International Journal of Systematic and Evolutionary Microbiology (2008), **58**, 149–159 **DOI** 10.1099/ijs.0.65224-0

A multilocus phylogeny of the *Streptomyces griseus* 16S rRNA gene clade: use of multilocus sequence analysis for streptomycete systematics

Yinping Guo,[1,2] Wen Zheng,[1] Xiaoying Rong[1,2] and Ying Huang[1]

Correspondence
Ying Huang
huangy@im.ac.cn

[1]State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, PR China

[2]Department of Biology, Graduate University of Chinese Academy of Sciences, Beijing 100049, PR China

Streptomycetes are a complex group of actinomycetes that produce diverse bioactive metabolites of commercial significance. Systematics can provide a useful framework for identifying species that may produce novel metabolites. However, previously proposed approaches to the systematics of *Streptomyces* have suffered from either poor interlaboratory comparability or insufficient resolution. In particular, the *Streptomyces griseus* 16S rRNA gene clade is the most challenging and least defined group within the genus *Streptomyces* in terms of phylogeny. Here we report the results of a multilocus sequence analysis scheme developed to address the phylogeny of this clade. Sequence fragments of six housekeeping genes, *atpD*, *gyrB*, *recA*, *rpoB*, *trpB* and 16S rRNA, were obtained for 53 reference strains that represent 45 valid species and subspecies. Analysis of each individual locus confirmed the suitability of loci and the congruence of single-gene trees for concatenation. Concatenated trees of three, four, five and all six genes were constructed, and the stability of the topology and discriminatory power of each tree were analysed. It can be concluded from the results that phylogenetic analysis based on multilocus sequences is more accurate and robust for species delineation within *Streptomyces*. A multilocus phylogeny of six genes proved to be optimal for elucidating the interspecies relationships within the *S. griseus* 16S rRNA gene clade. Our multilocus sequence analysis scheme provides a valuable tool that can be applied to other *Streptomyces* clades for refining the systematic framework of this genus.

INTRODUCTION

The genus *Streptomyces* encompasses a large group of micro-organisms that are ubiquitous in nature, and well-known producers of diverse bioactive compounds such as antibiotics. Actinomycetes make more than half of the known antibiotics derived from microbial sources, and amongst them some 75 % are made by streptomycetes (Berdy, 2005). Many members of this genus also play important roles in biodegradation and bioremediation (Ishiyama *et al.*, 2004), such as degrading lignin and recalcitrant aromatic compounds (Phelan *et al.*, 1979; Grund *et al.*, 1990). Notwithstanding these findings, it has been predicted that less than 10 % of streptomycete bioactive metabolites have been discovered (Watve *et al.*, 2001; Clardy *et al.*, 2006).

As a rich source of novel bioactive, commercially significant compounds, the genus *Streptomyces* has been subjected to intensive isolation and screening, resulting in more than 3000 species being described in the literature (including patents) from the 1970s. The large number of isolates and poor species definition caused overspeciation and taxonomic chaos within this genus, the largest genus in the prokaryotes. The International *Streptomyces* Project was launched to establish standard phenotypic criteria for *Streptomyces* species determination during 1964–1972. Several numerical taxonomic and chemotaxonomic studies followed, albeit with limited success (Anderson & Wellington, 2001). Since the 1980s, the advent of molecular techniques has provided a number of genotypic approaches to investigate the taxonomy of *Streptomyces*, including rRNA sequence comparison, DNA–DNA hybridization and DNA fingerprinting (Watve *et al.*, 2001; Lanoot *et al.*, 2004, 2005a; Song *et al.*, 2004). However, due to the respective

Abbreviations: K2P, Kimura two-parameter; ML, maximum-likelihood; MP, maximum-parsimony; NJ, neighbour-joining.

The GenBank/EMBL/DDBJ accession numbers for the sequence data reported in this paper are listed in Table 1.

A series of supplementary figures showing the chromosomal locations of the genes used, and phylogenetic trees for each protein-coding locus and for concatenated multiple genes are available with the online version of this paper.

drawbacks of each method, i.e. insufficient resolution, labour-intensive and high error rate, and low inter-laboratory reproducibility, all have their limitations in routine use.

Despite the current availability of almost all 16S rRNA gene sequences of *Streptomyces* type strains in public databases contributed by researchers from several countries, many species relationships within *Streptomyces* remain unclear. The *Streptomyces griseus* 16S rRNA gene clade has become one of the most taxonomically complex groups, and one which has proved most prolific in the production of bioactive secondary metabolites (Iwamoto *et al.*, 1990; Strohl, 2003; Piel, 2004). The majority of the members of this group share highly similar phenotypes and 16S rRNA sequences. However, reorganization of species within the *S. griseus* clade is a common occurrence. For example, based on DNA–DNA relatedness data, '*Streptomyces caviscabies*', '*Streptomyces setonii*' and '*Streptomyces argenteolus*' have been transferred to *S. griseus* (Liu *et al.*, 2005); '*S. griseus* subsp. *alpha*' and '*S. griseus* subsp. *cretosus*' have been transferred to *Streptomyces microflavus* (Lanoot *et al.*, 2005b); and '*Streptomyces citreofluorescens*', '*Streptomyces chrysomallus* subsp. *chrysomallus*' and '*Streptomyces fluorescens*' have been reassigned to *Streptomyces anulatus* (Lanoot *et al.*, 2005b).

It is important to unravel the taxonomic relationships of reference streptomycetes at species level to guide the species discrimination and the discovery of potentially novel species for ecological reasons and industrial purposes. MLST (Multilocus Sequence Typing) (Maiden *et al.*, 1998; Enright & Spratt, 1999; Urwin & Maiden, 2003) has been successfully applied to the unambiguous characterization of bacterial pathogens (Dingle *et al.*, 2001; Adiri *et al.*, 2003; Baldwin *et al.*, 2005; Coffey *et al.*, 2006), and to phylogenetic analysis of highly diverse bacterial groups, such as the genera *Mycobacterium* (Devulder *et al.*, 2005), *Pseudomonas* (Hilario *et al.*, 2004) and *Haemophilus* (Nørskov-Lauritsen *et al.*, 2005), because of its lab-to-lab portability and reproducibility, and the demonstrated efficiency of inter- and intra-species resolution. In the present investigation, a streptomycete multilocus sequence analysis scheme was developed to clarify the bona fide taxonomic structure of the *S. griseus* 16S rRNA gene clade. We selected six genes: *atpD*, *gyrB*, *recA*, *rpoB*, *trpB* and 16S rRNA, compared the phylogenetic trees derived from the sequence data, constructed a finer and more robust phylogeny, and set up a multilocus sequence database with open access on the internet. This study offers a primary multi-sequence-based framework for amending the systematics of *Streptomyces*, which facilitates our understanding of phylogeny and evolution of this genus.

