University of Groningen

Genome-based phylogenetic analysis of *Streptomyces* and its relatives

Alam, Mohammad Tauqueer; Merlo, Maria Elena; Takano, Eriko; Breitling, Rainer

Published in:
Molecular Phylogenetics and Evolution

DOI:
[10.1016/j.ympev.2009.11.019](https://doi.org/10.1016/j.ympev.2009.11.019)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2010

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Alam, M. T., Merlo, M. E., Takano, E., & Breitling, R. (2010). Genome-based phylogenetic analysis of *Streptomyces* and its relatives. *Molecular Phylogenetics and Evolution*, 54(3), 763-772.
<https://doi.org/10.1016/j.ympev.2009.11.019>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Genome-based phylogenetic analysis of *Streptomyces* and its relatives

Mohammad Tauqeer Alam^a, Maria Elena Merlo^{a,b}, Eriko Takano^b, Rainer Breitling^{a,*}

^a Groningen Bioinformatics Center, University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands

^b Department of Microbial Physiology, University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands

ARTICLE INFO

Article history:

Received 3 March 2009

Revised 18 November 2009

Accepted 19 November 2009

Available online 3 December 2009

Keywords:

Phylogeny

Streptomyces

Actinomycetales

Rank order

rRNA

ABSTRACT

Motivation: *Streptomyces* is one of the best-studied genera of the order *Actinomycetales* due to its great importance in medical science, ecology and the biotechnology industry. A comprehensive, detailed and robust phylogeny of *Streptomyces* and its relatives is needed for understanding how this group emerged and maintained such a vast diversity throughout evolution and how soil-living mycelial forms (e.g., *Streptomyces* s. str.) are related to parasitic, unicellular pathogens (e.g., *Mycobacterium tuberculosis*) or marine species (e.g., *Salinispora tropica*). The most important application area of such a phylogenetic analysis will be in the comparative re-annotation of genome sequences and the reconstruction of *Streptomyces* metabolic networks for biotechnology.

Methods: Classical 16S-rRNA-based phylogenetic reconstruction does not guarantee to produce well-resolved robust trees that reflect the overall relationship between bacterial species with widespread horizontal gene transfer. In our study we therefore combine three whole genome-based phylogenies with eight different, highly informative single-gene phylogenies to determine a new robust consensus tree of 45 *Actinomycetales* species with completely sequenced genomes.

Results: None of the individual methods achieved a resolved phylogeny of *Streptomyces* and its relatives. Single-gene approaches failed to yield a detailed phylogeny; even though the single trees are in good agreement among each other, they show very low resolution of inner branches. The three whole genome-based methods improve resolution considerably. Only by combining the phylogenies from single gene-based and genome-based approaches we finally obtained a consensus tree with well-resolved branches for the entire set of *Actinomycetales* species. This phylogenetic information is stable and informative enough for application to the system-wide comparative modeling of bacterial physiology.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Streptomyces species are among the best-studied and best-characterized bacteria due to their significant role for medical science, ecology and the biotech industry (Lechevalier and Lechevalier, 1967; Embley and Stackebrandt, 1994; Bentley et al., 2002). *Streptomyces* have a particularly complex secondary metabolism, which produces a large collection of biologically useful compounds. Most importantly, they are employed at large industrial scale in the production of most of the available antibiotics applied in human and veterinary medicine, as well as a large number of anti-parasitic agents, herbicides, immuno-suppressants and several enzymes important in the food and other industries (Bentley et al., 2002; Cerdano-Tarraga et al., 2003; Hopwood, 2007).

The genus *Streptomyces* is taxonomically located in the diverse bacterial order *Actinomycetales*. This group is characterized by an astonishing diversity in terms of morphology, ecology, pathogenicity, genome size, genomic G+C content, and the number of coding se-

quences in the genome (Embley and Stackebrandt, 1994; Hopwood, 2007; Ventura et al., 2007). Morphologically, some species are rod-shaped or coccoid, while others form fragmenting hyphae or a branched mycelium (Ventura et al., 2007). Spore formation is also very common in *Actinomycetales* but it is not ubiquitous (Ventura et al., 2007). Ecologically, some actinomycetes are soil-living bacteria, some are marine, are colonizing thermal springs (Barabote et al., 2009) or growing on gamma-irradiated surfaces (Phillips et al., 2002) or as plant root symbionts (Normand et al., 2007), and some are important pathogens of humans, animals and plants (Goodfellow and Williams, 1983; Castillo et al., 2002; Tokala et al., 2002). For instance, *Mycobacterium tuberculosis* causes tuberculosis (Cole et al., 1998), *Corynebacterium diphtheria* infection results in diphtheria (Cerdano-Tarraga et al., 2003), *Propionibacterium acnes* is the agent of acne (Leyden, 2001) and *Streptomyces scabies* causes potato scab (Takeuchi et al., 1996). *Streptomyces* species are found mostly in the soil where they live as saprophytes (Hopwood, 2007), but recently some species have been described in the rhizosphere of plant roots and in other plant tissue (Castillo et al., 2002; Tokala et al., 2002), isolated from leaf cutting ants (Kost et al., 2007) and also associated with marine sponge species (Zhang et al., 2008).

* Corresponding author. Fax: +31 (0) 50 363 7976.

E-mail address: r.breitling@rug.nl (R. Breitling).

Considering the importance of *Streptomyces* and its relatives in terms of both biological behavior and metabolic products, it becomes essential to understand its evolutionary relationships to other species in the diverse *Actinomycetales* order. On the one hand, an evolutionary study may help to explain how *Streptomyces* emerged and adapted to the soil environment. On the other hand, information obtained from a well-resolved phylogeny can be used for the comparison of genome sequences, comparative genome re-annotation, and genome visualization. A robust phylogeny is central for ongoing efforts in many groups to reconstruct system-wide metabolic models of *Streptomyces* and related species (Borodina et al., 2005), which are used for systematic strain-engineering in biotechnology.

