Evolution of the streptomycin and viomycin biosynthetic clusters and resistance genes

Paris Laskaris, B.Sc. (Hons.)

A thesis submitted to the University of Warwick for the degree of Doctor of Philosophy.

Department of Biological Sciences,

University of Warwick,

Coventry, CV4 7AL

September 2009

Contents

List of Figures

List of Tables

Abbreviations

AAC	Aminoglycoside ACetyltransferase
APH	Aminoglycoside PHosphotransferase
ATCC	American Type Culture Collection
BAC	Bacterial Artificial Chromosome
BAPS	Bayesian Analysis of Population Structure
BLAST	Basic Local Alignment Search Tool
BRAT	Bayesian Recombination Tracker
cDNA	complementary DNA
CDS	CoDing Sequence
DSM	German Collection of Microorganisms
FEL	Fixed Effects Likelihood
eBURST	electronic Based Upon Related Sequence Types
GARD	Genetic Algorithm Recombination Detection
HGT	Horizontal Gene Transfer
NRPS	NonRibosomal Peptide-Synthetase
NBRC	NITE Biological Resource Center
OD_{600}	Optical Density at 600 nm
ORF	Open Reading Frame
PAML	Phylogenetic Analysis by Maximum Likelihood
PCR	Polymerase Chain Reaction
REL	Random Effects Likelihood
RPM	Rounds Per Minute
RT	Reverse Transcription
RTPCR	Real-Time Polymerase Chain Reaction
SLAC	Single Likelihood Ancestor Counting
SNP	Single Nucleotide Polymorphism
ST	Sequence Type
^{T}	Type strain
TCDNA	Total Community DNA

Acknowledgements

I wish to thank my supervisor, Prof Elizabeth M.H. Wellington, for her guidance and the advice she provided for the development of this project.

I would also like to thank Dr L. A. Calvo-Bado, Dr W. H. Gaze, Dr A. Ul-Hassan, Dr. S. Tolba, Dr. F. Sweeney, Dr. A. Yaxley, Dr. M. Krsek, and the other members of Prof Wellington's research group for their technical advice and assistance.

I wish to thank Dr. D. Pearce, Dr. K. Hughes and Dr L. A. Calvo-Bado for providing the Antarctic soil DNA fosmid library.

I would like to thank Dr S. D. Bentley for arranging to have two of the fosmids sequenced at the Wellcome Trust Sanger Institute.

Finally I would like to thank my parents for all of their support.

Declaration

The results presented in this thesis are the work of the author unless otherwise specified in the text. None of this work has been previously submitted for a degree application. All sources of information presented in this thesis have been acknowledged by a reference.

Abstract

The distribution of the streptomycin (*strA*) and viomycin (*vph*) resistance genes was examined in *Streptomyces* isolates. It was hypothesised that non-antibiotic producers that are niche competitors with producers will need to possess resistance to the antibiotic and will thus have acquired resistance genes. A detailed phylogenetic study, utilizing a novel multilocus sequence typing (MLST) scheme, was made of a collection of isolates and types strains with a *Streptomyces griseus* phenotype in addition to type strains from known producers of streptomycin and related compounds. *strA* and *vph* were found either within a biosynthetic gene cluster or independently. *S. griseus* strains possessing the streptomycin cluster formed part of a clonal complex and have been readily isolated from soil originating in every continent except Antarctica. Few copies of *strA* were detected in soil total community DNA, none of which were identical to the gene from the streptomycin cluster. All *S. griseus* strains possessing solely *strA* belonged to two clades and were closely related to streptomycin producers. The *strA* in the resistance-only strains is likely to have originated from the self-resistance gene of another aminoglycoside cluster and arrived in those *S. griseus* strains via horizontal gene transfer. *S. griseus* strains with only *vph* also formed two clades and were more distantly related to the producers than to one another. The high sequence divergence of the viomycin resistance genes also suggests that the *vph* homologue arrived in these two groups from another peptide antibiotic cluster via horizontal gene transfer. The expression of the *strA* gene was constitutive in resistance-only strains from both subgroups whereas streptomycin producers showed peak *strA* expression in late log phase which correlates with the switch on of streptomycin biosynthesis. One example of horizontal gene transfer of the streptomycin cluster was discovered, to a *Streptomyces platensis* strain, which contained a cluster with 84% sequence identity and almost identical gene structure and arrangement to that of the *S. griseus* cluster. Its expression pattern was also highly similar to that of *S. griseus* producers, but at a much lower level. Whilst there is evidence that antibiotics have diverse roles in nature, this work clearly supports the co-evolution of resistance in the presence of antibiotic biosynthetic capability within closely related soil dwelling bacteria. This reinforces the view that, for some antibiotics at least, the primary role is one of antibiosis during competition in soil for resources.

1 Introduction

1.1 Antibiotic function and resistance

Antibiotics are secondary metabolites, i.e. metabolites not essential for growth, which can inhibit growth or cause mortality in microorganisms other than the producer. Antibiotics target a large number of processes that are essential for a cell's growth including cell wall synthesis, folic acid metabolism, DNA duplication, transcription, protein synthesis and the cytoplasmic membrane (Madigan & Martinko, 2005). The fact that many antibiotic compounds target structures specific to bacteria renders them very clinically useful.

The therapeutic utilization of antibiotics in the 20th century was however rapidly accompanied by the appearance of bacterial pathogens that showed resistance to them (Hawkey, 2008). Antibiotic resistance can not only increase the difficulty of treating infections but, since antibiotic stress can induce the horizontal transfer of virulence genes (Ubeda *et al.*, 2005), can also assist in the evolution of more virulent pathogens. Resistance to antibiotics can develop via three main mechanisms: the acquisition of resistance genes or gene complexes via plasmids and other transposable elements, recombination of DNA from other bacteria into the cell's chromosome by transformation or via spontaneous mutations in the organism's genome (Pope *et al.*, 2008). In addition, bacteria can also develop tolerance, i.e. instead of preventing the antibiotic from affecting its target they shut the target down and thus keep the antibiotic from killing them (Lewis, 2008). This can be done by entering into a persistent state, such as by the persister cells in biofilms (e.g. *Pseudomonas aeruginosa*) or by being sequestered into vacuoles (e.g. *Mycobacterium tuberculosis*) (Sacchettini *et al.*, 2008).

Four main mechanisms can provide antibiotic resistance: a) prevention of the antibiotic's arrival to the site of action by an increase in the efflux or decrease in the influx through the cell membrane; b) enzymatic degradation or alteration of the antimicrobial agent; c) alteration of the site of antimicrobial action, rendering the drug ineffective; and d) development of site-of-action bypass mechanisms (Shea, 2003). Mutations can provide resistance; it is estimated that approximately one in a million to one in a billion bacterial cells stochastically acquire mutations that make

them resistant towards one or more antibiotics (Khachatourians, 1998). Such mutations might function by: a) making it more difficult for antibiotics to cross the cell membrane, e.g. the membrane of *Mycobacterium smegmatis* porin mutants show a nine-fold reduction in cephalosporin permeability (Danilchanka *et al.*, 2008); b) improving the specificity of primary metabolism genes involved in hydrolysis, group transfer or redox reactions towards antibiotics (Wright, 2005), e.g. a peptidoglycan acetyltransferase provides gentamicin resistance in *Providencia stuartii* (Payie & Clarke, 1997); c) modifying target sites to make it impossible for antibiotics to bind to them, e.g. mutations in the β-subunit of RNA-polymerase in *Mycobacterium tuberculosis* that render it resistant to rifampicin (Lambert, 2005); d) utilizing an alternate enzyme that is not susceptible to the antibiotic, e.g. bacteria utilizing L,D-transpeptidases as peptidoglycan cross-linking enzymes demonstrate high level cross-resistance against glycopeptide and β-lactam antibiotics which bind specifically to D,D-transpeptidases (Magnet *et al.*, 2007).

Resistance arising via mutation is more common in non-recombinogenic bacteria such as mycobacteria; in other organisms resistance is more likely to appear due to horizontal gene transfer (HGT). Many cases of HGT have been documented; a family of extended spectrum β-lactamases found in certain *E. coli* strains appears to originate from another enterobacterium, *Kluyvera ascorbata* (Rodriguez *et al.*, 2004). "*Streptomyces coelicolor*" produces a number of antibiotics; however it also contains a cluster of seven genes that confers inducible, high level vancomycin resistance despite the fact that it cannot synthesize that antibiotic (Hong *et al.*, 2004). Homologues to the *otrA* and *otrB* tetracycline resistance genes present in tetracycline producer *Streptomyces rimosus* have been found in *Mycobacterium fortuitum* (Pang *et al.*, 1994). Since streptomycete plasmids can be acquired by *Mycobacterium* species via natural transformation (Bhatt *et al.*, 2002), it is likely that they were acquired from the producer. The *vanH*, *vanA*, and *vanX* resistance genes found in *Enterococcus* species have orthologues in vancomycin producers *Streptomyces toyocaensis* and *Amycolatopsis orientalis* (Marshall *et al.*, 1998). Resistance genes can be transferred between distantly related groups, including different orders, one example being the transfer of a streptothricin resistance cluster from *Staphylococcus* to *Enterococcus* (Werner *et al.*, 2001) or the transfer of a kanamycin resistance gene from *Streptococcus* to *Campylobacter* (Trieu-Cuot &

Courvalin, 1986). The fact that resistance genes can be transferred between such distant groups lends credence to the hypothesis that resistance genes found in pathogens may have originated in antibiotic producers (Benveniste & Davies, 1973), since the latter need to have resistance genes to accompany antibiotic production in order to avoid killing themselves. Presumably, the resistance genes were mobilized via plasmids or transposable elements to other soil-inhabiting organisms thus enabling them to survive in the presence of antibiotic compounds (Mindlin *et al.*, 2006).

Antibiotic resistance appears to be very widespread in soil; in a screen of 480 spore-forming soil bacteria (collected from various environments) against 21 antibiotics all the sampled bacteria were found to be multidrug resistant (D'Costa *et al.*, 2006). The organisms were on average resistant to 7-8 antibiotics and showed resistance to every compound used in the trial, including natural, semi-synthetic and synthetic antibiotics. This indicates that there is significant antibiotic production in soil, which is also supported by a study which found that *Streptomyces rochei* expresses spectinomycin biosynthesis genes when grown in soil or in the rhizosphere (Anukool *et al.*, 2004). This resistance is not solely due to anthropogenic inputs of antibiotics into the environment; large numbers of bacteria resistant to gentamicin, kanamycin, streptomycin, chloramphenicol or tetracycline have been isolated from multiple Siberian permafrost sediments dating back up to three million years before present, long before antibiotics were synthesized by humans (Mindlin *et al.*, 2008). That is unsurprising, as it has been estimated that the erythromycin, streptomycin and vancomycin biosynthetic pathways emerged over 880, 610 and 240 million years ago respectively (Wright, 2007), indicating that environmental bacteria have been under selective pressure to develop resistance for hundreds of millions of years.

1.2 Streptomycin

Streptomycin is a basic aminocyclitol aminoglycoside antibiotic. It was the first aminoglycoside to be discovered and the first antibiotic approved for the treatment of tuberculosis following human trials in 1947 (Alliance, 2008). Streptomycin is no longer widely used, and aminoglycosides as a whole account for only about 3% of the total of all antibiotics produced and used (Madigan & Martinko, 2005), due to its many toxic manifestations in the peripheral and central nervous system at higher

doses and hypersensitivity reactions (Kamal *et al.*, 2008), including ototoxicity and nephrotoxicity (Mingeot-Leclercq & Tulkens, 1999). Streptomycin is used in horticulture where it is primarily employed for control of *Erwinia amylovora* which causes fire blight in apple and pear trees, though it is also used to treat other bacterial infections (McManus *et al.*, 2002).

1.2.1 Streptomycin function

The effects of streptomycin on a bacterial cell can be divided into two-stages. In the first stage the positively charged streptomycin molecules bind in an energy-independent manner to the negatively charged moieties of phospholipids, lipopolysaccharides and outer membrane proteins in gram-negative bacteria or to the phospholipids and teichoic acids in gram-positive bacteria (Taber *et al.*, 1987). This displaces Mg^{2+} and Ca^{2+} ions that link adjacent lipopolysaccharide molecules, which destabilises the outer membrane and enhances its permeability (Hancock, 1984; Martin & Beveridge, 1986). This is followed by the energy-dependent uptake of streptomycin, which requires a threshold transmembrane potential generated by a membrane-bound respiratory chain. As a result anaerobic organisms or bacteria with malfunctioning electron-transport mechanisms can display resistance to streptomycin (Vakulenko & Mobashery, 2003). In *E. coli* streptomycin is taken up by the oligopeptide transport system, though other transport mechanisms may also be involved (Kashiwagi *et al.*, 1998). Uptake by a transport system is necessary, as streptomycin molecules are too large and polar to passively diffuse through the cell membrane (Mingeot-Leclercq *et al.*, 1999). The amount of antibiotic that can enter in this manner is too low to arrest protein synthesis. However streptomycin binds to the A-site (aminoacyl-tRNA binding site) on the 16S rRNA, causing a deformation in the 16S tertiary structure that significantly reduces the rate at which peptidyl-tRNA is translocated from the A- to the P-site (peptidyl-tRNA binding site) and destabilises the binding of peptidyl-tRNA to the P-site (Karimi & Ehrenberg, 1996). This induces a 25-fold or greater reduction in proof-reading accuracy (Karimi & Ehrenberg, 1994), which results in the production of defective proteins that are incorporated in the cell membrane leading to the loss of membrane integrity (Vakulenko & Mobashery, 2003). This results in the second, killing phase during which large amounts of the antibiotic cross the compromised cell membrane, which

accumulates rapidly within the cell as it binds electrostatically to anionic groups of macromolecules or is sequestered by the degradation products of mistranslated proteins (Piepersberg & Distler, 1997; Busse *et al.*, 1992). The antibiotic then irreversibly saturates all the cell's ribosomes, arresting protein synthesis and causing cell death (Vakulenko & Mobashery, 2003).

Streptomycin is also capable of interacting with RNA sequences other than the 16S rRNA such as the group I introns (Wallace & Schroeder, 1998). Group I introns are self-splicing ribozymes that catalyze their own excision from mRNA, tRNA and rRNA precursors and are found in bacteria and some eukaryotes (Stahley & Strobel, 2006). Group I intron splicing requires the binding of an exogenous guanosine via its guanidino group to the catalytic core of the intron. Streptomycin possesses a guanidino group and thus acts as a competitive inhibitor, preventing the ribozyme's activation (Wallace & Schroeder, 1998). In addition, streptomycin appears to also inhibit the ribozyme by binding to a RNA structural motif on the group I intron similar to the one it binds to on the 16S rRNA (von Ahsen & Noller, 1993). These structures are composed of a hairpin loop that is base-paired with an adjoining RNA loop and are termed pseudoknots (Powers & Noller, 1991; Piepersberg & Distler, 1997). Streptomycin is also capable of inhibiting eukaryotic nuclear pre-mRNA splicing *in vitro*. This appears to be due to the non-specific binding of streptomycin to the mRNA, which prevents the folding of the pre-mRNA into a splicing-compatible tertiary structure (Hertweck *et al.*, 2002)

There is also some preliminary evidence that streptomycin may regulate growth and peptidoglycan formation in some streptomycete strains, as it has been detected bound to a cell wall precursor unit (Szabo *et al.*, 1989) where it activated lytic enzymes in the cell wall (Szabo *et al.*, 1990).

1.2.2 Streptomycin production

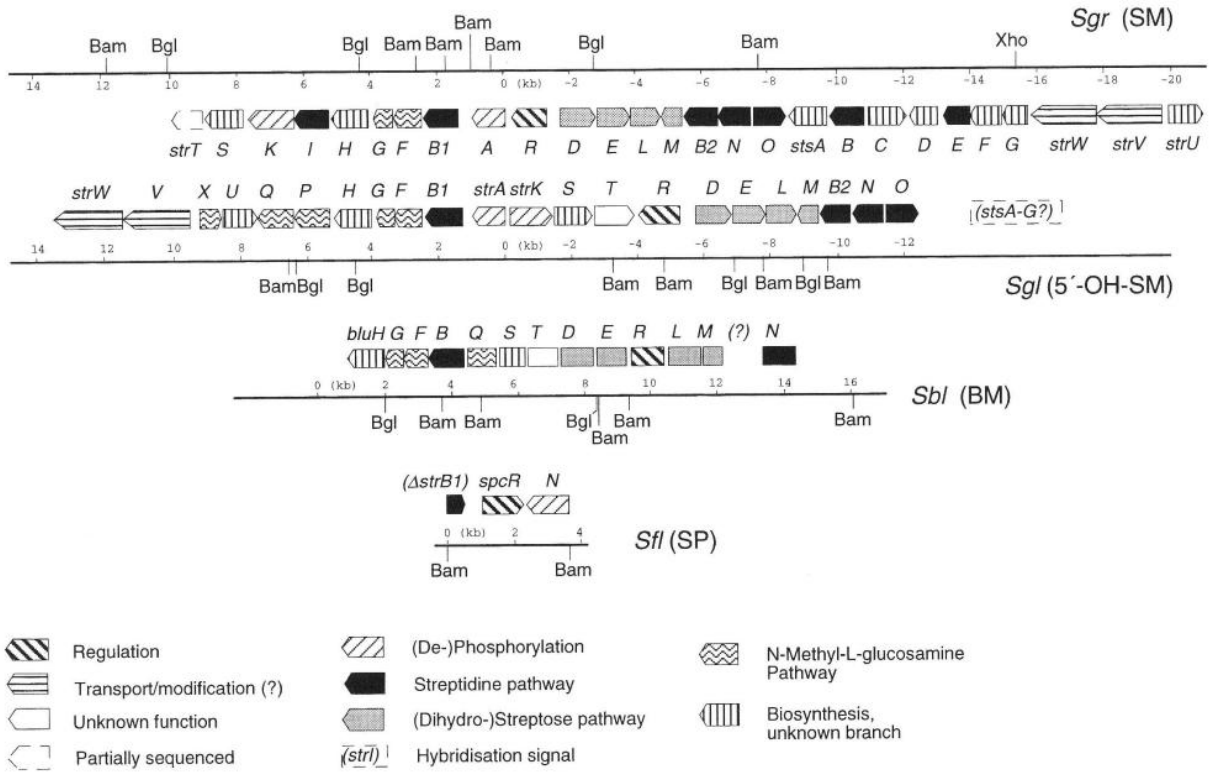


Figure 1.1. Diagram of the streptomycin (SM), hydroxy-streptomycin (5'-OH-SM), bluensomycin (BM) and biosynthetic gene clusters. After Piepersberg & Distler (1997).

The streptomycin gene cluster is composed of 27 genes and occupies a region 32.6 Kb in length (Figure 1.1) (Tomono *et al.*, 2005). The cluster genes encoding enzymes for the synthesis of the streptomycin subunits are not found in subpathway-specific operons; instead they are found in mixed operons which may indicate the need for strictly coordinated expression of the cluster genes (Piepersberg & Distler, 1997). The *Streptomyces glaucescens* hydroxy-streptomycin cluster has a number of genes that are homologous to the streptomycin gene cluster, however their gene sequence identiy ranges from 80% to less than 60% and the gene order of the two clusters differs as well (Distler *et al.*, 1992).