METHODS

Bacterial strains and culture conditions. Of the 55 strains used in this study (listed in Table 1), 53 were from CGMCC/AS (China General Microbiological Culture Collection Centre, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, PR China); the other two strains were from BCCM/LMG (Collection of the Laboratorium voor Microbiologie en Microbiele Genetica, Rijksuniversiteit, Ledeganckstraat 35, B-9000, Gent, Belgium) and JCM (Japan Collection of Microorganisms, Institute of Physical and Chemical Research (RIKEN), Wako, Saitama 351-0198, Japan), respectively. The majority were type strains. Strains were cultured in yeast extract/malt extract agar (ISP 2; Shirling & Gottlieb, 1966) at 28 °C.

Selection of gene loci and design of primers. In addition to 16S rRNA, five housekeeping genes, *atpD*, *gyrB*, *recA*, *rpoB* and *trpB*, were chosen. Locations of all these genes on the *Streptomyces coelicolor* A3(2) genome map are shown in Fig. S1 (available with the online version of this paper). Primers for *atpD*, *recA*, *rpoB* and *trpB* for amplification and sequencing were designed based on the corresponding sequences of *S. coelicolor* A3(2) (Bentley *et al.*, 2002) and *Streptomyces avermitilis* (Ikeda *et al.*, 2003), whose genomes are available in public databases. Software packages Primer premier 5.0 (Premier Biosoft International), Oligo 6.0 (Molecular Biology Insights) and SPCR 3.0 (http://moleco.sjtu.edu.cn/moleco/softwares.html) were used to design and evaluate the primers. For *gyrB*, amplification primers were derived from published primers PF-1 and PR-2, and sequencing primers from F-1 and R-4 (Hatano *et al.*, 2003). The information for loci and primers used in this scheme are described in Table 2.

DNA extraction, amplification and sequencing. Genomic DNA was extracted from cultures grown on ISP 2 using the method of Moller *et al.* (1992). Each 50 µl amplification reaction contained 1 µl template DNA (50–200 ng), 5 µl 10× PCR buffer, 1 µl each PCR primer (20 µM), 1 µl dNTP mix (10 mM), 6 µl $MgCl_2$ (25 mM), 2.5 U *Taq* DNA polymerase, 5 µl DMSO and 29 µl sterile MilliQ water. The reaction conditions were initial denaturation at 95 °C for 5 min, followed by 30 cycles of denaturation at 95 °C for 30 s, annealing for 30 s at the primer-pair-specific annealing temperature (Table 2) and extension at 72 °C for 90 s. A final extension was performed at 72 °C for 10 min. Reaction products were electrophoresed on a 1 % agarose gel and checked with ethidium bromide under UV light, and then purified and sequenced directly using a *Taq* DyeDeoxy Terminator Cycle Sequencing Kit and an ABI Prism 3730 automated DNA sequencer (Applied Biosystems). Both strands were sequenced as a cross-check by using forward and reverse sequencing primers.

Data analysis. Sequences of each locus were aligned using MEGA 3.1 software (Kumar *et al.*, 2004), and trimmed manually at the same position before being used for further analysis and to set up a database on the internet (http://pubmlst.org/streptomyces). Statistics for each locus, such as the number and proportion of polymorphic sites, mean G+C content, etc., were calculated using the START2 program (Jolley *et al.*, 2001). Pairwise distances between sequences of each locus were calculated by using the Kimura two-parameter (K2P) model (Kimura, 1980). The d*N*/d*S* ratios (the ratios of non-synonymous to synonymous polymorphisms) were calculated by using the method of Nei & Gojobori (1986) in START2.

To ensure the stability and reliability of phylogenetic relationships among strains used in this study, phylogenetic trees were constructed by using three different methods, neighbour-joining (NJ), maximum-parsimony (MP) and maximum-likelihood (ML). The NJ and MP methods were from the MEGA 3.1 package with the option of complete deletion of gaps, and the ML method was from PHYLIP package version 3.66 (Felsenstein, 2006). The K2P model was chosen as a substitution model for all NJ tree constructions. Single-locus and sequence-concatenated trees, including concatenation of sequences of three and four random protein-coding loci and all five protein-coding

Table 1. Strains used in this study, and GenBank accession numbers of alleles of loci