Currently available phylogenies of the group are based on 16S rRNA, individual genes, or comparative genomic approaches (Embley and Stackebrandt, 1994; Takeuchi et al., 1996; Stackebrandt et al., 1997; Anderson and Wellington, 2001; Egan et al., 2001; Gao and Gupta, 2005; Chater and Chandra, 2006; Manteca et al., 2006). Such reconstructions tend to be relatively unstable and

are not guaranteed to reflect the overall evolutionary history in a complex group with widespread horizontal gene transfer. To address this issue and to determine potential problematic areas in the single-gene phylogenies, we made use of whole-genome information. We not only reconstructed phylogenetic trees based on eight highly conserved single genes, including three rRNA sequences (5S, 16S, 23S rRNA) and five ubiquitous protein sequences (isoleucyl tRNA synthetase, ribosomal protein S1, SecY, GTPase, DNA topoisomerase), but also integrated three different phylogenetic reconstructions based on complete genome sequences, using gene content, gene order and gene concatenation analysis.

Our analysis is based on 45 species from eight different suborders of *Actinomycetales*, including four genome sequenced *Streptomyces* species (Table 1). *Escherichia coli* and *Bacillus subtilis* are used as outgroups, and two distantly related *Actinobacteria*, *Bifidobacterium longum* NCC2705 and *Bifidobacterium adolescentis* ATCC 15703, serve as sub-outgroups.

Our results showed that single-gene approaches did not yield a resolved phylogeny of *Actinomycetales*. Resolution is much im-

Table 1

List of species considered in the study.

Corynebacterineae	Corynebacteriaceae	<i>Corynebacterium jeikeium</i> K411	cjk
		<i>Corynebacterium glutamicum</i> ATCC 13032	cgl
		<i>Corynebacterium diphtheriae</i> NCTC 13129	cdi
		<i>Corynebacterium efficiens</i> YS-314	cef
		<i>Corynebacterium glutamicum</i> ATCC 13032_new	cgb
	Mycobacteriaceae	<i>Mycobacterium leprae</i> TN	mle
		<i>Mycobacterium tuberculosis</i> H37Ra	mra
		<i>Mycobacterium bovis</i> BCG str. Pasteur 1173P2	mbb
		<i>Mycobacterium tuberculosis</i> F11	mtf
		<i>Mycobacterium ulcerans</i> Agy99	mul
		<i>Mycobacterium gilvum</i> PYR-GCK	mgi
		<i>Mycobacterium vanbaalenii</i> PYR-1	mva
		<i>Mycobacterium</i> sp. KMS	mkm
		<i>Mycobacterium smegmatis</i> str. MC2 155	msm
		<i>Mycobacterium avium</i> 104	mav
		<i>Mycobacterium</i> sp. MCS	mmc
		<i>Mycobacterium avium</i> subsp. paratuberculosis K-10	map
		<i>Mycobacterium bovis</i> AF2122/97	mbo
		<i>Mycobacterium tuberculosis</i> CDC1551	mtc
		<i>Mycobacterium tuberculosis</i> H37Rv	mtu
		<i>Mycobacterium</i> sp. JLS	mjl
	Nocardiaceae	<i>Rhodococcus</i> sp. RHA1	rha
		<i>Nocardia farcinica</i> IFM 10152	nfa
Frankineae	Acidothermaceae	<i>Acidothermus cellulolyticus</i> 11B	ace
	Frankiaceae	<i>Frankia alni</i> ACN14a	fal
		<i>Frankia</i> sp. EAN1pec	fre
		<i>Frankia</i> sp. Cc13	fra
Kineosporiineae	Kineosporiaceae	<i>Kineococcus radiotolerans</i> SRS30216	kra
	Cellulomonadaceae	<i>Tropheryma whippelii</i> TW08/27	twc
Micrococcineae		<i>Tropheryma whippelii</i> str. Twist	twh
	Microbacteriaceae	<i>Clavibacter michiganensis</i> subsp. <i>michiganensis</i> NCPPB 382	cmi
		<i>Leifsonia xyli</i> subsp. <i>xyli</i> str. CTCB07	lxx
	Micrococcaceae	<i>Arthrobacter</i> sp. FB24	art
		<i>Arthrobacter aurescens</i> TC1	aau
		<i>Renibacterium salmoninarum</i> ATCC 3209	rsa
Micromonosporineae		<i>Salinispora arenicola</i> CNS-205	sar
		<i>Salinispora tropica</i> CNB-440	stp
Propionibacterineae		<i>Propionibacterium acnes</i> KPA171202	pac
		<i>Nocardioideis</i> sp. JS614	nca
Pseudonocardineae		<i>Saccharopolyspora erythraea</i> NRRL 2338	sen
	Streptomycineae	<i>Streptomyces avermitilis</i> MA-4680	sav
		<i>Streptomyces coelicolor</i> A3(2)	sco
		<i>Streptomyces scabies</i>	ssc
		<i>Streptomyces griseus</i>	sgr
Streptosporangineae		<i>Thermobifida fusca</i> YX	tfu
	Bifidobacteriales	<i>Bifidobacterium longum</i> NCC2705	blo
		<i>Bifidobacterium adolescentis</i> ATCC 15703	bad
Outgroup		<i>E. coli</i>	eco
		<i>B. subtilis</i>	bsu

Species in the same color belong to the same family or suborders in traditional taxonomies. *Bifidobacteriales* are close relatives of *Actinomycetales*, and also belong to the order *Actinobacteridae*. They are used as sub-outgroups, together with the more distantly related outgroup species *Escherichia coli* and *Bacillus subtilis*.

proved in the whole genome-based reconstructions: the phylogenetic trees are in good agreement with each other and better resolution of inner branches is achieved. We combine the information from the single-gene phylogenies and the whole genome-based approaches to reconstruct a final consensus tree. The resulting tree shows a detailed, well-supported phylogeny of the 45 *Actinomycetales* species with complete resolution of inner branches at species level, which was not achieved independently by any of the individual methods which generally show weakly resolved clusters. The consensus tree is in good agreement with the traditional taxonomic classification at the family and suborder level. The only major exception is *Kineococcus radiotolerans* SRS30216, which is placed in *Micrococcineae* rather than in *Frankineae* as reported previously (Lilburn and Garrity, 2004; Lee, 2006; Normand et al., 2007), in agreement with the most recent revision of the *Actinobacteria* (Zhi et al., 2009).

