

The genome of *Streptomyces rimosus* subsp. *rimosus* shows a novel structure compared to other *Streptomyces* using DNA/DNA microarray analysis

Ralph Kirby · Teng-Kuei Gan · Iain Hunter ·
Paul Herron · Emma Tilley

Received: 30 October 2007 / Accepted: 24 January 2008 / Published online: 27 February 2008
© Springer Science+Business Media B.V. 2008

Abstract DNA/DNA genome microarray analysis together with genome sequencing suggests that the genome of members of the genus *Streptomyces* would seem to have a common structure including a linear genomic structure, a core of common syntenous Actinomycete genes, the presence of species specific terminal regions and two intermediate group of syntenous genes that seem to be genus specific. We analyzed *Streptomyces* species using DNA/DNA microarray comparative genome analysis. Only *Streptomyces rimosus* failed to give a congruent genome pattern for the genes found in *Streptomyces coelicolor*. We expanded the analysis to include a number of strains related to the type strain of *S. rimosus* and obtained a similar divergence from the main body of *Streptomyces* species. These strains showed very close identity to the original strain with no gene deletion or duplication detected. The 16S

rRNA sequences of these *S. rimosus* strains were confirmed as very similar to the *S. rimosus* sequences available from the Ribosomal Database Project. When the SSU ribosomal RNA phylogeny of *S. rimosus* is analyzed, the species is positioned at the edge of the *Streptomyces* clade. We conclude that *S. rimosus* represents a distinct evolutionary lineage making the species a worthy possibility for genome sequencing.

Keywords *Streptomyces rimosus* · Actinomycetales · Genome comparison · Microarray · Novel lineage · Point mutations · Antibiotic production

Introduction

Streptomyces are a group of aerobic high %G+C Gram positive bacteria that undergo complex differentiation to form filamentous mycelium, aerial hyphae and spores. In addition, they produce a broad range of secondary metabolites including antibiotics, antiparasitic agents, herbicides, anti-cancer drugs and various enzymes of industrial importance. At present, two *Streptomyces* species have had their complete genome sequences published, namely the model organism *Streptomyces coelicolor* (%G+C=72.1) and avermectin producer *Streptomyces avermitilis* (%G+C=70.7) (Bentley et al. 2002; Ikeda et al.

Electronic supplementary material The online version of this article (doi:10.1007/s10482-008-9223-5) contains supplementary material, which is available to authorized users.

R. Kirby (✉) · T.-K. Gan
Department of Life Sciences and Institute of Genome
Sciences, National Yang-Ming University, Beitou, Taipei,
Taiwan
e-mail: rkirby@ym.edu.tw

I. Hunter · P. Herron · E. Tilley
Institute of Pharmacy & Biomedical Sciences,
Strathclyde, Glasgow, Scotland

2003), although a number of other sequences are close to publication. Two important aspects of the genomes structures of *Streptomyces* have been supported by sequence data. Firstly, that the genome size of *Streptomyces* is large compared to other bacteria; 8,667,507 basepairs for *S. coelicolor* (7825 protein coding genes) and 9,025,608 basepairs (7,577 protein coding genes) for *S. avermitilis*. Secondly, that the genomes of these two species are linear and both ends contain unique terminal inverted repeats that probably covalently bind a terminal protein. (Lin et al. 1993; Chen et al. 2002; Huang et al. 1998). Over 2500 *Streptomyces* strains are present in the Ribosomal Database Project (www.rdp.cme.msu.edu), over 1500 are available at the American Type Culture Collection (<http://www.atcc.org/>) and many more are held in both public and private culture collections throughout the world. Analysis of the small subunit ribosomal RNA gene sequences of *Streptomyces* confirms that they form a monophyletic clade, but one with considerable diversity. In addition, there is significant gene diversity at the interspecies level across the genomes of both completely sequenced *Streptomyces* with 2291 gene unique to *S. avermitilis* and 2307 genes unique to *S. coelicolor*. The comparative genomics of three species of the *Streptomyces* and one species of *Kitasatospora* have been analyzed by DNA/DNA microarray analysis and this showed that firstly the microarray data agreed well with the genome sequence data of *S. avermitilis* and secondly that the four species analyzed produced genome patterns that was congruent with each other (Hsiao and Kirby 2007). The available bioinformatics and microarray data suggest that *Streptomyces* and closely related genera have a common core region of the genome that is also present in many other Actinobacteria including *Mycobacterium* and *Corynebacterium* (Bentley et al. 2002). Outside of this core region there are two regions containing genus (or perhaps wider) specific genes. Finally, at the terminal regions of the linear chromosomes are species-specific regions that contain only a limited number of conserved genes, some of which are involved in terminal replication, including *tpgA* and *tapA*.

In this study we expanded the analysis to include a wider range of *Streptomyces* including *Streptomyces rimosus* (Petkovic et al. 2006) as well as two non-*Streptomyces* myceliate Actinomycetales,

Streptosporangium roseum and *Saccharomonospora viridis*. All tested species except for *Streptomyces rimosus* showed congruency of the microarray pattern, suggesting that *S. rimosus* may have a novel genome lineage compared to other *Streptomyces*.

Materials and methods

16S phylogeny

This was carried out on the small subunit 16S ribosomal RNA gene sequences (bp 91–447, *S. coelicolor* A3(2) 16S DNA sequence AL939108) obtained from Ribosomal Database Project-II Release 9 (<http://www.rdp.cme.msu.edu/index.jsp>) and PCR cloning/sequencing of various *S. rimosus* strains. These were aligned using CLUSTALX (Thompson et al. 1997). The analysis was carried out using Neighbor-Joining algorithm (Gascuel and Steel 2006) from the same program.

Arrays

PCR arrays covering about 97% of the complete genome of *Streptomyces coelicolor* A3(2) (www.surrey.ac.uk/SBMS/Fgenomics/Microarrays/index.html) were used in this study. The Surrey microarray is made up of 7758 unique PCR amplified sequences, 7563 from the chromosome and 195 from SCP1, the large linear plasmid found in *S. coelicolor* that encodes methylenomycin (Kirby et al. 1975; Kirby and Hopwood 1977; Bentley et al. 2004). There are an additional 376 non-unique, alternative and cross-hybridizing sequences that are also spotted on to the array together with no probe spots and control spots. The Surrey array does not include a number of transposition element related genes. The sequences of the PCR products are not available due to intellectual property protection requirements.