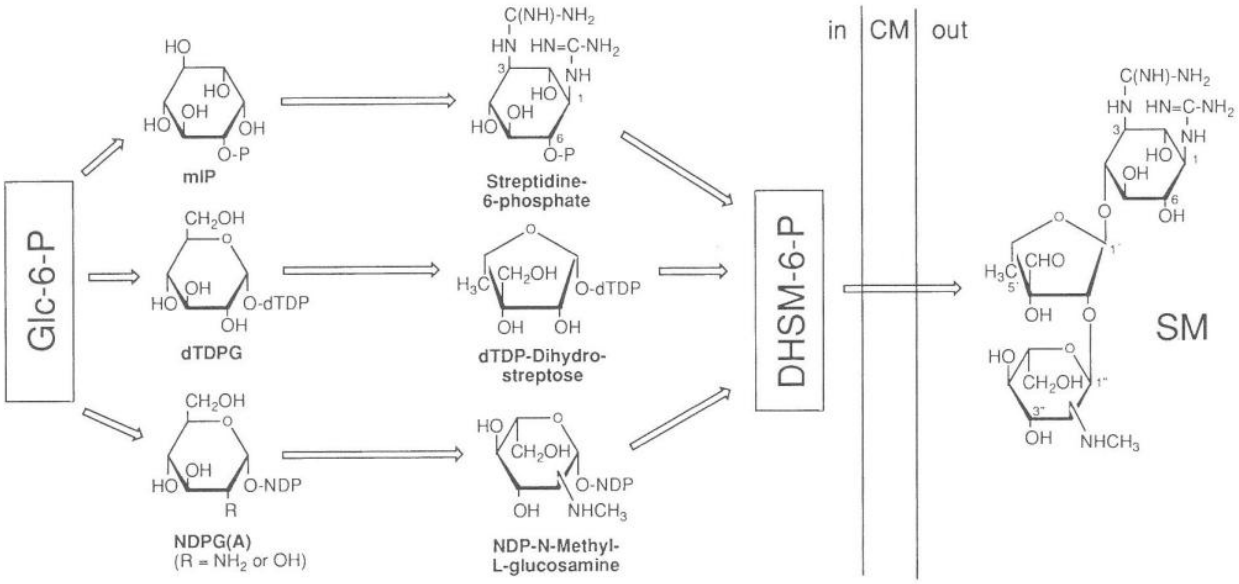


Figure 1.2. Outline of the streptomycin biosynthetic pathway. Glucose-6-phosphate (Glc-6-P) is used to synthesize myo-inosytolphosphate (mIP), deoxythymidinediphosphate-glucose (dTDPG) and nucleosidediphosphate-glucose (or glycosamine) (NDPG(A)) which are processed and condensed into dihydro-streptomycin-6-phosphate (DHSM-6-P), which is oxidized and dephosphorylated to streptomycin (SM) during or after transport through the cytoplasmic membrane (CM). After Piepersberg & Distler (1997).

Streptomycin is created by the condensation of three moieties that are all synthesized from glucose-6-phosphate (Figure 1.2). The enzymes StrO, StrI, StsC, StsE, StrB1, StsB, StsA and StrB1 or StrB2 produce streptidine-6-phosphate. StrN, StrD, StrE, StrM and StrL synthesize dTDP-Dihydrostreptose. StrN, StrQ, StrP, StrS, StsG, StrF, StrG and StrX produce NDP-N-Methyl-L-glycosamine (Flatt & Mahmud, 2007). The streptidine-6-phosphate and dTDP-Dihydrostreptose are condensed by StrH followed by NDP-N-Methyl-L-glycosamine to produce dihydro-streptomycin-6-phosphate (Flatt & Mahmud, 2007), a dehydrogenase (possibly StrU) then converts this molecule into streptomycin-6-phosphate which is then exported via a transmembrane complex formed by StrV and StrW. Once outside the cell membrane the inactive form of the antibiotic is dephosphorylated by StrK to produce streptomycin (Piepersberg & Distler, 1997).

1.2.3 Streptomycin resistance

Bacteria have been reported to develop streptomycin resistance using all the principal mechanisms except for site-of-action bypass, which is not encountered as it would require the replacement of ribosomes with an alternative protein synthesis mechanism.

The first resistance mechanism, prevention of the antibiotic's entry into the cell, requires the reduction or elimination of the electrical potential generated by a membrane-bound respiratory chain. As a result, bacteria with a deficient electron transport system can become resistant (Vakulenko & Mobashery, 2003). Reduced streptomycin uptake is mostly seen in *Pseudomonas* and other non-fermenting gram-negative bacilli, which may be due to membrane impermeabilisation though the molecular mechanisms are largely unknown (Mingeot-Leclercq *et al.*, 1999). *Pseudomonas aeruginosa* can display adaptive resistance to aminoglycosides which is not due to mutation but rather to upregulation of genes involved in anaerobic metabolism (Karlowsky *et al.*, 1997). A switch to anaerobic metabolism can provide protection against streptomycin as streptomycin can no longer enter the cell via energy-dependent uptake. It has been suggested that *Mycobacterium gordonae*, *M. szulgai*, and *M. avium* are resistant to streptomycin due to their cell envelope acting as a permeability barrier as they have no mutations in their *rpsL* gene (Honore & Cole, 1994), however there is no direct evidence supporting this hypothesis.

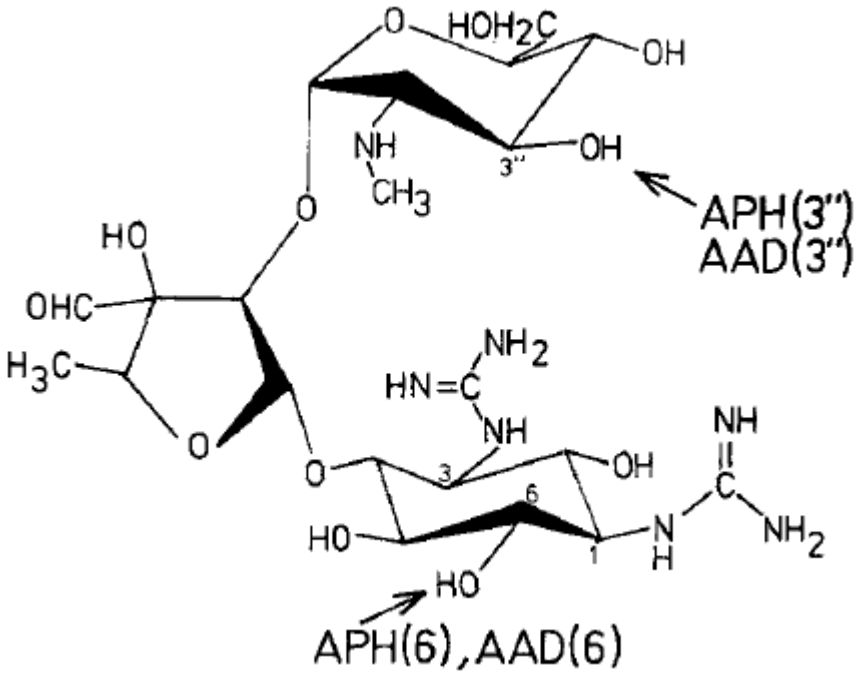


Figure 1.3. Structure of the streptomycin molecule showing the sites modified by adenyltransferase (AAD) and phosphotransferase (APH) enzymes. After Heinzel *et al.* (1988).

The second resistance mechanism, enzymatic modification of the antibiotic, can be performed by three groups of enzymes on aminoglycosides: acetyl CoA-dependent N-acetyltransferases, ATP-dependent O-adenyltransferases, and ATP-dependent O-

phosphoryltransferases (Wright & Thompson, 1999), though only the latter two have been documented to inactivate streptomycin (Figure 1.3) (Davies & Wright, 1997). The *Streptomyces griseus* streptomycin producers possess two resistance genes: *strA* (*aphD*), classified as APH(6)-Ia, which phosphorylates streptomycin on the hydroxyl group found on the 6' end of the molecule (Cundliffe, 1989), and *aphE* (APH(3")-Ia), which targets the 3" end hydroxyl group and which may have originated from *Streptomyces fradiae* (Heinzel *et al.*, 1988). The inactivated antibiotic can then be exported from the cell using the StrV/StrW ABC transporter. It is theorised that the APH resistance genes evolved from enzymes involved in antibiotic biosynthesis (Piepersberg & Distler, 1997). The streptomycin biosynthetic enzymes StrN and StsE display peptide motifs similar to those of phosphate transfer (Pissowotzki *et al.*, 1991) and may therefore share a common ancestor with StrA. The *sph* gene (APH(6)-Ib) from the hydroxystreptomycin gene cluster in *S. glaucescens* can also provide low level resistance against streptomycin due to the structural similarity of the two target molecules (Vogtli & Hutter, 1987). Genes homologous to *strA* are also present outside producers; the *str* (APH(6)-Ic) gene in the transposon Tn5 and the *strB* (APH(6)-Id) plasmid RSF1010, found in enterobacteria, may be derived from *strA* as they have a 51% amino acid similarity to the *S. griseus* and *S. glaucescens* resistance genes (Shaw *et al.*, 1993). The *strB* gene, located on plasmids such as RSF1010 next to an APH(3")-Ib streptomycin resistance gene that is confusingly also called *strA*, is widely distributed in the environment indicating that it is readily transferrable between organisms (Wright & Thompson, 1999; Sundin & Bender, 1996). *Mycobacterium fortuitum* also possesses a streptomycin resistance gene (APH(3")-Ic) that is homologous to the 3" APH genes from *S. griseus* and RSF1010, and which is assumed to have arrived in this species via HGT (Ramon-Garcia *et al.*, 2006).

Unlike phosphoryltransferases, adenyltransferases are not found in streptomycin producers. ANT(3")-I genes are very varied, displaying between 59% and 95% amino acid sequence identity to one another (Vakulenko & Mobashery, 2003), and are extremely widespread, having been found in plasmids, integrons and transposons from Gram negative (Chinault *et al.*, 1986; Levesque *et al.*, 1995; Fling *et al.*, 1985) as well as in Gram positive bacteria (Courvalin & Fiandt, 1980; Nesvera *et al.*, 1998). The ANT(3")-Ia gene was detected in more than 90% of streptomycin-

resistant clinical isolates (Shaw *et al.*, 1991). The *ant(6)-Ia* gene is also widespread, having been found in over 80% of enterococcal and staphylococcal clinical isolates tested in a European study (Ounissi *et al.*, 1990) and almost 50% of *Enterococcus faecalis* and *Enterococcus faecium* isolates in Japan (Kobayashi *et al.*, 2001). The gene *aadK*, a 6'-adenylyltransferase that provides low level streptomycin resistance present in the chromosome of *Bacillus subtilis* 168 (Noguchi *et al.*, 1993), has a 58% amino acid sequence identity to *ant(6)-Ia*, while a homologue from *Bacillus halodurans* had 41% sequence identity to *ant(6)-Ia* and even less to *aadK* (Vakulenko & Mobashery, 2003). The origin of O-adenyltransferases is uncertain, as they have no close homologues. A sequence that forms part of the ATP and Mg^{2+} binding sites is similar to that of enzymes that catalyze nucleoside-monophosphate transfer-generating pyrophosphates which suggests that ANTs may have evolved from existing metabolic enzymes such as DNA polymerases (Davies & Wright, 1997).

The third resistance mechanism, alteration of the site of antimicrobial action, can be achieved by mutation of *rrs*, encoding the 16S rRNA, *rpsL*, encoding the S12 ribosomal protein, or *rsmG*, encoding a 16S rRNA methyltransferase. Mutations that disrupt the 530 loop of the 16S rRNA in *E. coli* (Melancon *et al.*, 1988) or its homologue in other bacteria such as *Mycobacterium tuberculosis* (Meier *et al.*, 1994) or chloroplasts (Yeh *et al.*, 1994); or at position 912, part of the 900 stem region, of the *E. coli* 16S rRNA (Frattali *et al.*, 1990) or its homologue in other organisms (Honore & Cole, 1994; Gregory & Dahlberg, 2009) can generate a resistance phenotype. The 530 region is the most conserved area of the 16S rRNA (Noller, 1984); mutations at position 530 are lethal for the cell (Powers & Noller, 1990). Both the 530 and the 900 region interact with the S12 ribosomal protein (Stern *et al.*, 1988) and play a vital role in both translational accuracy control and tRNA binding (Wang *et al.*, 1999a). These two regions are located adjacent to one another in secondary structure models (Ramaswamy & Musser, 1998) and the resistance phenotype is due to these mutations disrupting the pseudoknot that streptomycin binds to (Moazed & Noller, 1987; Powers & Noller, 1991). Due to the presence of multiple *rss* copies in most bacterial species, a mutation in a single gene is unlikely to provide resistance as most of the cell's ribosomes will still be vulnerable. 16S rRNA mutations are therefore an important source of resistance in

organisms with one or two *rrs* genes, such as *Mycobacterium* species, but are unlikely to play a significant role in other bacteria (Musser, 1995).

All bacterial species however appear to have a single copy of the *rpsL* gene encoding for ribosomal protein S12 and its mutations are therefore more commonly responsible for streptomycin resistance. This highly conserved protein is part of the 30S ribosomal subunit and is located close to the codon–anticodon interaction site (Rodnina & Wintermeyer, 2001) where it is involved in determining the fidelity of protein synthesis (Yates, 1979). The most common mutation is at codon 43, where AAG changes to AGG (Lys to Arg) though an AAG to ACG (Lys to Thr) can also occur less frequently (Ramaswamy & Musser, 1998). Mutations in codon 88 such as AAG to AGG (Lys to Arg) or AAG to CAG (Lys to Gln) can also grant streptomycin resistance (Sreevatsan *et al.*, 1996). Mutations in the *rrs* (Honore *et al.*, 1995), the *rpsL* (Ito & Wittmann, 1973) or a combination of *rpsL* and *miaA* (leading to loss of a tRNA hypermodification) (Diaz *et al.*, 1986) genes can render the bacterium not only resistant but also dependant on streptomycin for its transcription. This is because proofreading becomes excessively intense and streptomycin is required to suppress the proofreading function of the ribosome. Without the increase in translational efficiency brought about by streptomycin, the dependent cells cannot manufacture proteins fast enough to survive (Diaz *et al.*, 1986). Mutations that inactivate the *rsmG* gene can provide *Streptomyces coelicolor* (Nishimura *et al.*, 2007) and *Thermus thermophilus* (Gregory *et al.*, 2009) with low level streptomycin resistance. The RsmG methyltransferase methylates base G527 of the 16S rRNA (Gregory *et al.*, 2009). As the streptomycin molecule comes into contact with the bases C526 and G527 (Carter *et al.*, 2000), that methyl group may be necessary for the efficient binding of the antibiotic.

1.3 Viomycin

Viomycin (tuberactinomycin B) is a side-chain-cyclised pentapeptide that belongs to the tuberactinomycin family of antibiotics (Figure 1.4) (Pittenauer *et al.*, 2006; Ju *et al.*, 2004; Morse *et al.*, 1997). Members of this family, such as capreomycin, are used to treat tuberculosis infections, though viomycin has never been used extensively as a therapeutic agent due to its toxicity (Yin *et al.*, 2003). It is a highly effective inhibitor of *Mycobacterium bovis* (MIC value of 0.32 mg/ml), however it

can also affect other microorganisms to a lesser extent, such as *Escherichia coli* (MIC 10 to 20), *Bacillus subtilis* (MIC 10) and *Staphylococcus aureus* (MIC 120) (Oliva *et al.*, 1998).

1.3.1 Viomycin function

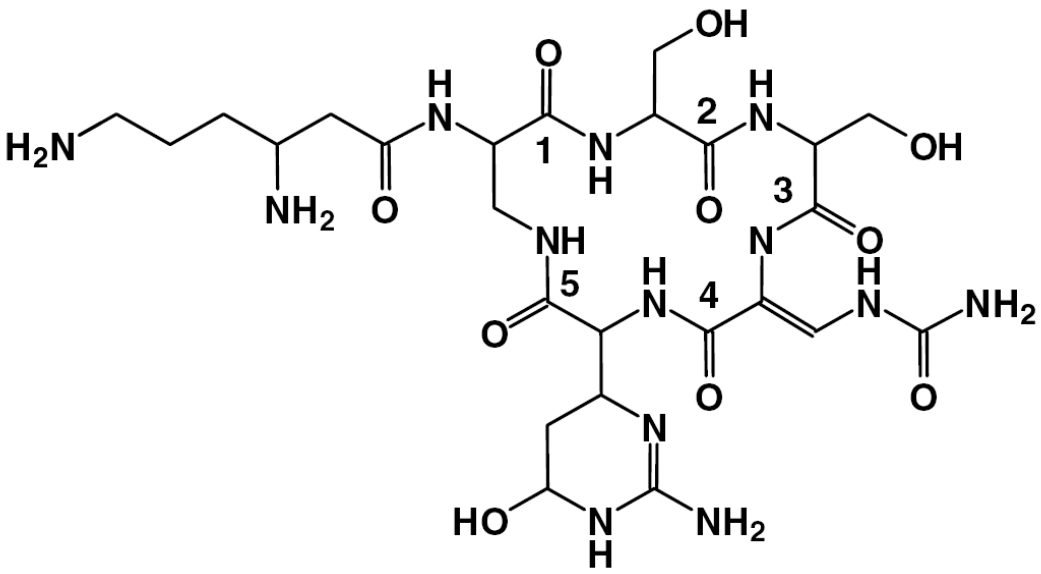


Figure 1.4. Chemical structure of viomycin. Elemental composition: $C_{25}H_{43}N_{13}O_{10}$, molecular weight: 685.33. After Pittenauer *et al.* (2006).

Tuberactinomycins inhibit protein synthesis with a mechanism to similar to that of aminoglycoside antibiotics (Schroeder *et al.*, 2000; Vos *et al.*, 2002). Viomycin achieves this by the inhibition of ribosomal translocation using two potential mechanisms: (i) the confinement of peptidyl-tRNA to the ribosomal acceptor site (A site) by interfering with its exit from the A site (Modolell & Vazquez, 1977) thus preventing the formation of the intermediate ribosomal complex (Pan *et al.*, 2007) and (ii) by blocking ribosomes in the hybrid, or ratcheted, state thus stabilising the translocational intermediate conformation of the ribosome where the tRNAs are in their hybrid A/P and P/E positions (Ermolenko *et al.*, 2007). Both mechanisms may lead to an inhibition of translocation by respectively stabilising the ground state or an otherwise transient intermediate of the pre-translocation complex (Lancaster *et al.*, 2008). Viomycin is more effective than all other studied translocation inhibitors, as it increases the affinity of tRNA to the A site by more than a 1000-fold and inhibits translocation by a factor of more than 10,000 which completely abolishes it *in vitro* (Peske *et al.*, 2004). Viomycin can additionally cause the misreading of the mRNA molecule, resulting in the incorporation of incorrect amino acids in the

resultant polypeptide which can produce an enzyme with reduced or no activity (Jerinic & Joseph, 2000; Szaflarski *et al.*, 2008).