2. Methods

In this study we combine single-gene and whole-genome approaches to obtain a completely resolved picture of the phylogeny of *Streptomyces* and their relatives.

2.1. Gene selection and single gene-based phylogeny

The majority of large scale phylogenetic reconstructions are currently based on rRNA sequences. Since rRNAs are essential, they are highly conserved throughout different species, lateral transfer is very rare, and their molecular size ensures they carry abundant evolutionarily informative sites. This makes rRNA extremely informative for phylogenetic analysis (Woese, 1987). For this study we included 3 different rRNA genes (5S, 16S and 23S) for single-gene analysis. For comparison, we included 5 large and broadly conserved protein sequences: *isoleucyl tRNA synthetase*, *ribosomal protein S1*, *DNA topoisomerase*, *SecY* and *GTPase*, involved in a variety of conserved cellular functions. The selection of these protein sequences was based on their level of conservation and sequence length. Since each sequence position contains information on a very narrow range of evolutionary time, a larger number of independently evolving positions leads to better phylogenetic resolution (Olsen and Woese, 1993). Reconstructions based on larger molecules are also less affected by local non-random re-arrangements (Snel et al., 2005; Kunisawa, 2007).

Sequences were aligned separately by ClustalW (Thompson et al., 1994), sites containing gaps were removed, and 100 bootstrap replicates of each individual sequences alignment were generated. We used neighbor-joining (Saitou and Nei, 1987), Fitch–Margoliash (Fitch and Margoliash, 1967), maximum parsimony (Sourdis and Nei, 1988) and maximum likelihood (Felsenstein, 1981), all implemented in the PHYLIP package (Felsenstein, 2007) as alternative approaches for phylogenetic inference. Final trees for all individual sequences were built by using the CONSENSUS program using the majority-rule consensus approach on all bootstrap results (Felsenstein, 2007). The resulting trees show a conservative picture of the phylogeny, including only the best-supported relationships as resolved.

2.2. Whole genome-based phylogeny

For the genome-based approach we collected all coding sequences from the genome annotations of 45 *Actinomycetales* species and 4 outgroup species (Table 1) available in the National Center for Biotechnology Information database (<http://www.ncbi.nlm.nih.gov/>). Homologs were assigned using a reciprocal best hit strategy. We found 155 broadly conserved proteins, which were present in

all species and were used for gene concatenation and gene order-based phylogenetic reconstruction. A third analysis was based on the entire gene content of the same 49 genomes.

2.2.1. Gene concatenation phylogeny

Phylogeny based on concatenated gene sequence data (Brown et al., 2001; Herniou et al., 2001; Ciccarelli et al., 2006) is an intuitive extension of the single-gene approach. Instead of focusing on one or a few genes, it uses the concatenated sequence of all conserved proteins. Hence, the approach is potentially more robust because of the greatly increased number of phylogenetic informative sites. The 155 ubiquitous protein sequences were individually aligned using ClustalW (Thompson et al., 1994) and, after removing all sites containing gaps, the 155 alignments were concatenated into one meta-alignment. The meta-alignment contained a total of 40333 phylogenetically informative sites. A complete list of all proteins included is provided in the [Supplementary material](#). For tree building we again used neighbor-joining, Fitch–Margoliash and maximum parsimony, with 100 bootstrap replicates. Due to computational constraints, maximum likelihood was not used for this part of the study.

2.2.2. Gene order phylogeny

Quite independent of the gene sequences, the physical order of genes along the genome is phylogenetically highly informative, and earlier studies have already confirmed that phylogenetically related genomes clearly have similar gene order (Koonin et al., 2000; Rokas and Holland, 2000; Snel et al., 2005; Kunisawa, 2007). An important advantage is that this approach is largely independent of sequence alignments and will not be affected by misalignments that often lead to wrong tree topologies. We again considered all 155 ubiquitous genes and then compared their left and right neighbors for all pairs of genomes (Fig. 1C). We calculate a similarity score for each species pair, determining how many neighbors are shared (maximum score = $2 * 155$). To build a tree from the resulting similarity matrix we used three different distance methods: neighbor-joining (Saitou and Nei, 1987), Fitch–Margoliash (Fitch and Margoliash, 1967) and Kitch (Felsenstein, 2007) with 100 bootstrap replicates.

2.2.3. Gene content phylogeny

The two genome-based approaches described so far consider only those genes that are universally conserved. Those molecules which are conserved only in sub-clades are ignored. To overcome this limitation, we used gene content analysis, which is a comprehensive way of phylogenetic inference which takes into account conservation distributed all over the genome (Fitz-Gibbon and House, 1999; Snel et al., 1999; Lin and Gerstein, 2000; Montague and Hutchison, 2000; Daubin et al., 2002; Huson and Steel, 2004; Henz et al., 2005). Here we use a very simple and robust way to implement this phylogenetic inference strategy: For each pair of species we calculated the percentage of genes that have a reciprocal best hit (homolog), relative to the number of genes in the longer of the two genomes.

The resulting similarity matrix was used for phylogenetic reconstruction using neighbor-joining, Fitch–Margoliash, and Kitch with 100 bootstrap replicates.