Strains and growth conditions

The following species were used for the interspecific comparative genomics part of this study: *Streptomyces coelicolor* A3(2) (SCP1⁺) 104, *Streptomyces antibioticus* (ATCC15848), *Streptomyces argenteolus* (ATCC11009), *Streptomyces avermitilis* (ATCC31267), *Streptomyces bikiniensis* (ATCC11062), *Streptomyces*

cattleya (ATCC35852), *Streptomyces clavuligerus* (ATCC27064), *Streptomyces lipmanii* (ATCC27357), *Streptomyces rimosus* subsp. *rimosus* G7 (type strain isolated by Pfizer Ltd and purportedly ATCC10970), *Streptomyces scabies* (BCRC13682), *Streptomyces tanashiensis* (ATCC23967), *Streptomyces virginiae* (ATCC12630), *Streptomyces aureofaciens* (ATCC 10762), *Streptosporangium roseum* (ATCC12428) and *Saccharomonospora viridis* (ATCC15345). For the intraspecific comparative genomics aspects of this study, the following *S. rimosus* strains were used, G37, 23383, M4018 and S15883, in addition to *S. rimosus* G7. These strains are all linear descendants of G7 over about 30 years and have undergone mutagenesis and selection affecting various aspects of oxytetracycline production and other phenotypic characteristics. Table 1 includes descriptions of all the Actinobacteria used in this study. Fresh spores were collected and mycelium cultured in TSB liquid medium with 0.5% glycine at 30°C for 3 days.

Preparation of labeled DNA

Genomic DNA from a stationary phase culture was purified by the salting out procedure (Pospiech and Neumann 1995) and had been sonicated to <2Kb. Four to six micrograms of sonicated genomic DNA were used as template and this was denatured in the presence of 12 µg of 72%-GC-content random hexamers in a total volume of 25 µl at 100°C for 10 min. The mixture was then snap-cooled on ice before adding the remaining reaction components: About 1.5 µl of Cy3-dCTP or Cy5-dCTP (Amersham Pharmacia Biotech), 4 µl Klenow fragment (NEB #212), 5 µl Klenow buffer, 0.5 µl dNTP (4mM dATP, 4 mM dTTP, 10 mM dGTP, and 0.2 mM dCTP), and 14 µl ddH₂O. The random primed labeling reaction was carried out for 2–3 h at 37°C. Buffer exchange, purification and concentration of the DNA products was accomplished by three cycles of diluting the reaction mixture in 0.5 ml TE buffer (10 mM Tris and 1 mM EDTA pH 8.0) and filtering through a Microcon-30 microconcentrators (Millipore).

Microarray hybridization and data analysis

In all cases array hybridizations were carried out in duplicate. The two DNA pools to be compared were

mixed and applied to an array in a hybridization mixture that contained 3.68X SSC, 0.18% SDS, and 1 µg yeast tRNA (total 16.3 µl), which had been heated at 100°C for 5 min before being applied to array. Hybridization took place under a glass coverslip sealed by glue in a humidified Omnislid (Thermo Hybaid) at 60°C for 12–14 h. The slides were washed, dried and scanned for fluorescence using a GenePix TM 4000B scanner (Axon instruments). Average signal intensity and local background measurements were obtained for each spot on each array using GenePixPro software. The dataset was screened for aberrant spots and these were eliminated from the analysis after manual checking. The signal from each gene spots was inputted into the computer program ScanAlyze (Eisen et al. 1998; Gollub et al. 2003). The data was then processed into a mean Log₂ Cy3/Cy5 ratio format. The dataset was normalized for each array separately and outputted to Excel where after checking the alignment of the datasets from each array, a mean signal for each common gene was calculated. Based on Bentley et al. 2002, the mean signal and standard deviation for the core region of genes from SCO2050 to SCO5800 was calculated. The standard deviation was used to set a cut-off for gene absence at 2SD below the core mean signal. The microarray data is presented relative to the *S. coelicolor* standard in two ways. This is either as a color plot of the genes where green presents a negative hybridization signal, black represents an equal hybridization signal and red indicates a positive hybridization signal using the program Treeview (Eisen et al. 1998), or as numeric values for the signal from each gene. The microarray data for the *S. rimosus* species described here can be accessed at NCBI Gene Expression Omnibus (GSE10126).

Genome comparison of the actinobacteria

The program Mauve (Darling et al. 2004; <http://www.gel.ahabs.wisc.edu/mauve>) is a system for efficiently constructing multiple genome alignments in the presence of large-scale evolutionary events such as rearrangement and inversion. Multiple genome alignment provides a basis for research into comparative genomics and the study of evolutionary dynamics. Mauve was used at its default settings to

Table 1 Species and strains of Actinobacteria used for bioinformatical and experimental purposes in this study

Species/strain	Bioinformatics	Experimental	Sources	Identification/description
<i>Actinomadura citrea</i>	Yes	No	RDP II R9	S000134253
<i>Actinomadura cremea rifamycini</i>	Yes	No	RDP II R9	S000437727
<i>Actinomadura echinospora</i>	Yes	No	RDP II R9	S000134120
<i>Amycolatopsis azurea</i>	Yes	No	RDP II R9	S000130089
<i>Amycolatopsis lactamdurans</i>	Yes	No	RDP II R9	S000012814
<i>Amycolatopsis orientalis</i>	Yes	No	RDP II R9	S000007274
<i>Bifidobacterium longum</i>	Yes	No	Genbank	AE014295
<i>Corynebacterium efficiens</i>	Yes	No	Genbank	BA000035
<i>Frankia alni</i>	Yes	No	Genbank	CT573213.2
<i>Kitasatospora cineracea</i>	Yes	No	RDP II R9	S000469278
<i>Kitasatospora phosalacinea</i>	Yes	No	RDP II R9	S000436292
<i>Kitasatospora viridis</i>	Yes	No	RDP II R9	S000504585
<i>Kitasatospora_azaticus</i>	Yes	No	Genbank	U93312
<i>Leifsonia xyli</i>	Yes	No	Genbank	AE016822
<i>Microtetraspora pusilla</i>	Yes	No	RDP II R9	S000435869
<i>Microtetraspora roseola</i>	Yes	No	RDP II R9	S000437705
<i>Microtetraspora salmonae</i>	Yes	No	RDP II R9	S000437706
<i>Mycobacterium bovis</i>	Yes	No	Genbank	BX248333
<i>Mycobacterium vanbaalenii</i>	Yes	No	Genbank	CP000511
<i>Nocardia farcinicia</i>	Yes	No	Genbank	AP006618
<i>Nocardiodes sp.</i>	Yes	No	Genbank	CP000509
<i>Nonomuraea africana</i>	Yes	No	RDP II R9	S000013571
<i>Nonomuraea longicatena</i>	Yes	No	RDP II R9	S000018932
<i>Nonomuraea roseola</i>	Yes	No	RDP II R9	S000007068
<i>Propionibacterium acnes</i>	Yes	No	Genbank	AE017283
<i>Rhodococcus sp. RHA1</i>	Yes	No	Genbank	CP000431
<i>Saccharomonospora azurea</i>	Yes	No	RDP II R9	S000003567
<i>Saccharomonospora thermoviridis</i>	Yes	No	RDP II R9	S000387332
<i>Saccharomonospora viridis</i>	Yes	Yes	RDP II R9/ATCC	S000002046
<i>Streptomyces acidiscabies</i>	Yes	No	RDP II R9	S000382780
<i>Streptomyces antibioticus</i>	Yes	Yes	RDP II R9/NRRL	S000581576/B-1701
<i>Streptomyces argenteolus</i>	Yes	Yes	RDP II R9/ATCC	S000011833/11009
<i>Streptomyces aureofaciens</i>	Yes	Yes	RDP II R9/ATCC	S000001105/10762
<i>Streptomyces bikiniensis</i>	Yes	Yes	RDP II R9/ATCC	S000381283/11062
<i>Streptomyces cattleya</i>	Yes	Yes	RDP II R9/NRRL	S000003379/8057
<i>Streptomyces clavuligerus</i>	Yes	Yes	RDP II R9/ATCC	S000004208/27064
<i>Streptomyces coelicolor A3(2)</i>	Yes	Yes	Genbank/Hopwood	AL645882/104
<i>Streptomyces griseocarneus</i>	Yes	No	RDP II R9	000581725
<i>Streptomyces griseus</i>	Yes	No	RDP II R9	S000326124
<i>Streptomyces hirosheimensis</i>	Yes	No	RDP II R9	S000544760
<i>Streptomyces lavendulae</i>	Yes	No	RDP II R9	S000001020
<i>Streptomyces lipmanii</i>	Yes	Yes	RDP II R9/ATCC	S000393821/27357
<i>Streptomyces longisporus</i>	Yes	No	RDP II R9	S000009302
<i>Streptomyces nanningensis</i>	Yes	No	RDP II R9	S000117241