While the inhibition of translation is viomycin's most well known and studied effect, this peptide has an effect in a number of RNA molecules in addition to rRNA. Viomycin can inhibit group I intron splicing and demonstrates 100 times greater activity than dGTP, arginine and streptomycin (Wank *et al.*, 1994). Viomycin acts as a competitive inhibitor, interacting with the G-binding site to prevent the binding of the guanidine cofactor that is necessary for group I intron splicing (Wank *et al.*, 1994). At subinhibitory concentrations viomycin induces the oligomerisation and circularization of group I introns, which may reflect an ancestral role of tuberactinomycins as mediators of RNA-RNA interactions (Wank & Schroeder, 1996). Viomycin is also capable of inhibiting the human hepatitis δ virus (HDV) ribozyme, which utilises a different cleavage mechanism than group I introns (Sharmeen *et al.*, 1988) and whose function is to process the multimeric molecules generated during rolling-circle replication of the virus (Rogers *et al.*, 1996). Viomycin demonstrates a high level of activity in this process, being capable of causing a 50% inhibition at a 50 μM concentration (Rogers *et al.*, 1996). It can also inhibit hammerhead-ribozymes (Jenne *et al.*, 2001), which self-cleave at a specific phosphodiester bond to produce 2',3' cyclic phosphate and 5' hydroxyl termini (Stage-Zimmermann & Uhlenbeck, 1998). Viomycin however does not solely act as an inhibitor; it can enhance self-cleavage of the *Neurospora crassa* VS ribozyme and stimulate a trans cleavage reaction by enhancing interactions between RNA molecules (Olive *et al.*, 1995). Viomycin is also capable of stabilising the 3EF4 and 3DEF4 subdomains of the hepatitis C virus internal ribosome entry site (IRES) (Vos *et al.*, 2002).

Viomycin is capable of associating with a diverse range of RNA molecules because it binds to conserved regions within them. Its association with a 14 bp conserved RNA sequence whose consensus is GCUGAAAGGAUCGC, which forms a pseudoknot to which the viomycin molecule binds to, has been extensively studied (Wallis *et al.*, 1997; Wank *et al.*, 1999; Wank *et al.*, 1994; Wank & Schroeder, 1996). However viomycin is capable of binding to RNA molecules that lack a

pseudoknot structure, such as the hepatitis C IRES, indicating that it has a broad target range and can bind to additional secondary RNA structures (Vos *et al.*, 2002).

1.3.2 Viomycin production

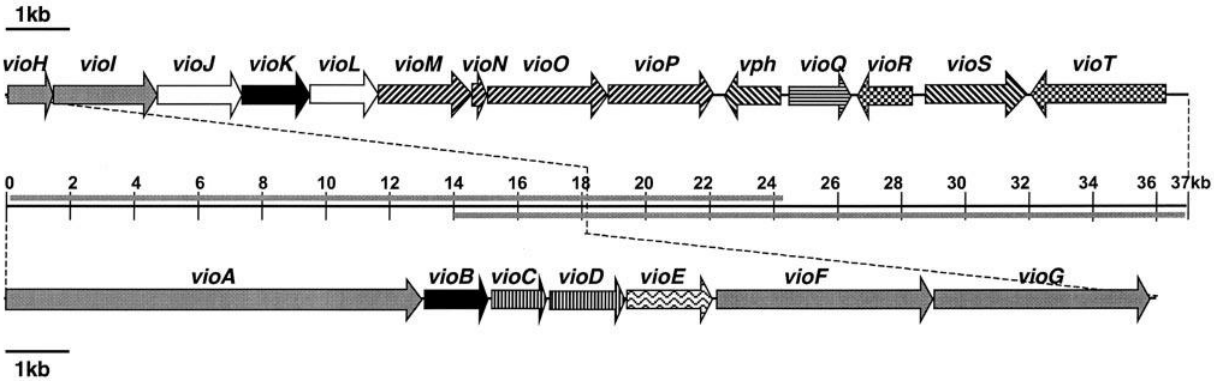


Figure 1.5. Diagram of the viomycin biosynthetic gene cluster. The centre line denotes the 37 kb encoding all ORFs involved in viomycin biosynthesis. Arrows above and below the centre line identify the direction of transcription of ORFs. Coding of ORF biosynthetic function is as follows: grey, NRPS; black, L-2,3-diaminopropionate; white, L-2,3-diaminopropionate→ β-ureidodehydroalanine; vertical lines, L-capreomycidine; horizontal lines, L-capreomycidine hydroxylation; right-slanted lines, β-lysine; left-slanted lines, resistance and activation; checkerboard pattern, regulation; waves, export. After Thomas *et al.* (2003).

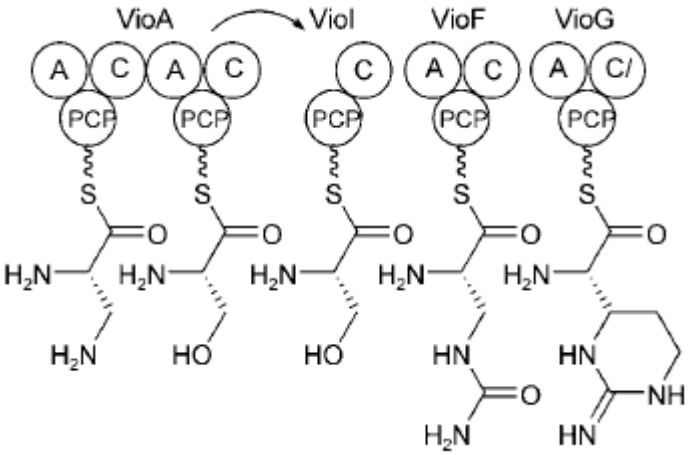


Figure 1.6. The synthesis of the central cyclic pentapeptide core of viomycin by a nonribosomal peptide synthetase composed of VioA, VioI, VioF and VioG. After Barkei *et al.* (2009).

The viomycin gene cluster is 36.3 Kb in size and contains 20 ORFs (Figure 1.5). To produce viomycin, VioB and VioK convert L-serine into L-2,3-diaminopropionate, VioC and VioD turn L-arginine to L-capreomycidine and VioP converts L-lysine to β-lysine (Thomas *et al.*, 2003). The central cyclic pentapeptide core of viomycin is assembled by a nonribosomal peptide synthetase (NRPS) consisting of VioA, VioI, VioF and VioG (Fig. 1.3) utilising two L-2,3-diaminopropionate, two L-serine and one (2S,3R)-capreomycidine molecule (Barkei *et al.*, 2009). The pentapeptide is

further modified by VioL, VioQ, VioO and VioM to generate the final product (Thomas *et al.*, 2003; Yin *et al.*, 2003; Fei *et al.*, 2007). Vph, a viomycin phosphotransferase, phosphorylates viomycin which is then exported from the cell by VioE and dephosphorylated once outside by VioS. VioR and VioT are two putative transcriptional regulators that are likely to control expression of the cluster (Thomas *et al.*, 2003), though it is not known what pathway regulates their expression.

1.3.3 Viomycin resistance

As with streptomycin, resistance to viomycin can either occur by the enzymatic modification of the antibiotic, mutations that affect the structure of the ribosome or a reduction in the cell's permeability to the antibiotic.

There is some evidence suggesting the existence of resistance via a reduction in the cell's permeability to viomycin. A locus in *Mycobacterium smegmatis* designated *vicC* may provide resistance via a change in permeability (Yamada *et al.*, 1985), but it has undergone little study.

The only enzyme known to inactivate viomycin via a structural modification is the viomycin phosphotransferase Vph that is used as a self-resistance enzyme by producers (Skinner & Cundliffe, 1980; Cundliffe, 1984). The *vph* gene has not been detected in non-producers, nor have any other enzymes that inactivate the antibiotic via alternate mechanisms such as acetylation.

Alterations in either the 30S or the 50S ribosomal subunit can provide resistance to viomycin (Yamada *et al.*, 1985; Choi *et al.*, 1979). The mutation from G to T at position 1484 of the *rrs* gene encoding the 16S rRNA in *Mycobacterium tuberculosis* (Maus *et al.*, 2005a) and mutations from G to A or T at position 1473 of the *rrs* gene in *Mycobacterium smegmatis* (Taniguchi *et al.*, 1997) can result in viomycin resistance. Mutations that inactivate the *tlyA* gene can also cause viomycin resistance in *Mycobacterium tuberculosis* (Maus *et al.*, 2005b). The *tlyA* gene encodes a 2'-O-methyltransferase that modifies nucleotide C1409 in helix 44 of 16S rRNA and nucleotide C1920 in helix 69 of 23S rRNA. Loss of these rRNA methylations confers resistance to viomycin (Johansen *et al.*, 2006); this has also been observed in *E. coli*, where an unmethylated guanine at position 745 of the 23S

rRNA, located in hairpin 35 of domain II of the 23S rRNA (Bujnicki *et al.*, 2002), results in viomycin resistance (Gustafsson & Persson, 1998).

1.4 *Streptomyces* genus

1.4.1 Streptomyces morphology and genetics

Both the streptomycin and the viomycin gene clusters are found in members of *Streptomyces*, a genus of Gram-positive, filamentous, chemoheterotrophic soil bacteria that belong to the class Actinobacteria (Madigan & Martinko, 2005). In addition to the two aforementioned antibiotics, streptomycetes synthesise some of the most economically important pharmaceutical products available today as secondary metabolites, including the antibiotics erythromycin and tetracycline, the anti-tumour agent daunorubicin, the immunosuppressant rapamycin, and the antihelminthic agent avermectin (Paradkar *et al.*, 2003). In addition to these, members of this genus synthesise thousands more secondary metabolites. It has been theorized that the large variety of secondary metabolites synthesized by this genus is due to the fact that they are non-motile and thus cannot avoid environmental stresses and because multiple metabolites may act synergistically or contingently against biological competitors (Challis & Hopwood, 2003). A screen of the Antibiotic Literature Database (ABL) discovered that *Streptomyces* strains were responsible for the biosynthesis of 32.1% of over 23,000 microbial products possessing some biological activity (Lazzarini *et al.*, 2000). The genus has been thoroughly studied both due to its importance to the pharmaceutical industry and due to its complex lifecycle (Figure 1.7). *Streptomyces* species show differentiation and grow in a manner similar to filamentous fungi. Streptomycete cells grow by forming a non-dividing, many-branched mycelium that extends mainly by tip growth (Bentley *et al.*, 2002). A colony propagates by the formation of spore chains on specialised aerial hyphae (Flardh, 2003).

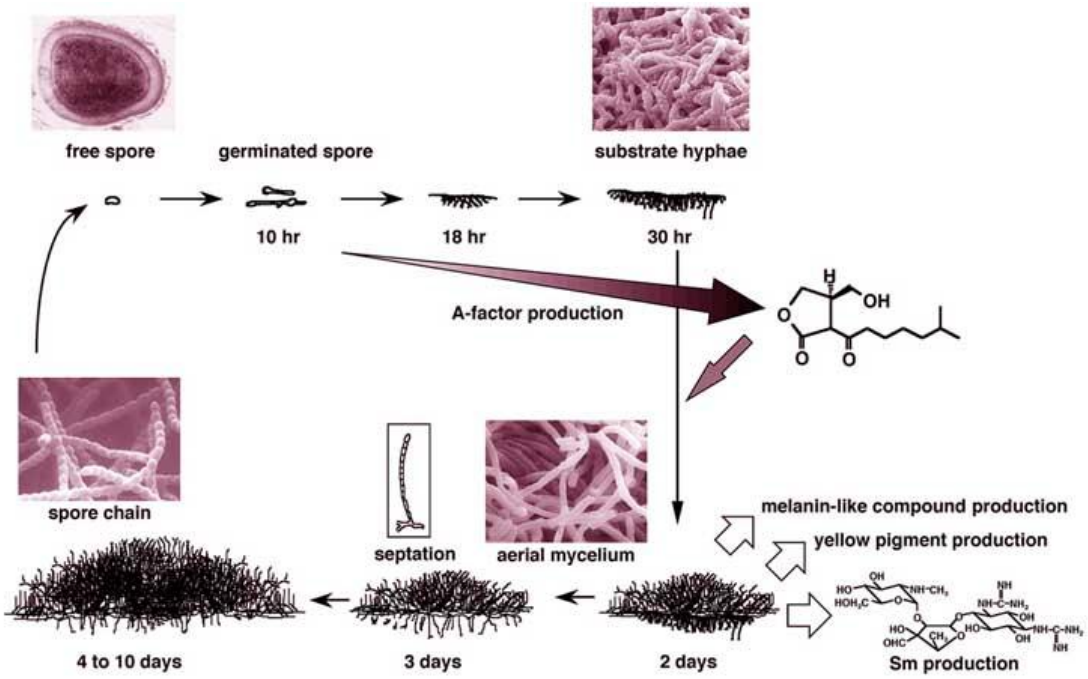


Figure 1.7. Life cycle of *S. griseus*. After (Horinouchi, 2002).

The genomes of *Streptomyces* species are also unusual among bacteria because they are linear and very large (Madigan & Martinko, 2005). Streptomycete genomes can range from 8 to 10 Mb in size and contain twice the coding capacity of *E. coli* and more open reading frames (ORFs) than the yeast *Saccharomyces cerevisiae* (Paradkar *et al.*, 2003). The streptomycete chromosome can be divided into a stable core region that contains the housekeeping genes of the cell and unstable arm regions that mainly contain genes involved in secondary metabolism (Bentley *et al.*, 2002). *Streptomyces* has one origin of replication (*oriC*) that in most species, with the exception of *Streptomyces rimosus*, is located at the centre of the linear chromosome close to the *dnaA-gyrB* region (Volff & Altenbuchner, 2000). The fact that the *oriC* is not found at the centre of the core region however indicates that the linearization of the genome did not occur opposite of *oriC*. Instead it is likely there was evolutionary pressure to have *oriC* at the centre, resulting in preferential insertions of new genes at the shorter arm of the chromosome (Chen *et al.*, 2002).

The arm regions of the *Streptomyces* genome are extremely unstable. It is estimated that the chromosomes of about 0.5% of germinating spores contain deletions of up to 25% of the genome. These deletions can therefore remove up to 2 Mb of DNA, which exceeds the size of a number of small bacterial genomes (Volff & Altenbuchner, 2000). A comparison of the *S. coelicolor* genome structure with other

Actinobacteria such as *Mycobacterium tuberculosis* or *Corynebacterium diphtheriae* reveals regions of synteny that cover the entire chromosome of *C. diphtheriae* and *M. tuberculosis*, but only the core region of *S. coelicolor* (Bentley *et al.*, 2002). This indicates that the arm regions are relatively recent additions to the chromosomes of *Streptomyces* species and can therefore be deleted without a loss of viability under laboratory conditions. The addition of the arm regions to the chromosome presumably took place because soil is very heterogeneous, necessitating the presence of a considerable number of genes to allow the organism to deal with the huge range of stresses and opportunities that it might encounter (Hopwood, 2003). These include a large number of secondary metabolic pathways which enable the bacteria to adapt to rapidly changing conditions and to intense competition. *S. coelicolor* for example is estimated to have 7825 genes, which is about twice the number of genes in the *M. tuberculosis* genome. Of these genes 12.3% are predicted to have a regulatory function, including 45 σ-factors and many two-component regulatory systems that can detect changes in the cell's environment, 7.8% are predicted to have a transport function, many of them belonging to the ABC transporter type, and 10.5% are thought to encode for secreted proteins (Bentley *et al.*, 2002). The chromosomes of *S. griseus* NBRC 13350 and *Streptomyces avermitilis* MA-4680 contain 7138 and 7583 protein-coding genes respectively (Ohnishi *et al.*, 2008), indicating that large gene numbers are a universal feature of streptomycetes.

1.4.2 Streptomycete ecology

Streptomyces species are found worldwide and their principal habitat is soil (McNeil & Brown, 1994). Streptomycetes favour soil that is alkaline or neutral and is well-drained (Madigan & Martinko, 2005). Similarly to fungi, streptomycetes exist in soil predominantly as spores, which are not evenly distributed in soil but cluster around areas of past or present growth (Mayfield *et al.*, 1972). Germination of spores only occurs when exogenous nutrients such as root fragments and dead fungal hyphae are available, making streptomycete growth sites extremely discontinuous in space and time (Goodfellow & Williams, 1983). Such discontinuity is unsurprising as soil is a very heterogeneous habitat characterised by not just nutrient availability variations but also water, and oxygen concentrations, as well as pH and the size of pores available for microbial habitation (Mummey *et al.*, 2006). The effect of oxygen availability can be seen in *Streptomyces* species by the fact that they only grow on

the peripheral areas of soil crumbs and their mycelia do not colonize the core of soil crumbs, as they cannot grow under anaerobic conditions (Mayfield *et al.*, 1972). In addition, the distribution of soil microorganisms is often spatially structured over distances of tens to hundreds of meters due to large-scale landscape gradients, such as soil carbon gradients and cultivation practices. On the scale of centimetres to meters, patterns are determined primarily by plant growth, and vary with plant size, growth form and spacing (Ettema & Wardle, 2002). This spatial separation not only promotes microhabitat diversity but also enables spatial separation of potentially competing soil organisms, which is one of the reasons soil has such a high species richness despite the low niche specialisation of constituent species (Ettema & Wardle, 2002).

Streptomycetes in soil play a very important role in the degradation of relatively complex, recalcitrant polymers occurring naturally in plant litter and soil such as lignocelluloses, which are the dominant kind of polymer present. However they can also degrade other polymers occurring in soil and litter including hemicelluloses, pectin, keratin, and chitin (McCarthy & Williams, 1991; Metcalfe *et al.*, 2002; Goodfellow & Williams, 1983). Streptomycetes are also present in the rhizosphere, which is rich in nutrient sources due to root exudates, secretions, mucigel/mucilage, lysates and sloughed root cells (Meharg & Killham, 1988). Some species, such as *S. scabies*, are pathogens that produce phytotoxins such as thaxtomin which allow their mycelia to penetrate plant roots (Loria *et al.*, 2003). However others play a number of other roles, such as that of biocontrol agents, promoters of the formation of symbioses between plant roots and microbes, and modulators of plant defences (Schrey & Tarkka, 2008).

Streptomyces species can be isolated from both fresh (Jiang & Xu, 1996) and sea water (Pimentel-Elardo *et al.*, 2009). Furthermore, streptomycete growth has been documented underwater (Aumen, 1980) and that of other actinomycetes in oceanic sediments (Mincer *et al.*, 2002), indicating that the isolates obtained from aquatic environments are not necessarily spores that have been washed out of soil and into fresh or saltwater habitats by the rain.

