

Conserved indels in protein sequences that are characteristic of the phylum *Actinobacteria*

Beile Gao and Radhey S. Gupta

Correspondence
Radhey S. Gupta
gupta@mcmaster.ca

Department of Biochemistry and Biomedical Science, McMaster University, Hamilton, Canada L8N 3Z5

Gram-positive bacteria with a high G + C content are currently recognized as a distinct phylum, *Actinobacteria*, on the basis of their branching in 16S rRNA trees. Except for an insert in the 23S rRNA, there are no unique biochemical or molecular characteristics known at present that can distinguish this group from all other bacteria. In this work, three conserved indels (i.e. inserts or deletions) are described in three widely distributed proteins that are distinctive characteristics of the *Actinobacteria* and are not found in any other groups of bacteria. The identified signatures are a 2 aa deletion in cytochrome-c oxidase subunit 1 (Cox1), a 4 aa insert in CTP synthetase and a 5 aa insert in glutamyl-tRNA synthetase (GluRS). Additionally, the actinobacterial specificity of the large insert in the 23S rRNA was also tested. Using primers designed for conserved regions flanking these signatures, fragments of most of these genes were amplified from 23 actinobacterial species, covering many different families and orders, for which no sequence information was previously available. All the 61 sequenced fragments, except two in GluRS, were found to contain the indicated signatures. The presence of these signatures in various species from 20 families within this phylum provides evidence that they are likely distinctive characteristics of the entire phylum, which were introduced in a common ancestor of this group. The absence of all four of these signatures in *Symbiobacterium thermophilum* suggests that this species, which is distantly related to other actinobacteria in 16S rRNA and CTP synthetase trees, may not be a part of the phylum *Actinobacteria*. The identified signatures provide novel molecular means for defining and circumscribing the phylum *Actinobacteria*. Functional studies on them should prove helpful in understanding novel biochemical and physiological characteristics of this group of bacteria.

INTRODUCTION

The *Actinobacteria* constitute one of the main phyla within the *Bacteria* (Boone *et al.*, 2001). This lineage is comprised of Gram-positive organisms with a high G + C content (greater than 55 mol%). The phylum encompasses genera covering a wide range of morphology: some species are coccoid (e.g. *Micrococcus*) or rod-coccoid (e.g. *Arthrobacter*) in shape, while others display fragmenting hyphal forms (e.g. *Nocardia*) or permanent and highly differentiated branched mycelium (e.g. *Streptomyces*) (Atlas, 1997). Previously, organisms producing a mycelium which resembles that of the unrelated fungi were classified as the actinomycetes

Abbreviations: Cox1, cytochrome-c oxidase subunit 1; GluRS, glutamyl-tRNA synthetase.

The GenBank/EMBL/DDBJ accession numbers for the novel sequences described in this paper are indicated in Figs 1–4.

Full Cox1, CTP synthetase and GluRS alignments and accession numbers of 16S rRNA gene and CTP synthetase gene sequences used are available as supplementary material in IJSEM Online.

(Embley & Stackebrandt, 1994; Atlas, 1997). Spore formation is common among the actinobacteria, although not ubiquitous, and spores range from motile zoospores to specialized propagules. They are also physiologically diverse bacteria, as evidenced by their production of numerous extracellular enzymes and by the thousands of metabolic products they synthesize and excrete (Schrempf, 2001), many of which are antibiotics (Lechevalier & Lechevalier, 1967). Actinobacteria, especially members of the *Streptomycetaceae*, are the major antibiotic producers in the pharmaceutical industry (Davies, 1996; Bentley *et al.*, 2002). However, a few actinobacteria are important human, animal and plant pathogens. For example, *Mycobacterium tuberculosis* infection results in tuberculosis, *Corynebacterium diphtheriae* causes diphtheria and *Propionibacterium acnes* is the causative agent of acne (Leyden, 2001). Actinobacterial species are widely distributed in both terrestrial and aquatic ecosystems, especially in soil (Goodfellow & Williams, 1983; Chun *et al.*, 2000). In nature, actinobacteria play an important role in decomposition and humus formation, which is an integral part of the recycling of

biomaterials (Lechevalier & Lechevalier, 1967; Goodfellow & Williams, 1983).

Due to their pharmaceutical, industrial and environmental importance, the taxonomy and phylogeny of the actinobacteria are of great interest (Embley & Stackebrandt, 1994; Stackebrandt *et al.*, 1997; Ahmad *et al.*, 2000; Stach *et al.*, 2003; Richert *et al.*, 2005; Bull *et al.*, 2005). Earlier attempts to determine actinobacterial phylogenies based on morphological and chemotaxonomic traits were found to be unreliable indicators of phylogenetic relationships above the family level (Embley & Stackebrandt, 1994). Hence, our current understanding of the taxonomy and evolutionary relationships of actinobacterial divisions is mainly based on the branching patterns of these species in 16S rRNA trees. Different species have been placed in this group based on 16S rRNA oligonucleotide catalogues and phylogenetic analysis based on full and partial 16S rRNA gene sequences, although some species do not possess a high G+C content (Woese, 1987; Stackebrandt *et al.*, 1997; Labeda & Kroppenstedt, 2000; Ludwig & Klenk, 2001). In 1997, a new taxonomic hierarchical classification for the actinobacteria was proposed by Stackebrandt *et al.* (1997), which recognized this group as a distinct class, *Actinobacteria*, within the Gram-positive bacteria. In the latest *Bergey's Manual*, the actinobacteria have been assigned the rank of a phylum, recognizing that the phylogenetic depth represented in this lineage is equivalent to that of existing phyla and that the group shows clear separation from the *Firmicutes* (Garrity & Holt, 2001). According to *Bergey's Manual* (Garrity & Holt, 2001), the phylum *Actinobacteria* comprises 39 families and 130 genera, making it one of the largest groups within the *Bacteria*. However, except for their distinct clustering in 16S rRNA trees, no other reliable biochemical or molecular characteristics are presently known which can clearly distinguish species belonging to the phylum *Actinobacteria* from other bacteria.

Presently, the genomes of at least 14 different actinobacteria (in some cases multiple strains of the same species) have been sequenced: *Mycobacterium leprae*, *Mycobacterium bovis*, *Mycobacterium tuberculosis*, *Mycobacterium avium*, *Corynebacterium diphtheriae*, *Corynebacterium efficiens*, *Corynebacterium glutamicum*, *Tropheryma whipplei*, *Leifsonia xyli* subsp. *xyli*, *Nocardia farcinica*, *Streptomyces avermitilis*, *Streptomyces coelicolor*, *Bifidobacterium longum* and *Symbiobacterium thermophilum* (Smith *et al.*, 1997; Bentley *et al.*, 2002; Schell *et al.*, 2002; Camus *et al.*, 2002; Fleischmann *et al.*, 2002; Garnier *et al.*, 2003; Ikeda *et al.*, 2003; Karlin *et al.*, 2003; Cerdano-Tarraga *et al.*, 2003; Raoult *et al.*, 2003; Ueda *et al.*, 2004; Bruggemann *et al.*, 2004; Monteiro-Vitorello *et al.*, 2004; Ishikawa *et al.*, 2004). This provides a valuable resource for discovering novel molecular characteristics that are useful for biochemical, taxonomic and phylogenetic purposes (Sutcliffe & Harrington, 2002; Karlin *et al.*, 2003). Our recent work has focused on identifying conserved inserts or deletions (i.e. indels) in widely distributed proteins that are characteristic

of the different groups of bacteria and are also helpful in understanding the interrelationships among them (Gupta, 1998, 2001) (see also <http://www.bacterialphylogeny.com>). We have previously identified a large number of conserved indels (or signature sequences) that provide distinctive molecular markers for the phyla *Proteobacteria*, *Chlamydiae*, *Cyanobacteria* and '*Deinococcus-Thermus*' (Gupta, 2000, 2004; Gupta *et al.*, 2003; Griffiths & Gupta, 2004a, b) (see also <http://www.bacterialphylogeny.com>). In the present work, we describe three novel signature indels in three highly conserved and widely distributed proteins [cytochrome-*c* oxidase subunit 1 (Cox1), glutamyl-tRNA synthetase (GluRS) and CTP synthetase] that are distinctive characteristics of the phylum *Actinobacteria* and are not found in any other bacteria. Sequence information for these proteins was previously available from only a limited number of actinobacteria, whose genomes have been sequenced. One possible signature for actinobacteria, consisting of a large insert in the 23S rRNA, has previously been described (Roller *et al.*, 1992). However, the validity and specificity of this signature also needs to be further tested using sequence data from additional species. In the present work, we have examined the presence of the newly identified signatures in Cox1, GluRS and CTP synthetase, and also the 23S rRNA signature, from a broad range of actinobacterial species by PCR amplification and sequencing of the corresponding gene fragments. The results of our studies show that the signatures in Cox1, CTP synthetase and 23S rRNA are present in all actinobacterial species, indicating that they are very likely distinctive characteristics of the entire phylum and might be used as molecular markers for this group of species. The signature for GluRS was found to be lacking in two actinobacterial species (*Thermobifida fusca* and *Propionibacterium acnes*), which could result from either selective loss or other non-specific mechanisms such as lateral gene transfer.