Table 1 continued

Species/strain	Bioinformatics	Experimental	Sources	Identification/description
<i>Streptomyces nodosus</i>	Yes	No	RDP II R9	S000429073
<i>Streptomyces olivochromogenes</i>	Yes	No	RDP II R9	S000248345
<i>Streptomyces platensis</i>	Yes	Yes	Sequencing/I. Hunter	23383
<i>Streptomyces rimosus</i>	Yes	Yes	Sequencing/I. Hunter	ATCC10970 (DSMZ)
<i>Streptomyces rimosus</i>	Yes	Yes	Sequencing/I. Hunter	G37
<i>Streptomyces rimosus</i>	Yes	Yes	Sequencing/I. Hunter	G7
<i>Streptomyces rimosus paromomycinus</i>	Yes	No	RDP II R9	S000271818
<i>Streptomyces rimosus</i>	Yes	Yes	Sequencing/I. Hunter	R6 554W
<i>Streptomyces rimosus</i>	Yes	Yes	Sequencing/I. Hunter	R6 dark
<i>Streptomyces rimosus rimosus</i>	Yes	No	RDP II R9	S000007214
<i>Streptomyces rimosus</i>	Yes	Yes	Sequencing/I. Hunter	S15883
<i>Streptomyces roseoverticullatus</i>	Yes	No	RDP II R9	S000581712
<i>Streptomyces scabiei</i>	Yes	Yes	RDP II R9/BSRC	S000439340/13682
<i>Streptomyces speibonae</i>	Yes	No	RDP II R9	S000394398
<i>Streptomyces subrutilus</i>	Yes	No	RDP II R9	S000381289
<i>Streptomyces tanashiensis</i>	Yes	Yes	RDP II R9/ATCC	S000652273/33159
<i>Streptomyces virginiae</i>	Yes	Yes	RDP II R9/ATCC	S000652521/19817
<i>Streptomyces_avermitilis</i>	Yes	Yes	Genbank/ATCC	BA000030.3/31267
<i>Streptosporangium album</i>	Yes	No	RDP II R9	S000005905
<i>Streptosporangium brasiliense</i>	Yes	No	RDP II R9	S000009296
<i>Streptosporangium roseum</i>	Yes	Yes	RDP II R9/ATCC	S000437720/12428
<i>Thermobifida fusca</i>	Yes	No	Genbank	CP000088.1
<i>Tropheryma whipplei</i>	Yes	No	Genbank	AE014184

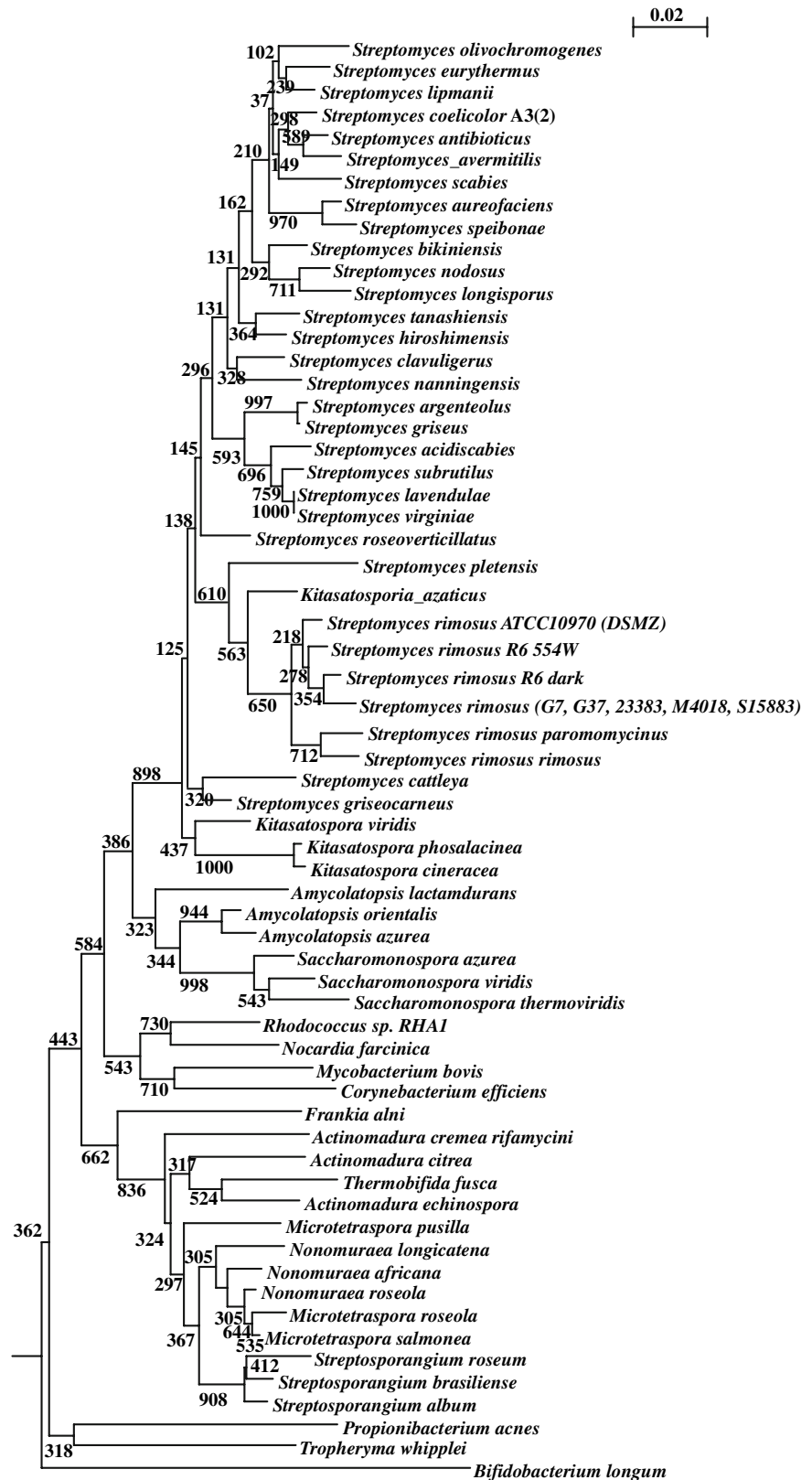
compare the genomes of selected Actinobacteria with complete genome sequences against the linear genome of *S. coelicolor*. This is done in order to confirm the presence of a core conserved syntenous region across most Actinobacteria. These Actinobacteria were selected based on the fact they had the largest genomes within a genus and that they covered a wide range across the Actinobacteria; they were *Bifidobacterium longum*, *Frankia alni*, *Leifsonia xyli*, *Mycobacterium vanbaalenii*, *Nocardia farcinicia*, *Nocardiodes* sp., *Rhodococcus* sp. RHA1 and *Thermobifida fusca*. The choice of the largest genome within a genus was in order to minimize any effects due to gene loss, particularly in pathogens and symbiotes. It should be noted that all the genomes analyzed in this part of the study are believed to be circular with the notable exception of *Rhodococcus* sp. RHA1 (McLeod et al. 2006).