Species	Strain no.	16S rRNA	*atpD*	*gyrB*	*recA*	*rpoB*	*trpB*
S. acrimycini	AS 4.1673^T	AY999889	EF031300	EF054987	EF055042	EF055097	EF055152
S. albovinaceus	AS 4.1631^T	AY999759	EF031298	EF054985	EF055040	EF055095	EF055150
S. alboviridis	AS 4.1627^T	AB184256	EF031296	EF054983	EF055038	EF055093	EF055148
S. anulatus	AS 4.1421^T	AB184875	EF031285	EF054974	EF055027	EF055082	EF055137
'*S. argenteolus*'	AS 4.1693^T	EU048540	EF031302	EF054989	EF055044	EF055099	EF055154
S. atroolivaceus	AS 4.1405^T	AJ781320	EF031281	EF054970	EF055023	EF055078	EF055133
S. aureus	AS 4.1833^T	AY094368	EF031309	EF054996	EF055051	EF055106	EF055161
S. badius	AS 4.1406^T	AB184114	EF031282	EF054971	EF055024	EF055079	EF055134
S. bobili	AS 4.1624^T	AB184328	EF031293	EF054981	EF055035	EF055090	EF055145
S. californicus	AS 4.570^T	AB184755	EF031273	EF054962	EF055015	EF055070	EF055125
'*S. caviscabies*'	AS 4.1836	AF112160	EF031310	EF054997	EF055052	EF055107	EF055162
S. cirratus	AS 4.1679^T	AY999794	EF031301	EF054988	EF055043	EF055098	EF055153
S. cremeus	AS 4.1625^T	AY999744	EF031294	ND	EF055036	EF055091	EF055146
S. cyaneofuscatus	AS 4.1612^T	AB184860	EF031291	EF054980	EF055033	EF055088	EF055143
S. cyaneus	AS 4.1671^T	AB184872	EF031299	EF054986	EF055041	EF055096	EF055151
S. erumpens	AS 4.1626^T	AB184654	EF031295	EF054982	EF055037	EF055092	EF055147
S. exfoliatus	AS 4.1407^T	AB184324	EF031283	EF054972	EF055025	EF055080	EF055135
S. fimicarius	AS 4.1629^T	AB184269	EF031297	EF054984	EF055039	EF055094	EF055149
S. finlayi	AS 4.1436^T	AY999788	EF031287	EF054976	EF055029	EF055084	EF055139
S. flavidofuscus	AS 4.1617^T	AB184655	EF031292	ND	EF055034	EF055089	EF055144
S. flavogriseus	AS 4.1884^T	AB184271	EF031318	EF055005	EF055060	EF055115	EF055170
S. floridae	AS 4.1972^T	AB184656	EF031324	EF055011	EF055066	EF055121	EF055176
S. fulvorobeus	JCM 9090^T	AB184711	EF031326	EF055013	EF055068	EF055123	EF055178
S. galilaeus	AS 4.1320	EU048542	EF031277	EF054966	EF055019	EF055074	EF055129
S. graminofaciens	AS 4.1359^T	AB184416	EF031279	EF054968	EF055021	EF055076	EF055131
S. griseinus	AS 4.1875^T	AB184205	EF031317	EF055004	EF055059	EF055114	EF055169
S. griseobrunneus	AS 4.1838^T	AY999888	EF031311	EF054998	EF055053	EF055108	EF055163
S. griseolus	AS 4.1864^T	AB184788	EF031315	EF055002	EF055057	EF055112	EF055167
S. griseoplanus	AS 4.1868^T	AY999894	EF031316	EF055003	EF055058	EF055113	EF055168
S. griseus subsp. *griseus*	AS 4.1419^T	AB184699	EF031284	EF054973	EF055026	EF055081	EF055136
'*S. griseus* subsp. *alpha*'	AS 4.1843	AY999869	EF031312	EF054999	EF055054	EF055109	EF055164
'*S. griseus* subsp. *cretosus*'	AS 4.1844	AY999897	EF031313	EF055000	EF055055	EF055110	EF055165
S. griseus subsp. *solvifaciens*	AS 4.1845^T	AY999871	EF031314	EF055001	EF055056	EF055111	EF055166
S. kanamyceticus	AS 4.1441^T	AB184388	EF031288	EF054977	EF055030	EF055085	EF055140
S. laceyi	AS 4.1832^T	AY094367	EF031308	EF054995	EF055050	EF055105	EF055160
S. luridiscabiei	LMG 21390^T	AF361784	EF031272	EF054961	EF055014	EF055069	EF055124
S. mauvecolor	AS 4.1997^T	AB184532	EF031325	EF055012	EF055067	EF055122	EF055177
S. mediolani	AS 4.1896^T	AB184674	EF031319	EF055006	EF055061	EF055116	EF055171
S. microflavus	AS 4.1428^T	AB184284	EF031286	EF054975	EF055028	EF055083	EF055138
S. mutomycini	AS 4.1747^T	AJ781357	EF031303	EF054990	EF055045	EF055100	EF055155
S. nojiriensis	AS 4.1897^T	AJ781355	EF031320	EF055007	EF055062	EF055117	EF055172
'*S. ornatus*'	AS 4.1321	AB184290	EF031278	EF054967	EF055020	EF055075	EF055130
S. peucetius	AS 4.1799^T	AB045887	EF031307	EF054994	EF055049	EF055104	EF055159
S. praecox	AS 4.1782^T	AB184293	EF031305	EF054992	EF055047	EF055102	EF055157
S. pulveraceus	AS 4.1928^T	AB184806	EF031323	EF055010	EF055065	EF055120	EF055175
'*S. setonii*'	AS 4.1367	AB184300	EF031280	EF054969	EF055022	EF055077	EF055132
S. sindenensis	AS 4.626^T	AB184759	EF031274	EF054963	EF055016	EF055071	EF055126
S. spiroverticillatus	AS 4.1749^T	AY999736	EF031304	EF054991	EF055046	EF055101	EF055156
S. spororaveus	AS 4.1926^T	AJ781370	EF031322	EF055009	EF055064	EF055119	EF055174
S. subrutilus	AS 4.1784^T	AB184372	EF031306	EF054993	EF055048	EF055103	EF055158
S. tanashiensis	AS 4.1924^T	AB184245	EF031321	EF055008	EF055063	EF055118	EF055173
S. venezuelae	AS 4.1526^T	AB184308	EF031289	EF054978	EF055031	EF055086	EF055141
S. vinaceus	AS 4.1305	EU048541	EF031276	EF054965	EF055018	EF055073	EF055128
S. virginiae	AS 4.1530^T	AB184175	EF031290	EF054979	EF055032	EF055087	EF055142
S. yanii	AS 4.1146^T	AB015854	EF031275	EF054964	EF055017	EF055072	EF055127
Mycobacterium tuberculosis	H37Rv	GeneID: 2700429	GeneID: 886932	GeneID: 887081	GeneID: 888371	GeneID: 888164	GeneID: 885297

ND, Not determined.

Table 2. Primers for amplification and sequencing

Gene	Function	Primer sequence (5′–3′)	Position	Amplicon size (bp)	Annealing temperature (°C)
gyrB	Amplification	*gyrB*PF GAGGTCGTGCTGACCGTGCTGCACGCGGGCGGCAAGTTCGGC	355–396	1305	65
		*gyrB*PR GTTGATGTGCTGGCCGTCGACGTCGGCGTCCGCCAT	1624–1659		
	Sequencing	*gyrB*F-1 GAGGTCGTGCTGACCGTGCTGCA	355–378	–	–
		*gyrB*R-4 CGCTCCTTGTCCTCGGCCTC	866–885		
rpoB	Amplification	*rpoB*PF GAGCGCATGACCACCCAGGACGTCGAGGC	1162–1190	994	65
		*rpoB*PR CCTCGTAGTTGTGACCCTCCCACGGCATGA	2126–2155		
	Sequencing	*rpoB*F1 TTCATGGACCAGAACAACC	1273–1291	–	–
		*rpoB*R1 CGTAGTTGTGACCCTCCC	2135–2152		
trpB	Amplification	*trpB*PF GCGCGAGGACCTGAACCACACCGGCTCACACAAGATCAACA	267–308	822	66
		*trpB*PR TCGATGGCCGGGATGATGCCCTCGGTGCGCGACAGCAGGC	1049–1088		
	Sequencing	*trpB*F GGCTCACACAAGATCAACAA	289–308	–	–
		*trpB*R TCGATGGCCGGGATGATGCC	1069–1088		
recA	Amplification	*recA*PF CCGCRCTCGCACAGATTGAACGSCAATTC	35–63	913	60
		*recA*PR GCSAGGTCGGGGTTGTCCTTSAGGAAGTTGCG	916–947		
	Sequencing	*recA*F ACAGATTGAACGGCAATTCG	45–64	–	–
		*recA*R ACCTTGTTCTTGACCACCTT	733–752		
atpD	Amplification	*atpD*PF GTCGGCGACTTCACCAAGGGCAAGGTGTTCAACACC	283–318	998	63
		*atpD*PR GTGAACTGCTTGGCGACGTGGGTGTTCTGGGACAGGAA	1243–1280		
	Sequencing	*atpD*F ACCAAGGGCAAGGTGTTCAA	295–314	–	–
		*atpD*R GCCGGGTAGATGCCCTTCTC	1027–1046		