METHODS

Bacterial strains and chromosomal DNA isolation. The various actinobacterial strains that were used in this study and their taxonomic positions, i.e. orders, suborders or families, within the phylum *Actinobacteria* are given in Table 1. This table also includes information for different actinobacterial species whose genomes have been sequenced and whose sequences were available to us. Together these strains cover a broad range of the diversity represented by this phylum. Various recently described type strains were purchased from the DSMZ (German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany) and cultured under the recommended conditions. Chromosomal DNA was purified by the following method. A 100 µl pellet was transferred to a microcentrifuge tube and washed three times with 0.5 ml 10.3% sucrose. The cell pellet was resuspended in 0.5 ml buffer containing 0.3 M sucrose, 25 mM Tris/HCl, 25 mM EDTA, pH 8.0, and 2 mg lysozyme ml⁻¹ and incubated for 2 h at 37 °C. Next, 50 µl 20% SDS was added and the cell suspension was incubated at 65 °C for 30 min. The cell lysate was extracted with chloroform twice and the aqueous layer was separated during centrifugation at 14 000 g for 15 min. The DNA was precipitated with 2 vols ethanol and dissolved in sterile water for PCR amplification. Chromosomal DNA of *Rhodococcus rhodochrous* and *Propionibacterium acnes* was generously

Table 1. Actinobacterial strains used in this study, including those with sequenced genomes

The taxonomy indicated is based on *Bergey's Manual* (Garrity & Holt, 2001).

Taxonomic position	Strain
Subclass II. <i>Rubrobacteridae</i>	<i>Rubrobacter radiotolerans</i> DSM 5868 ^T
Subclass V. <i>Actinobacteridae</i>	
Order I. <i>Actinomycetales</i>	
Suborder VI. <i>Micrococcineae</i>	
Family I. <i>Micrococcaceae</i>	<i>Arthrobacter nicotinovorans</i> DSM 420 ^T <i>Kocuria rhizophila</i> DSM 348
Family III. <i>Cellulomonadaceae</i>	<i>Cellulomonas fimi</i> DSM 20113 ^T <i>Oerskovia turbata</i> DSM 20577 ^T <i>Tropheryma whipplei</i> TW08/27 (genome) <i>Tropheryma whipplei</i> Twist ^T (genome)
Family VIII. <i>Microbacteriaceae</i>	<i>Microbacterium oxydans</i> DSM 20578 ^T <i>Clavibacter michiganensis</i> DSM 340 <i>Leifsonia xyli</i> subsp. <i>xyli</i> CTCB07 (genome)
Suborder VII. <i>Corynebacterineae</i>	
Family I. <i>Corynebacteriaceae</i>	<i>Corynebacterium diphtheriae</i> NCTC 13129 (genome) <i>Corynebacterium efficiens</i> YS-314 ^T (genome) <i>Corynebacterium glutamicum</i> ATCC 13032 ^T (genome)
Family III. <i>Gordoniaceae</i>	<i>Gordonia rubripertincta</i> DSM 43197 ^T
Family IV. <i>Mycobacteriaceae</i>	<i>Mycobacterium leprae</i> TN (genome) <i>Mycobacterium bovis</i> BCG (genome) <i>Mycobacterium tuberculosis</i> H37Rv (genome)* <i>Mycobacterium avium</i> subsp. <i>paratuberculosis</i> k10 (genome)
Family V. <i>Nocardiaceae</i>	<i>Nocardia corynebacterioides</i> DSM 20151 ^T <i>Rhodococcus rhodochrous</i> 116
Family VI. <i>Tsukamurellaceae</i>	<i>Tsukamurella paurometabola</i> DSM 20162 ^T
Family VII. 'Williamsiaceae'	<i>Williamsia muralis</i> DSM 44343 ^T
Suborder VIII. <i>Micromonosporineae</i>	<i>Micromonospora chersina</i> DSM 44151 ^T
Suborder IX. <i>Propionibacterineae</i>	
Family I. <i>Propionibacteriaceae</i>	<i>Propionibacterium acnes</i> AT1
Family II. <i>Nocardioideaceae</i>	<i>Nocardioides simplex</i> DSM 20130 ^T <i>Kribbella sandramycini</i> DSM 15626 ^T
Suborder X. <i>Pseudonocardineae</i>	<i>Pseudonocardia halophobica</i> DSM 43089 ^T <i>Saccharopolyspora erythraea</i> DSM 40517 ^T
Suborder XI. <i>Streptomycineae</i>	<i>Actinomadura glauciflava</i> DSM 44770 ^T ' <i>Trichotomospora caesia</i> ' DSM 43890 <i>Streptomyces avermitilis</i> MA-4680 ^T (genome) <i>Streptomyces coelicolor</i> A3(2) (genome)
Suborder XII. <i>Streptosporangineae</i>	<i>Streptosporangium roseum</i> DSM 43021 ^T <i>Microtetraspora niveoalba</i> DSM 43174 ^T <i>Planobispora rosea</i> DSM 43051 ^T <i>Thermobifida fusca</i> Tfus_25 (genome)
Order II. <i>Bifidobacteriales</i>	<i>Bifidobacterium longum</i> DJO10A (genome)

*We also examined *Mycobacterium tuberculosis* strain CDC1551, the genome of which has also been sequenced. However, because the sequence information for these proteins for the signature region was identical from these strains, information for only one of them is shown in various figures.

provided to us by Dr L. D. Eltis (University of British Columbia, Canada) (Warren *et al.*, 2004) and Dr Mark Farrar (Leeds University, UK) (Farrar *et al.*, 2000).

Identification of signature sequences. Multiple sequence alignments for a large number of proteins have been created in our

earlier work (Gupta, 2000, 2004; Gupta *et al.*, 2003; Griffiths & Gupta, 2004a). To search for actinobacteria-specific signatures, these alignments were inspected visually to identify any indel that was uniquely present in all available actinobacterial homologues and which was flanked by conserved sequences. Indels which were not flanked by conserved regions and/or which were not present in all

Table 2. PCR primers

N=A, T, C or G; Y=C or T; S=G or C; R=A or G; V=A, C or G; B=C, G or T; D=A, T or G; K=G or T; M=A or C; H=A, C or T.

Gene	Primer	Sequence (5'–3')	Fragment size (bp)
Cox1	Forward	TGGTTYTTYGGSCACCCYGARGT	581
	Reverse	CCVAVCCARTGCTGBAYSADRAA	
GluRS	Forward	ACBGCSCCTKTYAACTGG	773
	Reverse	AGRTARTTSARMAKRCCYTC	
CTP synthetase	Forward	AARACVAARCCVACHCAGCA	986
	Reverse	TCVGGRTGNGCCTGBGT	
23S rRNA insert	Forward	CCGANAGGCGTAGBCGATGG	361
	Reverse	CCWGWGTYGGTTTMSGTA	

actinobacterial species were omitted from further consideration. The specificity of potentially useful indels for actinobacteria was further evaluated by carrying out detailed BLAST searches on short sequence segments (usually between 60 and 100 aa) containing the indel and the flanking conserved regions. The purpose of these BLAST searches was to obtain sequence information from all available bacterial homologues to ensure that the identified signatures are only present in the actinobacterial homologues. Sequence information for representative species from different groups for various useful signatures was compiled into signature files such as those shown here. Detailed information for these signatures from all available species is provided as supplementary material in IJSEM Online.

PCR amplification and sequencing. Degenerate oligonucleotide primers, in opposite orientations, were designed for highly conserved regions that flanked the identified signatures in sequence alignments. The sequences of various PCR primers used in these studies are given in Table 2. The PCR was performed in 30 µl solution and all primer sets were optimized for Mg²⁺ concentration (in the range of 1.5 to 4 mM) for each strain tested. PCR amplification was carried out in a Techne Progene thermocycler. After an initial denaturation step at 97 °C, the DNA was amplified for 30 cycles (30 s at 97 °C, 30 s at 55 °C, 1 min at 72 °C). The last cycle was followed by a 15 min extension at 72 °C. Because all the genomic actinobacterial DNA we tested contained a high G+C content, the denaturation temperature was set to 97 °C so that the DNA was mostly denatured in a short time with less damage. DNA fragments of the expected size were purified from 0.8% (w/v) agarose gels and subcloned into the plasmid pDRIVE using a UA cloning kit (Qiagen). After transforming *Escherichia coli* JM109 cells with the plasmids, inserts from a number of positive clones were sequenced. Sequence information for various actinobacterial species has been deposited in GenBank and accession numbers are given in Figs 1–4.

Phylogenetic analysis. DNA sequences of newly sequenced fragments from different actinobacterial species were translated and added into the existing signature files as shown. Multiple alignments of CTP synthetase homologues and 16S rRNA gene sequences from different bacteria were created by using the CLUSTAL X program (Jeanmougin *et al.*, 1998). All the fragments were trimmed to the same length as the amplified fragments. Neighbour-joining distance trees showing branch lengths were constructed applying Kimura's correction (Kimura, 1980). The trees were produced by using the TREECON program (Van de Peer & De Wachter, 1994).

RESULTS

Description of novel actinobacteria-specific signatures in protein sequences and examination of their specificity

Our work has identified a number of useful signatures consisting of conserved inserts and deletions in protein sequences that are limited to actinobacterial species. Sequence information for most of these genes/proteins was mainly available only from those actinobacterial species whose genomes have been sequenced. Many of the sequenced species are closely related and some belong to the same genus (Table 1). Hence, to determine the actinobacterial specificities of the identified signatures, we have cultured and extracted chromosomal DNA of 23 actinobacterial type strains, covering a large number of orders and suborders (e.g. *Rubrobacterineae*, *Micrococccineae*, *Corynebacterineae*, *Micromonosporineae*, *Propionibacterineae*, *Pseudonocardineae*, *Streptomycineae* and *Streptosporangineae*) within this phylum (Table 1). Sequence information for the identified signatures was obtained from many of these species to confirm and validate the specificity of the signatures. A brief description of the newly identified signatures and the work that we have carried out on them is given below.