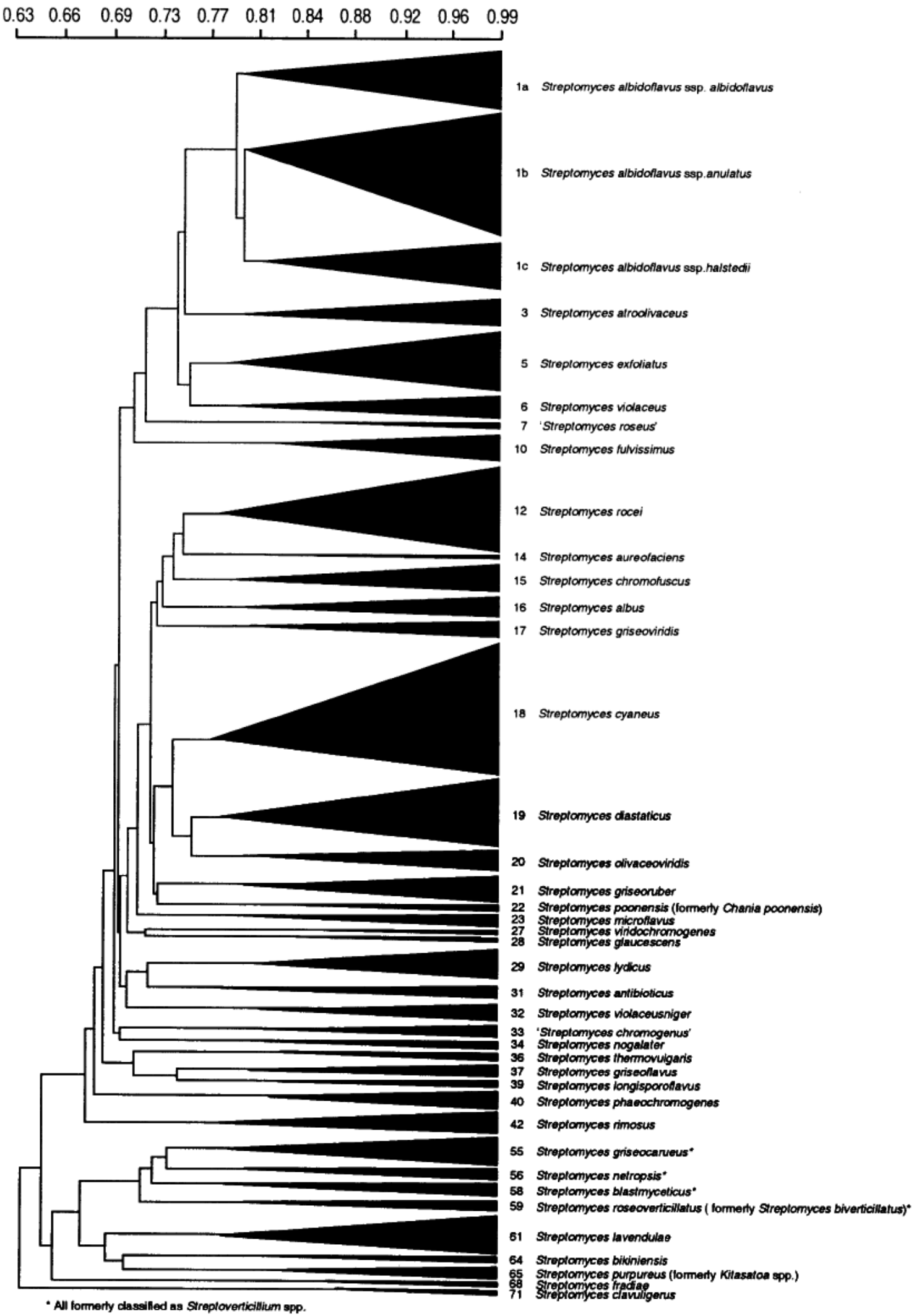
Some streptomycetes have succeeded in exploiting a completely different niche and become human and animal pathogens. *S. somaliensis*, which is a paraphyletic group,

can cause actinomycetoma, a localised chronic, destructive and progressive infection of skin, subcutaneous tissue and eventually bone which is common in certain tropical and subtropical regions (Trujillo & Goodfellow, 2003).

1.4.3 Streptomycete phylogeny

Streptomyces is the largest bacterial genus, containing 606 species and subspecies (DSMZ, 2009). A review by Anderson and Wellington (Anderson & Wellington, 2001) examines the history of *Streptomyces* classification and a number of large-scale studies have been undertaken to study the structure of this genus. One, utilising a large number of phenetic characters, included 394 *Streptomyces* strains and representatives from 14 related genera, divided the genus into 19 major and 40 minor clusters, with 18 strains recovered as single member clusters (Figure 1.8) (Williams *et al.*, 1983). The ribosomal AT-I30 protein was used to classify 81 *Streptomyces* species and separated them into six species clusters which correlated well with the 16S rRNA data but less so with the numerical phenetic survey (Ochi, 1995). A study using the *rpoB* gene sequence also produced a phylogenetic tree that was similar to that of 16S rDNA (Kim *et al.*, 2004). The α-hypervariable region of the 16S rRNA has been used to classify *Streptomyces* and 60 of the 89 strains examined were segregated into seven clusters (Kataoka *et al.*, 1997). Another large-scale screening of 463 *Streptomyces* and *Kitasatospora* type strains using 16S-ITS RFLP fingerprinting revealed 59 clusters which correlated with the 16S classification but provided greater phylogenetic resolution (Lanoot *et al.*, 2005). There have also been studies on the phylogeny of specific streptomycete clades including *S. griseus* (Guo *et al.*, 2008; Rong & Huang, 2009), *Streptomyces violaceusniger* (Goodfellow *et al.*, 2007), *Streptomyces albidoflavus* (Hain *et al.*, 1997) and *Streptomyces avermitilis* (Kim & Goodfellow, 2002).

Figure 1.8. A simplified dendrogram depicting the taxonomic clusters within the *Streptomyces* genus produced by (Williams *et al.*, 1983) using numerical phenetic taxonomy. After Anderson *et al.* (2001).



* All formerly classified as *Streptoverticillium* spp.

The genus *Streptomyces* along with the genera *Kitasatospora* and *Streptacidiphilus* form the family Streptomycetaceae (Zhi *et al.*, 2009), which is the sole member of the suborder Streptomycineae (Stackebrandt *et al.*, 1997). Members of the Streptomycetaceae can be differentiated from other bacteria by the presence of L-DAP in their cell walls, which can be detected along with glycine in whole-cell hydrolysates (Dietz, 1986). A 16S rRNA phylogenetic tree presents the Streptomycineae as a sister clade to the Pseudonocardineae and Corynebacterinae (Zhi *et al.*, 2009), though the bootstrap values of these nodes are too low to draw any firm conclusions concerning the interrelationship of these suborders. This is supported by the fact that in a previous 16S rRNA tree the Streptomycineae instead formed a sister clade to the Corynebacterinae, Micrococcineae, Actinomycineae, Propionibacterineae, Streptosporangineae and Glycomycineae (Stackebrandt *et al.*, 1997). The Streptomycineae along with 12 other subclades form the order Actinomycetales which belongs to the class and phylum Actinobacteria, as do the orders Bifidobacteriales, Acidimicrobiales, Coriobacteriales and Rubrobacterales (Zhi *et al.*, 2009). The phylum Actinobacteria is one of the largest bacterial taxonomic units and contains Gram-positive bacteria with a high GC content in their DNA, ranging from 51% to more than 70%; the only exception being the obligate pathogen *Tropheryma whipplei* which has a GC content of less than 50% (Ventura *et al.*, 2007). Members of the phylum have four documented unique indels (inserts or deletions) that serve to both distinguish its members from other bacteria and reinforce the findings from 16S rRNA trees that Actinobacteria are a monophyletic group. These consist of a 90 to 100 nucleotide insertion in the 23S rRNA, a two amino acid residue deletion in cytochrome-c oxidase subunit 1, a four amino acid residue insert in CTP synthetase and a five amino acid residue insert in glutamyl-tRNA synthetase (Gao & Gupta, 2005). In addition 233 proteins unique to actinobacteria have been identified; however the fact that they are not present in every species limits their taxonomic usefulness (Gao *et al.*, 2006). A 16S rRNA tree suggests that Actinobacteria are a sister clade to the low GC Gram positive bacteria, the spirochetes and the proteobacteria (Stackebrandt *et al.*, 1997), however phylogenetic analysis based on gene order of sequenced genomes suggests that Actinobacteria are a sister clade to the low GC Gram positive bacteria instead (Kunisawa, 2003).

1.5 Aims and objectives

The hypothesis of this project is that non-antibiotic producers who are niche competitors with producers will need to possess resistance to the antibiotic and will therefore have acquired resistance genes.

Aims:

1. To establish the phylogenetic relationship between antibiotic producing and resistant strains in *S. griseus*.
2. To determine if evidence of coevolution between closely related producers and non-producers exists.
3. To perform a phylogenetic study to elucidate the ancestry of the identified resistance genes and discover information on the evolution of resistance genes not linked to antibiotic production in soil.

Objectives:

1. To define the structure of streptomycin and viomycin gene clusters within a range of *Streptomyces* strain isolates.
2. To determine the phylogenetic distribution of the streptomycin and viomycin gene clusters and resistance genes.
3. To explore the diversity of streptomycin and viomycin resistance genes in soil via culture-independent methods.
4. To elucidate the mechanisms for regulation of the streptomycin resistance gene and cluster in a diverse range of strains and establish how the *strA* resistance gene has achieved constitutive regulation in some strains in contrast to that exemplified in *S. griseus* antibiotic producers where it is regulated by the A-factor cascade.

2 Materials and Methods

2.1 Media and Buffers

All media and buffers were made with distilled water, with the exception of Mannitol-Soya Agar for which tap water was used, and autoclaved at 121°C for 15 min (Table 2.1).

Table 2.1. Media and buffers used in this study.

Media/Buffers	Constituents
Luria Broth (LB)	Peptone 10 g/l (MERCK 1.07213) Yeast Extract 5 g/l (MERCK 1.03753) NaCl 10 g/l Adjusted to pH 7.0
LB Agar	Peptone 10 g/l (MERCK 1.07213) Yeast Extract 5 g/l (MERCK 1.03753) NaCl 10 g/l Bacto Agar 15 g/l (BD 214010)
Nutrient Broth	Nutrient Broth 8 g/l (Difco 0003-17)
Nutrient Agar	Nutrient Agar 8 g/l (Difco 0003-17) Bacto Agar 15 g/l (BD 214010)
Oatmeal Agar	Oatmeal 20 g/l Yeast Extract 5 g/l (MERCK 1.03753) Bacto agar 15 g/l (BD 214010)
RASS Agar	L-arginine 0.1 g/l K_2HPO_4 1.0 g/l NaCl 1 g/l $MgSO_4 \cdot 7H_2O$ 0.5 g/l 1% $Fe_2(SO_4)_3 \cdot 6H_2O$ 1 ml/l 0.1% $CuSO_4 \cdot 6H_2O$ 1 ml/l 0.1% $ZnSO_4 \cdot 7H_2O$ 1 ml/l 0.1% $MnSO_4 \cdot 4H_2O$ 1 ml/l Soluble Starch 12.5 g/l Bacto Agar 15 g/l

Mannitol-Soya Agar	Mannitol 20 g/l Soya flour 20 g/l Bacto Agar 15 g/l (BD 214010)
Tryptone Soya Broth (TSB)	Tryptone Soya Broth powder 30 g/l (Oxoid CM0129)
Yeast Extract-Malt Extract Medium (YEME)	Difco yeast extract 3 g/l Difco Bacto-peptone 5 g/l Oxoid malt extract 3 g/l Glucose 10 g/l Sucrose 340 g/l $MgCl_2.6H_2O$ (2.5 M) 2 ml/l (added after autoclaving) Glycine (20%) 25 ml/l (added after autoclaving)
2 M Tris-HCl (pH 8) buffer	Tris(hydroxymethyl)methylamine 403.8 g/l (BDH 103156X) Adjusted to pH 8.0 with HCl
Unbuffered 3 M Sodium Acetate	Sodium acetate 408.24 g/l
TE	10 mM TrisHCl pH 8.0 1.0 mM EDTA pH 8.0
$T_{20}E_{50}$	20 mM Tris-HCl pH 8.0 50 mM EDTA pH 8.0
$T_{10}E_{10}$	10 mM Tris-HCl pH 8.0 10 mM EDTA pH 8.0
$TE_{25}S$	25 mM Tris-HCl pH 8.0 25 mM EDTA pH 8.0 0.3 M Sucrose
BDS	10 mM Tris-HCl pH 9.0 500 mM EDTA pH 8.0 1% N-Lauroylsarcosine
2x Kirby Mix	N-Lauroylsarcosine 20 g/l (Sigma L-9150) Sodium 4-Aminosalicylate 120 g/l (Sigma A3505) 2 M Tris-HCl (pH 8) 50 ml/l

	Phenol Solution 60 ml/l (Sigma P4557)
Phenol:Chloroform	Phenol:Chloroform:Isoamyl alcohol (25:24:1) (Sigma P2069)
50 x TAE	Tris base 242 g/l Glacial acetic acid 57.1 g/l 0.5 M EDTA (pH 8.0) 100 ml/l
Culture suspension buffer	0.4 M NaCl 50 mM $MgSO_4$ 50 mM PIPES-NaOH (pH 7.0) 1% Porcine skin gelatine (type A)

2.2 Growth, Storage and DNA acquisition from *Streptomyces* strains

2.2.1 Preparation and storage of spore suspensions

Streptomyces spores were streaked across either Oatmeal Agar or Mannitol-Soya Agar plates, which were then incubated at 30°C for 5 days. 5 ml of 10% glycerol (v/v) were added to each plate and the spores suspended in it using sterile inoculation loops. The solution was filtered through autoclaved cotton using a syringe to eliminate mycelia, transferred to cryovials and stored at -20°C. A 10^{-4}, 10^{-6}, 10^{-8} and 10^{-9} dilution series was generated from the spore suspension. 10 µl of each dilution were spread onto Mannitol-Soya Agar plates that were incubated at 30°C for 5 days. The plates were examined for possible contaminants and a single colony was picked whose spores were used to inoculate two new Mannitol-Soya Agar plates. They were left to grow at 30°C for 5 days and 10% glycerol spore suspensions were prepared from them.

2.2.2 Determination of antibiotic resistance level

10 µl of spore suspension of the isolates used were streaked across nutrient agar plates containing 0, 10 and 50 µg/ml of streptomycin or viomycin. They plates were incubated at 30°C for 48 h and examined for growth. If there was no, or almost no growth, in relation to the plate with no antibiotic the strain was classified as being sensitive. If there was a greater amount of growth the strain was classified as being resistant to that level of antibiotic.

2.2.3 Determination of phenotype

10 µl of spore suspensions from the isolates used were streaked across RASS (Reduced Arginine-Soluble Starch) Agar plates as well as from a sample of *S. griseus* DSM 40236^{T}(type strain). The plates were incubated at 30°C for 5 days and their colony morphology and spore colour were compared to that of the *S. griseus* type strain.

2.2.4 Growth and extraction of DNA

10 µl of spore suspension were used to inoculate 40 ml of sterile liquid broth (TSB) in baffled 250 ml Erlenmeyer flasks. The flasks were incubated in orbital shakers at 30°C and 150 rpm for 72 hours. Cells were harvested by transferring the culture into opaque 40 ml Oak Ridge tubes and centrifuging them at 5000 RPM for 15 min at 4°C. The pellet was resuspended in 10 ml 10 µM EDTA (pH 8.0) to wash away extracellular material. The cells were pelleted again by centrifugation at 5000 RPM for 15 min at 4°C and stored at -20°C until needed. The pellets were thawed, 3 ml $TE_{25}S$ buffer and 100 µl lysozyme (100 ng/ml) solution was added and they were then incubated at 37ºC for 60 min in a rotating incubator. 4 ml of 2 x Kirby mix was added and the mixture was gently agitated. 6 ml of phenol/chloroform/isoamyl alcohol was added and the tubes were shaken until there were no longer two phases visible. The resulting emulsion was centrifuged at 3000 RPM for 10 min at 4°C. The upper aqueous phase was transferred to a 15 ml Light Phase Lock Gel tube (5 Prime, 2302840) to which 3 ml of phenol/chloroform/isoamyl alcohol and 600 µl 3M unbuffered sodium acetate was added. The mixture was vortexed for 1 min and centrifuged at 3500 RPM for 20 min at 4°C. The layer above the phase lock gel was transferred to a clear Oak Ridge tube to which 0.6 volumes of isopropanol was added and the mixture was gently agitated until it had become homogeneous. The tubes were centrifuged at 13000 RPM for 30 min at 4°C and 5 ml of 70% ethanol was added to the pellet. The tubes were agitated until the pellets detached and were then centrifuged at 13000 RPM for 10 min at 4°C. The pellets were allowed to dry at room temperature for 30 min and 500 µl of TE buffer containing 10 µg/ml RNase A was added. The pellets were left to dissolve overnight at 4°C and the resultant DNA solution's concentration was determined using a Nanodrop spectrophotometer (ND-

1000, Nanodrop Technologies) and stored at -20°C. DNA dilutions with a 100 µg/ml concentration were used in downstream applications.

2.2.5 Agarose Gel Electrophoresis

Chromosomal DNA, plasmid DNA and PCR products were analysed by gel electrophoresis. The agarose gels were 1.0% (w/v) in 1 x TAE containing 0.5 µg/ml ethidium bromide. Electrophoresis was performed at 120 volts for 150 min. Gels were visualised on a short wavelength UV transilluminator (UVP Inc.) and photographed (Gene Flash).

2.3 Gene amplification and sequencing

2.3.1 Polymerase Chain Reaction (PCR)

The PCR primers used in this study and their annealing temperatures are listed in the relevant results chapters. Temperature gradient PCRs with an annealing temperature range between 50°C and 65°C were conducted to determine the optimal annealing temperature for each primer set. PCR amplifications were performed in 50 µl reactions and each reaction consisted of 25 µl PCR Master Mix (Promega), 2.5 µl DMSO, 1.0 µl Forward Primer (100 pM), 1.0 µl Reverse Primer (100 pM), primer, 5 µl template DNA, 15.5 µl nuclease-free water. PCR amplification conditions were identical for all primer pairs except for the annealing temperature. The amplifications were carried out as follows; 95°C for 5 min, then 35 cycles of 95°C for 60 sec, TA for 45 sec and 72°C for 90 sec followed by a final extension step of 72°C for 10 min. PCR products were visualised on a 1% agarose gel.

2.3.2 Purification of PCR products

PCR products underwent gel electrophoresis and DNA bands were excised and purified using Qiaquick gel extraction kit (QIAGEN) as per manufacturer's instructions with the following modifications: After the removal of the wash buffer, the filters were left to dry for 5 min in a 50°C oven to remove residual ethanol. The DNA-containing eluate was dialysed by pipetting it on a 25 mm nitrocellulose filter (VSWP02500, Millipore) with 0.025 µm pores floating on a Petri dish filled with sterile water for 20 min to remove salts that might interfere with a sequencing reaction and a Nanodrop spectrophotometer (ND-1000, Nanodrop Technologies) was then used to determine the DNA concentration.