Results and discussion

Phylogeny of *Streptomyces rimosus*

Figure 1 shows a SSU rRNA gene phylogenetic tree of the species that were analyzed by DNA/DNA genomic microarray analysis in this study together with a range of Actinobacteria that have complete genome sequences and a number of other selected Actinobacteria including other *Streptomyces* spp. It is clear that the *Streptomyces* as a genus form a distinct clade with *Kitasatospora* species with a good bootstrap value of 898/1000. This clade includes both various *Streptomyces rimosus* subsp. *rimosus* strains, *Streptomyces rimosus* subsp. *paromomycinus* and *Streptomyces platensis*, all of which are closely related to *Streptomyces rimosus* G7 and its derivatives. These strains also form a clade with quite good

Fig. 1 SSU rRNA gene phylogenetic tree of selected Actinobacteria including those analyzed in this study by microarray comparative genomics. The sequences were aligned using CLUSTALX and the tree produced by Neighbor-Joining using the same program



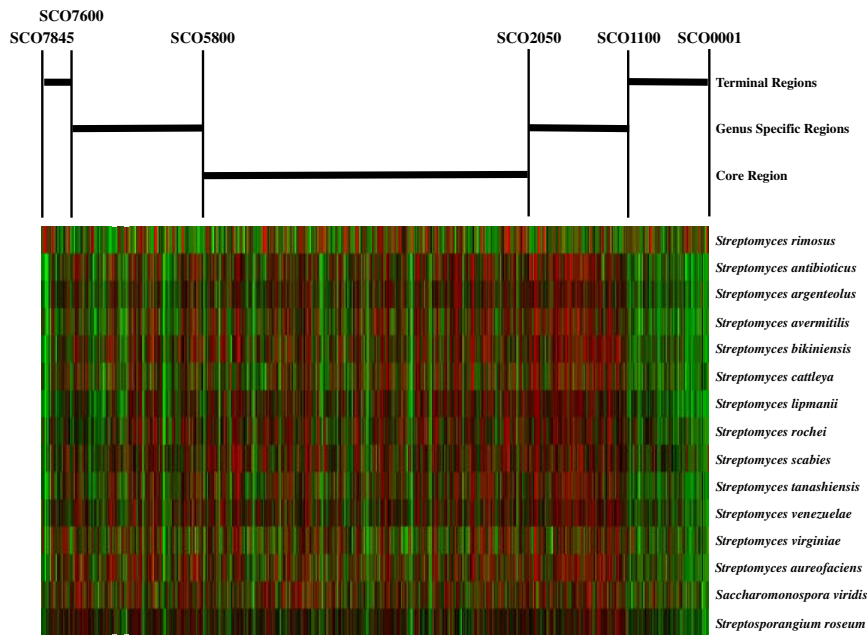


Fig. 2 Analysis of the genomes by comparative whole genome DNA microarray analysis of *Streptomyces antibioticus*, *Streptomyces argenteolus*, *Streptomyces aureofaciens*, *Streptomyces avermitilis*, *Streptomyces bikiniensis*, *Streptomyces cattleya*, *Streptomyces clavuligerus*, *Streptomyces lipmanii*, *Streptomyces rimosus* G7, *Streptomyces scabies*, *Streptomyces tanashiensis*, *Streptomyces virginiae*, *Streptosporangium roseum* and *Saccharomonospora viridis* using a *Streptomyces*

coelicolor genome microarray against *Streptomyces coelicolor* A3(2) 104. The arrays are normalized using the core region of *Streptomyces coelicolor*; green shows a lack of hybridization, black shows equivalent hybridization and red shows a stronger average hybridization. The five regions of the *Streptomyces coelicolor* genome described in Hsiao and Kirby 2007 are also indicated

bootstrap support (610/1000) with only the presence of *Kitasatospora azaticus* disrupting the group. Furthermore, the situation is made more complex by the presence of two further clades near the root of the *Streptomyces/Kitasatospora* clade. These are a *Kitasatospora* subclade, which supports the distinctness of *Kitasatospora*, and another containing *Streptomyces cattleya* and *Streptomyces griseocarneus*. Together with *Streptomyces roseovorticillatus* on the other side of the *S. rimosus* group, these suggest that there may be something novel about the species close to the base of the *Streptomyces/Kitasatospora* clade. In terms of the organisms included in this study, *Streptosporangium roseum* and *Saccharomonospora viridis* are both outside of the clade containing the *Streptomyces* and *Kitasatospora*. When the whole tree is examined, *Saccharomonospora viridis* groups with *Amycolatopsis azureus* in a sister clade to *Streptomyces/Kitasatospora* with good bootstrap support and *Streptosporangium roseum* is found in a quite separate clade with *Thermobifidia fusca* and *Frankia alni* at

some distance from the *Streptomyces/Kitasatospora* clade, also with good bootstrap support. The *Streptomyces* species analyzed in this study by microarray genotyping are spread evenly across the *Streptomyces/Kitasatospora* clade. Thus, while SSU rRNA gene phylogenetic analysis places *S. rimosus* at the edge of the *Streptomyces* clade, *S. cattleya*, does branch outside of *S. rimosus*, retains the genome profile found in the other *Streptomyces* found in the main part of the clade. The molecular phylogeny of *S. cattleya* conforms well with the phylogeny proposed earlier based on phenotypic characters (Williams et al. 1983a, b). Furthermore, this phylogeny also places *S. cattleya* and *S. rimosus* at the boundary of the *Streptomyces* clade.