One of the actinobacteria-specific signatures that we have identified is present in the Cox1 protein. Cytochrome-c oxidase is an intrinsic membrane protein, composed of three subunits, that functions as the terminal enzyme of the respiratory electron transport chain (Michel *et al.*, 1998). In the Cox1 subunit, a 2 aa gap is present in a conserved region (boxed in Fig. 1) that is unique to various actinobacterial species, and is not seen in any other bacteria. Because this sequence gap is absent in Cox1 homologues from all other groups of bacteria, it likely constitutes a deletion in the actinobacterial homologues. By means of PCR amplification, we were successful in obtaining sequence information for the Cox1 gene from 22 additional actinobacterial species belonging to different orders and families. All of

		326	368
Proteobacteria	<i>Escherichia coli</i>	NP_752476	LSFIVVLLHHFFTMG AG ANVNAFFGITMI IAIPTGVKIFNWL
	<i>Pseudomonas aeruginosa</i>	ZP_00138941	-G-T-----S- GD--G--VA--L-S-----L-----
	<i>Vibrio vulnificus</i>	NP_762523	---V--A--M--T- MP VFAEL--MYC--M--V-----VA
	<i>Yersinia pseudotuberculosis</i>	YP_069492	-----S- -----A--S-----
	<i>Agrobacterium tumefaciens</i>	NP_353177	--YL-----S- -----S--A-----
	<i>Caulobacter crescentus</i>	NP_420580	--Y-----S- -----S--A-----
	<i>Bordetella parapertussis</i>	NP_886335	---L--A--M--T- IP VVGQL--MYA--L-S-----VA
	<i>Burkholderia fungorum</i>	ZP_00032266	---M--A--M--AT- MP VTGQL--MYA--L--V-----V--VA
	<i>Geobacter sulfurreducens</i>	NP_951280	AGSL--G--M--S- MS DTAVLV-SFLSF-V---SAI-V---IS
	<i>Desulfovibrio vulgaris</i>	YP_011033	AGSL--A--M--S- MS DTAVLV-SFL-FVV---SAI-V---VS
Other bacteria	<i>Aquifex aeolicus</i>	NP_214504	V--FL-I--M-VS- VP NWTRVL-SY--LL--V---I---ML
	<i>Cytophaga hutchinsonii</i>	ZP_00309629	-----A--M-VT- MN FFLGSI-MFL-L--V-SA--A--YIA
	<i>Leptospira interrogans</i>	NP_710424	V--L--G--M-VS- QS EFAGVL-SF---LVGV--AI-L---IS
	<i>Prochlorococcus marinus</i>	NP_895169	-GLV--A--MF-S- TP PWMRL--T-A-SF--V---I-F---A
	<i>Gloeobacter violaceus</i>	NP_925109	-GFL--A--M--S- TP DWLRM--MV-SFL--V---I-V-S--G
	<i>Nostoc punctiforme</i>	ZP_00110624	-GLI--A--M--S- IP GWLRM--M-----V---I---S--A
	<i>Chloroflexus aurantiacus</i>	ZP_00355855	-G-L--G--M-VS- QS VYAGLI-SFI--LV---SAI-V---TA
	<i>Thermus thermophilus</i>	YP_005640	-GTM--A--M--V- ES TLFQIA-AFF-AL--V---L--IIG
	<i>Deinococcus radiodurans</i>	NP_296339	V-C-----M-AV- IP EAWQIA-M-S-L-V-V-----LIG
	<i>Bacillus cereus</i>	NP_830510	-----V- -- PA--S--S--S--A--S-----
Firmicutes	<i>Exiguobacterium</i> sp.	ZP_00182647	-GFM--V--M--V- L PVA--I--AVA--A--V-----
	<i>Geobacillus kaustophilus</i>	YP_149311	-----V- -- PA--SA--S--A-----
	<i>Oceanobacillus iheyensis</i>	NP_693175	-----V- -- Q--LT--SI--S--A--V---I---L
	<i>Listeria innocua</i>	NP_469361	-----V- -- S--L--S--S---M-----I-----
	<i>Staphylococcus aureus</i>	YP_040448	-----L-V----- N--LI--S--S--S--L-G-----L---L
	<i>Symbiobacterium thermophilum</i>	YP_075926	MG-T--S--M--V- M- PV--SI--SL--A--V-----S
	<i>Mycobacterium leprae</i>	NP_302190	--VA--A--M--AT- -VLLP--SFM-YL--V---F--VG
	<i>Mycobacterium tuberculosis</i>	NP_217559	--VA--A--M--AT- -VLLP--SFM-YL--V---I-F--IG
	<i>Corynebacterium efficiens</i>	NP_739028	--MAV--A--MFVT- -VLLP--SFM-FL-SV---F--VG
	<i>Corynebacterium glutamicum</i>	NP_601724	--MA--A--M-VT- -VLLP--SFM-FL-SV---F--VG
Actinobacteria	<i>Nocardia corynebacterioides</i>	AY876134	--IA--A--MYAT- -VLLP--SFM-FL--V---F--IG
	<i>Rhodococcus erythropolis</i>	NP_898717	--IA--A--MYAT- -VLLPY-SFM-FL--V---F--IG
	<i>Rhodococcus rhodochrous</i>	AY876126	--IA--A--MYAT- -VLLPY-SFM-FL--V---F--IG
	<i>Gordonia westfalica</i>	NP_954783	--VA--A--MYVT- -VLLP--SFM-FL--V---F--IG
	<i>Gordonia rubripertincta</i>	AY876122	--VA--A--MYVT- -VLLP--SFM-FL--V---F--IG
	<i>Tsukamurella paurometabola</i>	AY876131	--VA--A--MYAT- -VLLP--SFM-FL--V---F--I
	<i>Williamsia muralis</i>	AY876139	--VA--A--MYAT- -VLLP--SFM-F--V---F--IG
	<i>Streptomyces avermitilis</i>	NP_827224	--VT--A--MYVT- -VLLP--SFM-FL--V---F--IG
	<i>Actinoadura glauciflava</i>	AY876130	--IT--A--M-VT- QVLLP--SFM-FL--V---F--VG
	' <i>Trichotomospora caesia</i> '	AY876119	--VV--A--M-AT- -VLLP--SVLSFL--V---F--AG
	<i>Streptosporangium roseum</i>	AY876128	--IT--A--M-PT- QVLLP--SFM-FL--V---F--IG
	<i>Microtetraspora niveoalba</i>	AY876123	--MT--A--M-AT- -ALLP--SMLSFL--V---I-F--TG
	<i>Planobispora rosea</i>	AY876132	--VA--A--M-PT- QVLLP--SFM-FL--V---F--IG
	<i>Thermobifida fusca</i>	ZP_00057854	--MT--A--M-PT- -VLLP--SFM-FL--V---F--IG
	<i>Micromonospora chersina</i>	AY876125	--MS--A--M-AT- QVLLP--SFLSYL--V---M-F-S-IG
	<i>Propionibacterium acnes</i>	AY876127	--VS--A--M-VT- -VSLP--SFMFTT--V---F--IG
	<i>Pseudonocardia halophobica</i>	AY876129	--AA--A--MYAT- -VLL--SF--LL--V---I-FV--IG
	<i>Saccharopolyspora erythraea</i>	AY876135	--VV--A--MYAT- -VLLP--AF--FL--V---M-F--IG
	<i>Nocardioides simplex</i>	AY876138	--VA--A--M-VT- -VNL--SGM-FL--V---F--IG
	<i>Kribbella sandramycini</i>	AY876137	--VA--A--I-VT- -MNL--SFM-FL--V---F--IG
<i>Tropheryma whipplei</i>	NP_787372	--VT--A--MYVT- -VLLP--SFM--L--V-----VG	
<i>Leifsonia xyli</i>	YP_062417	--IT--A--MYVT- SVLLPW-SLM--L--V-----IG	
<i>Cellulomonas fimi</i>	AY876120	--VT--A--RMVVT- SVLLP--AFM--L--V---F--IG	
<i>Oerskovia turbata</i>	AY876140	--V--A--MYVT- -VLLP--AFM--L--V---F--IG	
<i>Kocuria rhizophila</i>	AY876121	--VT--A--MYVT- -VALG--SFM--M--V---F--IG	
<i>Arthrobacter nicotinovorans</i>	AY876124	--VT--A--MYVT- SVLLP--AFM--L--V---F--IG	
<i>Clavibacterium michiganensis</i>	AY876133	--VT--A--MYVT- SVLLP--SLM--L--V---F--IG	
<i>Microbacterium oxydans</i>	AY876136	--VA--A--MHVT- SVLLP--ALM--L--V-----IG	

Fig. 1. Partial alignment of Cox1 sequences depicting a signature consisting of 2 aa deletion (underneath the boxed region) that is specific for actinobacteria. Dashes indicate identity with the amino acid on the top line. Accession numbers are shown in the second column. Sequences in Figs 1–4 whose accession numbers begin with the letters 'AY' were obtained in this work. Only representative sequences from different groups are shown. Sequence information for all available species is presented in Supplementary Fig. S1 in IJSEM Online.

