avoided. Instead, this order should be split into several monophyletic orders: one that would keep the name Micrococcaceae with the genera Micrococcus, Kocuria, Rothia, Arthrobacter, Tropheryma, Microbacterium, Clavibacter and Leifsonia; one that could be designated Cellulomonadales and would contain the genera Beutenbergia, Cellulomonas, Xylanimonas, Jonesia and Sanguibacter; and finally the orders Brachybacteriales (Brachybacteriaceae) and Dermacoccales (Kytococcus and Intrasporangium); these latter two orders being difficult to defend at this time, because only one or two genomes have been analysed so far.

Actinobacteria have colonized several ecological niches beyond the soil, where they are known to be enriched. Among the 100 genomes analysed here, less than a third (28) are considered as soil dwellers, while another third (33) have been isolated from mammalian hosts. There is certainly a bias in these data due to the economic and social importance of disease-causing microbes, but this can modify the perspective on soil as the emblematic biotope of actinobacteria. Soil actinobacteria can nevertheless be credited as possessing relatively large genomes, slow growth rates, rich secondary metabolism and high rates of gene exchanges.

In the concatenated-RAxML tree, it was found that many of the species sharing the same biotope were grouped. For instance, all Frankia species which were considered plant-related were together. In case of the order ‘Micromonosporales’, two water-inhabiting species, Salinispora tropica and Verrucosimpora maris, were phylogenetically close together with two other soil-inhabiting species, i.e. Actinoplanes missouriensis and Micromonospora aurantiaca. Similarly, in case of the order ‘Pseudonocardiales’, out of five species studied, one water-inhabiting (Pseudonocardia dioxanivorans) and one mammal-inhabiting (Saccharomonospora viridis) were distantly positioned in two parts of the branch. In the case of the order ‘Corynebacteriales’, out of seven Corynebacterium species, except Corynebacterium variabile (which is a plant inhabitant), all species were mammal-inhabiting and Corynebacterium variabile clearly appears as a side branch. Likewise, in case of the order Micrococcales I, ‘Tropherymona whipplei’, which is a mammal-inhabiting species, appears clearly as a distant species whereas plant-inhabiting Microbacterium testaceum, Clavibacter michiganensis and Leifsonia xylei were clustered. Lastly in the case of the order Bifidobacteriales, out of six Bifidobacterium species, five were mammal-inhabiting along with Gardnerella vaginalis, and Bifidobacterium asteroides was arthropod-related. In the concatenated tree, all the mammal-inhabiting Bifidobacterium species grouped along with Gardnerella vaginalis, and clearly positioned at the root of the order.

From a technical point of view, it may be mentioned that out of two multi-gene concatenated trees (concatenated-RAxML tree and concatenated-MEGA tree), the concatenated-RAxML tree reconstruction approach has been found to perform better in terms of LGT as evident from analysis of the Prunier data. It may therefore be concluded that the RAxML tree is most accurate among all trees studied at least from the Prunier test.

The status of the families Cryptosporangiaceae and Sporichthyaceae in the absence of genome data is uncertain within the class Actinobacteria and will have to wait. There is little doubt that the rate of acquisition of genome sequences from isolates and of metagenomes from various biotopes will continue unabated in the coming years, and given the diminishing costs of sequence acquisition and ease of use of computer analytical tools, all described species will soon have a genome sequence available in the databases. This should result in a generalization of approaches such as the Prunier (Abby et al., 2010) to buttress phylogeny and identify the details of genes’ evolution. Actinobacteria have been postulated to have been the originators of innovations such as synthesis of cholesterol (Lamb et al., 1998), the ability to recycle proteins through a proteasome (De Mot, 2007), and the presence of suspended metabolism structures such as exosomes, which together have led to the suggestion that actinobacteria were both more ancient than generally considered and possible ancestors of archaea and eukaryotes (Cavalieri-Smith, 2002). If the present study does not touch on these aspects, it nevertheless supports a position for the phylum Actinobacteria as ancient, diversified and prone to frequent gene exchange.

Description of Frankiales ord. nov.

Frankiales [Frank.i.a’les. N.L. masc. n. Frank a German botanist (1839–1900) who studied root symbioses; suff. -ales ending to denote order; N.L. fem. pl. n. Frankiales the Frankia order].

Cells are Gram-positive, non-motile, form multi-locular sporangia, most form thick-walled vesicles, fix nitrogen and establish root nodules on actinorrhizal plants. A member of the class Actinobacteria. Contains a single family, Frankiaceae (Stackebrandt et al., 1997). The type genus is Frankia. Phylogenetic analyses have been published previously (Normand & Bousquet, 1989; Normand et al., 1996). The name of the order including families Frankiaceae, Geodermatophilaceae, Nakamuraellaceae, Sporichthyaceae, Acidothermaceae and Cryptosporangiaceae was effectively published in the latest Bergey’s Manual (Normand & Benson, 2012b).

Description of Geodermatophilales ord. nov.

Geodermatophilales (Ge.o.der.ma.to.phil.a’les. N.L. fem. pl. n. Geodermatophilus type genus of the order; suff. -ales ending to denote order; N.L. fem. n. Geodermatophilales the Geodermatophilus order].

Cells are Gram-positive, aerobic, have motile and non-motile forms, inhabit stone surfaces and interiors, and are pigmented. A member of the class Actinobacteria. Contains
a single family, *Geodermatophilaceae* (Normand, 2006). The type genus is *Geodermatophilus*, there are two other genera, *Modestobacter* and *Blastococcus*. A phylogenetic analysis has been published previously (Normand et al., 1996).

**Description of Acidothermales ord. nov.**

*Acidothermales* (Acido.therm.a.les. N.L. masc. n. Acidothermus type genus of the order; suff. -ales ending to denote order; N.L. fem. pl. n. Acidothermales the Acidothermus order).

Cells are Gram-positive, non-motile and thermo-resistant (55 °C). A member of the class *Actinobacteria*. Contains a single family, * Acidothermaceae* (Stackebrandt et al., 1997). A phylogenetic analysis has been published previously (Barabote et al., 2009).

**Description of Nakamurellales ord. nov.**

*Nakamurellales* (Na.ka.mu.rel.la’les. N.L. fem. n. Nakamurella type genus of the order; suff. -ales ending to denote order; N.L. fem. pl. n. Nakamurellales the Nakamurella order).

Cells are Gram-positive, non-spore-forming, coccus-shaped, non-motile and inhabit soil. A member of the class *Actinobacteria*. Contains a single family, *Nakamurellaceae* (Stackebrandt et al., 1997). The type genus is *Nakamurella*; there are two other genera, *Humicoccus* and *Saxebacter*. A phylogenetic analysis has been published previously (Kim et al., 2012).

**ACKNOWLEDGEMENTS**

A S acknowledges the receipt of the DBT-CREST award. Thanks are expressed to the French Research Agency ANR (Sesam, ANR-10-BLAN-1708) and to the France-Berkeley-Fund (collaborative grant to A. M. B. and P. N.). This work was partially funded by a UGC project (BLAN-1708) and to the France-Berkeley-Fund (collaborative grant to A. M. B. and P. N.). This work was partially funded by a UGC project (BLAN-1708) and to the France-Berkeley-Fund (collaborative grant to A. M. B. and P. N.). This work was partially funded by a UGC project (BLAN-1708) and to the France-Berkeley-Fund (collaborative grant to A. M. B. and P. N.).

**REFERENCES**