2.6. Rank order

A phylogenetic tree visualizes the evolutionary relationships among organisms by grouping them in different clades. However, for some applications, it is more convenient to also obtain a linear ordering of species according to their similarity to a target species. Such a ranking of organisms is particularly useful for genome visualization, but also for determining the proper weighting of species

Fig. 1. Cartoon of the independent sources of information used to resolve the phylogeny of 45 *Actinomycetales* species. Six different genomes are illustrated in the cartoon. Each box represents a gene and boxes with the same color represent homologous genes in different genomes. Four different independent sources of information have been used for the phylogenetic reconstruction: (a) Single-gene phylogeny considers only individual, highly conserved genes in different genomes. The blue boxes represent one selected universally conserved gene, while crossed boxes indicate those genes which are not considered in the tree reconstruction. (b) Gene concatenation phylogeny uses the concatenated sequence of all the universally conserved proteins in the genomes; crossed boxes indicate genes that are not taken into consideration because they are not ubiquitous. (c) Gene order phylogeny takes into account the physical order of genes: by comparing the identity (not the sequence) of the left and the right neighbors of each ubiquitous gene, a similarity matrix is obtained and used to build the phylogenetic tree; crossed boxes indicate genes that are neither ubiquitous nor neighbors of ubiquitous genes. (d) Gene content phylogeny takes into account conservation through all genomes: by calculating the percentage of homologs between pairs of species relative to the total number of genes, we obtain a similarity matrix which is used for phylogenetic reconstruction. Crossed boxes indicate that genes that occur only in single species. (For interpretation of color mentioned in this figure the reader is referred to the web version of the article.)

in comparative metabolic network reconstruction. We implemented three different ways of ranking the species, based on average rank, median rank and rank product (Breitling et al., 2004). The ranking of species based on their phylogenetic distances to *Streptomyces coelicolor*, our main target species, was obtained from all eight single gene-based phylogenies and from all three genome-based methods.

3. Results and discussion

3.1. Single gene-based phylogeny

As expected, all single-gene trees are rather poorly resolved. The tree based on 5S rRNA gene gives the most unresolved evolutionary picture, due to the short sequence: most of the organisms of *Actinomycetales* belonging to the same family are grouped in one cluster, but at the suborder level no resolution is shown (Fig. 2A). The phylogenetic trees based on 16S rRNA and 23S rRNA sequence genes are more resolved at the suborder level [Supplementary data: Phylogenetic trees with bootstrap values], as is the case for the highly conserved protein trees [Supplementary data: Protein

sequence trees]. While the overall evolutionary structure is the same for all 8 consensus trees, there are a number of poorly resolved relationships, where various single-gene trees disagree.

For instance, in the 23S rRNA tree two families of the suborder *Frankineae* (*Acidothermaceae* and *Frankiaceae*) are clustered with the suborder *Streptomycineae* and a member of *Kineosporiaceae*, *Kineococcus radiotolerans* SRS30216, is merged with the suborder *Micrococcineae*. In the protein trees, *Kineococcus radiotolerans* SRS30216 is placed in the suborder *Micrococcineae* or unresolved. In the 16S rRNA tree the suborders *Frankineae* and *Streptomycineae* are shown in two unresolved independent branches. In contrast, in some of the protein trees, *Frankineae* is clustered with *Streptomycineae*.

Thermobifida fusca YX of the suborder *Streptosporangineae* is clustered with two families of *Frankineae* in the 16S rRNA tree, but with *Saccharopolyspora erythraea* NRRL 2338 of the suborder *Pseudonocardineae* in the 23S rRNA tree.

The placement of the suborder *Micromonosporineae* is not clearly resolved in the rRNA-based trees: in the 5S rRNA and 23S rRNA tree their placement is unresolved, while in 16S rRNA *Micromonosporineae* are with suborder *Frankineae*.

Fig. 2. Consensus trees based on rRNA sequences, (A) 5S rRNA, (B) 16S rRNA, (C) 23S rRNA. Initial trees were reconstructed using four tree inference approaches (NJ, Fitch, ML, PARS) with 100 bootstrap replicates each and combined by majority-rule consensus to include only well-supported branches. Organisms having the same colors are members of the same suborder in the taxonomical classification of NCBI. (For interpretation of color mentioned in this figure the reader is referred to the web version of the article.)

The families *Corynebacteriaceae*, *Mycobacteriaceae* and *Nocardia-ceae* of suborder *Corynebacteriineae* which should be grouped in a monophyletic cluster, as indicated by Casanova and Abel (2002) fall in three separate unresolved branches instead in the rRNA trees, while the five protein sequence trees show *Corynebacteriaceae* plus *Mycobacteriaceae* plus *Nocardia-ceae* as a monophyletic group as expected (Casanova and Abel, 2002), while other subgroups mostly remain unresolved.

Most species of *Mycobacteriaceae* are well resolved, but the phylogeny among the six pathogenic species within the *Mycobacterium* suborder [*Mycobacterium tuberculosis* F11, *Mycobacterium tuberculosis* H37Rv, *Mycobacterium tuberculosis* H37Ra, *Mycobacterium tuberculosis* CDC1551, *Mycobacterium bovis* BCG str. Pasteur 1173P2, *Mycobacterium bovis* AF2122/97] is very poorly resolved: the six species form a single, completely unresolved branch in all rRNA consensus trees and in four protein-based trees.

When the tree is rooted on *E. coli* and *B. subtilis*, the sub-outgroups *Bifidobacterium longum* NCC2705 and *Bifidobacterium adolescentis* ATCC 15703 are placed on an unresolved branch among the *actinomycetes*.

Thus, from the analysis of three rRNA genes and eight proteins we could not obtain a completely resolved, well-supported picture

of *Actinomycetales* phylogeny. To overcome this problem we moved onto whole genome-based phylogenetic reconstruction methods.

3.2. Gene concatenation phylogeny

The phylogenetic tree based on the concatenation of 155 protein sequences highly conserved among actinomycetes gives better resolution than single molecule based phylogeny (Fig. 3A). It shows excellent congruence with the 16S and 23S rRNA trees (Fig. 2A–C) for all major taxonomic groupings, confirming the validity of the method. One of the more resolved branches contains the families *Corynebacteriaceae*, *Mycobacteriaceae* and *Nocardia-ceae*, which are now clustered in a monophyletic group as expected (Koonin et al., 2000). The cluster containing the six pathogenic species of *Mycobacterium*, which was completely unresolved in the single gene-based phylogeny, is instead well resolved here in agreement with the taxonomic classifications in the NCBI database. Three families of the suborder *Micrococcineae* [*Cellulomonadaceae*, *Microbacteriaceae*, *Micrococcaceae*] are grouped together in one branch and individual species are correctly related to their traditional taxonomic family members. When the tree is rooted on *E. coli* and *B. subtilis*, *Bifidobacterium longum* NCC2705 and *Bifidobacterium*

Fig. 3. Phylogenetic trees based on whole-genome approaches. (A) Gene concatenation phylogeny based on the concatenation of 155 ubiquitously conserved proteins. (B) Gene order phylogeny based on conserved neighbor relationships along the genome. (C) Gene content phylogeny based on comparison of the set of homologous proteins shared between pairs of species. Coloring as in Fig. 2. (For interpretation of color mentioned in this figure the reader is referred to the web version of the article.)

adolescentis ATCC 15703 are clearly present as well-resolved sub-outgroups.