these new fragments were also found to contain this 2 aa deletion at the same position. Sequence information for this region from various actinobacterial species (including those sequenced in the present work) and some representatives from other groups of bacteria is presented in Fig. 1. A complete alignment of all available Cox1 homologues in GenBank, which includes 37 sequences from actinobacterial species and 117 sequences from other bacterial groups such as *Firmicutes*, 'Deinococcus–Thermus', *Cyanobacteria*, *Chlamydiae*, the *Cytophaga–Flavobacterium–Bacteroides*–green sulfur bacteria (CFBG), *Spirochaetes*, *Aquifex* and *Proteobacteria*, is provided in Supplementary Fig. S1 available in IJSEM Online. As seen, the observed 2 aa deletion is

unique to actinobacteria, and no exceptions were observed. The shared presence of this 2 aa deletion in various actinobacteria strongly indicates that it was introduced only once in a common ancestor of the actinobacteria and passed on to all their descendants. Because of its unique presence in various actinobacteria, this signature provides a good molecular marker for distinguishing actinobacterial species from other bacterial phyla. In the structure of the Cox1 protein from *Paracoccus denitrificans*, the region where this deletion is present (residues 332–333) is located on the periplasmic surface (Michel *et al.*, 1998), but its functional significance remains to be determined. It should be mentioned that the 2 aa deletion in Cox1 is absent in

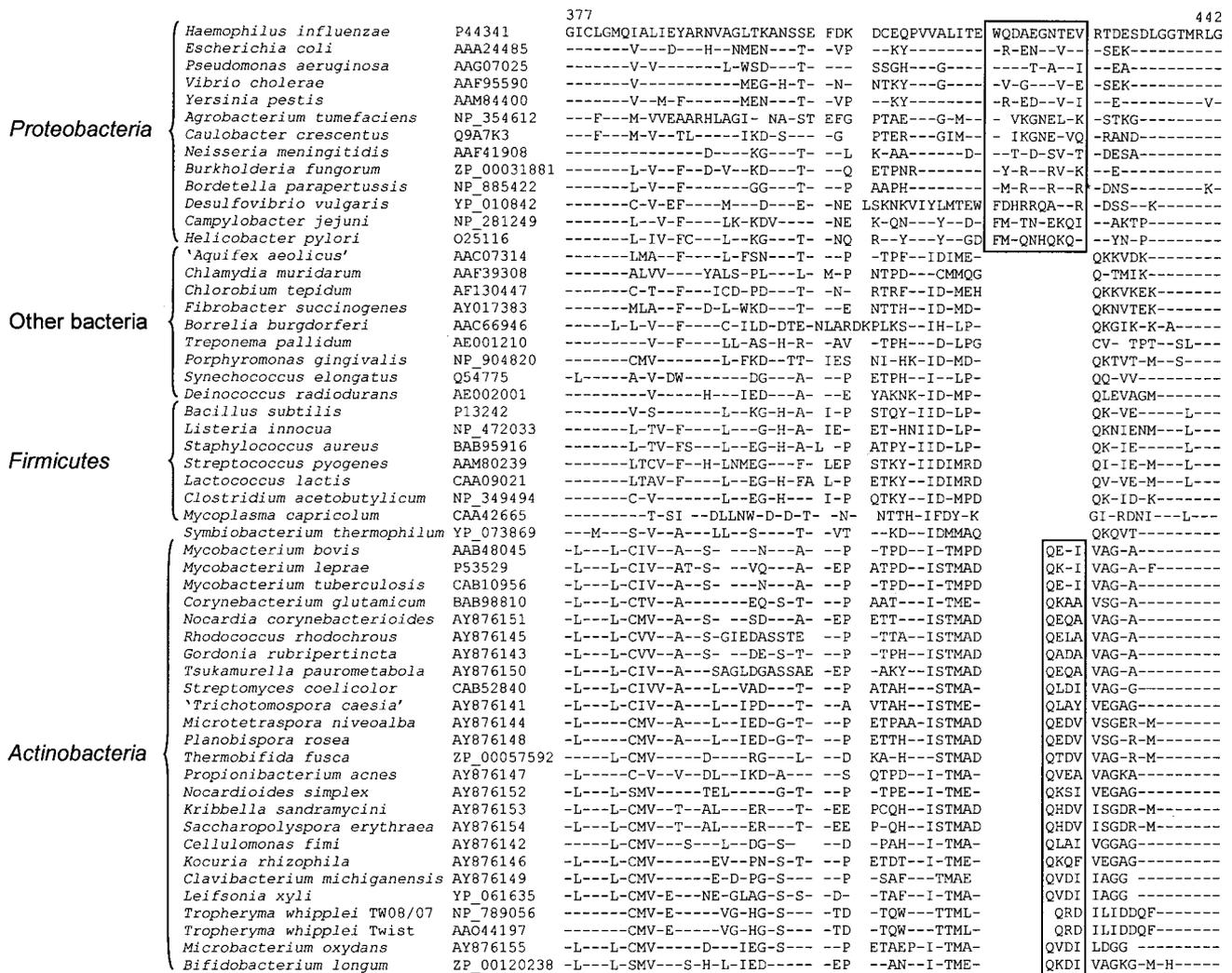


Fig. 2. Partial alignment of CTP synthetase sequences showing two different signatures. The 10 aa insert is specific for the proteobacteria, whereas the smaller, 4 aa insert is a characteristic of all actinobacteria. Sequence information for other available species is presented in Supplementary Fig. S2.

Symbiobacterium thermophilum, which is presently grouped within the *Actinobacteria*. However, recent genomic analyses indicate that this species is much more closely related to bacilli and clostridia than to actinobacteria (Ueda *et al.*, 2004) (see Discussion).

Another signature for actinobacteria is present in the enzyme CTP synthetase, which catalyses the conversion of UTP into CTP by transferring an amino group to the 4-oxo group of the uracil ring (Endrizzi *et al.*, 2004). Except for the mycoplasma species, the gene encoding CTP synthetase is present in all other microbial genomes and has only one copy in the genome. A 10 aa insert which is distinctive of various proteobacterial species has previously been identified in this protein (Gupta, 2000). Interestingly, in the same position where this proteobacterial insert is present, a smaller, 4 aa insert is found in all actinobacterial species

except *Tropheryma whippelii*, which contains a 3 aa insert in this position (Fig. 2). This insert is again highly specific for actinobacteria and is not present in the CTP synthetase homologues from any other bacteria whose sequences are available in the databases (see Supplementary Fig. S2). In the present work, we have successfully amplified CTP synthetase gene fragments from 15 additional actinobacterial strains and all of these were found to contain this 4 aa insert in the same position (Fig. 2). We were not able to amplify sequences for other actinobacterial strains, including that for *Rubrobacter radiotolerans* DSM 5868^T, using these primer sets. Based on its shared presence in various actinobacterial species, the identified insert in CTP synthetase was likely introduced in a common ancestor of the phylum *Actinobacteria*, independently of the insert in the proteobacteria, and it provides a specific and useful molecular marker for the actinobacterial group. The CTP synthetase