Genome analysis by DNA/DNA microarray hybridization

The initial species analyzed by DNA/DNA genomic microarray analysis in this study include 12

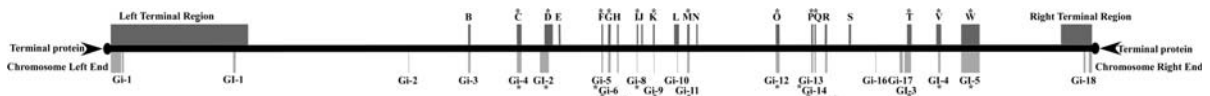


Fig. 3 A scale analysis of the open reading frames of *S. coelicolor* showing the horizontally transferred regions or HTRs identified by Hsiao and Kirby (2007) (HTR B–W; upper dark grey) and the large and small genetic islands identified by Jayapal et al. (GI 1–5 and GI 1–18, respectively; lower light grey). Note that the terminal regions defined in this paper using the larger dataset than Hsiao and Kirby includes HTR A within the terminal region. Furthermore, the chromosome ends defined by Jayapal et al. are not the same as the terminal regions defined by Hsiao and Kirby. The latter were defined on

Streptomyces species, one of which is *Streptomyces rimosus* subsp. *rimosus* G7, type strain of this species obtained from Pfizer & Co. Inc. and purported to be the origin of ATCC10970; in addition a *Streptosporangium* species and a *Saccharomonospora* species were included. Figure 2 shows a pseudo-color display of the microarray data covering 7083 genes from the *S. coelicolor* genome sequence. It is clear from this that all the species analyzed except for *S. rimosus* show a relatively similar pattern of DNA conservation. This include *S. cattleya* and the two none *Streptomyces* species, *Saccharomonospora* and *Streptosporangium*, although the overall signal for *Streptosporangium* is distinctly weaker than for the other species. Bentley et al. (2002) identified using the *S. coelicolor* genome sequence that there was a conserved core of genes within the *S. coelicolor* chromosome from about SCO2050 to SCO5800. These genes are present in a number of other Actinobacteria genomes. Supplementary Fig. 1 shows analysis of a wider range of Actinobacteria genomes than that available to Bentley et al. analyzed at the DNA using the program Mauve and this confirms that this core region is quite ubiquitous across many Actinobacteria. In terms of the phylogenetic tree in Fig. 1, this core structure only seems to break down once the clades containing *Bifidobacterium longum* is reached and includes both linear and circular genomes. Thus, the result whereby the majority of species analyzed by microarray comparison in this study have a conserved genome pattern is not that unexpected and it is suggested that a core genome structure in terms of gene synteny may be the norm in the Actinobacteria. It should be noted that microarray analysis does not imply that these species have a similar linear genome structure to *S. coelicolor*,

lack of gene signal across four *Streptomyces*, while the former show that the terminal regions of *S. coelicolor* resemble those of *S. lividans* much more closely than those present in the four Hsiao and Kirby strains. The “*” indicates a HTR identified in Bentley et al. (2002) and all are confirmed by Hsiao and Kirby and by Jayapal et al. There are some discrepancies in start positions, end position and size between the HTRs and the GIs/Gis. This is to be expected, not least because the hybridization conditions used for the two microarray analysis datasets were not identical

S. avermitilis, *S. scabies* and *Rhodococcus* sp. RHA1, but the results do suggest this as a first approximation if linear replication associated genes seem to be present. Another specific characteristic of linear Actinobacteria genomes would seem to be the absence in the resolution of circular dimers and *recB*. However, the comparative genomic approach used here cannot detect genes that are already absent in *S. coelicolor*. Thus it is not unexpected that *S. cattleya*, *Saccharomonospora* and *Streptosporangium* retain a conserved pattern by microarray analysis. What is unexpected is that *Streptomyces rimosus* G7, the type strain for *S. rimosus*, does not.

An earlier study by our laboratory using DNA/DNA genomic microarray analysis of four species confirmed that the terminal regions of these species seemed to be poorly conserved (SCO0001 to SCO01100 and SCO7600 to SCO7850). Furthermore, between the core regions and these terminal regions, two regions of conservation within the *Streptomyces*, perhaps with even higher conservation than the core region, could be detected and these regions show synteny when the genome sequences of other *Streptomyces* are compared to *S. coelicolor* (Chater, KF, unpublished and analysis of *S. coelicolor*, *S. avermitilis* and *S. scabies* genome sequences by Artemis (<http://www.sanger.ac.uk/Software/Artemis>)). If we examine these genus specific regions in Fig. 2, it will be seen that the results for most strain support a higher level of gene conservation in these two regions as expressed by a red shift, this is most clear in the left region. This structural pattern is also present *S. cattleya*, *Saccharomonospora viridis* and *Streptosporangium roseum*, although the divergence of *Streptosporangium roseum* has

Table 2 Microarray genotyping data for *Streptosporangium roseum*, *Saccharomonospora viridis*, *Streptomyces rimosus* G7 and *Streptomyces cattleya*