2.3.3 Cloning of PCR products

If the sequencing reaction indicated the presence of multiple PCR products they were cloned using the PGEM T-easy vector system and plasmid DNA extracted using the QIAGEN Mini-prep kit as per the manufacturer's instructions to isolate single amplicons for sequencing.

2.3.4 DNA Sequencing and Analysis

All PCR products were sequenced with both the forward and the reverse primers to guard against sequencing errors. pGEM-T plasmids were sequenced with both the SP6 and the T7 primers. The sequencing was done in-house by Molecular Biology Services, University of Warwick. The sequences were examined in Sequence Scanner version 1.0 (Applied Biosystems) to look for possible errors or overlapping peaks that were manually corrected. The two reads for each amplicon were aligned in Seqman (DNASTAR), checked for contradictions between the two reads and the two ends of the sequence were trimmed. All the sequences of a gene were added in a FASTA file and imported into Jalview (www.jalview.org) where they were aligned using the MUSCLE algorithm. The ends of the sequences were trimmed to ensure they were all of equal length and all single nucleotide polymorphisms, insertions or deletions present were confirmed to be genuine by re-examining the original sequencing reads. Phylip version 3.67 (http://evolution.genetics.washington.edu/phylip.html) was used for the statistical analysis of the sequences and for the construction of phylogenetic trees. The trees were rooted using the nearest non-streptomycete homologue to a gene. A distance matrix was generated in DNADIST using the Kimura-2-parameter model of sequence evolution and was used in NEIGHBOR to make a neighbour-joining tree. For maximum likelihood trees, DNAMLK was used. Confidence of the groupings were calculated in SEQBOOT, which was used to generate 10000 trees that whose consensus tree was generated using CONSENSE. Only bootstrap values higher than 50 were included on the nodes. The correct reading frame was determined using the ExPASy Proteomics server (www.expasy.org) and the sequences were translated using BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). For amino acid sequences the distance matrix was calculated in PRODIST (using the Jones-Taylor-

Thornton model) with the remaining steps the same as for trees based on nucleotide sequence.

2.4 Fosmid library construction and screening

2.4.1 Isolation of high molecular weight *Streptomyces* DNA

10 µl of spore suspension were inoculated into flasks containing 50 ml of YEME in baffled 250 ml Erlenmeyer flasks. The flasks were incubated in orbital shakers for 3 days at 30°C in a shaking incubator. The liquid cultures were transferred into opaque 50 ml Oak Ridge tubes, taking care to remove most of the mycelia, centrifuged for 20 min at 5000 RPM at 4˚C and the supernatant was decanted. The pellets were resuspended in 10 ml of 10 mM EDTA (pH 8.0) and the tubes centrifuged for 20 min at 5000 RPM at 4˚C and stored at -20˚C until needed. The cell pellets were resuspended into 1% PFGE agarose, using 1 g per 0.5 g of cells. The agarose was poured into 100 µl Plug Moulds (Biorad) and left to set. The plugs were removed and incubated for 2 h at 37°C in $TE_{25}S$ buffer containing 2mg/ml of lysozyme. The plugs were incubated for 16 h at 50°C in BDS containing 1 mg/ml proteinase K. They were washed for 1 h at 4°C in $T_{20}E_{50}$ and then washed twice for 1 h in $T_{10}E_{10}$ containing 1mM PMSF to inactivate proteinase K. They were washed again for 1 h at 4°C in $T_{20}E_{50}$ and then incubated overnight at 4°C in $T_{20}E_{50}$ containing 1% blocking reagent. The plugs were stored at 4°C in $T_{20}E_{50}$ buffer containing 1% blocking reagent until needed.

2.4.2 Ligation into vector and transformation

The plugs were loaded on a PFGE gel (1% Certified Low Melting Point agarose (Bio-Rad) in 0.5 TBE). The PFGE settings were 0.1s to 40s ramp pulse time at 6V/cm for 16 hours at 14°C in 0.5x TBE. A gel slice containing DNA fragments of approximately 40 kb in size, determined by the position of the Fosmid Control DNA band from the EpiFOS Fosmid Library Production Kit (Epicentre), was excised. The DNA was extracted using the GELase Agarose Gel-Digesting Preparation (Epicentre) as per the manufacturer's instructions. The DNA was then pipetted 30 times in order to fragment it and it underwent PFGE again under the same conditions. The DNA in the 40 kb region was cut out and extracted using GELase. The End-It DNA End-Repair Kit (Epicentre) was used to create blunt ends on all the

DNA fragments. The DNA was then used with the EpiFOS Fosmid Library Production Kit (Epicentre) to generate a fosmid library as per the manufacturer's instructions.

2.4.3 Screening of library and sequencing

A Falcon 96 well tissue culture plate (Becton Dickinson Labware) was loaded with 200 µl LB broth containing 12.5 mg/ml chloramphenicol in each well with a multichannel pipette. Using a pipette tip, a colony of the transformed *E. coli* cells was touched and then the tip was shaken in the LB broth of a well to inoculate it. The process was repeated for each well, then the sides of the plate were taped shut to prevent desiccation and it was loaded on a 37°C shaking incubator (150 RPM) overnight. Using a multichannel pipette, 10 µl from each of 48 wells were transferred into a reservoir. The mixed culture was transferred into a 500 µl Eppendorf tube and centrifuged at 13000 RPM for 5 min. The supernatant was discarded and the pellet resuspended into 30 µl of sterile water. A PCR (Section 2.3.1) using 5 µl of the suspension were used to check for genes belonging to the streptomycin cluster. The genes used to screen the library were *strA*, *strT* and *strW*. The half-plates that got a positive result had identical PCRs performed for each individual well to identify the transformant containing the fosmid of interest. The transformant was then grown in LB broth containing 12.5 mg/ml chloramphenicol and 2 µl/ml of CopyControl Fosmid Autoinduction Solution (Epicentre). The fosmid was extracted using the FosmidMAX DNA Purification Kit (Epicentre).

Two fosmids underwent shogun sequencing at the Wellcome Trust Sanger Institute. The third fosmid was partially sequenced using primer walking by the Molecular Biology Services in the University of Warwick.

2.4.4 Creation and screening of orgy prep

140mm Petri Dishes (Fisher) with LB agar containing 12.5 mg/ml chloramphenicol and 2 µl/ml of CopyControl Fosmid Autoinduction Solution (Epicentre) were prepared and stored at 4°C until use. A Falcon 96 well tissue culture plate (Becton Dickinson Labware) containing transformed *E. coli* was thawed and a 48 pin multi-blot replicator (V&P Scientific, inc) was used to transfer the 96 cultures onto the 14 cm plate. Between culture transfers the replicator was immersed in ethanol, flamed

and allowed to cool to prevent contamination. The LB plate was incubated overnight at 37°C and the colonies were resuspended into 5 ml of sterile water using a sterile spreader. The suspension was pipetted into a 15 ml Greiner centrifuge tube (Sigma-Aldritch) and centrifuged for 20 min at 4000 RPM at room temperature. The supernatant was discarded and the tube stored at -20°C until needed.

The pellet was thawed and sterile water added until the volume on the Greiner tube reached 6 ml. The pellet was resuspended and 3 ml of that was used for fosmid extraction using the FosmidMAX DNA Purification Kit (Epicentre). The protocol for extraction from a 1.5 ml sample was followed with the exception that all the amounts of reagent used were doubled. A PCR (Section 2.3.1) using 5 µl of the extracted fosmid DNA was used to determine whether genes of interest were present. Any bands had their DNA extracted (Section 2.3.2) and sequenced (Section 2.3.4). If the sequencing reaction indicated the presence of multiple products then the amplicons were cloned (Section 2.3.3) and the inserts of 5 plasmids were sequenced.

2.5 Antibiotic production assay

10 µl of spore suspensions were used to inoculate 40 ml of sterile liquid broth (TSB) in baffled 250 ml Erlenmeyer flasks. The flasks were incubated in orbital shakers at 30°C and 150 rpm. 1 ml was taken from each culture at 24, 48 and 72 hours, centrifuged and the supernatant stored at -20°C. 25 µl of supernatant were pipetted on 6 mm diameter Whatman paper discs and air dried for 20 min. The discs were placed on LB agar plates inoculated with an overnight culture of *B. subtillis* DSM 347. The plates were kept at 4°C for one hour then incubated overnight at 37°C and examined for zones of clearance. This assay was repeated three times for each isolate.

2.6 Gene detection from soil samples

2.6.1 Estimation of water content

Two glass beakers per sample were incubated at a 105°C oven for 40 min, removed with forceps and immediately weighed. Approximately 10 g of soil was added to two beakers and they were then weighted. The beakers with the soil were then incubated in the 105°C oven for 16 h. The lower half of the desiccator with the silica

crystals was placed in the oven until the crystals turned from pink to dark blue (approximately 1 h). The samples were placed in the desiccator using forceps and left to cool for 1 h. The beakers were weighted and the dry weight was subtracted from the wet weight to estimate the water content of the soil.

2.6.2 Preparation of soil microcosms

Approximately 10 g of soil was placed in a 50 ml Greiner centrifuge tube (Sigma-Aldritch) and enough water was added to bring the soil's water content to 16%. 1% of the soil's weight in chitin (powdered α chitin from crab shells) and 1% in soluble starch (Analar) was added. The tube was shaken until any clumps of soil had been broken up and the water, starch and chitin were mixed homogeneously. The tubes were placed in a 30°C incubator with their lids somewhat loose to allow the exchange of air and left to grow for 7 days. A bowl of water was placed in the bottom of the incubator to reduce soil desiccation.

2.6.3 Preparation of seeded soil microcosms

Approximately 10 g of soil was placed in a glass universal bottle and autoclaved twice (on two consecutive days). Any clumps of soil were broken down. A *Streptomyces* spore suspension was diluted down to approximately 10^7 spores/ml and 1 ml of the suspension was added. An additional 0.5 ml of sterile water was added to bring the soil's water content to 16% and the universal bottle was mixed until its contents were homogeneous. The seeded universal bottles were placed in a 30°C incubator with their lids somewhat loose to allow the exchange of air and left to grow for 7 days. A bowl of water was placed in the bottom of the incubator to reduce soil desiccation.

2.6.4 DNA extraction from soil

DNA was extracted from soil or soil microcosms using the UltraClean Soil DNA Isolation Kit (MO-BIO) as per the manufacturer's instructions. To check the quality and amount of DNA, 5 µl of the eluate was electrophoresed (Section 2.2.4) and the eluate's concentration was determined using a Nanodrop spectrophotometer (ND-1000, Nanodrop Technologies) and stored at -20°C.

2.6.5 Amplification and cloning of target gene

The PCRs were performed as described in section 2.3.1, with the exception that 2 µl of BSA (bovine serum albumin) was added to each reaction. Any bands of the correct size had their DNA extracted (Section 2.3.2), cloned (Section 2.3.3) and sequenced (Section 2.3.4).

2.7 RT real-time PCR

2.7.1 Preparation of high-density spore suspension

10 mannitol-soya agar plates were plated with 50 µl of spore suspension to ensure the formation of a thick lawn and incubated for 5 days at 30°C. 5 ml of sterile water was added to each plate and the spores suspended in it using sterile inoculation loops. All 10 spore suspensions were filtered through autoclaved cotton wool using a syringe to eliminate mycelia, placed into a transparent 40 ml Oak Ridge tube and centrifuged at 13000 RPM at room temperature for 10 min. The supernatant was discarded and the spores resuspended into 5 ml of 10% glycerol (v/v), divided into 500 µl aliquots to prevent frequent freezing and thawing which reduces spore viability and stored at -20°C. A 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}, 10^{-8} and 10^{-9} dilution series was generated from the spore suspension and 10 µl of each were plated out on two LB agar plates. The plates were incubated at 30°C for 48 h and the colony number was used to calculate the number of viable spores per ml in the original spore suspension.

2.7.2 Growth and sampling

Enough spore suspension was added into 150 ml of TSB to produce a spore concentration of approximately 10^7 spores/ml. The broth was mixed well and 50 ml were placed into 3 previously autoclaved 250 ml Erlenmeyer flasks containing precisely 26 cm of stainless steel spring in order to generate 3 biological replicates. The flasks were grown in a shaking incubator at 28°C and 220 RPM.

Optical readings were taken at 10 h past the inoculation and every 2 h onwards. 100 µl of culture was added into 900 µl of 4°C culture suspension buffer, as the buffer was more viscous at a lower temperature, and 200 µl of culture into 800 µl of buffer for the 0.1 and 0.2 dilutions respectively. The 1.5 ml Eppendorf tubes containing them were vortexed and the samples were transferred into 1 ml disposable Fisherbrand semi-micro cuvetts (Fisher Scientific). A cuvette containing suspension

buffer with an appropriate amount of LB was used as a blank for the spectrophotometer (Ultrospec 3000 *pro*) and the readings were taken at a wavelength of 600 nm. Before being measured, the contents of the cuvette were pipetted a few times in order to ensure that the bacteria were evenly suspended.

Samples for RNA extraction were taken at 12 h and every 4 h afterwards. 500 µl samples were taken at 12 h due to the low growth, 200 µl at 16 and 100 µl from 20 h onwards. The sample was added to a 1.5 ml Eppendorf tube containing 2x the sample volume of RNAprotect Bacteria Reagent (QIAGEN), vortexed the culture for 5 sec, incubated it at room temperature for 5 min and vortexed it at 13000 RPM for 10 min. The supernatant was removed and the pellets stored at -80°C until they were needed.

After the end of the protocol, a sterile loop was used to streak out culture from each flask onto LB agar plates that were incubated at 37°C for 72 h in order to confirm the absence of contamination.

2.7.3 RNA extraction and DNA removal

The RNA extraction was performed using the protocol #5, employing both enzymatic lysis and proteinase K digestion, from the RNAprotect Bacteria Reagent manual. The manufacturer's protocol was followed with the exceptions that 20 µl of QIAGEN Proteinase K (QIAGEN) and 100 µl of TE buffer with lysozyme were used and the lysozyme incubation step was 15 min instead of 10 min on a shaking incubator. Protocol #7, purification of total RNA from bacterial lysate using the RNeasy mini kit, from the RNAprotect Bacteria Reagent manual was performed next including the optional on-column DNase digestion using the RNAase-free DNase set (QIAGEN) step. The manufacturer's protocol was followed with the exception that a 30 min instead of 10 min incubation step was used for the DNA digestion step.

A second DNA digestion step was employed using the TURBO DNA-free Kit (Ambion). The RNA-containing eluate underwent a 1 in 4 dilution and a 50 µl reaction volume was prepared using the rigorous DNase treatment from the kit manual. A PCR (Section 2.3.1) with the *hrdB* primer set was used to confirm the absence of DNA contamination. If DNA was still present, an additional DNA

digestion step was performed using the TURBO DNA-free Second Digest Protocol with rigorous DNase treatment and the absence of DNA was then confirmed via PCR. The RNA solution's concentration was determined using a Nanodrop spectrophotometer (ND-1000, Nanodrop Technologies) and stored at -80°C.

2.7.4 Reverse transcription

The High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) was used to generate cDNA as per the manufacturer's instructions. The reactions utilised the optional RNase Inhibitor (Applied Biosystems). The cDNA solution's concentration was determined using a Nanodrop spectrophotometer (ND-1000, Nanodrop Technologies) and stored at -20°C. A PCR (Section 2.3.1) with the *hrdB* primer set was used to confirm the success of the reverse transcription reaction.

2.7.5 Real-time PCR

The SYBR Green PCR Master Mix (Applied Biosystems) was used for the real-time PCR reactions. Primers were generated via the GeneFisher2 PCR Primer Design website (http://bibiserv.techfak.uni-bielefeld.de/genefisher2/) with the following criteria (Dorak, 2006):

- Primer length 18-24 nucleotides
- Amplicon length <250 bp (ideally <150 bp)
- Both sense and anti-sense primers should have a Tm ≤2°C of each other
- Primer Tm 50-60°C (ideal range 55-59°C)
- No consecutive runs of the same nucleotide more than 6 times
- No runs of more than three consecutive Gs
- % GC content of primers ~50% (no less than 35% and no more than 65%)
- No 3' GC clamp on primers (i.e. GG, CC, CG or GC)
- ≤2 GC in the last five nucleotides of the 3' end of the primer

An attempt was made to generate primers based on the consensus sequence of all available genes, but that was not always possible. All primers were tested on conventional PCR to ensure the presence of only a single band and that band was then extracted and sequenced to confirm the presence of only a single product. The effect of using 10, 20 or 30 pM of primer on the real-time PCR efficiency was tested

on the DNA standards but no differences to the reactions' CT values were found and it was thus decided to use a concentration of 10 pM for all the primers.

To reduce pipetting errors the cDNA underwent a 1 in 4 dilution and 4 µl of the dilution were used per reaction. A 10 ng/µl genomic DNA solution from the strain being examined was diluted to 10^{0}, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5} in Nuclease-Free Water (Ambion) and this dilution series was used as the real-time PCR standards. All reactions were done in triplicate following the manufacturer's instructions and each plate included 6 no template controls. The 25 µl reaction mixes were loaded on MicroAmp 96-Well Reaction Plate (Applied Biosystems) which was sealed with a MicroAmp Optical Adhesive Film (Applied Biosystems) and centrifuged for 1 min at 1250 RPM. The amplifications were carried out on a 7500 Fast Real-Time PCR System (Applied Biosystems) using the Standard 7500 program which consisted of 50°C for 2 min, 95°C for 10 min, then 40 cycles of 95°C for 15 sec and 60°C for 60 sec, followed by a dissociation step of 95°C for 15 sec, 60°C for 60 sec, 95°C for 15 sec and 60°C for 60 sec.

2.7.6 Data analysis

The real-time PCR method used was a standard curve (absolute quantification) assay. The number of gene copies present in the standards was estimated using the dsDNA copy number calculator (http://www.uri.edu/research/gsc/resources/cndna.html). The calculator estimated that 10 ng of DNA with a sequence of 8,545,929 bp in length, the size of the *S. griseus* NBRC 13350 chromosome, would contain 1.08×10^{6} gene copies. Due to the fact that the genome size of the isolates used in this study was not known, and 8.5 Mb is close to the average size of a *Streptomyces* chromosome (Paradkar *et al.*, 2003), the same gene copy number was used for the standards of all the strains in this study. The results were analyzed using the 7500 Fast System SDS Software (Applied Biosystems). In the Analysis Settings, the Auto Ct option was used to calculate the CT values. Each triplicate was examined and if one of the three reactions had a CT value that differed by more than 1 compared to the other two it was omitted from the study. The software produced the means of the number of cDNA copies initially present in each technical triplicate based on the standards. The results were exported as CSV files and processed using Microsoft Excel 2007. Each

mean cDNA copy number from the genes being studied was divided by the mean number of *hrdB* cDNA copies from the same sample to calculate the ratio of the two which indicated that gene's expression level. The mean and standard deviation of these ratios was then calculated from the three biological replicates and used in this study.