loci, were inferred including coding positions (1st + 2nd + 3rd + non-coding). Concatenated sequences of all five protein-coding loci were joined head-to-tail in-frame. The bootstrap technique (Felsenstein, 1985) was employed to evaluate the tree topologies by resampling the sequence alignment 1000 times. *Mycobacterium tuberculosis* H37Rv was used as the outgroup.

RESULTS

Sequence attributes

Amplification products from all 55 strains were obtained and each locus was subjected to sequencing. However, the PCR products of *gyrB* from two strains (*Streptomyces flavidofuscus* and *Streptomyces cremeus*) could not be sequenced because the reverse primer for sequencing did not specifically match the template DNA of these strains. In the end, 496 bp of *atpD*, 423 bp of *gyrB* (with gaps), 504 bp of *recA*, 540 bp of *rpoB* and 573 bp of *trpB* (with gaps) were obtained. 16S rRNA gene sequences of the strains retrieved from the GenBank database and those determined in this study were aligned and trimmed to 459 bp (with gaps), corresponding to positions 59–485 of the 16S rRNA gene (*rrnB*) sequence of *S. coelicolor* A3(2), containing the species-specific variable γ region (158–203) (Anderson & Wellington, 2001). GenBank accession

numbers of the sequences used in this study are listed in Table 1.

The features of each locus are displayed in Table 3. The proportion of variable sites in the alleles varied from 20.4 % (16S rRNA allele) to 48 % (*gyrB*), in contrast with 15.5 % variability in full-length 16S rRNA sequences. The higher percentage of variable sites led to wider ranges of distance and higher mean K2P distances, which made it possible to obtain finer relationships between close strains. Among the five protein-coding genes, *gyrB* had the largest mean K2P distance (0.140), significantly larger than that of the full-length 16S rRNA sequence; *atpD* and *recA* shared a similar distance range and mean K2P distance, as did *rpoB* and *trpB*. With the exception of 16S rRNA, the G+C contents of the loci were close to the genomic G+C content range of the genus *Streptomyces*, 67–78 mol%. The d*N*/d*S* ratio is used to estimate the degree of selection operating on each locus. All of the d*N*/d*S* values shown in Table 3 were much less than 1, indicating that most of the sequence variability identified is selectively neutral at protein level, validating their suitability for this multilocus sequence analysis scheme.

16S rRNA gene trees

The alignment and trimming of 16S rRNA gene sequences produced 1406 nt with gaps. Of these, 218 (15.5 %) sites were variable. The overall mean distance of 55 sequences was 0.020. Comparing the trees constructed using the three treeing algorithms, we found poor congruence of the topology between each tree (Fig. 1). Bootstrap support for the trees was low: nodes with values less than 60 % accounted for 21 out of 40 nodes (52.5 %) in the NJ tree, 41 out of 53 nodes (77.4 %) in the MP tree and 100 % in the ML tree; in addition it cannot be neglected that 10 nodes in the MP tree have a bootstrap value of 0 %.

Looking at the NJ tree more closely, it is divided into two parts by the *Streptomyces yanii* and *Streptomyces pulveraceus* branch. All 29 strains constituting the compact and cohesive upper part (without *S. yanii* and *S. pulveraceus*) share high 16S rRNA gene sequence identities of 99.1–100 %, and contain four branches designated by four coloured clusters (Fig. 1); within each of these clusters the strains group together without divergence, sharing completely identical 16S rRNA gene sequences. *S. griseus* subsp. *griseus* forms cluster IV (red) with *Streptomyces erumpens* and '*Streptomyces ornatus*', while '*S. griseus* subsp. *alpha*' and '*S. griseus* subsp. *cretosus*' belong to cluster III (green). It is interesting that *S. griseus* subsp. *solvifaciens* forms a distinct cluster with '*S. argenteolus*', *Streptomyces galilaeus* and *Streptomyces vinaceus* at the bottom of the tree, and shares an equally low sequence identity of 95.4 % with *S. griseus* subsp. *griseus*, *S. griseus* subsp. *alpha* and *S. griseus* subsp. *cretosus*. However, the levels of identity between *S. griseus* subsp. *griseus* and *S. griseus* subsp. *alpha* and *S. griseus* subsp. *cretosus* are 99.9 % in both cases.

Strains in the lower part of the tree were generally well separated, except *Streptomyces spororaveus* and *Streptomyces nojiriensis*, and *S. galilaeus*, *S. griseus* subsp. *solvifaciens* and *S. vinaceus*. 16S rRNA gene sequence identities between strains in this part ranged from 87.1 to 100 %, with a mean distance of 0.033. However, without *S. flavidofuscus*, the range of sequence identities narrowed to 95.0–100 %.

Single and concatenated protein-coding-gene trees

Phylogenetic trees for each protein-coding locus were constructed by using the three different methods (Supplementary Figs S2–S6, available with the online version of this paper). These trees, based on different loci, show general congruence. Clusters of strains indicated by different colours in the 16S rRNA gene tree (Fig. 1) are largely maintained in each single-gene tree, but with better discrimination. However, none of these genes alone contains enough phylogenetic information to reliably discriminate all species in the trees. The congruence among single-gene trees allows concatenation of multigene sequences for global analyses.

Sequences of three, four and five protein-coding genes were concatenated for phylogenetic analysis. It can be seen from

Table 3. Properties of loci

Locus	Allele length (bp)	No. of alleles	No. of polymorphic sites	Percentage of polymorphic sites	Mean G+C content (mol%)	Distance range*	Mean K2P distance†	d*N*/d*S*
atpD	496	46	147	29.6	64.5	0–0.205	0.081	0.1444
gyrB	423	45	203	48.0	67.2	0–0.244	0.140	0.2635
recA	504	44	168	33.3	68.9	0–0.204	0.082	0.0677
rpoB	540	46	197	36.5	67.4	0–0.258	0.090	0.2357
trpB	573	51	232	40.5	72.6	0–0.304	0.090	0.2647
16S rRNA (allele)	442	37	90	20.4	59.7	0–0.180	0.028	–
16S rRNA (full-length)	1406	40	218	15.5	58.5	0–0.129	0.017	–

*Pairwise distance calculated by using the K2P substitution model.
†Distance calculated by using the K2P substitution model.