<i>Streptomyces coelicolor</i> gene	<i>Streptosporangium roseum</i>	<i>Saccharomonospora viridis</i>	<i>Streptomyces rimosus</i> G7	<i>Streptomyces cattleya</i>
SCO0436 ribosomal protein L32	0.33	-0.68	0.70	0.08
SCO0569 ribosomal protein L36	-1.08	-1.25	-1.39	0.34
SCO1150 ribosomal protein L31	0.03	0.74	-1.02	-0.73
SCO1505 ribosomal protein S4	-0.05	0.36	0.21	-0.26
SCO1598 ribosomal protein L20	0.02	-0.53	0.94	0.81
SCO1599 ribosomal protein L35	-0.04	0.28	0.12	0.21
SCO1998 ribosomal protein S1	0.54	1.89	1.40	0.83
SCO2563 ribosomal protein S20	0.21	1.29	-1.87	-0.67
SCO2596 ribosomal protein L27	0.52	-0.17	-0.17	0.64
SCO2597 ribosomal protein L21	1.14	-0.31	0.42	-0.17
SCO3124 ribosomal L25	0.03	0.71	1.42	0.46
SCO3427 ribosomal protein L31	0.82	0.92	2.93	0.71
SCO3428 ribosomal protein L33	-0.04	-0.47	-0.09	0.07
SCO3429 ribosomal protein L28	-0.15	-1.69	-0.18	0.42
SCO3430 ribosomal protein S14	0.61	0.11	0.06	-0.73
SCO3880 ribosomal protein L34	0.63	0.26	-1.18	-0.54
SCO3906 ribosomal protein S6	0.03	0.27	0.31	1.70
SCO3909 ribosomal protein L9	0.37	0.18	0.62	0.02
SCO4648 ribosomal protein L11	-0.50	0.28	1.71	0.44
SCO4649 ribosomal protein L1	-0.54	0.42	0.99	0.94
SCO4652 ribosomal protein L10	0.12	0.80	-0.59	-0.54
SCO4653 ribosomal protein L7/L12	0.24	1.82	1.32	1.33
SCO4659 ribosomal protein S12	0.03	1.77	-0.79	0.93
SCO4660 ribosomal protein S7	0.43	1.07	-0.20	-0.35
SCO4701 ribosomal protein S10	0.50	1.96	0.39	1.23
SCO4702 ribosomal protein L3	-0.65	-0.08	0.53	-0.33
SCO4703 ribosomal protein L4	0.22	1.40	1.52	1.00
SCO4704 ribosomal protein L23	0.04	1.60	0.84	1.60
SCO4705 ribosomal protein L2	0.41	1.46	1.34	-0.23
SCO4706 ribosomal protein S19	-0.22	-0.74	-1.12	0.57
SCO4707 ribosomal protein L22	0.70	1.98	-0.36	0.28
SCO4708 ribosomal protein S3	0.23	1.53	0.76	0.15
SCO4709 ribosomal protein L16	0.48	1.81	0.72	1.13
SCO4710 ribosomal protein L29	-0.34	-0.13	-1.25	-0.64
SCO4711 ribosomal protein S17	1.28	0.72	-0.32	1.29
SCO4712 ribosomal protein L14	-0.17	0.29	0.02	0.49
SCO4713 ribosomal protein L24	0.65	1.86	0.42	1.02
SCO4714 ribosomal protein L5	0.45	1.43	1.09	0.92
SCO4715 ribosomal protein S14	0.03	0.27	-1.69	-1.06
SCO4716 ribosomal protein S8	-0.48	-0.39	-0.11	0.16
SCO4717 ribosomal protein L6	-0.02	1.18	0.48	0.20
SCO4718 ribosomal protein L18	1.04	0.38	-0.95	0.31
SCO4719 ribosomal protein S5	0.33	1.77	0.39	1.25

Table 2 continued

<i>Streptomyces coelicolor</i> gene	<i>Streptosporangium roseum</i>	<i>Saccharomonospora viridis</i>	<i>Streptomyces rimosus</i> G7	<i>Streptomyces cattleya</i>
SCO4720 ribosomal protein L30	−0.48	−0.39	1.01	0.28
SCO4721 ribosomal protein L15	0.53	1.03	1.15	0.57
SCO4726 ribosomal protein L36	0.39	0.2	−0.66	−0.81
SCO4727 ribosomal protein S13	0.13	1.63	−0.62	−0.15
SCO4728 ribosomal protein S11	0.84	2.01	−0.02	0.77
SCO4730 ribosomal protein L17	0.36	0.31	0.11	0.10
SCO4734 ribosomal protein L13	−0.34	−0.53	0.81	0.52
SCO4735 ribosomal protein S9	−0.22	−0.30	−0.91	−0.41
SCO5359 ribosomal protein L31	−1.09	0.18	0.64	0.22
SCO5564 ribosomal protein L28	0.29	1.45	2.49	−0.25
SCO5591 ribosomal protein S16	−0.20	0.13	0.18	−0.35
SCO5595 ribosomal protein L19	−0.40	0.34	1.32	1.07
SCO5624 ribosomal protein S2	0.15	2.04	−1.49	0.83
SCO5736 ribosomal protein S15	0.22	0.74	−3.06	−0.34
Total of Genes scored as absent	2	2	11	4

The signal levels are the Log_2 signal ratio for that species against *Streptomyces coelicolor* DNA. The data points marked in bold indicate a level below 2 standard deviations from the mean signal for that species and are suggestive of gene absence or a high level of sequence divergence

weakened the effect in this species. This further supports a similar genome structure across these strains.

Gene presence and absence

As a result of the fact that the one clear exception to this congruent pattern in Fig. 2 is *S. rimosus*, the major horizontally transferred regions (HTRs) within the *S. coelicolor* genome postulated by Bentley et al. (2002) and expanded by Hsiao and Kirby (2007) were examined. These regions were also identified by microarray analysis between *S. coelicolor* and the very closely related species *Streptomyces lividans* by Jayapal et al. (2007). Taking the slightly different criteria used to define an HTR in the three publications, there is almost a 100% agreement between them. Hsiao and Kirby (2007) and Jayapal et al. identify more HTRs than Bentley et al. (2002), but with excellent inclusivity and excluding the terminal regions, only 4/22 unique to their study HTRs are postulated by Hsiao and Kirby and 3/21 by Jayapal et al. (Fig. 3). Based on this good agreement, Supplementary Fig. 2 shows examples of some of the HTRs (the possible HTRs in *S. coelicolor* may be as

many as 27 or more when the larger microarray dataset used here is examined). In all eight cases, the genes within the HTR are absent from most or all of the species analyzed with the exception of *S. rimosus*. *S. rimosus*, on the other hand, shows high signal levels within the HTR regions for more but by no means all genes, with perhaps HTR-L, HTR-T and HTR-V being the best examples. This implies the presence of some genes that are present in *S. coelicolor* but seem to be generally absent from most *Streptomyces*. The implication of this is not that the HTRs as a whole are present in *S. rimosus*, but that unlike the main group of *Streptomyces*, *S. rimosus* contains some genes that have been only found as yet in the *S. coelicolor* HTRs, which may have origins outside of the main group of *Streptomyces*. These genes represent common ancestral genes that have been lost by most *Streptomyces*, but have re-enter *S. coelicolor* as HTRs. Notwithstanding what their origin may be in either *S. coelicolor* or *S. rimosus*, the absence of clearly defined HTRs found by both Hsiao and Kirby and Jayapal et al. further supports the novel genome structure for *S. rimosus* earlier proposed based on the overall microarray hybridization pattern.