3 Distribution of the streptomycin cluster, *strA* resistance gene and other secondary metabolite genes in the genus *Streptomyces*

3.1 Introduction

The ability to produce streptomycin *in vitro* is widespread within actinobacteria but it is predominantly an antibiotic biosynthesized by streptomycetes (Tohyama *et al.*, 1987). Approximately 1% of randomly screened soil actinobacteria can synthesize streptomycin (Baltz, 2006). Few antibiotics can match its prevalence; only streptothricin is found at a frequency of 10^{-1}, while actinomycin D and tetracycline are in the 10^{-2} to 10^{-3} range. Most antibiotic producers are much less prevalent; for example, daptomycin production required the screening of nearly 10^{7} actinobacteria to be discovered (Baltz, 2008). There are exponentially more antibiotics present at lower frequencies; in one study about 200 were discovered at frequencies of around $4x10^{-7}$, about 800 at around $2x10^{-7}$ and more than 1000 at around $1x10^{-7}$ (Baltz, 2005).

The production of streptomycin is not only common it is also very widespread, as it appears to have an almost global distribution. Producers belonging to the *S. griseus* species have been isolated from soils in diverse countries such as the USA, Mexico, Japan (Gordon & Horan, 1968), Germany (Tolba *et al.*, 2002) and Brazil (Huddleston *et al.*, 1997). The wide distribution and high frequency of streptomycin production indicates that this phenotype provides a significant advantage to the organism possessing it even when compared to most other antibiotics whose producers are normally found in vastly lower numbers in soil.

The widespread prevalence of streptomycin producers must have applied significant evolutionary pressure on other soil-inhabiting microorganisms. A number of resistance mechanisms do exist (Davies & Wright, 1997), however this project focused on resistance due to genes that are orthologous to the *strA* resistance gene found in producers. Divergent *strA* genes have been found in streptomycin-resistant streptomycete soil isolates from Brazil (Wiener *et al.*, 1998), the UK and Germany (Tolba *et al.*, 2002). The biosynthetic gene *strB1* that flanks *strA* could not be amplified via PCR in a number of isolates from the UK and Germany (Tolba *et al.*, 2002). This indicates that these strains may have partial or truncated streptomycin

clusters. Alternative resistance mechanisms found in streptomycetes include substitutions on the 43rd and 88th positions of the S12 ribosomal protein (Musser, 1995), mutations in the 530 loop of the 16S rRNA sequence (Springer *et al.*, 2001) and presence of the *aphE* gene which encodes a streptomycin-3"-phosphotransferase (Trower & Clark, 1990).

3.2 Aims

In this chapter is to examine the distribution of streptomycin and viomycin resistance and biosynthetic genes in uncharacterised soil isolates and in *Streptomyces* type strains is considered.

The hypotheses of this chapter are that (i) antibiotic production provides a selective advantage in soil and therefore it can be expected to be conserved once it has evolved and if mobile retained because it grants a selective advantage and (ii) that there is selection for antibiotic resistance in soil due to the production of antibiotics in it.

3.3 Results

3.3.1 Screening of eight previously isolated strains containing *strA* for other streptomycin genes

Eight strains previously isolated by Tolba (2004) possessing *strA* were screened for streptomycin cluster genes other than *strA* via PCR and sequencing in order to establish the structure of the streptomycin clusters within them. Of the eight strains examined, two (DW15, DW21) were isolated from pastureland in Droitwich, UK with sewage sludge injection, one (AR23) from an agricultural site in Dossenheim, Germany where plantomycin was used, two (CR13, CR50) from a control site in Dossenheim where no plantomycin was used, two (CW12, CW45) from limestone-based soil in Cotswolds, UK and one (666) was part of Prof. Wellington's culture collection. These sites are described in (Tolba *et al.*, 2002).

The presence of *str* genes was examined using PCR and sequencing. The primers for *strA* and *strR* were designed by Tolba *et al.* (2002). The other *str* primers (Table 3.1) used were designed based on the consensus sequences of homologues from the *S.*

griseus streptomycin cluster and the *Streptomyces glaucescens* hydroxystreptomycin cluster. As a result this study was limited to genes that are shared between these two biosynthetic clusters, as a set of primers based solely on the sequences from *S. griseus* genes would have been too specific to detect homologues with significant sequence divergence and the 14 genes examined were dispersed across the entire cluster and would have sufficed to detect deletions of large sections of the *str* cluster (Figure 3.1).

The PCR master mix and cycling parameters detailed in Section 2.3.1 were used, however a low annealing temperature of 50°C was employed to assist in amplifying genes with divergent sequences. All DNA gel bands of approximately the correct size or ones that had high intensity when viewed under UV light were excised and sequenced to ensure that versions of the gene with insertions or deletions would not go undetected. The *S. griseus* DSM 40236^{T} (the ^{T} denotes it is the type strain) was also included in this study as a positive control for all the PCR and sequencing reactions.

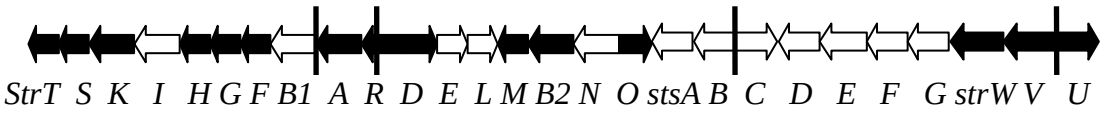


Figure 3.1. The *S. griseus* DSM 40236^{T} streptomycin gene cluster. Genes screened for by PCR are coloured black.

Table 3.1. Primers used to screen strains for the presence of streptomycin biosynthetic cluster genes

Gene	Protein function	Primer	Sequence
strA	streptomycin 6-phosphotransferase	strA_F	GCG GCT GCT CGA CCA CGA C
		strA_R	CCG TCC TCG ATG TCC CAC AGG G
strB2	putative scyllo-inosamine-4-phosphate amidinotransferase	strB2_F	CTG GTG TCC GAC AGC GGC AA
		strB2_R	CAG AAC CGG TAC GWC TCC AG
strD	putative glucose-1-phosphate thymidylyltransferase	strD_F	CAT GTA TCT CGG CGA CAA TT
		strD_R	GCC STG CTG TGG TCG CCG AG
strF	StrF protein	strF_F	GTT CAC CAT GCG CAA CCA GC
		strF_R	TGC CAG CGC AGT TCG TTG T
strG	StrG protein	strG_F	CCA CCT TCC GGG TSC ACC TG
		strG_R	GTC CCA GTA CTC CCC GAG C
strH	putative dTDP-dihydrostreptose--streptidine-6-phosphate dihydrostreptosyltransferase	strH_F	GGT CGG CAG CCA CGA GAT CT
		strH_R	CTC CAG CAG CAG CGC GTC GA
strK	putative streptomycin-6-phosphate phosphatase	strK_F	TCG GTG ACG GCA TGG GCG AC
		strK_R	GCG AAC AGC CCG AGS ACC GG
strM	putative dTDP-4-dehydrorhamnose 3,5-epimerase	strM_F	GCC AGG CCA AGT ACG TCA C
		strM_R	GGG CCA GGC GAT GCC SAG G

strO	putative glucose-1-phosphate thymidylyltransferase	strO_F	GAC GGC AGC CCG GTC AGC G
		strO_R	CGT CCG CCS CGT ACA CSA GC
strR	streptomycin biosynthesis operon regulator	strR_F	GGA TCG GGA GGG ACG GGC GG
		strR_R	ACA GCC CCG GGCCAG CTC GG
strS	putative N-methyl-L-glucosamine biosynthetic aminotransferase	strS_F	CCC ACA CCT TCA TCG GCT C
		strS_R	GCG GGC TGC TTG TGC AGC GG
strT	putative oxidoreductase	strT_F	CGA CGC CTA CTG GGA GGG CG
		strT_R	CCT GGC TCG CCC AGG GGT AC
strU	putative NAD(P)-dependent oxidoreductase	strU_F	GGT CAC CGA GGA CGC CTT CG
		strU_R	CCG AGC TTG TAC AGC TCC AG
strV	putative ABC-type multidrug transport system ATPase and permease component	strV_F	GCG TCG GCA CCT CGG ACA TG
		strV_R	GGA AGG CCG CGT AGG CGT AG
strW	putative ABC-type multidrug transport system ATPase and permease component	strW_F	CGA GAA GGT CAC CGC CGA GG
		strW_R	CGT CCA GCG GGA AGA TCA GC

Table 3.2. Streptomycin cluster genes detected via sequencing in selected strains. Boxes shaded in gray indicate the detection of the relevant gene in that strain. The percentage included in these boxes is the sequence identity of a gene to its homologue in the *S. griseus* type strain.

Genes	666	DW15	DW21	CR13	CR50	AR23	CW12	CW45	*S. griseus*
strA	75%	72%	75%	100%	83%	75%	73%	73%	100%
strB2				100%					100%
strD				100%					100%
strF				100%	93%				100%
strG				100%	88%				100%
strH				100%	87%				100%
strK				100%	85%				100%
strM				100%	92%				100%
strO				100%	73%				100%
strR				100%					100%
strS				100%	88%				100%
strT				100%	90%				100%
strU				100%	88%				100%
strV				100%	87%				100%
strW				100%	86%				100%

As Table 3.2 indicates, CR13 had a streptomycin cluster that is identical to that of the *S. griseus* type strain. CR50 also had a cluster, however it had diverged from that of *S. griseus* and three of its genes could not be amplified via PCR. The remaining strains had divergent *strA* genes and none of the biosynthetic or the regulatory genes of the cluster were amplified. This implies that they possessed solely the resistance gene and none of the remaining cluster.

3.3.2 Screening of other streptomycete isolates for streptomycin resistance

This study was expanded by the inclusion of soil isolates from Cayo Blanco, Cuba (labelled CB), various sites in the UK (E or RB), isolates from potato scabs (Scab), a set of *S. coelicolor* isolates from a chemostat (BTG) and strains either belonging or related to *S. griseus* and strains with SARC (Streptomycin And Related Compounds) clusters from the German Resource Centre for Biological Material (DSM). Strains were grown on nutrient agar plates with 10 µg/ml, 50 µg/ml and no streptomycin to determine their resistance levels (Section 2.2.2) and on RASS plates to compare their phenotype (spore colour and colony morphology) with the *S. griseus* type strain (Section 2.2.3) in order to search for *S. griseus* strains that had lost the *strA* resistance gene.

Table 3.3. Resistance profiles of streptomycete isolate groups plated on RASS agar supplemented with 0, 10, 50 µg/ml streptomycin.

Isolates	Origin	0 µg/ml	10 µg/ml	50 µg/ml	*S. griseus* spore colour	*S. griseus* colony morphology
Tolba	Various	9	6	1	1	1
CB	Cuba	207	32	27	18	5
E	UK	58	10	7	7	1
RB	UK	15	1	1	0	0
Scab	Various	7	3	1	0	0
DSM	Various	54	25	17	14	10
BTG	Chemostat	8	0	0	0	0
Total	Various	339	82	67	40	17

From the Cuban samples (CB), 15.5% of streptomycete isolates had low level resistance (10 µg/ml) and 13.0% had high level resistance (50 µg/ml). The UK isolates (E and RB) had very similar resistance levels, with 15.1% having low level and 11.0% having high level resistance (Table 3.3). There was no significant difference between the two soils either at the low (χ^2 p = 0.9147) or high (χ^2 p = 0.8114) levels of resistance. A higher percentage of the DSM strains were resistant (14.8% low and 31.5% high level resistance, χ^2 p = 0.9324 for Cuban soils and p = 0.8876 for UK soil for low resistance p = 0.0024 for Cuban soils and p = 0.0082 for UK soil for high resistance), however that was because *S. griseus* strains were screened preferentially to look for unusual phenotypes within that species, which resulted in an overrepresentation of resistant strains. All soil isolates were selected

for belonging to the genus *Streptomyces* as a whole and thus were not biased toward streptomycin resistance.

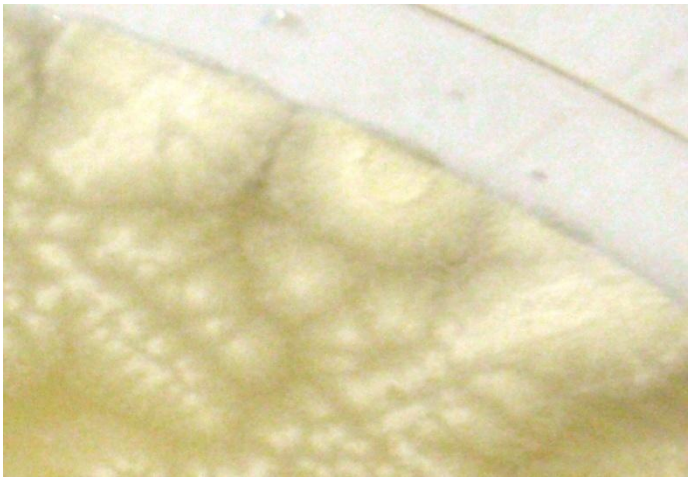


Figure 3.2. Colouration and morphology of *S. griseus* DSM 40236^{T} colonies. Colonies had a beige colour and looked flat and powdery.

The spore coloration and colony morphology are unreliable phylogenetic markers, however they have the advantage that they are easy to examine and can thus provide a simple preliminary method of screening strains for potential membership of the *S. griseus* species and their exact phylogenetic position can later be determined via sequencing of housekeeping genes. From the Cuban samples, 8.70% had coloration that matched the *S. griseus* type strain and 2.42% had identical colony morphology. The UK isolates again had very similar ratios to the Cuban isolates, with 9.59% displaying *S. griseus* spore coloration and 1.37% displaying *S. griseus* colony morphology. A large number of DSM strains belonged to the *S. griseus* species, which is why a high percentage of strains matched that species' coloration (34.1%) and morphology (24.4%).

3.3.3 Screening of streptomycete isolates for streptomycin cluster genes

Strains displaying resistance to streptomycin or a phenotype similar to that of the *S. griseus* type strain, as well as a set of Zambian isolates from bee hives and the water and soil surrounding them (Z) were screened via PCR for the presence of *strA*, *strR* and *strW*. *strA* was selected because it is the resistance gene and most widespread gene of the cluster, *strR* because it regulates expression of the streptomycin cluster and must therefore be present if the *strA* gene is regulated in the same manner as in *S. griseus* streptomycin producers and *strW* because it is a biosynthetic gene located

far from *strA* and *strR* (Figure 3.1), enabling the detection of potential deletions of the 3' end of the cluster.

Table 3.4. Presence of *str* genes in streptomycete isolate groups. Screened denotes the number of strains that underwent PCR and *strA, strR, strW* denotes the number of strains from which *strR* was amplified or *strA* and *strW* were successfully sequenced.

Isolates	Origin	Screened	*strA*	*strR*	*strW*
Tolba	Various	9	8	1	2
CB	Cuba	36	9	2	2
E	UK	10	3	0	0
RB	UK	1	1	0	0
Scab	Various	1	0	0	0
DSM	Various	54	25	12	15
Z	Zambia	44	1	1	1
BTG	Chemostat	8	0	0	0
Total	Various	132	47	16	20

Only some of the strains screened possessed *strA* (Table 3.4). This included 20 of the 46 strains with high level resistance (43.5%), 14 of the 21 strains with low level resistance (66.7%) and 12 of the 58 strains with no resistance (20.7%). This indicates that almost half the strains (47.8%) rely on mechanisms other than StrA for streptomycin resistance.

In total 4.3% of Cuban isolates possessed *strA* and 0.97% possessed *strW* and *strR*, while 5.5% of UK isolates possessed *strA* and 0% possessed *strW* and *strR*. 2.2% of Zambian strains possessed *strA*, *strW* and *strR*. There was no significant difference in the number of strains possessing the resistance or biosynthetic genes from Cuba, the UK or Zambia (χ^2 p = 0.9237, 0.8229 and 0.7012 for *strA* and χ^2 p = 0.9698, 0. 9344 and 0. 8266 for *strW* and *strR*). 46.3% of the DSM strains possessed *strA* and 22.2% possessed *strR* and 37.0% possessed *strW*.

The *strR* and *strW* genes were found only in strains with high streptomycin resistance (Table 3.5, Table 3.6). This is to be expected, as their presence indicates that the entirety of the cluster is present and streptomycin producers need to have high level resistance. The only exception was *S. glaucescens*, which has a divergent cluster and thus does not require high level streptomycin resistance to survive. The fact that *strR* failed to amplify from certain strains possessing *strW* may be due to the fact that the PCR primers used were too stringent and could only amplify the

sequence found in the *S. griseus* streptomycin cluster. CR50 possesses a copy of *strR* (Section 4.3.3) but the primers used in this study failed to amplify it.

Table 3.5. Strains from culture collections screened for *str* cluster genes. On the phenotype column C denotes a match for spore colour, M a match for colony morphology to the *S. griseus* type strain and –ve an absence of both. In the 'genes present' column *A* denotes *strA*; *R, strR*; *W, strW*; *B, vioB*; *D, vioD*. Boxes that are greyed-out denote that these strains were not screened for viomycin resistance. A ^{T} by the accession number denotes a type strain.