Fig. 1. Phylogenetic relationships among 55 *Streptomyces* strains based on 16S rRNA gene sequences. The tree was constructed using the NJ method. *Mycobacterium tuberculosis* H37Rv was used as the outgroup. Numbers at nodes represent levels (%) of bootstrap support from 1000 resampled datasets. L and P indicate branches that were also recovered using ML and MP methods, respectively. The bar indicates 1 % estimated sequence divergence. Strains of clusters I, II, III and IV are highlighted with blue, yellow, green and red backgrounds, respectively.

the resultant trees (Supplementary Figs S7–S9, available with the online version of this paper) that, with the addition of each gene, the discriminatory power and the robustness of the trees increases slowly and regularly. Compared to the three- and four-gene trees, the five-gene tree holds bootstrap support that is slightly increased in the upper part, but greatly increased for the lower part, especially for the higher nodes, and gives enough resolution to separate closely related *Streptomyces* species.

Six-gene trees

The concatenated alignment of six loci contains 2900 nt with a mean G+C content of 67 mol% and a mean pairwise distance of 0.079. The sequence identities ranged from 88.3 to 100 %. Two-section division and a similar tendency of congruence and incongruence of each section were also observed for six-gene trees constructed by different methods (Fig. 2). The sequence identity range

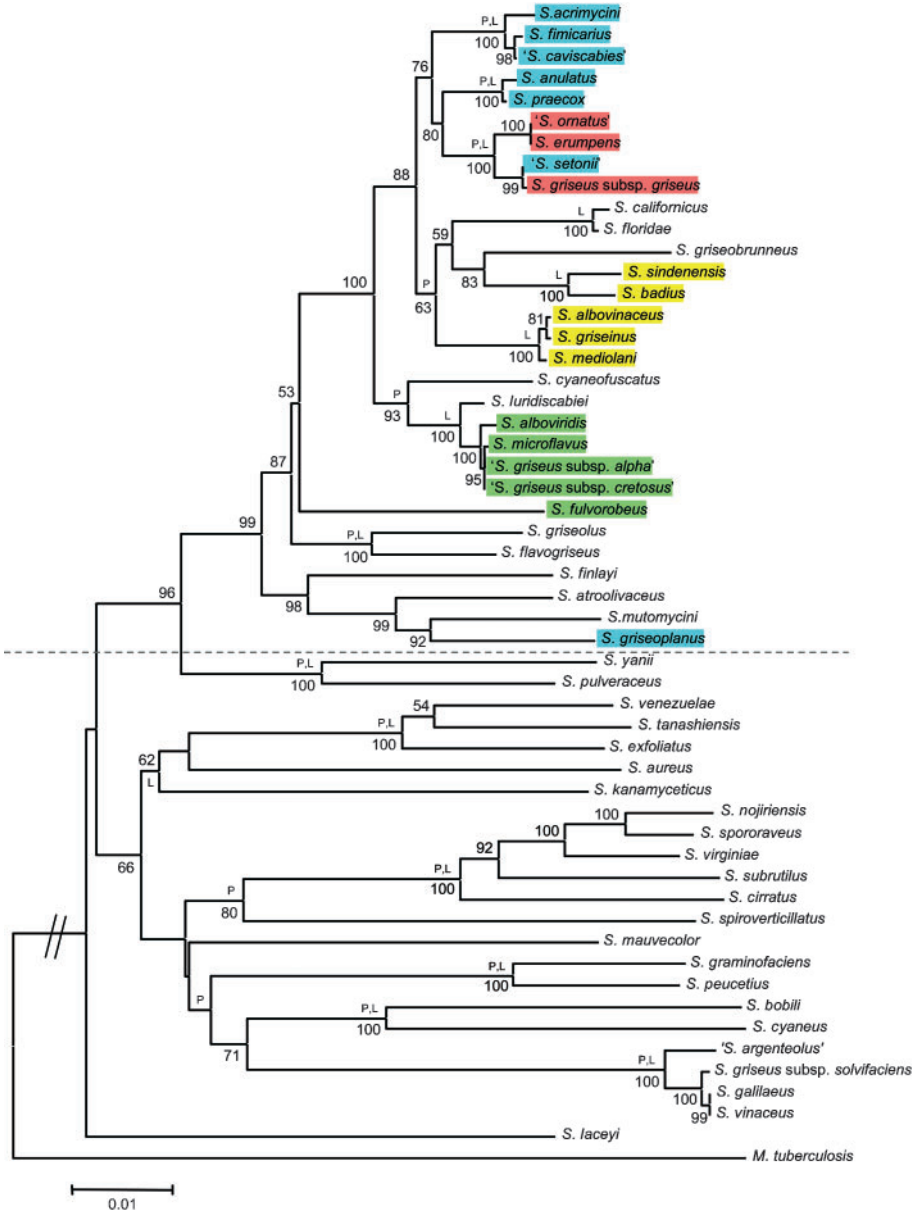

Fig. 2. Phylogenetic relationships among 53 *Streptomyces* strains based on six-gene concatenated sequences. The tree was constructed using the NJ method. *Mycobacterium tuberculosis* H37Rv was used as the outgroup. Numbers at nodes represent levels (%) of bootstrap support from 1000 resampled datasets. L and P indicate branches that were also recovered using ML and MP methods, respectively. The bar indicates 1 % estimated sequence divergence. Strains of clusters I, II, III and IV are highlighted with blue, yellow, green and red backgrounds, respectively.

was 91.2–100 % for the upper part and 88.3–99.7 % for the lower part.

Strains of cluster I (blue), sharing 98.9–99.9 % identities (excluding *Streptomyces griseoplanus*) in six-gene concatenated sequences, were reliably differentiated as distinct entities. *S. griseoplanus*, placed in cluster I in the 16S rRNA gene tree with zero distance, consistently falls into a distant branch in all multigene trees and was loosely related to *Streptomyces atroolivaceus* and *Streptomyces mutomycini* in the upper part neighbouring the boundary branch. Strains of cluster II (yellow), with identities of 97.2–99.9 %, were separated into two branches supported by high bootstrap values (100 %): *Streptomyces sindenensis* and *Streptomyces badius* were in a deep branch loosely related to *Streptomyces griseobrunneus*, while *Streptomyces albovinaceus*, *Streptomyces griseinus* and *Streptomyces mediolani* were in a higher branch as independent entities. All strains, with the exception of *Streptomyces fulvorobeus* in cluster III (green), were closely related within a branch, reflected by a

sequence identity range of 99.8–100 %. '*S. griseus* subsp. *alpha*', '*S. griseus* subsp. *cretosus*' and *S. microflavus* consistently formed a closely related branch with zero distance, representing an *S. microflavus* complex. '*S. setonii*' shifted from cluster I to cluster IV (red), which was separated into two stable branches with 99.3–100 % sequence identities and shared a higher node with members of cluster I.