		194		241
Proteobacteria	<i>Escherichia coli</i>	NP_416899	ITHVIRGEDHINNTPRQINILKAL	KAPV PVYAH VSMINGDDGKLSKR
	<i>Haemophilus somnus</i>	ZP_00122037	-----V-----	G--I -T--- -----Q-----
	<i>Pseudomonas aeruginosa</i>	NP_251824	----L---EWLPSA-KL-KLYEYF	GWEQ -QLCYMPLLRNP-KS ----
	<i>Agrobacterium tumefaciens</i>	NP_356183	----A---EWLASV-KH-LLYRYF	GWDO -TFM-L-LMRNA-KS ----
	<i>Caulobacter crescentus</i>	NP_420712	V-----D--L-AA--TL-YQ-M	DWA- -AF--IPL-H--P--A----
	<i>Bordetella pertussis</i>	NP_881555	-----L-D-V-----R--	G-TL -E-G--P--L--P--E----
	<i>Burkholderia mallei</i>	YP_103237	-----D-V-----R--	GGE- -L-T--N-Q-E-M----
	<i>Neisseria meningitidis</i>	NP_273069	V-----D--V-----K----	D-NL -E-G-LP-L-NEQ---I----
	<i>Geobacter sulfurreducens</i>	NP_952272	--T---D-----LLYE--	GY-- -RF--P--L--A-KTR----
	<i>Desulfovibrio vulgaris</i>	YP_010911	V-Q-V--A----C----L---SL-	G-QP -A-L- --LL-DAT-ER-A--
Other bacteria	<i>Campylobacter jejuni</i>	NP_282434	VSD--D--LS---K--VLYE--	GFKI -KFF--A--H--E-----
	<i>Helicobacter pylori</i>	NP_207274	--D---D--LS---K--VLY--	NFKI -NFF--P--L-NEE-Q-----
	<i>'Aquifex aeolicus'</i>	NP_213835	-----P---K--L-YE--	GF-- -KFF- LPV-L-E-RS----
	<i>Chlamydia muridarum</i>	AAF39540	-----L---EWLSS--KHLLLYE-F	GWEP -QF--MPLLRNP-GS ----
	<i>Cytophaga hutchinsonii</i>	ZP_00118857	-----EWLPSA-LHVLLYRF-	GWEDTM-QF--LPLLLKP-GNG----
	<i>Leptospira interrogans</i>	NP_714370	-----VG-LS-----LLYE--	GYNI -EF-- A-E-V-M-----
	<i>Notoc punctiforme</i>	ZP_00105808	-----A---AK--LLYE-M	G-KI -EFS- TPL-LNME-R-----
	<i>Synechococcus sp.</i>	CAD29422	-S-----A--AK--LLYE--	G-A- -EF-- TPL-LNKE-R-----
	<i>Prochlorococcus marinus</i>	ZP_00104468	-----V-----S--AK--L-Y--	NFKL -IF-- TPL-LNSE-----
	<i>Thermus thermophilus</i>	CAA45854	V-D--A-EWLVS--IHVLLYR-F	GWEA -RFY-MPLLRNP-KT -I----
Firmicutes	<i>Deinococcus radiodurans</i>	NP_294208	V--V-A-EW-TS--IHVLLYR-F	GW-E --F--MPLLRNA-KS -I----
	<i>Thermotoga maritima</i>	NP_229152	-----L---L--LALYE-F	EKAP --F-- -T-L-----
	<i>Bacillus subtilis</i>	NP_387973	M--L---I-S---K--M-YQ-F	GWDI -QFG- MTL-VNESR-----
	<i>Listeria monocytogenes</i>	NP_463768	-S--L--D--S---K--L-YN-F	GWEP -IFG- MTL-VNESR-----
	<i>Staphylococcus aureus</i>	AAW37684	-SD--D--S---K--M-YE-F	GWEP -RFG-M-L-VNEER-----
	<i>Lactobacillus johnsonii</i>	NP_964422	-----L-D-VA---K-LVYYE--	GWEP -KFG-MTL-INSET-----
	<i>Enterococcus faecalis</i>	NP_813853	-S--L--D--A--K--LM-YE-F	EWTP --FG-MTL-INSET-----
	<i>Streptococcus pyogenes</i>	NP_663974	-S-----D---A--K--LMVYE--	GWEA -EFG-MTL--SET-----
	<i>Lactococcus lactis</i>	NP_268210	-S-----D---A--K-LVYYD--	GWEA -QFG-MTL-INSET-----
	<i>Clostridium acetobutylicum</i>	NP_347626	-----V--NEYLSSA-KYNRLYE-F	GW-N -I-V- CPP-MK-AHS-----
Actinobacteria	<i>Mycoplasma penetrans</i>	NP_758418	-S-I---E-LS---Y--A-NE--	DINKN-I -I-FG-L-V-I-DET-----
	<i>Symbiobacterium thermophilum</i>	YP_073845	-----L---G--P---V--L-YQ--	GF-- -EFG- LGHMTNPERG-----
	<i>Mycobacterium leprae</i>	NP_302160	-----L---LLSS---VALYQ--	IRIGM AERI -EFG- FFSVL-EGT-----
	<i>Mycobacterium tuberculosis</i>	NP_217508	-----L---LLPS---LALHQ--	IRIGV AERI -KF-- LPTVL-EGT-----
	<i>Mycobacterium avium</i>	NP_961963	-----L---LLPS---LALYQ--	IRIGV AERI -QF-- LPTVL-EGT-----
	<i>Corynebacterium diphtheriae</i>	CAE49635	-----L---LLPS---LALYE--	KVIGV AQQT -EFG- LPFVM-EGN-----
	<i>Corynebacterium efficiens</i>	NP_737999	-----L---LLPS---ALYE--	KRIGV AKQT -EFG- LPFVM-EGN-----
	<i>Corynebacterium glutamicum</i>	BAB98686	V---L---LLPS---LALYE--	KRIGV AKAT -AFG- LPFVM-EGN-----
	<i>Nocardia corynebacterioides</i>	AY876165	-----L---LLSS---APYE--	VRIGV ADR- -EFG- LPFVM--GN-----
	<i>Rhodococcus rhodochrous</i>	AY876163	-----L---LLSS---LALYA--	QRIGV TDFT -EFG- LPFVM-QGN-----
Actinobacteria	<i>Streptomyces avermitilis</i>	NP_823867	-----L---LLSS---VALYR--	IELGI AKS- -EFG- LPYVM-EGN-----
	<i>'Trichotomospora caesia'</i>	AY876161	-----L---LLSS---ALYR--	IELGI ATA- -EFG- LPYVM-EGN-----
	<i>Pseudonocardia halophobica</i>	AY876160	-----L---LLSS---AL-E--	QRVGI GNGP F---- LPLVT-EGNR-----
	<i>Nocardioides simplex</i>	AY876162	-----L---LLSS---ALYA--	EEVGI AKAT -QFG- FPYVM-RGH-----
	<i>Kribbella sandramycini</i>	AY876164	-----L---LLSS---ALYA--	AEIGI GSGRT -RFG- LPMVMGPEGN-----
	<i>Kineococcus radiotolerans</i>	ZP_00228423	-----L---LLSS---ALHE--	VDIGV SQA- -LFG- LPYVT-EGS-----
	<i>Oerskovia turbata</i>	AY876156	-----L---LLSS---VVLYR--	LELGV AHVM -EFG- LPYVM-EGN-----
	<i>Cellulomonas fimi</i>	AY876166	-----L---LLSS---VVLYR--	LDLGV ATVM -EFG- MPYVM-EGN-----
	<i>Kocuria rhizophila</i>	AY876157	-Q-L---LLSS---ALYR--	VAIGV AEFI -EFG- LPYVM-EGN-----
	<i>Clavibacterium michiganensis</i>	AY876158	-----L---LLSS---ALYR--	YAIGV AEYM -EFG- LPYVM-QGN-----
Actinobacteria	<i>Arthrobacter nicotinovorans</i>	AY876159	-----L---LLSS---ALYR--	YAIGV AEYM -EFG- LPYVM-QGN-----
	<i>Tropheryma whippelii</i>	NP_787337	-----L---LLSS---ALYK--	ITIGI TDY- -RFG- LPYVM-EGNR-----
	<i>Leifsonia xyli</i>	YP_062250	V---L---LLSS---ALYH--	IETGV TTF- -EFG- LPYVM-EGN-----
	<i>Bifidobacterium longum</i>	ZF_00121389	-----L---LLSS---VLYRY-	MELGI AKEM -LFG- MPYVM-QGN-----
	<i>Thermobifida fusca</i>	ZF_00057289	-----V---E-LS---K-QLLWE-	GQTP --W-- LPV-VNEKRO-----
	<i>Propionibacterium acnes</i>	YP_056550	-DT-V---EW-SS--KHLLLY-W	GWEA -KF-- MPLLRNT-KS-I----
	<i>Rubrobacter xylanophilus</i>	ZF_00186416	-----D---S-----M-HR--	GHEL -AF-- -PQVL-P-R-----

Fig. 3. Partial alignment of GluRS sequences showing a 5 aa insert (boxed) which is specific for actinobacteria. Sequence information for other available species is presented in Supplementary Fig. S3.

homologue from *Symbiobacterium thermophilum* again did not contain the 4 aa insert, supporting the inference from the Cox1 signature and other studies that this species is distinct from other actinobacteria. The enzyme CTP synthetase consists of a single polypeptide containing two domains: the C-terminal glutamine amide transfer (GAT) domain catalyses the hydrolysis of glutamine, whereas the N-terminal synthase domain is responsible for the amination of UTP (Endrizzi *et al.*, 2004). The observed inserts are located in the GAT domain of the enzyme and they are indicated to be present on the outer surface of the protein. However, their functional significance for these groups of bacteria remains to be determined.

In the enzyme glutamyl-tRNA synthetase (GluRS), which plays an essential role in protein synthesis by charging the

glutamyl-tRNA with its cognate amino acid (Woese *et al.*, 2000), a 5 aa insert is present in almost all actinobacterial species, with the exception of *Thermobifida fusca*, *Propionibacterium acnes* and *Rubrobacter xylanophilus*, but not found in any other bacterial groups (Fig. 3). Sequence information is presently available from 182 species belonging to other bacterial groups, and they are all found to lack this insert (see Supplementary Fig. S3). We were successful in PCR-amplifying and sequencing GluRS gene fragments from 11 additional actinobacterial strains, and all of them were found to contain this insert (Fig. 3). Sequences for other actinobacterial strains could not be amplified using the primers employed in this work (Table 2). Of the three actinobacterial species that are lacking this insert, *Rubrobacter xylanophilus* is one of the deepest branching species based on the 16S rRNA tree and other

biochemical traits, whereas the other two species belong to the families *Nocardiopsaceae* (*Thermobifida fusca*) and *Propionibacteriaceae* (*Propionibacterium acnes*) (Embley & Stackebrandt, 1994; Stackebrandt *et al.*, 1997; Bruggemann *et al.*, 2004). Other species belonging to these latter families contain this insert. Hence, the most parsimonious explanation for these results is that the insert in GluRS was introduced after the divergence of *Rubrobacter xylanophilus*, and that *Thermobifida fusca* and *Propionibacterium acnes* have lost this insert as a result of either lateral gene transfer or some other means. The insert in GluRS is also lacking in the homologue from *Symbiobacterium thermophilum*, supporting the inference from other signatures. In the 3D structure of GluRS, the identified insert is present in a loop between an α -helix and a β -sheet (Sekine *et al.*, 2003). Although the flanking conserved regions of this insert are known to be functionally important, the possible consequences of the presence of this insert on GluRS function remain to be determined.