Ain	Species	S. griseus phenotype	Streptomycin resistance (µg/ml)	Streptomycin genes present	Viomycin resistance (µg/ml)	Viomycin genes present
1326	*S. lividans*	-ve	0	-	0	-
DSM 40005^{T}	*S. cinnamoneus*	-ve	0	*A*	50	-
DSM 40041^{T}	*S. platensis*	-ve	0	-	0	-
DSM 40049^{T}	*S. violaceoruber*	-ve	0	-	0	-
DSM 40058^{T}	*S. californicus*	C M	10	-	50	*vph, B, D*
DSM 40066^{T}	*S. griseobrunneus*	C M	0	-	50	*SGR421*
DSM 40070	*S. microflavus*	C	0	-	0	-
DSM 40083^{T}	*S. puniceus*	-ve	10	-	50	*vph, B, D*
DSM 40093	*S. netropsis*	-ve	50	*A*	0	-
DSM 40131^{T}	*S. limosus*	C	0	-	0	-
DSM 40135^{T}	*S. acrimycini*	C M	0	-	0	-
DSM 40136^{T}	*S. albovinaceus*	C M	0	*A*	0	-
DSM 40155^{T}	*S. glaucescens*	-ve	10	*A, W*	0	-
DSM 40211^{T}	*S .olivoviridis*	M	0	-	0	-
DSM 40221^{T}	*S. mashuensis*	-ve	10	*A, W*	0	-
DSM 40233^{T}	*S. coelicolor*	-ve	0	-	0	-
DSM 40236^{T}	*S. griseus*	C M	50	*A, R, W*	50	*SGR421*
DSM 40257	*S. vinaceus*	-ve	0	-	50	*vph, B, D*
DSM 40263^{T}	*S. humidus*	-ve	0	*A*	0	-
DSM 40265	*S. anulatus*	C M	0	*A*	0	-
DSM 40281^{T}	*S. griseoruber*	-ve	0	-	0	-
DSM 40320	*S. sp*	C M	0	-	0	-
DSM 40326^{T}	*S. alboviridis*	C M	0	-	0	-
DSM 40419	*S. sp*	-ve	0	-	0	-
DSM 40421^{T}	*S. coelescens*	-ve	0	-	0	-
DSM 40438^{T}	*S. violaceolatus*	-ve	0	-	0	-
DSM 40459	*S. microflavus*	C M	0	-	0	-
DSM 40463	*S. sp*	C M	0	-	50	*SGR421*
DSM 40480	*S. galbus*	-ve	0	*A, W*	0	-
DSM 40505^{T}	*S. tuirus*	-ve	0	-	0	-
DSM 40541	*S. sp*	-ve	50	-		
DSM 40561	*S. microflavus*	C	0	-	0	-
DSM 40598^{T}	*S. bacillaris*	M	0	-	50	*SGR421*
DSM 40627	*S. griseus*	C M	50	*A, R, W*	50	*SGR421*
DSM 40653	*S. griseus*	C M	50	*A, R, W*	50	*SGR421*
DSM 40654	*S. griseus*	C M	50	*A, R, W*	50	*SGR421*
DSM 40657	*S. griseus*	C M	50	*A, R, W*	50	*SGR421*
DSM 40658	*S. griseus*	M	50	*A, R, W*	50	*SGR421*
DSM 40659	*S. griseus*	C	50	*A, R, W*	50	*SGR421*
DSM 40660	*S. griseus*	M	50	*A, R, W*	50	*SGR421*
DSM 40670	*S. griseus*	C M	50	*A, R, W*	50	*SGR421*
DSM 40707	*S. griseus*	C M	0	*A*	10	-
DSM 40757	*S. griseus*	C	0	*A*	0	-
DSM 40759	*S. griseus*	C	50	*A, R, W*	50	*SGR421*
DSM 40776	*S. sp*	-ve	10	*A*	0	-

DSM 40783	*S. violaceoruber*	-ve	0	-	0	-
DSM 40817	*S. griseus*	-ve	10	*A*	10	-
DSM 40855	*S. griseus*	M	50	*A, R, W*	50	*SGR421*
DSM 40864	*S. sp*	M	0	-	50	*SGR421*
DSM 40878	*S. griseus*	C	50	*A, R, W*	50	*SGR421*
DSM 40931	*S. griseus*	C M	0	-	0	-
DSM 40932	*S. griseus*	C M	50	-	10	-
DSM 40938^{T}	*S. floridae*	-ve	10	-	50	*vph, B, D*
DSM 40939	*S. griseus*	-ve	10	*A*	0	-
DSM 46445	*S. scabiei*	C	50	-		
M110	*S. coelicolor*	-ve	50	-	0	-
NRRL 12000	*S. coelicolor*	-ve	0	-	0	-

Table 3.6. Isolates screened for *str* cluster genes. On the phenotype column S denotes a match for spore colour, C a match for colony morphology to the *S. griseus* type strain and –ve an absence of both. In the 'genes present' column *A* denotes *strA*; *R, strR*; *W, strW*; *B, vioB*; *D, vioD*. Boxes that are greyed-out denote that these strains were not screened for viomycin resistance.

Isolate	Species	*S. griseus* phenotype	Streptomycin resistance (µg/ml)	Streptomycin genes present	Viomycin resistance (µg/ml)	Viomycin genes present
651	*S. sp*	-ve	0	-	0	-
666	*S. sp*	-ve	0	*A*	0	-
AR23	*S. sp*	-ve	10	*A*	0	-
BTG 4-723	*S. sp*	-ve	0	-	0	-
BTG 4-738	*S. sp*	-ve	0	-	0	-
BTG 4-758	*S. sp*	-ve	0	-	0	-
BTG 4-759	*S. sp*	-ve	0	-	0	-
BTG 6-708	*S. sp*	-ve	0	-	0	-
BTG 6-715	*S. sp*	-ve	0	-	0	-
BTG 717 I	*S. sp*	-ve	0	-	0	-
BTG 723 I	*S. sp*	-ve	0	-	0	-
CB141	*S. sp*	C M	50	-	0	-
CB142	*S. sp*	C M	10	*A*	10	-
CB143	*S. sp*	C M	50	-	0	-
CB144	*S. sp*	C	0	-	0	-
CB145	*S. sp*	C	10	*A*	10	-
CB146	*S. sp*	C	0	-	0	-
CB147	*S. sp*	C	0	*A*	0	-
CB148	*S. sp*	C	0	-	0	-
CB149	*S. sp*	C	0	-	0	-
CB150	*S. sp*	C	0	-	0	-
CB151	*S. sp*	C	0	-	0	-
CB152	*S. sp*	C	0	-	0	-
CB153	*S. sp*	C	50	*A*	0	-
CB154	*S. sp*	C	0	*A*	0	-
CB155	*S. sp*	-ve	50	-	0	-
CB156	*S. sp*	-ve	50	-	10	-
CB157	*S. sp*	-ve	10	-	50	*vph, B, D*
CB158	*S. sp*	-ve	50	*A*	50	-
CB159	*S. sp*	-ve	50	-	10	-
CB160	*S. sp*	-ve	50	-		
CB161	*S. sp*	-ve	50	-		
CB162	*S. sp*	C	50	*A, R, W*	10	*SGR421*
CB163	*S. sp*	C	50	*A, R, W*	10	*SGR421*
CB164	*S. sp*	-ve	0	-		
CB165	*S. sp*	-ve	50	-		
CB166	*S. sp*	-ve	50	-		

CB167	*S. sp*	-ve	50	-		
CB168	*S. sp*	-ve	50	-		
CB169	*S. sp*	-ve	50	-	10	-
CB170	*S. sp*	-ve	50	-	10	-
CB171	*S. sp*	C	50	-	0	-
CB172	*S. sp*	C	0	*A*	0	-
CB184	*S. sp*	C M	50	-		
CB185	*S. sp*	C M	50	-		
CB206	*S. sp*	-ve	50	-	0	-
CB218	*S. sp*	-ve	0	-	0	-
CR13	*S. sp*	-ve	50	*A, R, W*	50	*SGR421*
CR50	*S. sp*	-ve	10	*A, W*	0	-
CW12	*S. sp*	-ve	10	*A*	10	-
CW45	*S. sp*	-ve	10	*A*	10	-
DW15	*S. sp*	-ve	10	*A*	10	-
DW21	*S. sp*	-ve	10	*A*	0	-
E948	*S. sp*	C	50	-	0	-
E953	*S. sp*	C	50	-	0	-
E956	*S. sp*	C	0	-	0	-
E961	*S. sp*	C	0	-	0	-
E984/1	*S. sp*	-ve	50	-		
E989	*S. sp*	-ve	10	*A*	10	-
E996	*S. sp*	C	0	*A*	0	-
E1002	*S. sp*	C	0	-	0	-
E1013	*S. sp*	M	10	*A*	10	-
E1024	*S. sp*	-ve	50	-		
RB063	*S. sp*	-ve	50	*A*	10	-
SCAB3	*S. sp*	-ve	50	-		
Z34	*S. sp*	C M	50	*A, R, W*	50	*SGR421*

S. griseus DSM 40817 due to its phylogenetic position (Section 5.3.3) appeared likely to possess *strA.* However, while PCRs using the strA_F/R primer set managed to produce a band of the correct size, it proved impossible to purify an *strA* sequence from it. For that reason a number of other primers were designed to amplify *strA*. The primers strA_B1_F (CCG CAC ACT GGC TTC GA) and strA_B1_R (GAC CTG CGC CAG CAT CT) generated a 117 bp amplicon that was successfully sequenced and proven to be part of *strA.*

3.3.4 Analysis of the phylogeny of partial *strA* nucleotide sequences

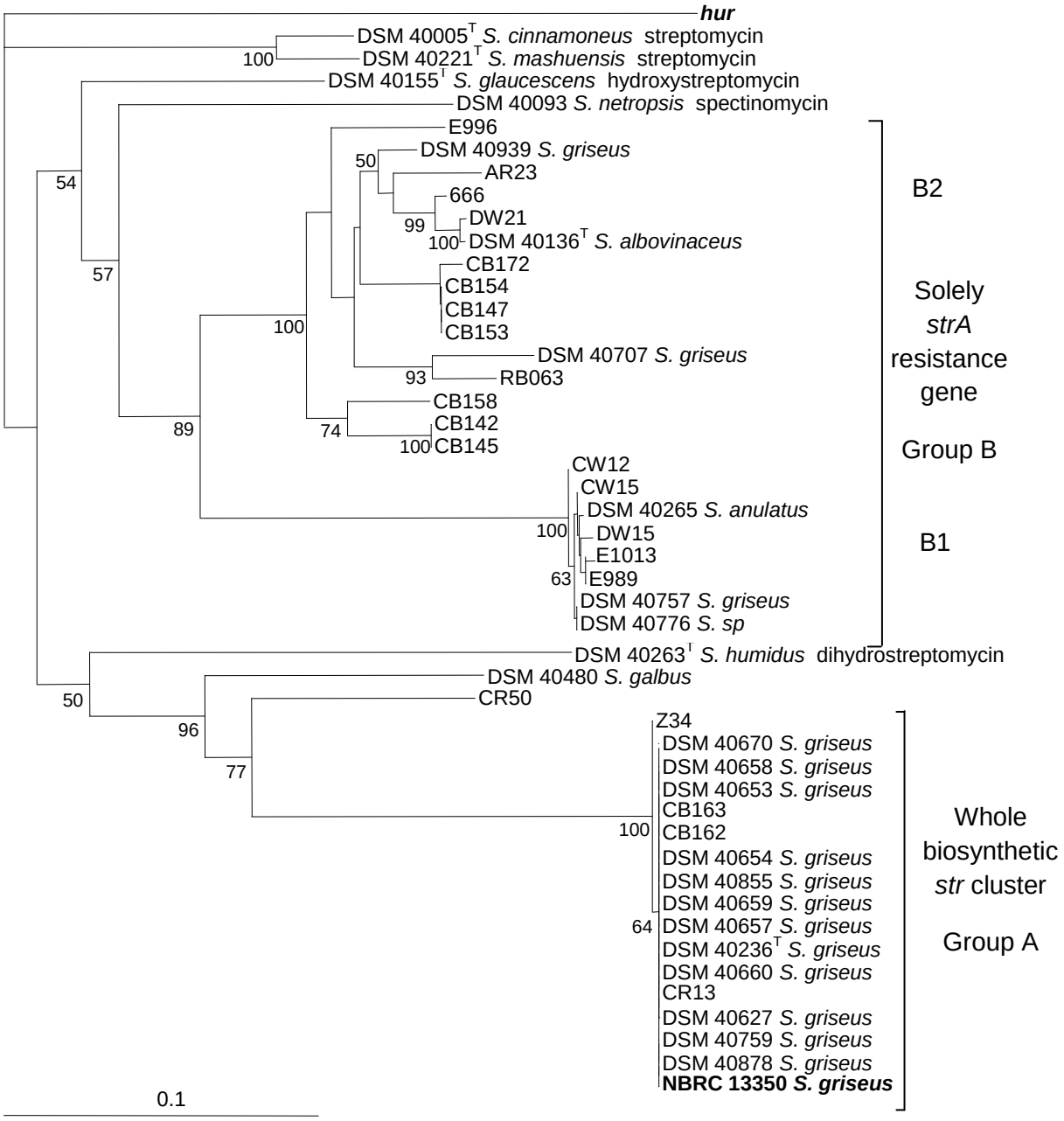


Figure 3.3. The phylogeny of the streptomycin resistance gene *strA*, revealing two major clades for resistance versus production. Sequences in bold were acquired from GenBank. The tree was constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% nucleotide dissimilarity.

The majority of *strA* sequences obtained fell into two distinct clades (Figure 3.3); Group A were from streptomycin producers, containing the entire streptomycin biosynthetic cluster, and Group B were from non-streptomycin producing strains that were missing *strR* and *strW*. The remaining genes were from clusters producing

compounds related to streptomycin. The streptomycin producers had invariant *strA* genes with the exception of strain Z34, whose *strA* had one synonymous single nucleotide polymorphism (SNP), resulting in a 99.8% identity to the other sequences in that clade. Strain CR50 contained a more diverse *strA* sequence with 83.3% homology to Group A, though it appeared to be related to the Group A *strA* gene.

Group B contained two subgroups, B1 and B2, with sequences that were considerably more diverse than those of Group A. Both subgroups formed very robust clades with high bootstrap values; however Group B in its entirety was weaker and thus may not be a genuine clade. Group B1 sequences were very closely related to one another, having an average sequence homology of 99.6% and a minimum sequence identity of 99.3%. Group B2 sequences were more diverse, with an average sequence identity of 93.4% to one another and a minimum sequence identity of 90.3%. Group B1 *strA* had a sequence identity that ranged from 72.5% to 73.0% and an average of 72.7% to the Group A. Group B2 *strA* was more diverse, with a sequence identity that ranged from 73.2% to 76.6% and an average of 75.5% sequence identity to Group A (Figure 10.1).

Group B *strA* genes appeared to be more closely related to the *sph* gene encoding a (5'-hydroxy)-streptomycin-6-phosphotransferase from *S. glaucescens* and the *aph(6)* gene encoding an aminoglycoside phosphotransferase from *Streptomyces netropsis*, while Group A strains clustered with the *strA* homologue of the dihydrostreptomycin gene cluster found in *Streptomyces humidus*. However, the very low bootstrap values of these nodes mean that such inferences are very tenuous.

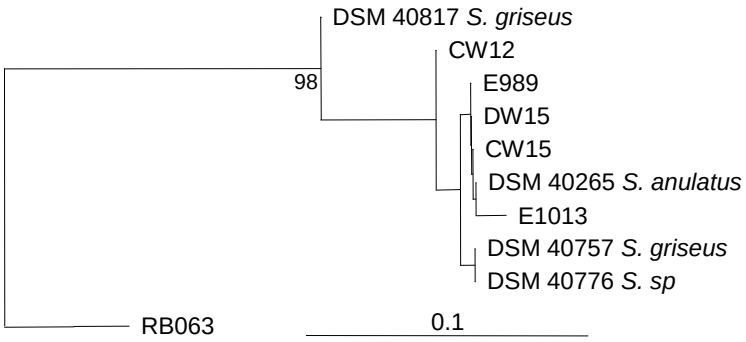


Figure 3.4. Segment of the phylogenetic tree of the 117 bp *strA* fragment amplified from *S. griseus* DSM 40817. Sequences in bold were acquired from GenBank. The tree was constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% nucleotide dissimilarity.

The *strA* from *S. griseus* DSM 40817 clustered with Group B1, though it was considerably more divergent than other members of that clade (Figure 3.4).

3.3.5 Analysis of the phylogeny of partial *strA* translated nucleotide sequences

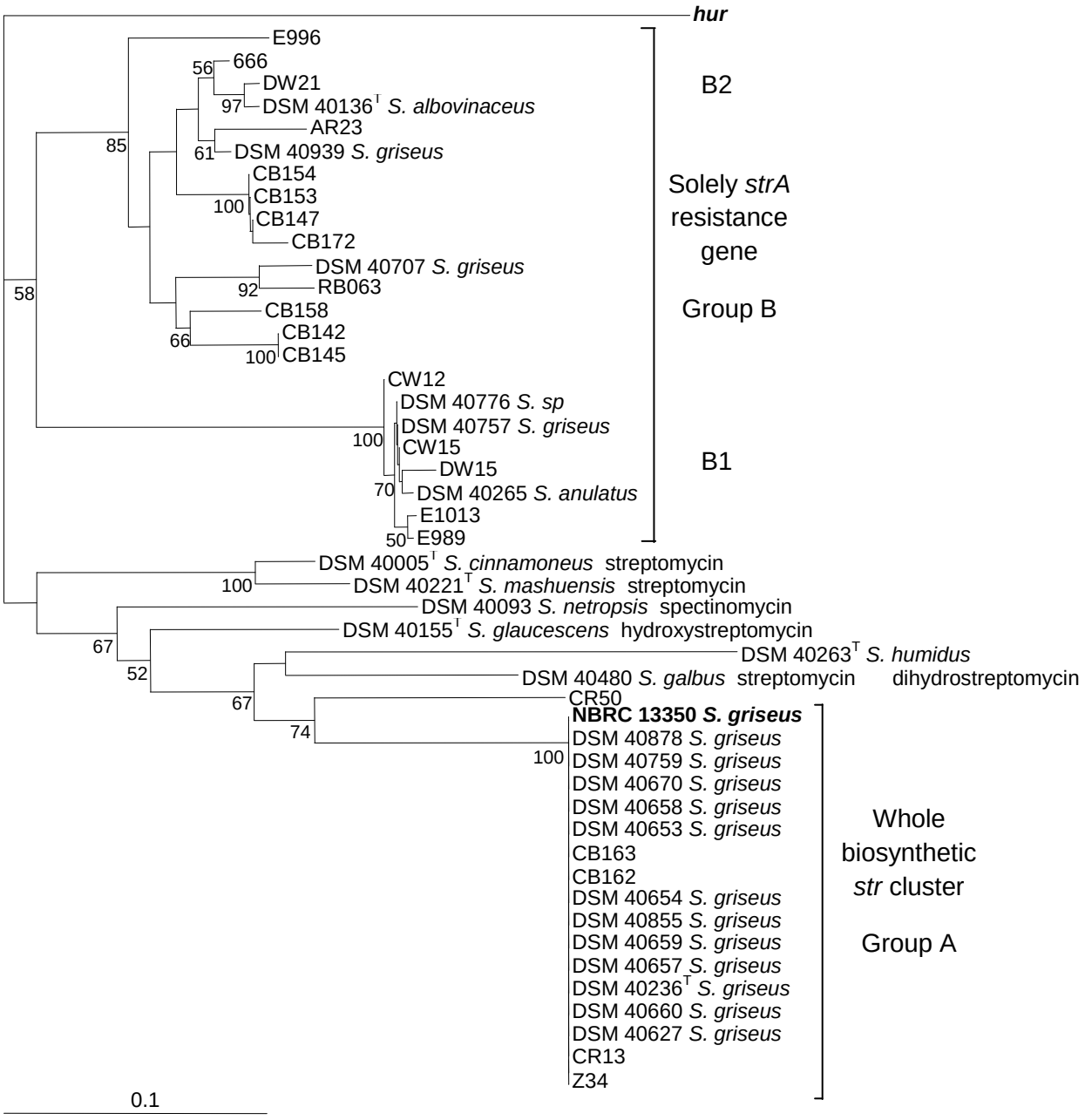


Figure 3.5. A phylogenetic tree of *strA* constructed using translated protein sequences. Sequences in bold were acquired from GenBank. The tree was constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% amino acid dissimilarity.

The overall structure of the amino acid tree (Figure 3.5) is the same as that of the nucleotide tree (Figure 3.3). The principal differences between the two trees is that

most bootstrap values are lower, this is to be expected since the information provided by silent mutations is lost, and that the *strA* homologues from *S. glaucescens* and *S. netropsis* form a clade with Group A instead of Group B sequences, which reinforces the fact that the phylogenetic positions of these genes are uncertain.