In the lower part of the tree, bootstrap values were generally higher than those in five-gene trees, as demonstrated by 80 % of nodes holding values over 60 % in the NJ tree. The topological congruence between the three trees constructed by different methods is better than that between trees of five loci: high bootstrap values support the resemblance between deep branches of each tree, while the topology of some higher nodes remains unstable.

The six-gene tree proved to be the most robust and reliable phylogenetic tree that differentiated most strains in this study well, with the exceptions of *S. griseus* subsp. *griseus* and '*S. setonii*', '*S. ornatus*' and *S. erumpens*, *S. microflavus*, '*S. griseus* subsp. *alpha*' and '*S. griseus* subsp. *cretosus*', and *S. galilaeus* and *S. vinaceus*.

DISCUSSION

Gene selection and primers

Genes used in molecular systematics should be evaluated for their phylogenetic performance from historical (e.g. ease of amplification, previous analyses of related taxa) and analytical (e.g. copy number, rate of evolution) aspects (Hillis *et al.*, 1996). The five protein-coding genes we selected are commonly used in the identification of bacteria, and some of them have been used individually for phylogenetic analyses of *Streptomyces* (Egan *et al.*, 2001; Hatano *et al.*, 2003; Kim *et al.*, 2004). Each gene has only one copy on the *S. coelicolor* A3(2) genome, performs a different function in primary metabolism and is evolutionarily conserved. Primers designed for this multilocus sequence analysis scheme gave good results, but the sequencing primers of *gyrB* derived from the literature (Hatano *et al.*, 2003) performed inconsistently, although sequences were eventually obtained. New sequencing primers should be designed for future phylogenetic research using *gyrB*.

Combinability of loci

The suitability of the loci for this study was validated by their remote positions on the chromosome, low d*N*/d*S* ratios and high content of informative sites (Table 3). The minimum distance between two loci (*rrnE* and *recA*) among the six distant genes is over 30 kb (Fig. S1), which means there is no genetic linkage between these loci theoretically. The low *dN/dS* ratios (far below 1) indicated that the five housekeeping genes are not subjected to selection pressure for amino acid changes – they are selectively neutral. These properties guarantee the independence of each locus as they underwent evolution. Moreover, the variability of the loci, as demonstrated by the high proportion of polymorphic sites, and larger distance range and larger mean distance for each locus than for full-length 16S rRNA gene sequences, implied that they would give a higher resolution for the phylogeny of the *S. griseus* clade than the 16S rRNA gene. The widespread congruence between phylogenies recovered from individual genes indicated that the phylogenetic signals would not be counteracted when the loci are concatenated, thus making the combined analysis of different loci possible in this study.

Phylogenetic resolution and reliability of single and combined datasets

The 16S rRNA gene tree is unreliable based on the conflicting topologies and the low bootstrap support for the trees generated by different algorithms (Fig. 1), pointing to a significant probability of support for incorrect relationships for the strains included. The relationships of strains in the upper section are poorly resolved, especially strains in the coloured clusters. In contrast, strains in the lower section which share low sequence identities are largely well distinguished. So we can assume that the 16S rRNA gene is more appropriate for discrimination of (moderately) distantly related streptomycetes, but is not efficient for closely related strains such as those in clusters I–IV.

We found all five protein-coding loci to be informative, but not powerful or reproducible for all reference strains; our effort to construct well-resolved, deep phylogenies of *Streptomyces* was plagued by rate heterogeneity of different loci and weak phylogenetic signals in single genes. Therefore, we used multiple genes and, as expected, the effect of these problems was reduced considerably. Considering the potential systematic biases in different datasets, concatenation of the six genes would be optimal in resolving the relationships among all the strains. The strains that could not be discriminated in the six-gene tree belong to same species (Fig. 2). Lanoot *et al.* (2005b) transferred '*S. griseus* subsp. *alpha*' and '*S. griseus* subsp. *cretosus*' into *S. microflavus* because of their 100 % DNA–DNA relatedness. '*S. setonii*' has been transferred into *S. griseus* based on DNA–DNA relatedness (Liu *et al.*, 2005) as well. In our study, all trees except the 16S rRNA gene tree reached a consensus that '*S. setonii*' falls consistently into the closely related branch with *S. griseus* subsp. *griseus*, which is in agreement with the result of DNA–DNA relatedness. '*S. ornatus*' and *S. erumpens* have identical six-gene concatenated sequences, with which *S. griseus* subsp. *griseus* shares 99.3 % sequence identity. DNA–DNA relatedness data from previous studies (Okanishi, 1972; Liu *et al.*, 2005) support '*S. ornatus*' belonging to the same genomic species as the type strain of *S. griseus*. Based on

results of present and previous studies, we propose that '*S. ornatus*', *S. erumpens*, '*S. setonii*', and *S. griseus* subsp. *griseus* should be considered as members of the same genomic species, *S. griseus*, which however, should not encompass '*S. caviscabies*' based on their well-separated locations in the multigene trees. The two remaining strains, *S. galilaeus* and *S. vinaceus*, which share identical sequences of all six genes, do not merit separate species status, because their 16S rRNA gene sequences (EU048542 and EU048541) determined in this study are evidently different from those (AB045878 and AB184394) of the respective type strains (with distances of 0.042–0.044), and they should be assigned to a single species with a different name.

In light of the clear differentiation of all strains at species level by the six-gene tree and the accordance between data of this study and DNA–DNA relatedness studies, our multilocus sequence analysis scheme based on a combined analysis of the six loci gives credible and appropriate phylogenetic resolution at species level for the *S. griseus* 16S rRNA gene clade.

Comparison between 16S rRNA gene and concatenated trees

The upper parts of NJ trees include almost the same closely related strains except *S. griseobrunneus*, which failed to appear in the same position as in the 16S rRNA gene tree and moved to the upper part in the concatenated trees, with *S. badius* and *S. sindenensis* of cluster II as its consistent neighbours. Another strain with noticeably different positions within the upper part of the trees is *S. griseoplanus* (cluster I), as mentioned above. To check the incongruence observed for these two strains between the 16S rRNA gene and concatenated trees, we sequenced their 16S rRNA gene again, but obtained identical sequences as those retrieved from GenBank, which we had used. This result reveals the misleading probability of using the 16S rRNA gene solely for phylogenetic relationships.