Examination of the actinobacterial specificity of 23S rRNA signature

Roller *et al.* (1992) have previously described a large insert of about 100 nt in the 23S rRNA that was present in a large number of actinobacterial species, but not found in other groups of bacteria. The original paper describing this signature included 64 high-G+C-content Gram-positive strains from the currently recognized phylum *Actinobacteria*. However, many of these strains were from the same genus, even the same species, and they represented only 22 genera. To examine the specificity of this signature further, we have successfully amplified and sequenced the 23S rRNA insert region from 13 additional actinobacterial species representing 13 additional genera, covering all of the major groups within this phylum. Sequence information for these sequences as well as various other actinobacterial species and a few other bacterial groups is presented in the partial sequence alignment shown in Fig. 4. As seen, an insert of between 90 and 100 bp is present in all of the actinobacterial species that we sequenced but it is not found in any other bacterial groups (Fig. 1). Sequence information for many other actinobacterial species which has become available in recent years is also included in Fig. 4. Of these species, smaller inserts were present in this position in *Tropheryma whipplei* (79 bp) and *Microbispora bispora* (30 bp) and *Rubrobacter xylanophilus* was found to be lacking the insert entirely. Sequence information for these three strains was available in GenBank. We have also confirmed the absence of this insert in another *Rubrobacter* species, *Rubrobacter radiotolerans*, by amplifying and sequencing of the insert from the type strain of this species. Because *Rubrobacter* represents a deep branch within the phylum *Actinobacteria* (Stackebrandt *et al.*, 1997; Ludwig & Klenk, 2001), similar to the insert in GluRS, the large insert in 23S rRNA gene was very likely introduced in an actinobacterial ancestor, after the divergence of *Rubrobacter*.

Phylogenetic analyses of actinobacteria based on 16S rRNA and CTP synthetase sequences

In order to see the inter-relationships among different subgroups within the phylum *Actinobacteria*, we constructed neighbour-joining trees based on 16S rRNA gene and CTP synthetase protein sequences. Although there are many phylogenetic trees drawn for actinobacteria based on 16S rRNA gene sequences, they generally do not show any bootstrap scores or other measures for the reliability of the branching order (Embley & Stackebrandt, 1994; Stackebrandt *et al.*, 1997; Ludwig & Klenk, 2001). Fig. 5 shows a phylogenetic tree based on 16S rRNA gene sequences from different actinobacterial species for which sequence information is available for different signatures, as well as representative species from other main groups of bacteria. This tree was rooted using the 16S rRNA gene sequence from *Archaeoglobus fulgidus*. The bootstrap scores for different nodes that were >50 are marked on this tree. As seen, in the 16S rRNA gene tree, most of the actinobacterial species formed a well-defined cluster, branching together in 100% of the bootstrap replicates. A clade consisting of *Rubrobacter radiotolerans* and *Symbiobacterium thermophilum*, which was separated from other actinobacteria by a long branch length, formed the outgroup of the main actinobacterial cluster. A number of subgroups within the *Actinobacteria* are clearly identified in this tree, including *Streptosporangineae*, *Streptomyceinae*, the *Corynebacterium–Mycobacterium–Nocardia* cluster, *Propionibacterineae*, *Pseudonocardineae* and *Micrococccineae*. However, the inter-relationships among these subgroups were not resolved, as seen by the lower (<50) bootstrap scores of the corresponding nodes.

A phylogenetic tree for the *Actinobacteria* was also constructed using CTP synthetase sequences for which the amplified fragment was sufficiently large (330 aa) to carry out phylogenetic analysis (Fig. 6). The overall topology of the CTP synthetase tree was very similar to that seen in the 16S rRNA gene tree, with the exception of deep branching of *Tropheryma whipplei*. Phylogenetic studies based on 16S rRNA gene sequences and several other conserved protein sequences indicated that *Tropheryma whipplei* is closely related to species of the *Micrococccaceae*. Hence, its deep branching in the CTP synthetase tree is most likely due to its accelerated rate of evolution, which is commonly observed for many genes in intracellular bacteria. No sequence information was available for CTP synthetase from *Rubrobacter radiotolerans*. However, *Symbiobacterium thermophilum* was distantly related to the *Actinobacteria*, as seen in the 16S rRNA gene tree.

DISCUSSION

In this work, we describe three novel molecular signatures consisting of conserved inserts and deletion in widely distributed proteins that are present only in actinobacteria. These signatures (a 2 aa deletion in Cox1, a 4 aa insert in CTP synthetase and a 5 aa insert in GluRS) are found only in actinobacterial homologues and are not found in

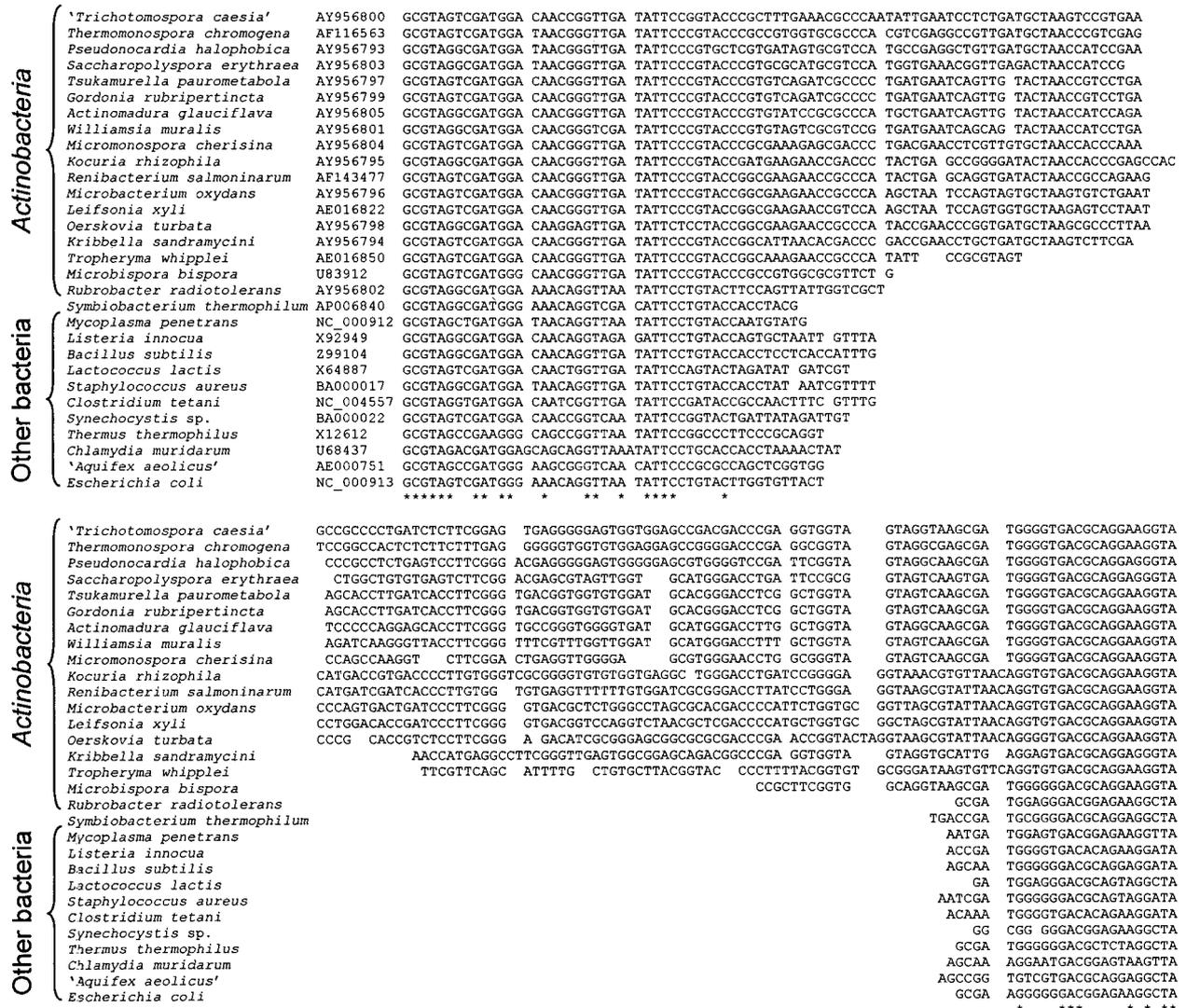


Fig. 4. Partial alignment of 23S rRNA gene sequences showing a large insert (99–110 nt) that is specific for actinobacteria. Sequences with accession numbers starting with ‘AY’ were obtained in this work. Positions which are completely conserved are identified by asterisks (*).

homologues from any other groups of bacteria (> 150) for which extensive sequence information is now available (see Supplementary Figs S1–S3). In addition, we have also tested the actinobacterial specificity of a large insert in the 23S rRNA that was previously described by Roller *et al.* (1992). We have tested the hypothesis that these signatures are distinctive characteristics of actinobacteria by obtaining sequence information from a large number of actinobacterial species, covering a wide range of families and suborders within this phylum, for which no sequence information was available. All of these species were found to contain the indicated signatures, confirming that they are distinctive characteristics of this group.