Table 3 Microarray data for the genes *tpgA*, *tapA* and *ttrA* for *Streptomyces antibioticus*, *Streptomyces argenteolus*, *Streptomyces aureofaciens*, *Streptomyces avermitilis*, *Streptomyces bikiniensis*, *Streptomyces cattleya*, *Streptomyces clavuligerus*, *Streptomyces lipmanii*, *Streptomyces rimosus* G7, *Streptomyces scabies*, *Streptomyces tanashiensis*, *Streptomyces virginiae*, *Streptosporangium roseum* and *Saccharomonospora viridis*

	<i>S. rimosus</i>	<i>S. antibioticus</i>	<i>S. argenteolus</i>	<i>S. aureofaciens</i>	<i>S. avermitilis</i>	<i>S. bikiniensis</i>	<i>S. cattleya</i>	<i>S. lipmanii</i>
SCO7733								
tapA	0.138	0.359	−0.019	−0.084	0.175	0.037	0.310	−0.404
SCO7734								
tpgA	−0.519	−0.148	−0.188	−0.333	0.663	−0.478	−0.576	−0.444
SCO0002								
ttrA	0.974	0.461	0.071	0.255	0.435	0.585	0.536	−0.135
	<i>S. rochei</i>	<i>S. scabies</i>	<i>S. tanashiensis</i>	<i>S. venezuelae</i>	<i>S. virginiae</i>	<i>Saccharomonospora viridis</i>	<i>Streptosporangium roseum</i>	
SCO7733								
tapA	−0.398	0.137	0.1741	−0.320	0.759	0.27	−0.104	
SCO7734								
tpgA	0.054	−0.256	−0.375	−0.121	0.028	−0.737	−0.203	
SCO0002								
ttrA	0.250	−1.160	0.194	0.887	−1.712	0.142	−1.129	

Bold values indicate that the signal for that gene is more than 2SD below the mean core signal for that species and such a value is suggestive of either gene absence or very low similarity

In order to avoid the possibility of a systematic error, the SSU rRNA gene sequence of *S. rimosus* G7 and its selected derivatives G37, 23383, M4018 and S15883, were amplified by PCR and sequenced. The sequence was 99.5% identical to the sequence available from Ribosomal Database Project-II Release 9 over about 360bp. The initial analysis using two microarrays of *S. rimosus* G7 was then repeated using freshly isolated DNA from a new single colony by a different scientist. The results were 97% in agreement with the previous microarray data. The above *S. rimosus* strains were derived by mutation and selection from *S. rimosus* G7 and represent 30 years of selection. When these strains were compared to *S. rimosus* G7 by genotyping, they gave almost identical microarray profiles (Supplementary Fig. 3). Variations in signal for individual genes was less than 3% and no cases of gene deletion or gene duplication were found between these five strains.

The dataset from two specific groups of genes were analyzed in more detail. The first was the ribosomal proteins in order to confirm the validity of the dataset. Table 2 shows that Log₂ signal levels for

these genes and it is clear that they show a high level of conservation as would be expected. The exceptions are few in number. Two each are scored as absent in *Streptosporangium* and *Saccharomonospora*, four in *S. cattleya*, but eleven in *S. rimosus*. This supports the divergent nature of the *S. rimosus* genome compared to even certain non-*Streptomyces* mycelate Actinomycetales. One particular ribosomal protein, SCO0569 (L36) is known to have a duplicated elsewhere on the *S. coelicolor* chromosome and therefore is probably a pseudogene in *S. coelicolor*. SCO0569 is scored as absent in many *Streptomyces* species. Interestingly, as shown in Table 2, the main copy of L36, SCO4726 is scored as absent in *S. cattleya*, but SCO0569 is scored as present in *S. cattleya*, which contrasts with *Streptosporangium*, *Saccharomonospora* and *S. rimosus*; this suggests perhaps that SCO0569 is a horizontally transferred pseudogene from a distant species.

Certain genes can be specifically associated with the *Streptomyces* linear chromosomal model and these include *TapA*, *TpgA* and *TtrA*. *TapA* and *TpgA* have been shown to be involved in terminal capping of the linear chromosome and linear replication (Bao

and Cohen 2001; Bey et al. 2000; Yang et al. 2002), while *TtrA* seems to have a function associated with genetic exchange of linear DNA between *Streptomyces* strains (Bey et al. 2000). Table 3 shows the microarray data for these three genes across the species studied here and almost all the species show a signal level that implies all three genes are present; this includes *S. rimosus*. This cutoff is based on 2SD below the mean core signal level for a strain (Hsiao and Kirby 2007). Interestingly, however, the two *Streptomyces* at the extreme edge of the *Streptomyces* clade, *S. cattleya* and *S. rimosus* do show the lowest signal for *tpgA*, and this is very close to the boundary for 2SD. The signals for *tapA* are significantly higher implying the low *tpgA* signal may be due to evolutionary divergence. The exception to the conservation of *tpgA* and *tapA* is *Saccharomonospora viridis*, where the signal suggests that *tpgA* is absent but *tapA* is present; in this case too, gene divergence might be the explanation. In *S. scabies*, *S. virginiae* and *Streptosporangium roseum*, the signals suggest that *ttrA* might be absent. This is especially unexpected for *S. scabies* as this gene seems to be present based on Blast analysis of the preliminary genome sequence from http://www.sanger.ac.uk/Projects/S_scabies/. However, the *S. scabies* strain used in this study is not the same as used for genome sequencing and therefore a difference is not inexplicable, particularly when the diversity of the *S. scabies*-like pathogens is taken into account. It is also important to note that *ttrA* is found at the extreme end of the linear chromosome of all known chromosomal terminal regions, namely *S. coelicolor*, *Streptomyces lividans*, *S. avermitilis*, *S. scabies* and *S. ambofaciens* and therefore is particularly vulnerable to circularization, which can occur at relatively high frequencies in *Streptomyces* (Lin and Chen 1997; Choulet et al. 2006a; Choulet et al. 2006b).

Failure to detect major differences after significant mutation and selection over 30 years

The lack of divergence between the *S. rimosus* strains analyzed in this study is most interesting and unexpected (Supplementary Fig. 3). These strains have undergone mutation and selection for tetracycline production over a period of about 30 years and show phenotype variation associated with this selection (data not shown). Although microarray

genomotyping of the type used here does not allow study of genes novel to the strain being analyzed (*S. rimosus* in this case and an example would be the oxytetracycline biosynthesis cluster), the lack of changes in genes conserved between *S. coelicolor* and *S. rimosus* suggests that point mutations may play an important role in any changes to central metabolism induced during selection for higher antibiotic production. In other words, while it is possible that gene duplication/deletion does occur with respect to metabolic pathway not detected by the microarray approach, changes to core metabolic functions would seem to be less drastic, even though changes in the metabolic flux within the cell are probably essential to provide enough appropriate metabolites for high antibiotic production.

Conclusions

Based on the microarray results presented in this study, we suggest that the genome of *S. rimosus* ATCC10970 has a different structure from the majority of *Streptomyces*. Interestingly, the authors were unable to find any pattern to the genes that seemed to be conserved between *S. coelicolor* and *S. rimosus*, suggesting that the genome differences are major and not just vertical evolutionary divergence, the possibility remains that, based on the evidence for conservation of *tapA*, *tpgA* and *ttrA*, the chromosome of *S. rimosus* is linear like the chromosomes of other *Streptomyces* that have been investigated, but that the genes making up the chromosome are highly divergent from those of other *Streptomyces*. In addition, it is likely that the synteny found across the Actinobacterial core region, as well as the *Streptomyces* genus specific regions defined in Hsiao and Kirby (2007), are not present in *S. rimosus*. Alternatively, if the Actinobacterial core region is present, it would seem to be highly disrupted and/or highly divergent from other *Streptomyces* that have been analyzed. We conclude that *S. rimosus* may represent a distinct evolutionary lineage making it a worthy and important candidate for genome sequencing in the future. Furthermore, based on the phylogenetic clustering in Fig. 1, an extension of the DNA/DNA microarray genomotyping to various *Kitasatospora* spp., *Streptomyces platensis*, *Streptomyces griseocarnus*, *Streptomyces roseovertillatus* and *Streptomyces rimosus paromomycinus* would help to define how wide spread this structure might be.