Three species with abnormal taxonomic positions were observed in both the 16S rRNA gene and concatenated trees. One is '*S. argenteolus*', which always clusters with *S. griseus* subsp. *solvifaciens*, *S. galilaeus* and *S. vinaceus* in the lower part of all trees, supported by high bootstrap values. Since it is separated from *S. griseus* subsp. *griseus* in the 16S rRNA gene tree with a phylogenetic distance of 0.046, and in the six-gene tree with a significant distance of 0.103, we suggest that '*S. argenteolus*' AS 4.1693^T merits species status distinct from *S. griseus*, rather than being a member of the latter species as proposed by Liu *et al.* (2005). The 16S rRNA gene sequence (EU048540) obtained for this strain in our study is markedly divergent (with a distance of 0.051) from those (AB045872 and AB184187) obtained previously for '*S. argenteolus*' JCM 4623^T (=NBRC 12841^T). The second is *S. griseus* subsp. *solvifaciens* that was located unusually far from *S. griseus* subsp. *griseus* in all trees. Combined with its genotypical heterogeneity with other members of *S. griseus* as reported by Lanoot *et al.* (2005b), we propose that *S. griseus* subsp. *solvifaciens* should be removed from *S. griseus*. Additional analyses with more *Streptomyces* species are needed to clarify its taxonomic status. The third is *S. flavidofuscus* that was consistently located at the periphery of the other streptomycetes selected in this study. Further examination demonstrated that this type strain had sequences very different from that of other strains at all loci, but shared high 16S rRNA gene sequence similarities with strains of *Nocardiopsis* [e.g. 99.5 % identity with *Nocardiopsis synnemataformans* DSM 44143^T (Y13593)] as determined by a BLAST search against the GenBank database. Therefore, we propose that *S. flavidofuscus* should be transferred to the genus *Nocardiopsis*. Therefore, *S. griseobrunneus*, *S. griseoplanus*, '*S. argenteolus*', *S. griseus* subsp. *solvifaciens* and *S. flavidofuscus*, whose taxonomic positions have significantly changed, may be worth looking at again for potentially useful metabolites.

Several points can be obtained from comparison of the concatenated six-gene and 16S rRNA gene trees. First, the phylogenetic relationships among most strains in this study are generally congruent between the two trees. Second, the six-gene tree shows a much higher power of discrimination in that most species are clearly discriminated from each other, especially the closely related species in clusters I–IV, and the overall mean distance of the six-gene tree (0.079) is significantly greater than that of the 16S rRNA gene tree (0.020). Third, the topological structure of the six-gene tree, which is supported by noticeably higher bootstrap values, is much more stable than that of the 16S rRNA gene tree. These points emphasize the fact that the six-gene tree is obviously superior to the 16S rRNA gene tree in both resolution power and topological stability. Moreover, the concatenation of a sufficient number of genes overwhelms possible conflicting phylogenetic signals in different genes.

In conclusion, we have developed a pilot multilocus sequence analysis scheme for *Streptomyces*, and have shown its promising potential for refining the phylogeny of this genus. Our scheme, based on combined analysis of sequence variations at six different loci, can discriminate and define phylogenetic relationships among diverse and closely related species of the *S. griseus* 16S rRNA gene clade. We believe that this multilocus sequence analysis scheme can be applied further to more taxonomically diverse actinomycete taxa where identification and discrimination are valuable tools in the discovery of novel and commercially important metabolites.

ACKNOWLEDGEMENTS

We are grateful to Dr Keith Jolley (University of Oxford) for his excellent technical assistance in the construction of the multilocus sequence internet database and to Professor John Mekalanos (Harvard Medical School) for his help in editing the manuscript. We also thank CGMCC/AS, BCCM/LMG and JCM for providing reference strains. This work was funded by the Natural Science Foundation of China (NSFC, No. 30670002) and by the Knowledge Innovation Project of Chinese Academy of Sciences (No. KSCX2-YW-Z-042).

REFERENCES

Adiri, R. S., Gophna, U. & Ron, E. Z. (2003). Multilocus sequence typing (MLST) of *Escherichia coli* O78 strains. *FEMS Microbiol Lett* **222**, 199–203.

Anderson, A. S. & Wellington, E. M. (2001). The taxonomy of *Streptomyces* and related genera. *Int J Syst Evol Microbiol* **51**, 797–814.

Baldwin, A., Mahenthiralingam, E., Thickett, K. M., Honeybourne, D., Maiden, M. C., Govan, J. R., Speert, D. P., Lipuma, J. J., Vandamme, P. & Dowson, C. G. (2005). Multilocus sequence typing scheme that provides both species and strain differentiation for the *Burkholderia cepacia* complex. *J Clin Microbiol* **43**, 4665–4673.

Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H. & other authors (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141–147.

Berdy, J. (2005). Bioactive microbial metabolites. *J Antibiot* **58**, 1–26.

Clardy, J., Fischbach, M. A. & Walsh, C. T. (2006). New antibiotics from bacterial natural products. *Nat Biotechnol* **24**, 1541–1550.

Coffey, T. J., Pullinger, G. D., Urwin, R., Jolley, K. A., Wilson, S. M., Maiden, M. C. & Leigh, J. A. (2006). First insights into the evolution of *Streptococcus uberis*: a multilocus sequence typing scheme that enables investigation of its population biology. *Appl Environ Microbiol* **72**, 1420–1428.

Devulder, G., Pérouse de Montclos, M. & Flandrois, J. P. (2005). A multigene approach to phylogenetic analysis using the genus *Mycobacterium* as a model. *Int J Syst Evol Microbiol* **55**, 293–302.

Dingle, K. E., Colles, F. M., Wareing, D. R., Ure, R., Fox, A. J., Bolton, F. E., Bootsma, H. J., Willems, R. J., Urwin, R. & Maiden, M. C. (2001). Multilocus sequence typing system for *Campylobacter jejuni*. *J Clin Microbiol* **39**, 14–23.

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.

Enright, M. C. & Spratt, B. G. (1999). Multilocus sequence typing. *Trends Microbiol* **7**, 482–487.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

Felsenstein, J. (2006). PHYLIP (phylogeny inference package). Version 3.66. Department of Genome Sciences, University of Washington, Seattle, USA.

Grund, E., Knorr, C. & Eichenlaub, R. (1990). Catabolism of benzoate and monohydroxylated benzoates by *Amycolatopsis* and *Streptomyces* spp. *Appl Environ Microbiol* **56**, 1459–1464.