An interesting agreement with the single-gene phylogenies is the grouping of the two families *Frankineae* and *Streptomyceae* with a member of *Streptosporangineae*, *Thermobifida fusca* YX. Surprisingly, *Kineococcus radiotolerans* SRS30216 still does not branch with any *Frankineae* species, in contrast to traditional phylogenies (Garrity et al., 2004; Lee, 2006; Normand et al., 2007).

3.3. Gene order phylogeny

The tree based on gene order is independent from gene sequences and multiple sequence alignments. This tree again shows very good resolution and good overall agreement with single-gene trees (Fig. 3B). One important improvement is the finding of the three families *Corynebacteriaceae*, *Mycobacteriaceae* and *Nocardia-ceae* (*Corynebacterineae*) in one group with strong bootstrap

support. The internal branch of *Mycobacteriaceae* containing the six pathogenic species is again more resolved than in the single-gene phylogenies, and the branching mostly agrees with the gene concatenation tree. On the other hand, relationships among the different families of *Frankineae*, *Streptosporangineae* and *Streptomyceae* are less defined here. While all four *Streptomyces* species are places in a single cluster, the three families of *Frankineae* are scattered and their phylogeny unresolved.

3.4. Gene content phylogeny

The tree based on gene content (Fig. 3C), which uses a third independent source of phylogenetic evidence, is congruent with the 16S and 23S rRNA (Fig. 2B and C) trees and with two other whole genome-based trees (Fig. 3A and B) for all major branches. *Frankineae*, *Streptomyceae* and *Thermobifida fusca* YX are again clustered in one group. Also, once again *Kineococcus radiotolerans*

SRS30216 is unexpectedly clustered with *Micrococcineae*. The three families *Corynebacteriaceae*, *Mycobacteriaceae* and *Nocardiaceae* of the suborder *Corynebacterineae* fall in one monophyletic group. Branching of the six pathogens of *Mycobacteriaceae* is fully resolved and agrees with previous taxonomic classifications. The phylogeny of the three families of the suborder *Micrococcineae* [*Cellulomonadaceae*, *Microbacteriaceae*, *Micrococcaceae*] is completely resolved and branching follows traditional groupings.

3.5. Rank order

The rank order of species, using *Streptomyces coelicolor* as our focus species, shows a close relationship between *Streptomycineae*, *Frankineae*, and *Streptosporangineae* (Table 1). Among the *Streptomyces* species, *Streptomyces coelicolor* is most closely related to *Streptomyces avermitilis*, followed by *Streptomyces scabies* and *Streptomyces griseus*.

3.6. Consensus tree

As we have shown, all individual tree reconstruction methods show rather low resolution, but good agreement among each other and with classical taxonomical groupings [Supplementary data: Phylogenetic trees with bootstrap values]. We also saw that whole genome-based reconstructions generally show a notably improved resolution. Can we use all this information to finally build one, well-supported consensus tree that fully resolves the relationships among our species of interest? To make sure that the final consensus tree contains only branches that are strongly supported by the available evidence, we combined the information from all single-gene and whole genome-based using a majority-rule consensus (Fig. 4). The resulting tree shows a completely resolved phylogeny of all 45 *Actinomycetales* species, a result that was not achieved independently by any of the individual approaches, whether single gene or genome-based.

Fig. 4. Fully resolved final consensus tree of *Streptomyces* and its relatives, reconstructed from all the independent sources of evolutionary information (gene sequences, gene order, gene content) collected in this study.

The advantage of our integrative approach is most easily seen in the family *Mycobacteriaceae*: the rRNA-based reconstructions (Fig. 2), the species are only divided into broad clusters with poor resolution of inner branches. In the genome-based approaches the resolution improves notably, but the branching in the different trees is not always consistent (Fig. 3). In the final consensus tree, the subdivisions of the *Mycobacteriaceae* species are fully resolved (Fig. 4), despite using a strict majority-rule consensus approach, which includes only the best-supported phylogenetic relationships.

In the final consensus reconstruction (Fig. 4) the four species of *Streptomyces* are grouped consistently in a single cluster. The closest organism is *Thermobifida fusca* YX (suborder *Streptosporangineae*), in agreement with the rank order analysis (Table 2). These five species

are shown to be the sister group of the genus *Frankineae*, and closely related to the large cluster containing *Micromonosporineae* and *Corynebacterineae*. The suborders *Pseudonocardineae* and *Corynebacterineae* are phylogenetically very close to each other as seen already in two of the genome-based approaches (gene concatenation and gene order) and in several single gene-based trees (*ribosomal protein S1*, *SecY*, *GTPase*, *DNA topoisomerase*). Remarkable is also the placement of *Kineococcus radiotolerans* SRS30216: previous classifications placed this species close to *Frankineae* (Garrity et al., 2004; Lee, 2006; Normand et al., 2007), while in the consensus tree in Fig. 4 the species is placed among the *Micrococcineae*, in agreement with the most recent phylogeny of the phylum *Actinobacteria* (Zhi et al., 2009).

Table 2

Consensus ranking of species based on their distance from our focus species *Streptomyces coelicolor*.