The placement of a novel bacterial species into the phylum *Actinobacteria* is at present based solely on the branching

pattern in 16S rRNA gene trees. Based on phylogenetic trees, it has proven difficult to circumscribe a given phylum reliably (Ludwig & Klenk, 2001; Gupta & Griffiths, 2002; Oren, 2004). The problems that one faces in these regards are illustrated by the 16S rRNA gene tree shown in Fig. 5. In this tree, most of the actinobacterial species formed a well-defined cluster, branching together in 100% of the bootstrap replicates. A clade consisting of *Rubrobacter radiotolerans* and *Symbiobacterium thermophilum*, which is separated by a large genetic distance from the other actinobacterial species, forms the outgroup of this cluster. As seen by the bootstrap score (< 50%), the two species that form this clade are not specifically related to each other, and the clade itself is only weakly related to other actinobacteria. However, in the current taxonomy based on 16S rRNA trees, both of these species are recognized as part of the phylum

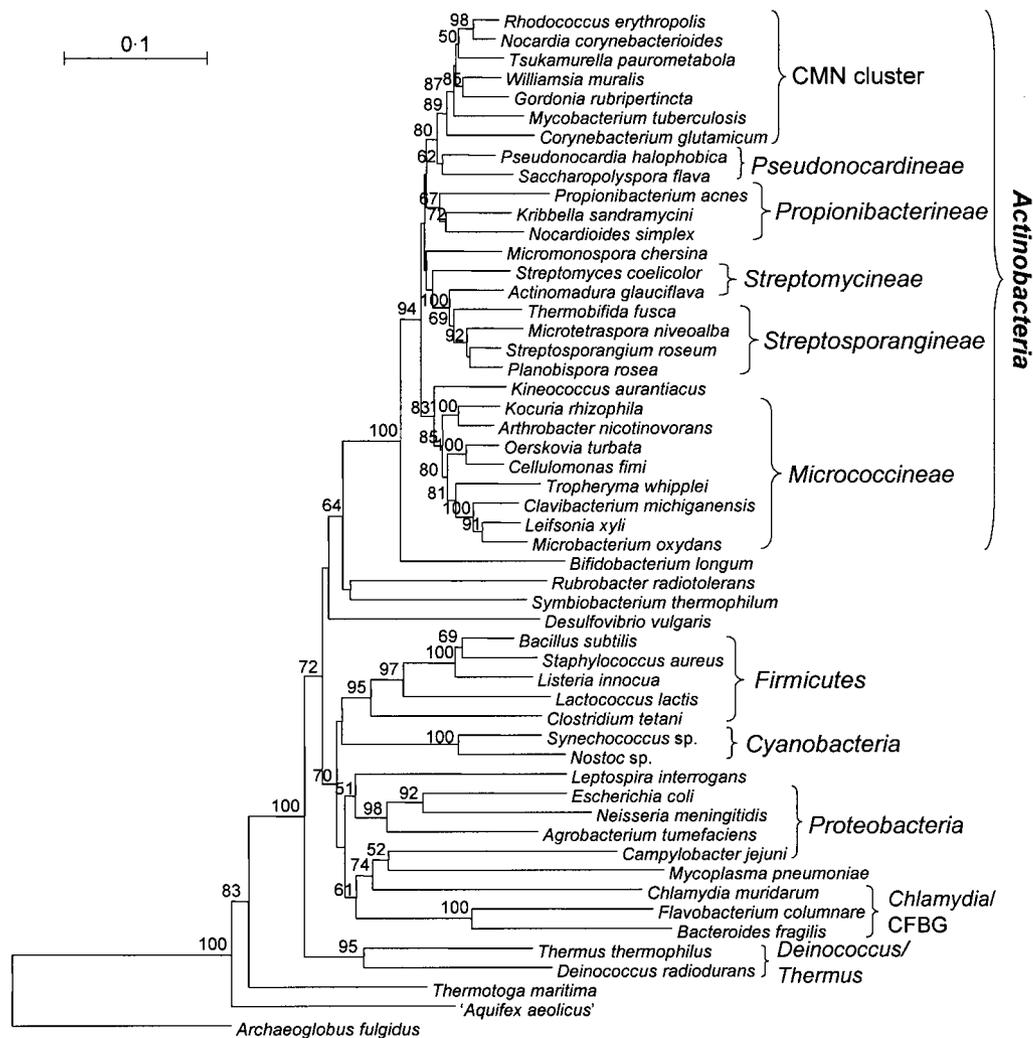


Fig. 5. Neighbour-joining distance tree based on 16S rRNA gene sequences. The tree is based on 1400 nucleotides excluding all indels. Accession numbers of various sequences used for this work are given in Supplementary Table S1 in IJSEM Online. Bootstrap scores greater than 50 are indicated.

Actinobacteria, and the outer boundary of this phylum is placed outside this clade (Ludwig & Klenk, 2001; Maidak *et al.*, 2001). This is despite the fact that analysis of various genes in the *Symbiobacterium thermophilum* genome indicates that the species is most closely related to members of the *Firmicutes* rather than the *Actinobacteria* (Ueda *et al.*, 2004).

In our analyses based on signature sequences, all four signatures (Cox1, CTP synthetase, 23S rRNA and GluRS) that are present in most other actinobacteria were found to be lacking in *Symbiobacterium thermophilum*, indicating strongly that this species should not be placed in the phylum *Actinobacteria* despite its high G+C content. For *Rubrobacter radiotolerans*, the signatures in 23S rRNA as well as the GluRS protein, which are present in most other actinobacteria (two exceptions seen in the case of

GluRS), were found to be lacking in this species. Information for Cox1 and CTP synthetase signatures is at present lacking for this species as we were not able to amplify the corresponding gene fragments. However, based on the absence of signatures in the 23S rRNA and GluRS, and the fact that this species is distantly related to other actinobacteria in the 16S rRNA tree (Fig. 5), it is suggested that the species of this genus are not typical actinobacterial species. However, if the future work reveals other reliable characteristics (including shared presence of indels in the Cox1 and CTP synthetase) which are uniquely shared by *Rubrobacter* species and other actinobacteria, this would validate the placement of this genus in the phylum *Actinobacteria*.

Based on the distribution of different signatures and phylogenetic studies presented here, it can be inferred that

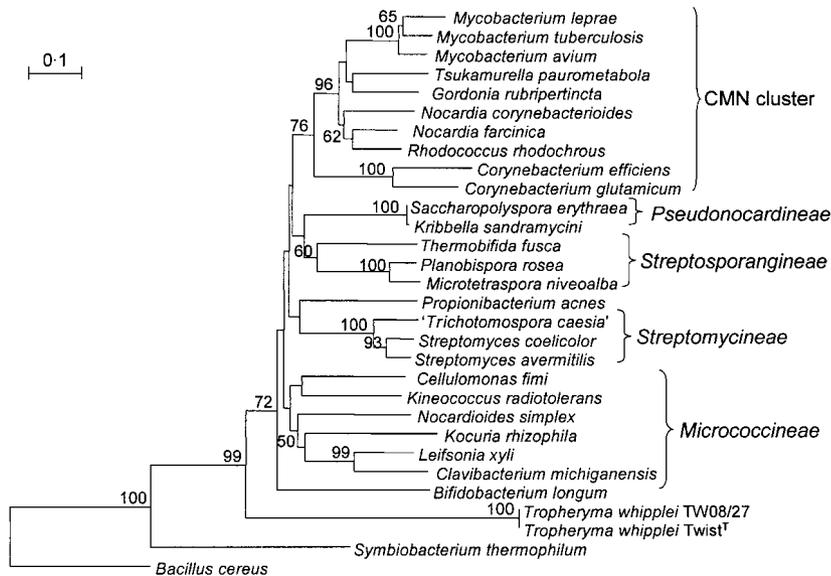


Fig. 6. Neighbour-joining distance tree based on CTP synthetase sequences. The tree is based on 330 aligned amino acid positions excluding all indels. The tree was arbitrarily rooted using a CTP synthetase sequence from a firmicute species (*Bacillus cereus*). Bootstrap scores greater than 50 are indicated. Accession numbers of sequences used for this purpose are given in Supplementary Table S2.

all of these signatures (a 2 aa deletion in Cox1, 4 aa insert in CTP synthetase and a 5 aa insert in GluRS) were introduced in a common ancestor of the typical actinobacteria. There were only two instances in a single protein (absence of the GluRS insert in *Thermobifida fusca* and *Propionibacterium acnes*) where these signatures were found to be missing from an actinobacterial species. The other three signatures were present in all of the newly sequenced as well as previously available actinobacterial homologues. These novel molecular signatures should prove helpful in circumscribing the actinobacterial phylum and for the placement of deep branching species within this group.

It is also of much interest to understand the biological and functional significance of these conserved genetic changes in important proteins that are limited to actinobacterial species. Because most of these indels have not been lost from any of the species belonging to this phylum, it is reasonable to assume that they play important biological roles in these organisms. Hence, studies aimed at understanding their functional significance should be of much interest.

ACKNOWLEDGEMENTS

This work was supported by research grants from the Canadian Institute of Health Research and the National Science and Engineering Research Council of Canada.