Hatano, K., Nishii, T. & Kasai, H. (2003). Taxonomic re-evaluation of whorl-forming *Streptomyces* (formerly *Streptoverticillium*) species by using phenotypes, DNA–DNA hybridization and sequences of *gyrB*, and proposal of *Streptomyces luteireticuli* (ex Katoh and Arai 1957) corrig., sp. nov., nom. rev. *Int J Syst Evol Microbiol* **53**, 1519–1529.

Hilario, E., Buckley, T. R. & Young, J. M. (2004). Improved resolution on the phylogenetic relationships among *Pseudomonas* by the combined analysis of *atpD*, *carA*, *recA* and 16S rDNA. *Antonie Van Leeuwenhoek* **86**, 51–64.

Hillis, D. M., Moritz, C. & Mable, B. K. (1996). *Molecular Systematics*. Sunderland, MA: Sinauer Associates.

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.

Ishiyama, D., Vujaklija, D. & Davies, J. (2004). Novel pathway of salicylate degradation by *Streptomyces* sp. strain WA46. *Appl Environ Microbiol* **70**, 1297–1306.

Iwamoto, T., Tsujii, E., Ezaki, M., Fujie, A., Hashimoto, S., Okuhara, M., Kohsaka, M., Imanaka, H., Kawabata, K. & other authors (1990). FR109615, a new antifungal antibiotic from *Streptomyces setonii*. Taxonomy, fermentation, isolation, physico-chemical properties and biological activity. *J Antibiot* **43**, 1–7.

Jolley, K. A., Feil, E. J., Chan, M. S. & Maiden, M. C. (2001). Sequence type analysis and recombinational tests (START). *Bioinformatics* **17**, 1230–1231.

Kim, B.-J., Kim, C.-J., Chun, J., Koh, Y.-H., Lee, S.-H., Hyun, J.-W., Cha, C.-Y. & Kook, Y.-H. (2004). Phylogenetic analysis of the genera *Streptomyces* and *Kitasatospora* based on partial RNA polymerase β-subunit gene (*rpoB*) sequences. *Int J Syst Evol Microbiol* **54**, 593–598.

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.

Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* **5**, 150–163.

Lanoot, B., Vancanneyt, M., Dawyndt, P., Cnockaert, M., Zhang, J., Huang, Y., Liu, Z. & Swings, J. (2004). BOX-PCR fingerprinting as a powerful tool to reveal synonymous names in the genus *Streptomyces*. Emended descriptions are proposed for the species *Streptomyces cinereorectus*, *S. fradiae*, *S. tricolor*, *S. colombiensis*, *S. filamentosus*, *S. vinaceus* and *S. phaeopurpureus*. *Syst Appl Microbiol* **27**, 84–92.

Lanoot, B., Vancanneyt, M., Hoste, B., Vandemeulebroecke, K., Cnockaert, M. C., Dawyndt, P., Liu, Z., Huang, Y. & Swings, J. (2005a). Grouping of streptomycetes using 16S-ITS RFLP fingerprinting. *Res Microbiol* **156**, 755–762.

Lanoot, B., Vancanneyt, M., Van Schoor, A., Liu, Z. & Swings, J. (2005b). Reclassification of *Streptomyces nigrifaciens* as a later synonym of *Streptomyces flavovirens*; *Streptomyces citreofluorescens*, *Streptomyces chrysomallus* subsp. *chrysomallus* and *Streptomyces fluorescens* as later synonyms of *Streptomyces anulatus*; *Streptomyces chibaensis* as a later synonym of *Streptomyces corchorusii*; *Streptomyces flaviscleroticus* as a later synonym of *Streptomyces minutiscleroticus*; and *Streptomyces lipmanii*, *Streptomyces griseus* subsp. *alpha*, *Streptomyces griseus* subsp. *cretosus* and *Streptomyces willmorei* as later synonyms of *Streptomyces microflavus*. *Int J Syst Evol Microbiol* **55**, 729–731.

Liu, Z., Shi, Y., Zhang, Y., Zhou, Z., Lu, Z., Li, W., Huang, Y., Rodriguez, C. & Goodfellow, M. (2005). Classification of *Streptomyces griseus* (Krainsky 1914) Waksman and Henrici 1948 and related species and the transfer of '*Microstreptospora cinerea*' to the genus *Streptomyces* as *Streptomyces yanii* sp. nov. *Int J Syst Evol Microbiol* **55**, 1605–1610.

Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K. & other authors (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* **95**, 3140–3145.

Moller, E. M., Bahnweg, G., Sandermann, H. & Geiger, H. H. (1992). A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies, and infected plant tissues. *Nucleic Acids Res* **20**, 6115–6116.

Nei, M. & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**, 418–426.

Nørskov-Lauritsen, N., Bruun, B. & Kilian, M. (2005). Multilocus sequence phylogenetic study of the genus *Haemophilus* with description of *Haemophilus pittmaniae* sp. nov. *Int J Syst Evol Microbiol* **55**, 449–456.

Okanishi, M. (1972). An evaluation of taxonomic criteria in streptomycetes on the basis of deoxyribonucleic acid homology. *J Gen Microbiol* **72**, 49–58.

Phelan, M. B., Crawford, D. L. & Pometto, A. L., III (1979). Isolation of lignocellulose-decomposing actinomycetes and degradation of specifically 14C-labeled lignocelluloses by six selected *Streptomyces* strains. *Can J Microbiol* **25**, 1270–1276.

Piel, J. (2004). Metabolites from symbiotic bacteria. *Nat Prod Rep* **21**, 519–538.

Shirling, E. B. & Gottlieb, D. (1966). Methods for characterization of *Streptomyces* species. *Int J Syst Bacteriol* **16**, 313–340.

Song, J., Lee, S. C., Kang, J. W., Baek, H. J. & Suh, J. W. (2004). Phylogenetic analysis of *Streptomyces* spp. isolated from potato scab lesions in Korea on the basis of 16S rRNA gene and 16S–23S rDNA internally transcribed spacer sequences. *Int J Syst Evol Microbiol* **54**, 203–209.

Strohl, W. R. (2003). Antimicrobials. In *Microbial Diversity and Bioprospecting*, pp. 336–355. Edited by A. T. Bull. ASM Press.

Urwin, R. & Maiden, M. C. (2003). Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol* **11**, 479–487.

Watve, M. G., Tickoo, R., Jog, M. M. & Bhole, B. D. (2001). How many antibiotics are produced by the genus *Streptomyces? Arch Microbiol* **176**, 386–390.