Organism	Product rank
<i>Streptomyces coelicolor</i> _A3(2)	1
<i>Streptomyces avermitilis</i> _MA_4680	2.629032212
<i>Streptomyces scabies</i> _strain_8722	2.727747733
<i>Streptomyces griseus</i> _strain_IFO13350	4.100117402
<i>Thermobifida fusca</i> _YX	9.663449262
<i>Nocardioideis</i> _sp_JS614	11.08997499
<i>Acidothermus cellulolyticus</i> _11B	11.39609635
<i>Frankia alni</i> _ACN14a	12.06983368
<i>Frankia</i> _sp_Ccl3	12.07899215
<i>Salinispora tropica</i> _CNB_440	12.33425514
<i>Kineococcus radiotolerans</i> _SRS30216	12.49817983
<i>Frankia</i> _sp_EAN1pec	13.186375
<i>Salinispora arenicola</i> _CNS_205	13.69372661
<i>Saccharopolyspora erythraea</i> _NRRL_2338	16.38845893
<i>Nocardia farcinica</i> _IFM_10152	17.18458994
<i>Arthrobacter aureus</i> _TC1	18.39612703
<i>Clavibacter michiganensis</i> _subsp._michiganensis_NCPPB_382	18.45680603
<i>Rhodococcus</i> _sp_RHA1	18.83692805
<i>Propionibacterium acnes</i> _KPA171202	19.40448096
<i>Arthrobacter</i> _sp_FB24	20.27505199
<i>Leifsonia xyli</i> _subsp._xyli_str_CTCB07	20.46342192
<i>Renibacterium salmoninarum</i> _ATCC_33209	21.4049949
<i>Mycobacterium</i> _sp_KMS	24.40733018
<i>Mycobacterium</i> _sp_MCS	24.44682101
<i>Mycobacterium smegmatis</i> _str_MC2155	24.52484051
<i>Mycobacterium vanbaalenii</i> _PYR-1	24.72413092
<i>Mycobacterium</i> _sp_JLS	25.83612758
<i>Mycobacterium avium</i> _subsp._paratuberculosis_str_k10	26.4275277
<i>Mycobacterium gilvum</i> _PYR-GCK	26.5103845
<i>Mycobacterium tuberculosis</i> _H37Rv	26.7345383
<i>Mycobacterium tuberculosis</i> _CDC1551	26.9326884
<i>Mycobacterium tuberculosis</i> _H37Ra	27.18827008
<i>Mycobacterium bovis</i> _subsp._bovis_AF2122_97	27.212368
<i>Mycobacterium bovis</i> _BCG_Pasteur_1173P2	27.93566467
<i>Mycobacterium tuberculosis</i> _F11	28.13822088
<i>Tropheryma whipplei</i> _TW08_27	28.8506511
<i>Mycobacterium avium</i> _104	28.95281287
<i>Tropheryma whipplei</i> _str_Twist	29.5717043
<i>Corynebacterium jeikeium</i> _K411	29.92280148
<i>Mycobacterium ulcerans</i> _Agy99	29.96251954
<i>Mycobacterium leprae</i> _TN	30.30757177
<i>Corynebacterium diphtheriae</i> _NCTC_13129	30.48200786
<i>Corynebacterium efficiens</i> _YS-314	30.66788683
<i>Corynebacterium glutamicum</i> _ATCC_13032	33.63478094
<i>Corynebacterium glutamicum</i> _ATCC_13032_new	33.67214223
<i>Bifidobacterium longum</i> _NCC2705	38.99584717
<i>Bifidobacterium adolescentis</i> _ATCC_15703	39.88074983
<i>Bacillus subtilis</i>	48.08436608
<i>Escherichia coli</i> _K12	48.36125292

The rank product ranking (Breitling et al., 2004) combines the similarity metrics based on all sources of phylogenetic information (gene sequences, gene order, gene content) into a consensus linear ordering of species. Rankings based on average or median rank yield very similar results [Supplementary information: Average and Median rank]. This arrangement can be used, for instance, for arranging species in genome visualization or for determining the weight of species in comparative metabolic modeling.

4. Conclusion

We aimed to determine a comprehensive, detailed and robust phylogeny of *Streptomyces* and its closest relatives in *Actinomycetales* by using single gene-based and genome-based phylogenies. The results from the different approaches were combined in a consensus tree that shows a completely resolved, well-supported phylogeny of the 45 *actinomycetes* with high resolution of all inner branches.

Overall, the high-level branchings in our consensus tree agree well with traditional taxonomic subdivisions, grouping together those species which belong taxonomically to the same family or the same suborder. The only major exception is *Kineococcus radiotolerans* SRS30216, which is placed in *Micrococcineae* rather than in *Frankineae* as reported previously (Lilburn and Garrity, 2004; Lee, 2006; Normand et al., 2007). This confirms the general validity of our strategy.

The relationships among the *Streptomyces* species and their relatives are fully resolved only in the final consensus tree (Fig. 4). The species closest to *Streptomyces* is *Thermobifida fusca* YX, a moderately thermophilic soil bacterium from the suborder *Streptosporangineae*. The *Streptomyces* cluster is furthermore closely related to *Frankineae* species, a group of nitrogen-fixing, nodule-forming bacteria. This finding will be useful for future computational studies exploring the evolution of metabolic capacities in *Streptomyces* species by quantitative genome-based modeling (Price et al., 2004; Breitling et al., 2008).

For the gene concatenation analysis, the reconstruction of phylogenies at the species level could be affected by lateral gene transfer (LGT). Genes that are likely to be subject to LGT should be excluded from the analysis. In our case, the focus on a closely related group of species makes the reliable identification of such LGT candidates impossible. However, the majority of proteins included in the concatenation are encoded by single-copy genes in the majority of genomes, making them unlikely to be subject to widespread LGT. Moreover, we focus on the most highly conserved subset of 155 genes (those present in all species studied), which are least likely to be affected by LGT. This is confirmed by the fact that for all resolved branches the gene concatenation phylogeny and the gene order phylogeny show complete agreement.