REFERENCES

- Ahmad, S., Selvapandiyar, A. & Bhatnagar, R. K. (2000). Phylogenetic analysis of Gram-positive bacteria based on *grpE*, encoded by the *dnaK* operon. *Int J Syst Evol Microbiol* **50**, 1761–1766.
- Atlas, R. M. (1997). *Principles of Microbiology*. New York: WCB McGraw-Hill.
- Bentley, S. D., Chater, K. F., Cerdano-Tarraga, A. M. & 40 other authors (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141–147.
- Boone, D. R., Castenholz, R. W. & Garrity, G. M. (editors) (2001). *Bergey's Manual of Systematic Bacteriology*, 2nd edn, vol. 1. New York: Springer.
- Bruggemann, H., Henne, A., Hoster, F., Liesegang, H., Wiezer, A., Strittmatter, A., Hujer, S., Durre, P. & Gottschalk, G. (2004). The complete genome sequence of *Propionibacterium acnes*, a commensal of human skin. *Science* **305**, 671–673.
- Bull, A. T., Stach, J. E., Ward, A. C. & Goodfellow, M. (2005). Marine actinobacteria: perspectives, challenges, future directions. *Antonie van Leeuwenhoek* **87**, 65–79.
- Camus, J. C., Pryor, M. J., Medigue, C. & Cole, S. T. (2002). Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* **148**, 2967–2973.
- Cerdano-Tarraga, A. M., Efstratiou, A., Dover, L. G. & 23 other authors (2003). The complete genome sequence and analysis of *Corynebacterium diphtheriae* NCTC13129. *Nucleic Acids Res* **31**, 6516–6523.
- Chun, J., Bae, K. S., Moon, E. Y., Jung, S. O., Lee, H. K. & Kim, S. J. (2000). *Nocardioopsis kunsanensis* sp. nov., a moderately halophilic actinomycete isolated from a saltern. *Int J Syst Evol Microbiol* **50**, 1909–1913.
- Davies, J. (1996). Origins and evolution of antibiotic resistance. *Microbiology* **12**, 9–16.
- Embley, T. M. & Stackebrandt, E. (1994). The molecular phylogeny and systematics of the actinomycetes. *Annu Rev Microbiol* **48**, 257–289.
- Endrizzi, J. A., Kim, H., Anderson, P. M. & Baldwin, E. P. (2004). Crystal structure of *Escherichia coli* cytidine triphosphate synthetase, a nucleotide-regulated glutamine amidotransferase/ATP-dependent amidoligase fusion protein and homologue of anticancer and antiparasitic drug targets. *Biochemistry* **43**, 6447–6463.
- Farrar, M. D., Ingham, E. & Holland, K. T. (2000). Heat shock proteins and inflammatory acne vulgaris: molecular cloning, over-expression and purification of a *Propionibacterium acnes* GroEL and DnaK homologue. *FEMS Microbiol Lett* **191**, 183–186.
- Fleischmann, R. D., Alland, D., Eisen, J. A. & 23 other authors (2002). Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J Bacteriol* **184**, 5479–5490.
- Garnier, T., Eiglmeier, K., Camus, J. C. & 19 other authors (2003). The complete genome sequence of *Mycobacterium bovis*. *Proc Natl Acad Sci U S A* **100**, 7877–7882.

- Garrity, G. M. & Holt, J. G. (2001).** The road map to the *Manual*. In *Bergey's Manual of Systematic Bacteriology*, 2nd edn, vol. 1, pp. 119–166. Edited by D. R. Boone, R. W. Castenholz & G. M. Garrity. New York: Springer.
- Goodfellow, M. & Williams, S. T. (1983).** Ecology of actinomycetes. *Annu Rev Microbiol* **37**, 189–216.
- Griffiths, E. & Gupta, R. S. (2004a).** Distinctive protein signatures provide molecular markers and evidence for the monophyletic nature of the *Deinococcus-Thermus* phylum. *J Bacteriol* **186**, 3097–3107.
- Griffiths, E. & Gupta, R. S. (2004b).** Signature sequences in diverse proteins provide evidence for the late divergence of the order *Aquificales*. *Int Microbiol* **7**, 41–52.
- Gupta, R. S. (1998).** Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaeobacteria, eubacteria, and eukaryotes. *Microbiol Mol Biol Rev* **62**, 1435–1491.
- Gupta, R. S. (2000).** The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. *FEMS Microbiol Rev* **24**, 367–402.
- Gupta, R. S. (2001).** The branching order and phylogenetic placement of species from completed bacterial genomes, based on conserved indels found in various proteins. *Int Microbiol* **4**, 187–202.
- Gupta, R. S. (2004).** The phylogeny and signature sequences characteristics of *Fibrobacteres*, *Chlorobi*, and *Bacteroidetes*. *Crit Rev Microbiol* **30**, 123–143.
- Gupta, R. S. & Griffiths, E. (2002).** Critical issues in bacterial phylogenies. *Theor Popul Biol* **61**, 423–434.
- Gupta, R. S., Pereira, M., Chandrasekera, C. & Johari, V. (2003).** Molecular signatures in protein sequences that are characteristic of cyanobacteria and plastid homologues. *Int J Syst Evol Microbiol* **53**, 1833–1842.
- Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. & Omura, S. (2003).** Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol* **21**, 526–531.
- Ishikawa, J., Yamashita, A., Mikami, Y., Hoshino, Y., Kurita, H., Hotta, K., Shiba, T. & Hattori, M. (2004).** The complete genomic sequence of *Nocardia farcinica* IFM 10152. *Proc Natl Acad Sci U S A* **101**, 14925–14930.
- Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. & Gibson, T. J. (1998).** Multiple sequence alignment with CLUSTAL X. *Trends Biochem Sci* **23**, 403–405.
- Karlin, S., Mrázek, J. & Gentles, A. J. (2003).** Genome comparisons and analysis. *Curr Opin Struct Biol* **13**, 344–352.
- Kimura, M. (1980).** A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111–120.
- Labeda, D. P. & Kroppenstedt, R. M. (2000).** Phylogenetic analysis of *Saccharothrix* and related taxa: proposal for *Actinosynnemataceae* fam. nov. *Int J Syst Evol Microbiol* **50**, 331–336.
- Lechevalier, H. A. & Lechevalier, M. P. (1967).** Biology of actinomycetes. *Annu Rev Microbiol* **21**, 71–100.
- Leyden, J. J. (2001).** The evolving role of *Propionibacterium acnes* in acne. *Semin Cutan Med Surg* **20**, 139–143.
- Ludwig, W. & Klenk, H.-P. (2001).** Overview: a phylogenetic backbone and taxonomic framework for prokaryotic systematics. In *Bergey's Manual of Systematic Bacteriology*, 2nd edn, vol. 1, pp. 49–65. Edited by D. R. Boone, R. W. Castenholz & G. M. Garrity. New York: Springer.
- Maidak, B. L., Cole, J. R., Lilburn, T. G. & 7 other authors (2001).** The RDP-II (Ribosomal Database Project). *Nucleic Acids Res* **29**, 173–174.
- Michel, H., Behr, J., Harrenga, A. & Kannt, A. (1998).** Cytochrome *c* oxidase: structure and spectroscopy. *Annu Rev Biophys Biomol Struct* **27**, 329–356.
- Monteiro-Vitorello, C. B., Camargo, L. E., Van Sluys, M. A. & 41 other authors (2004).** The genome sequence of the gram-positive sugarcane pathogen *Leifsonia xyli* subsp. *xyli*. *Mol Plant Microbe Interact* **17**, 827–836.
- Oren, A. (2004).** Prokaryote diversity and taxonomy: current status and future challenges. *Philos Trans R Soc Lond B Biol Sci* **359**, 623–638.
- Raoult, D., Ogata, H., Audic, S., Robert, C., Suhre, K., Drancourt, M. & Claverie, J. M. (2003).** *Tropheryma whipplei* Twist: a human pathogenic actinobacteria with a reduced genome. *Genome Res* **13**, 1800–1809.
- Richert, K., Brambilla, E. & Stackebrandt, E. (2005).** Development of PCR primers specific for the amplification and direct sequencing of *gyrB* genes from microbacteria, order *Actinomycetales*. *J Microbiol Methods* **60**, 115–123.
- Roller, C., Ludwig, W. & Schleifer, K. H. (1992).** Gram-positive bacteria with a high DNA G+C content are characterized by a common insertion within their 23S rRNA genes. *J Gen Microbiol* **138**, 167–175.
- Schell, M. A., Karmirantzou, M., Snel, B. & 9 other authors (2002).** The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *Proc Natl Acad Sci U S A* **99**, 14422–14427.
- Schrempf, H. (2001).** Recognition and degradation of chitin by streptomycetes. *Antonie van Leeuwenhoek* **79**, 285–289.
- Sekine, S., Nureki, O., Dubois, D. Y., Bernier, S., Chenevert, R., Lapointe, J., Vassilyev, D. G. & Yokoyama, S. (2003).** ATP binding by glutamyl-tRNA synthetase is switched to the productive mode by tRNA binding. *EMBO J* **22**, 676–688.
- Smith, D. R., Richterich, P., Rubenfield, M. & 22 other authors (1997).** Multiplex sequencing of 1.5 Mb of the *Mycobacterium leprae* genome. *Genome Res* **7**, 802–819.
- Stach, J. E., Maldonado, L. A., Ward, A. C., Goodfellow, M. & Bull, A. T. (2003).** New primers for the class *Actinobacteria*: application to marine and terrestrial environments. *Environ Microbiol* **5**, 828–841.
- Stackebrandt, E., Rainey, F. A. & Ward-Rainey, N. L. (1997).** Proposal for a new hierarchic classification system, *Actinobacteria* classis nov. *Int J Syst Bacteriol* **47**, 479–491.
- Sutcliffe, I. C. & Harrington, D. J. (2002).** Pattern searches for the identification of putative lipoprotein genes in Gram-positive bacterial genomes. *Microbiology* **148**, 2065–2077.
- Ueda, K., Yamashita, A., Ishikawa, J., Shimada, M., Watsuji, T., Morimura, K., Ikeda, H., Hattori, M. & Beppu, T. (2004).** Genome sequence of *Symbiobacterium thermophilum*, an uncultivable bacterium that depends on microbial commensalism. *Nucleic Acids Res* **32**, 4937–4944.
- Van de Peer, Y. & De Wachter, R. (1994).** TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comput Appl Biosci* **10**, 569–570.
- Warren, R., Hsiao, W. W., Kudo, H. & 23 other authors (2004).** Functional characterization of a catabolic plasmid from polychlorinated-biphenyl-degrading *Rhodococcus* sp. strain RHA1. *J Bacteriol* **186**, 7783–7795.
- Woese, C. R. (1987).** Bacterial evolution. *Microbiol Rev* **51**, 221–271.
- Woese, C. R., Olsen, G. J., Ibba, M. & Soll, D. (2000).** Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol Mol Biol Rev* **64**, 202–236.