Finally, point mutations between the various high production strains derived from *S. rimosus* G7 would seem to be the most probable source of variation in the core metabolic genes that allow increased oxytetracycline biosynthesis.

Acknowledgements Special thanks goes to Carton Chen and the rest of the *Streptomyces* research team at Yang-Ming without whose help and support this work would not have been possible and to everyone in the Department at Strathclyde for their hospitality and help. Financial support is acknowledged from National Yang-Ming University, Taiwan, the National Science Council, Taiwan (Grants 93-2320-B-010-041 and 94-2320-B-010-037) and the Royal Society (London).

References

- Bao K, Cohen SN (2001) Terminal proteins essential for the replication of linear plasmids, and chromosomes in *Streptomyces*. *Genes Dev* 15:1518–1527
- Bentley SD, Brown S, Murphy LD, Harris DE, Quail MA, Parkhill J, Barrell BG, McCormick JR, Santamaria RI, Losick R, Yamasaki M, Kinashi H, Chen CW, Chandra G, Jakimowicz D, Kieser HM, Kieser T, Chater KF (2004) SCP1, a 356,023 bp linear plasmid adapted to the ecology and developmental biology of its host, *Streptomyces coelicolor* A3(2). *Mol Microbiol* 51:1615–1628
- Bentley SD, Chater KF, Cerdeño-Tárraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabinowitz E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417:141–147
- Bey SJ, Tsou MF, Huang CH, Yang CC, Chen CW (2000) The homologous terminal sequence of the *Streptomyces lividans* chromosome and SLP2 plasmid. *Microbiology* 146:911–922
- Chen CW, Huang CH, Lee HH, Tsai HH, Kirby R (2002) Once the circle has been broken: dynamics, and evolution of *Streptomyces* chromosomes. *Trends Genet* 18:522–529
- Choulet F, Aigle B, Gallois A, Mangenot S, Gerbaud C, Truong C, Francou FX, Fourrier C, Guérineau M, Decaris B, Barbe V, Pernodet JL, Leblond P (2006a) Evolution of the terminal regions of the *Streptomyces* linear chromosome. *Mol Biol Evol* 23:2361–2369
- Choulet F, Gallois A, Aigle B, Mangenot S, Gerbaud C, Truong C, Francou FX, Borges F, Fourrier C, Guérineau M, Decaris B, Barbe V, Pernodet JL, Leblond P (2006b) Intraspecific variability of the terminal inverted repeats of the linear chromosome of *Streptomyces ambifaciens*. *J Bacteriol* 188:6599–6610
- Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394–1403
- Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis, and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95:14863–14868
- Gascuel O, Steel M (2006) Neighbor-joining revealed. *Mol Biol Evol* 23:1997–2000
- Gollub J, Ball CA, Binkley G, Demeter J, Finkelstein DB, Hebert JM, Hernandez-Boussard T, Jin H, Kaloper M, Matese JC, Schroeder M, Brown PO, Botstein D, Sherlock G (2003) Microarray database: data access, and quality assessment tools. *Nucleic Acids Res* 31:94–96
- Hsiao N-H, Kirby R (2007) Comparative genomics of *Streptomyces avermitilis*, *Streptomyces cattleya*, *Streptomyces maritimus* and *Kitasatospora aureofaciens* using a *Streptomyces coelicolor* microarray system. *Antonie van Leeuwenhoek International Journal of General and Molecular Microbiology* ISSN 0003–6072 (Print) 1572–9699 (Online)
- Huang CH, Lin YS, Yang YL, Huang SW, Chen CW (1998) The telomeres of *Streptomyces* chromosomes contain conserved palindromic sequences with potential to form complex secondary structures. *Mol Microbiol* 28:905–916
- Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, Sakaki Y, Hattori M, Omura S (2003) Complete genome sequence comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nature Biotechnol* 21:526–531
- Jayapal KP, Lian W, Glod F, Sherman DH, Hu WS (2007) Comparative genomic hybridizations reveal absence of large *Streptomyces coelicolor* genomic islands in *Streptomyces lividans*. *BMC Genomics* 8:229–241
- Kirby R, Hopwood DA (1977) Genetic determination of methylenomycin synthesis and resistance by the SCP1 plasmid of *Streptomyces coelicolor* A3(2). *J Gen Microbiol* 98:239–252
- Kirby R, Wright LF, Hopwood DA (1975) Plasmid-determined antibiotic synthesis and resistance in *Streptomyces coelicolor*. *Nature (London)* 254:265–267
- Lin YS, Chen CW (1997) Instability of artificially circularized chromosomes of *Streptomyces lividans*. *Mol Microbiol* 26:709–719
- Lin YS, Kieser HM, Hopwood DA, Chen CW (1993) The chromosomal DNA of *Streptomyces lividans* 66 is linear. *Mol Microbiol* 10:923–933
- McLeod MP, Warren RL, Hsiao WW, Araki N, Myhre M, Fernandes C, Miyazawa D, Wong W, Lillquist AL, Wang D, Dosanjh M, Hara H, Petrescu A, Morin RD, Yang G, Stott JM, Schein JE, Shin H, Smailus D, Siddiqui AS, Marra MA, Jones SJ, Holt R, Brinkman FS, Miyauchi K, Fukuda M, Davies JE, Mohn WW, Eltis LD (2006) The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. *Proc Natl Acad Sci USA* 103:15582–15587
- Petkovic H, Cullum J, Hranueli D, Hunter IS, Peric-Concha N, Pigac J, Thamchaipenet A, Vujaklija D, Long PF (2006) Genetics of *Streptomyces rimosus*, the oxytetracycline producer. *Microbiol Mol Biol Rev* 70:704–728

- Pospiech A, Neumann B (1995) A versatile quick-prep of genomic DNA from Gram-positive bacteria. *Trends Genet* 11:217–218
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acid Res* 24:4876–4882
- Williams ST, Goodfellow M, Alderson G, Wellington EM, Sneath PH, Sackin MJ (1983a) Numerical classification of *Streptomyces* related genera. *J Gen Microbiol* 129:1743–1813
- Williams ST, Goodfellow M, Wellington EM, Vickers JC, Alderson G, Sneath PH, Sackin MJ, Mortimer AM (1983b) A probability matrix for identification of some *Streptomyces*. *J Gen Microbiol* 129:1815–1830
- Yang CC, Huang CH, Li CY, Tsay YG, Lee SC, Chen CW (2002) The terminal proteins of linear *Streptomyces* chromosomes plasmids: a novel class of replication priming proteins. *Mol Microbiol* 43:297–305