In this study we have shown that single-gene analysis can lead to poorly resolved or even misleading phylogenies, and demonstrated a way of using whole-genome analysis to overcome this limitation. We demonstrate that only by a combination of methodologies it is possible to achieve a fully resolved phylogeny with detailed and well-supported inner branches. It will be interesting to apply this method to larger groups of organisms (for instance, all bacteria); the small number of universally conserved protein-coding genes will, however, necessitate modifications in the approach, so that phylogenetic signal from widespread, but non-universal genes can also be included in the analysis.

The phylogenetic reconstructions from the different sources (single gene or whole genome-based phylogenies) agree at the family level with the traditional taxonomy, but at species level they show a much richer, fine-grained picture of the relationships.

The resulting consensus tree, which is supported by a variety of independent sources of evidence (gene sequences, gene order, gene content), will form a solid basis for future comparative genome annotation and across-species modeling of metabolic pathways. This will be an essential resource for the further exploitation of *Streptomyces* species in biotechnology and systems biology research.

Acknowledgments

M.E.M. was funded by a 4 × 4 Ubbo Emmius scholarship from the University of Groningen. E.T. was funded by a Rosalind Franklin Fellowship from the University of Groningen. R.B. is supported by an NWO-Vidi fellowship.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2009.11.019.

References

- Anderson, A.S., Wellington, E.M., 2001. The taxonomy of *Streptomyces* and related genera. *Int. J. Syst. Evol. Microbiol.* 51, 797–814.
- Barabote, R.D., Xie, G., Leu, D.H., Normand, P., Necseula, A., Daubin, V., Medigue, C., Adney, W.S., Xu, X.C., Lapidus, A., Parales, R.E., Detter, C., Pujic, P., Bruce, D., Lavire, C., Challacombe, J.F., Brettin, T.S., Berry, A.M., 2009. Complete genome of the cellulolytic thermophile *Acidothermus cellulolyticus* 11B provides insights into its ecophysiological and evolutionary adaptations. *Genome Res.* 19, 1033–1043.
- Bentley, S.D., Chater, K.F., Cerdano-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neill, S., Rabinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J., Hopwood, D.A., 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417, 141–147.
- Borodina, I., Krabben, P., Nielsen, J., 2005. Genome-scale analysis of *Streptomyces coelicolor* A3(2) metabolism. *Genome Res.* 15, 820–829.
- Breitling, R., Armengaud, P., Amtmann, A., Herzyk, P., 2004. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett.* 573, 83–92.
- Breitling, R., Vitkup, D., Barrett, M.P., 2008. New surveyor tools for charting microbial metabolic maps. *Nat. Rev. Microbiol.* 6, 156–161.
- Brown, J.R., Douady, C.J., Italia, M.J., Marshall, W.E., Stanhope, M.J., 2001. Universal trees based on large combined protein sequence data sets. *Nat. Genet.* 28, 281–285.
- Casanova, J.L., Abel, L., 2002. Genetic dissection of immunity to mycobacteria: the human model. *Annu. Rev. Immunol.* 20, 581–620.
- Castillo, U.F., Strobel, G.A., Ford, E.J., Hess, W.M., Porter, H., Jensen, J.B., Albert, H., Robison, R., Condrón, M.A., Teplov, D.B., Stevens, D., Yaver, D., 2002. Munumbicins, wide-spectrum antibiotics produced by *Streptomyces* NRRL 30562, endophytic on *Kennedia nigricans*. *Microbiology* 148, 2675–2685.
- Cerdano-Tarraga, A.M., Efstratiou, A., Dover, L.G., Holden, M.T., Pallen, M., Bentley, S.D., Besra, G.S., Churcher, C., James, K.D., De Zoysa, A., Chillingworth, T., Cronin, A., Dowd, L., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Moule, S., Quail, M.A., Rabinowitsch, E., Rutherford, K.M., Thomson, N.R., Unwin, L., Whitehead, S., Barrell, B.G., Parkhill, J., 2003. The complete genome sequence and analysis of *Corynebacterium diphtheriae* NCTC13129. *Nucleic Acids Res.* 31, 6516–6523.
- Chater, K.F., Chandra, G., 2006. The evolution of development in *Streptomyces* analysed by genome comparisons. *FEMS Microbiol. Rev.* 30, 651–672.
- Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B., Bork, P., 2006. Toward automatic reconstruction of a highly resolved tree of life. *Science* 311, 1283–1287.
- Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry 3rd, C.E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S., Barrell, B.G., 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537–544.
- Daubin, V., Gouy, M., Perriere, G., 2002. A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. *Genome Res.* 12, 1080–1090.
- Egan, S., Wiener, P., Kallifidas, D., Wellington, E.M., 2001. Phylogeny of *Streptomyces* species and evidence for horizontal transfer of entire and partial antibiotic gene clusters. *Antonie Van Leeuwenhoek* 79, 127–133.
- Embley, T.M., Stackebrandt, E., 1994. The molecular phylogeny and systematics of the actinomycetes. *Annu. Rev. Microbiol.* 48, 257–289.
- Felsenstein, J., 2007. Phylogeny Inference Package Version 3.67. University of Washington, Seattle, WA. pp. 98195–5065.
- Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17, 368–376.
- Fitch, W.M., Margoliash, E., 1967. Construction of phylogenetic trees. *Science* 155, 279–284.
- Fitz-Gibbon, S.T., House, C.H., 1999. Whole genome-based phylogenetic analysis of free-living microorganisms. *Nucleic Acids Res.* 27, 4218–4222.
- Gao, B., Gupta, R.S., 2005. Conserved indels in protein sequences that are characteristic of the phylum Actinobacteria. *Int. J. Syst. Evol. Microbiol.* 55, 2401–2412.
- Garrity, G.M., Bell, J.A., Lilburn, T.G., 2004. Taxonomic outline of the prokaryotes. In: *Bergey's Manual of Systematic Bacteriology*, second ed., release 5.0.
- Goodfellow, M., Williams, S.T., 1983. Ecology of actinomycetes. *Annu. Rev. Microbiol.* 37, 189–216.
- Henz, S.R., Huson, D.H., Auch, A.F., Nieselt-Struwe, K., Schuster, S.C., 2005. Whole-genome prokaryotic phylogeny. *Bioinformatics* 21, 2329–2335.
- Herniou, E.A., Luque, T., Chen, X., Vlak, J.M., Winstanley, D., Cory, J.S., O'Reilly, D.R., 2001. Use of whole genome sequence data to infer baculovirus phylogeny. *J. Virol.* 75, 8117–8126.
- Hopwood, D.A., 2007. *Streptomyces* in Nature and Medicine: The Antibiotic Makers. Oxford University Press.
- Huson, D.H., Steel, M., 2004. Phylogenetic trees based on gene content. *Bioinformatics* 20, 2044–2049.
- Koonin, E.V., Aravind, L., Kondrashov, A.S., 2000. The impact of comparative genomics on our understanding of evolution. *Cell* 101, 573–576.
- Kost, C., Lakatos, T., Bottcher, I., Arendholz, W.R., Redenbach, M., Wirth, R., 2007. Non-specific association between filamentous bacteria and fungus-growing ants. *Naturwissenschaften* 94, 821–828.
- Kunisawa, T., 2007. Gene arrangements characteristic of the phylum Actinobacteria. *Antonie Van Leeuwenhoek* 92, 359–365.
- Lechevalier, H.A., Lechevalier, M.P., 1967. Biology of actinomycetes. *Annu. Rev. Microbiol.* 21, 71–100.
- Lee, S.D., 2006. *Kineococcus marinus* sp. nov., isolated from marine sediment of the coast of Jeju, Korea. *Int. J. Syst. Evol. Microbiol.* 56, 1279–1283.
- Leyden, J.J., 2001. The evolving role of *Propionibacterium acnes* in acne. *Semin. Cutan. Med. Surg.* 20, 139–143.
- Lilburn, T.G., Garrity, G.M., 2004. Exploring prokaryotic taxonomy. *Int. J. Syst. Evol. Microbiol.* 54, 7–13.
- Lin, J., Gerstein, M., 2000. Whole-genome trees based on the occurrence of folds and orthologs: implications for comparing genomes on different levels. *Genome Res.* 10, 808–818.
- Manteca, A., Pelaez, A.I., Zardoya, R., Sanchez, J., 2006. Actinobacteria cyclophilins: phylogenetic relationships and description of new class- and order-specific paralogues. *J. Mol. Evol.* 63, 719–732.
- Montague, M.G., Hutchison 3rd, C.A., 2000. Gene content phylogeny of herpesviruses. *Proc. Natl. Acad. Sci. USA* 97, 5334–5339.
- Normand, P., Lapierre, P., Tisa, L.S., Gogarten, J.P., Alloisio, N., Bagnarol, E., Bassi, C.A., Berry, A.M., Bickhart, D.M., Choise, N., Couloux, A., Cournoyer, B., Cruveiller, S., Daubin, V., Demange, N., Francino, M.P., Goltsman, E., Huang, Y., Kopp, O.R., Labarre, L., Lapidus, A., Lavire, C., Marechal, J., Martinez, M., Mastrorunzio, J.E., Mullin, B.C., Niemann, J., Pujic, P., Rawnsley, T., Rouy, Z., Schenowitz, C., Sellstedt, A., Tavares, F., Tomkins, J.P., Vallenet, D., Valverde, C., Wall, L.G., Wang, Y., Medigue, C., Benson, D.R., 2007. Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. *Genome Res.* 17, 7–15.
- Olsen, G.J., Woese, C.R., 1993. Ribosomal RNA: a key to phylogeny. *FASEB J.* 7, 113–123.
- Phillips, R.W., Wiegand, J., Berry, C.J., Fliermans, C., Peacock, A.D., White, D.C., Shimkets, L.J., 2002. *Kineococcus radiotolerans* sp. nov., a radiation-resistant, gram-positive bacterium. *Int. J. Syst. Evol. Microbiol.* 52, 933–938.
- Price, N.D., Reed, J.L., Palsson, B.O., 2004. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.* 2, 886–897.
- Rokas, A., Holland, P.W., 2000. Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol.* 15, 454–459.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Snel, B., Bork, P., Huynen, M.A., 1999. Genome phylogeny based on gene content. *Nat. Genet.* 21, 108–110.
- Snel, B., Huynen, M.A., Dutilh, B.E., 2005. Genome trees and the nature of genome evolution. *Annu. Rev. Microbiol.* 59, 191–209.
- Sourdis, J., Nei, M., 1988. Relative efficiencies of the maximum parsimony and distance-matrix methods in obtaining the correct phylogenetic tree. *Mol. Biol. Evol.* 5, 298–311.
- Stackebrandt, E., Rainey, F.A., Ward-Rainey, N.L., 1997. Proposal for a new hierarchical classification system, Actinobacteria classa nov. *Int. J. Syst. Bacteriol.* 47, 479–491.
- Takeuchi, T., Sawada, H., Tanaka, F., Matsuda, I., 1996. Phylogenetic analysis of *Streptomyces* spp. causing potato scab based on 16S rRNA sequences. *Int. J. Syst. Bacteriol.* 46, 476–479.

- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- Tokala, R.K., Strap, J.L., Jung, C.M., Crawford, D.L., Salove, M.H., Deobald, L.A., Bailey, J.F., Morra, M.J., 2002. Novel plant–microbe rhizosphere interaction involving *Streptomyces lydicus* WYEC108 and the pea plant (*Pisum sativum*). *Appl. Environ. Microbiol.* 68, 2161–2171.
- Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G.F., Chater, K.F., van Sinderen, D., 2007. Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol. Mol. Biol. Rev.* 71, 495–548.
- Woese, C.R., 1987. Bacterial evolution. *Microbiol. Rev.* 51, 221–271.
- Zhang, H., Zhang, W., Jin, Y., Jin, M., Yu, X., 2008. A comparative study on the phylogenetic diversity of culturable actinobacteria isolated from five marine sponge species. *Antonie Van Leeuwenhoek* 93, 241–248.
- Zhi, X.Y., Li, W.J., Stackebrandt, E., 2009. An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class Actinobacteria, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. *Int. J. Syst. Evol. Microbiol.* 59, 589–608.