The SNP of Z34 was a silent mutation and its nucleotide sequence was thus identical to the other producers. CR50 had a sequence identity of 83.7%, which was only slightly higher than that of the nucleotide sequence. The Group B1 sequences had an average sequence identity of 99.6% and a minimum sequence identity of 97.8%. Group B2 sequences were more diverse, with an average sequence identity of 93.7% to one another and a minimum sequence identity of 89.4%. Group B1 *strA* had a sequence homology that ranged from 69.7% to 70.8% and an average of 70.4% to Group A. Group B2 *strA* was again more diverse, with a sequence identity that ranged from 72.5% to 75.3% and an average of 74.4% sequence identity to the Group A. All Group B amino acid sequences had slightly lower sequence identities than their nucleotide counterparts, indicating the presence of a large number of missense mutations (Figure 10.2).

3.3.6 Analysis of the amino acid sequence of StrA

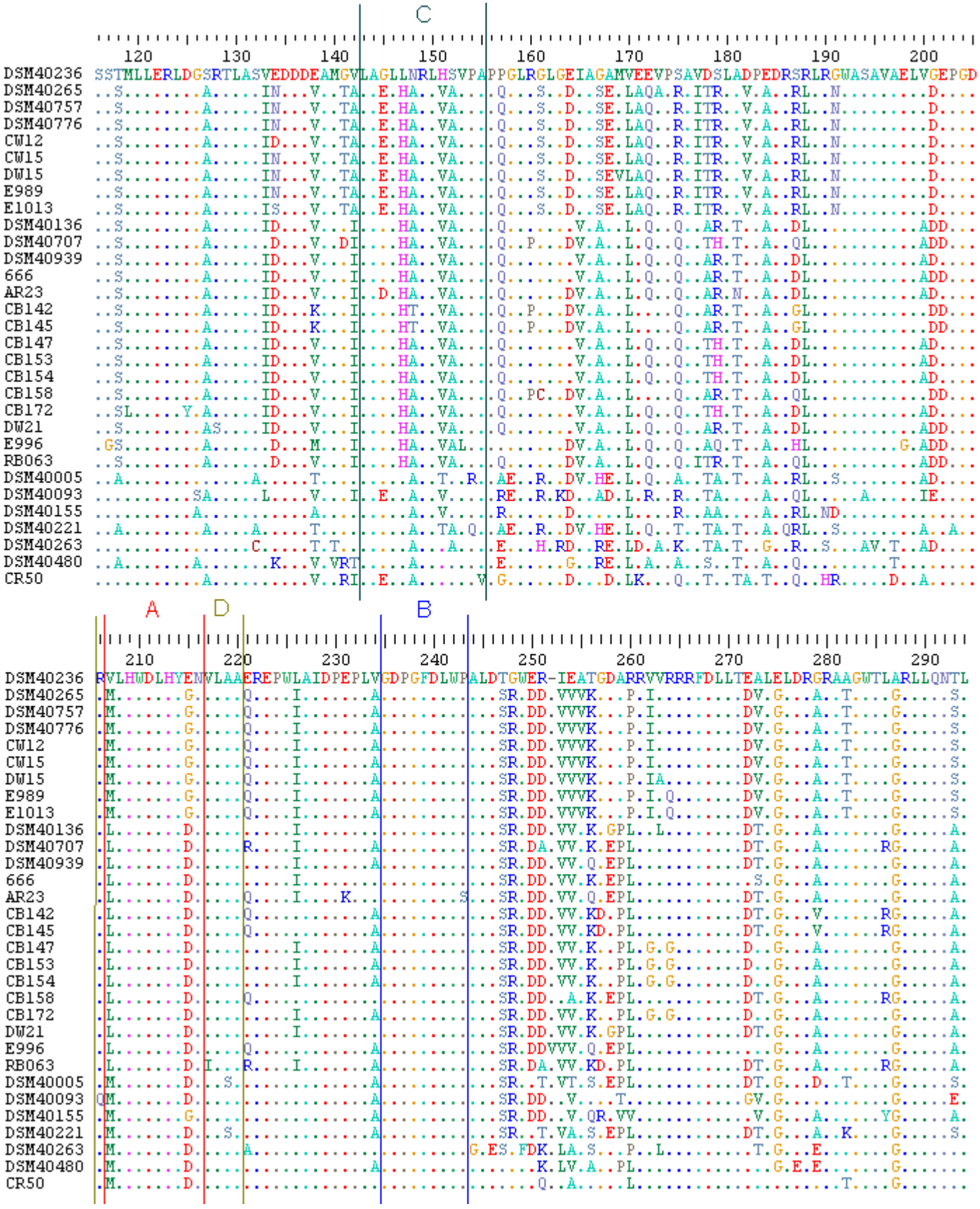


Figure 3.6. Amino acid alignment of *strA* gene. All sequences from Group A other than DSM 30236 have been omitted because their sequences are identical. Residues with similar physico-chemical characteristics are coloured alike. Residue grouping colours are as specified in BioEdit software package. The APH

universal catalytic site is flanked by red lines (A), the Mg^{2+} binding motif is flanked by blue lines (B), the putative nucleotide-binding domain is flanked by green lines (C) and the putative phosphorylation domain is flanked by brown lines (D).

The aligned amino acid sequences were displayed in the BioEdit sequence alignment editor (Wright, 2007) which uses the same colour for residues with similar physico-chemical characteristics (Figure 3.6). The PCR and sequencing reactions generated a sequence that could be translated into a 178 amino acid sequence out of a total of 307 for the full enzyme. Amongst all sequences there was only a single insertion, a valine at position 252 in strain E996, and no deletions. There was variation in the amount of sequence conservation across the protein; some of the regions, e.g. 164 to 188 vere higly diverse, while others, e.g. 227 to 247, had very few substitutions. This may be due to the fact that some regions are important for the functioning of the enzyme and mutations there will not be tolerated to the same extent as mutations to non-essential parts of the enzyme.

There are three conserved motifs in the aminoglycoside phosphotransferase protein family (Shaw *et al.*, 1993). The first is V--HGD----N, in Group A sequences it is VLHWDLHYEN and occupies the area 206 to 216 (Anderson *et al.*, 2002), which may be involved in the catalytic transfer of the terminal phosphate upon ATP catalysis. There are only two amino acid substitutions; Val-207 has been replaced by a methionine in Group B1 and most resistance genes in non-streptomycin biosynthetic clusters and a leucine in Group B2. Both these amino acids have similar physico-chemical characteristics, indicating that these substitutions have not altered the functioning of the enzyme. Glu-215 has been replaced with glycine in Group B1 and *S. glaucescens* and with aspartic acid in Group B2 and the cluster-associated genes. Aspartic acid has very similar physico-chemical characteristics to glutamic acid; however glycine is nonpolar rather than polar and has a neutral rather than positive side chain charge. This mutation may have affected the specificity of StrA in Group B1 strains. His-209 and Asp-211 however are invariably found in all functional members of this family; their presence in all the translated sequences implies that this domain is functional in all the examined proteins.

The second conserved motif is G--D-GR-G, in Group A sequences it is GDPGFDLWP and occupies the area 235 to 243, is a glycine-rich flexible loop, which is involved with Mg^{2+} binding and may be part of the nucleotide-binding site,

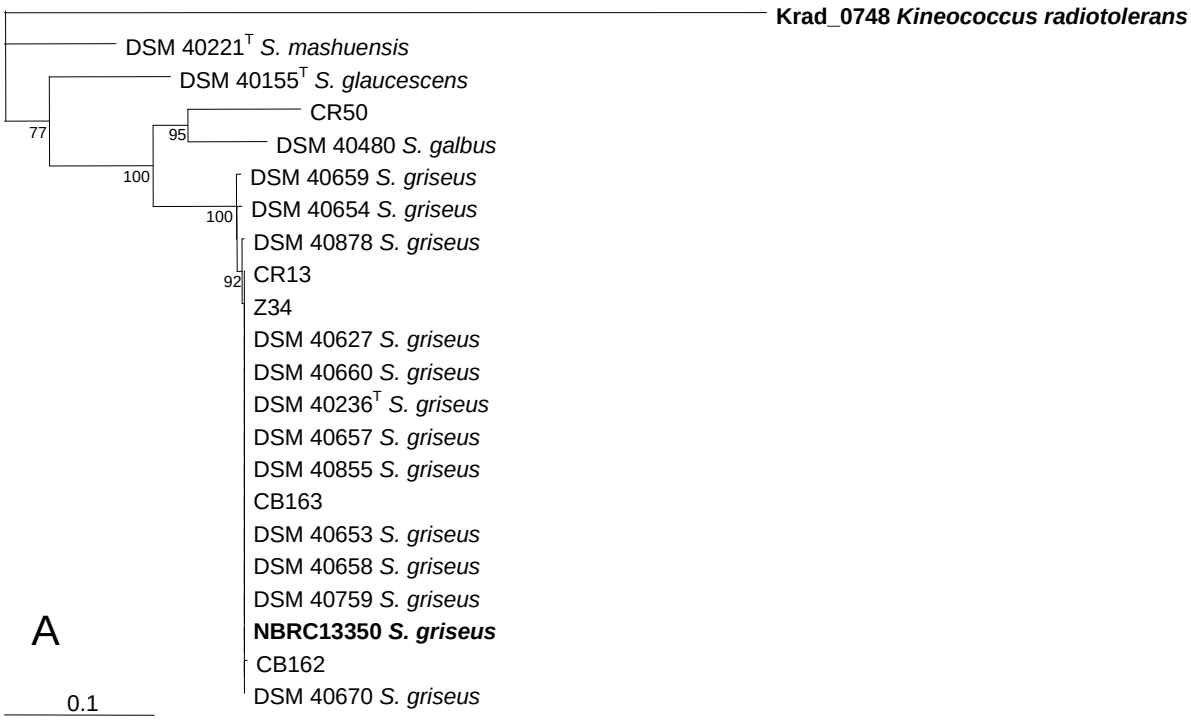
forming a ternary complex composed of the enzyme, ATP and Mg^{2+} (Shaw *et al.*, 1993). This motif is completely conserved in all isolates except for AR23, where Pro-241 has been replaced with a serine. The physico-chemical characteristics of these two residues differ; while both have neutral side chains, serine is polar and proline is nonpolar. The fact that Arg-240 is present, which is the most important residue of this motif, implies that this domain is also functional in all the examined proteins.

The third conserved motif is D--R/K--F/Y---LDE, which may be involved ATP hydrolysis and/or in a conformational change in the enzyme-aminoglycoside complex (Shaw *et al.*, 1993). No homologous region was found in the sequences used in this study. Furthermore, examination of the full StrA sequence also failed to reveal any such conserved regions. The original study which identified this motif (Martin *et al.*, 1988) only identified it in APH(3") rather than APH(6') sequences. Aligning the APH(3") sequences possessing this motif failed to reveal any homology of that region to the full *S. griseus* StrA, which may indicate an absence of that domain from APH(6') enzymes.

A previous study found two conserved regions on StrA (Distler *et al.*, 1987). The first was a putative nucleotide-binding domain whose sequence in *S. griseus* is LAGLLNRLHSVPA and occupies the area from the 143rd to 155th amino acid residue. This region is highly divergent in Group B sequences, having an average of only 65.9% sequence identity to Group A compared to the 80.0% of the APH phosphorylation site or the 99.5% of the glycine-rich loop. The amino acid substitutions also have very different physico-chemical characteristics, indicating that there is little conservation of function. The second domain is the APH universal catalytic site that was also identified by Shaw (1993), however Distler (1987) includes an additional amino acid in the 5' of the motif and four on the 3', considering RVLHWDLHYENVLAA to be the conserved region. These additional amino acids are also highly conserved; the only exception being the replacement of Val-217 with an isoleucine residue in RB063, though since both are non-polar and hydrophobic the functioning of the enzyme is presumably unaffected. The fact that they are conserved implies that these additional amino acids are also important.

3.3.7 Phylogeny of the *strW* biosynthetic gene

Only strains belonging to Group A as well as CR50, *S. glaucescens* DSM 40155^T, *Streptomyces mashuensis* DSM 40221^T and *Streptomyces galbus* DSM 40480 possessed *strW*, indicating that strains in Group B lacked the ability to synthesize streptomycin. *Streptomyces cinnamoneus* DSM 40005^T and *S. humidus* DSM 40263^T are known producers of aminoglycosides, and *S. netropsis* DSM 40093 produces an aminocyclitol antibiotic which is closely related to the aminoglycosides, however no *strW* was amplified from them. This is not necessarily a limitation of the primers used, as the entire spectinomycin biosynthetic cluster of *S. netropsis* DSM 40093 is available and attempts to align *strW* with the cluster fail to reveal any homologous genes despite it possessing an *strA* homologue. Presumably the clusters lacking *strW* rely on a non-homologous gene that serves to export the synthesised antibiotic. The *strW* biosynthetic gene tree (Figure 3.7) showed significant similarity to the tree of *strA* (pairwise analysis r=0.997), indicating that the evolutionary history of biosynthetic genes is the same as for resistance in producers and that therefore there has been no separate HGT events within the cluster in the strains examined.



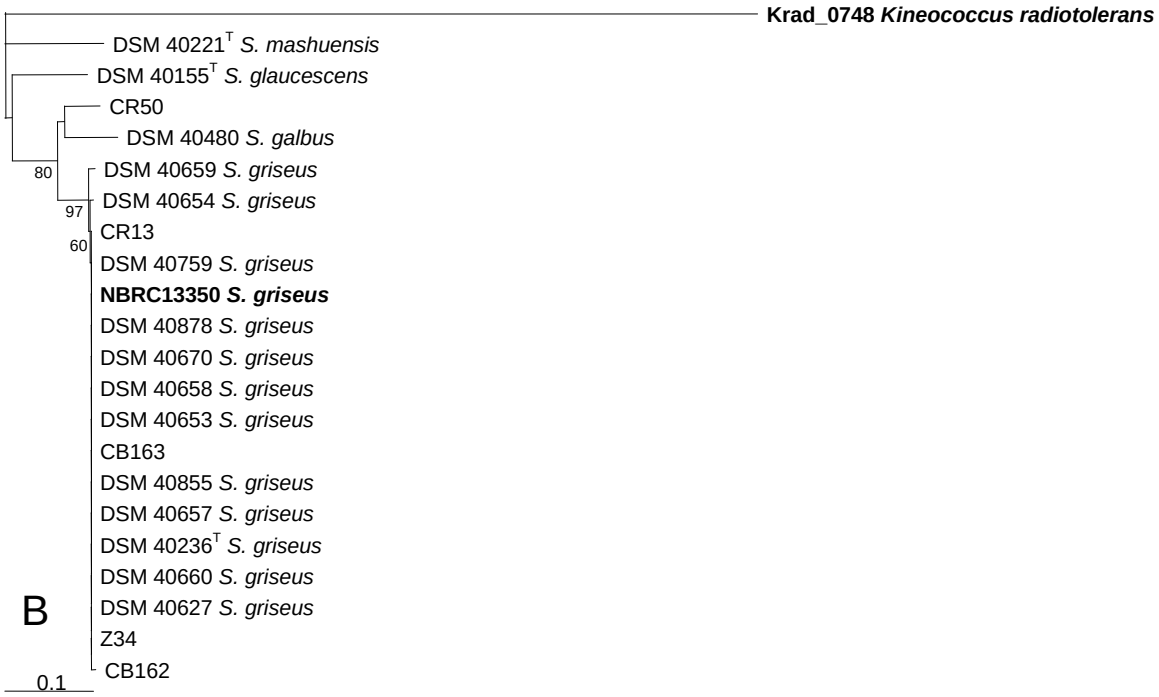


Figure 3.7. An (A) nucleotide (B) amino acid sequence phylogenetic tree of the streptomycin biosynthetic gene *strW*. Sequences in bold were acquired from GenBank. The tree was constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% nucleotide dissimilarity.

3.3.8 Screening for alternate resistance mechanisms

3.3.8.1 Mutations in the rpsL gene

Conserved primers for *rpsL* were designed based on the consensus sequence from GenBank sequences (rpsL_F: GGC AGG ACA AGG TCG AGA AG, rpsL_R: TCC TTC TTG GCG CCG TAG C). The area amplified included the 129th nucleotide position where a point mutation can generate resistance. The *rpsL* genes from all the strains screened for *strA* were sequenced (nucleotide tree: Figure 10.17 amino acid tree: Figure 10.18). The nucleotide sequences were then translated to find non-synonymous substitutions. Lys-43 was replaced with an Asn in M110 and a Thr in CB148, Lys-88 was replaced with an Arg in CB141 and an additional Arg was inserted between Gly-85 and Arg-86 in CB158 (Figure 3.8).

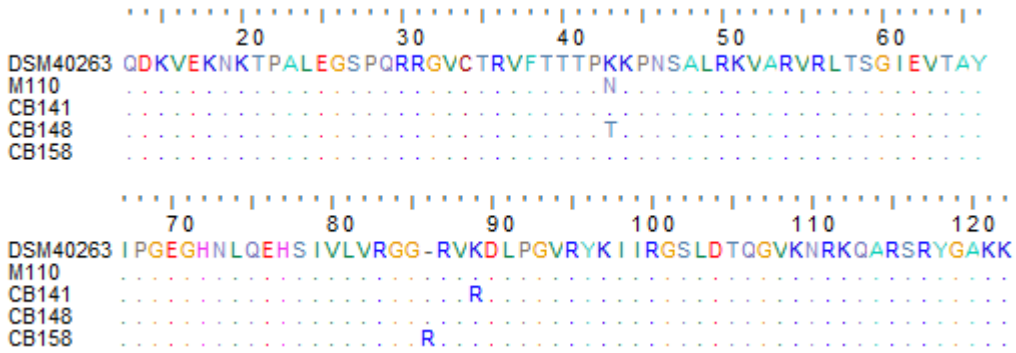


Figure 3.8. Alignment of the translated *rpsL* sequences. All strains not included had sequences that were identical to that of *S. griseus* DSM 40236^{T}.

3.3.8.2 Mutations in the rrnA gene

The 530 loop region on the 16S rRNA gene was also examined for mutations using the pA/pH 16S primer set (Edwards *et al.*, 1989) (16S phylogenetic analysis is in section 5.3.1). None of the strains included in this study had any mutations in the 530 loop region of the 16S rRNA.

3.3.8.3 Presence of aphE

Attempts to construct primers from the consensus sequence of *aphE* and *MAB2385*, a putative streptomycin phosphotransferase from *Mycobacterium abscessus* failed to generate any primers that could be used to amplify *aphE*. As there were no sequences with a greater identity than that of *MAB2385*, the primers used were created using only the *S. griseus aphE* sequence as a template. A slightly divergent version of *aphE* was amplified with these primers, which was then used to design the consensus primers used (aphE_d2F: ACC CGG TAC GCC AAG TGC G, aphE_d3R: CCC AGG TCG ATG AAG CCC G, 50°C annealing).

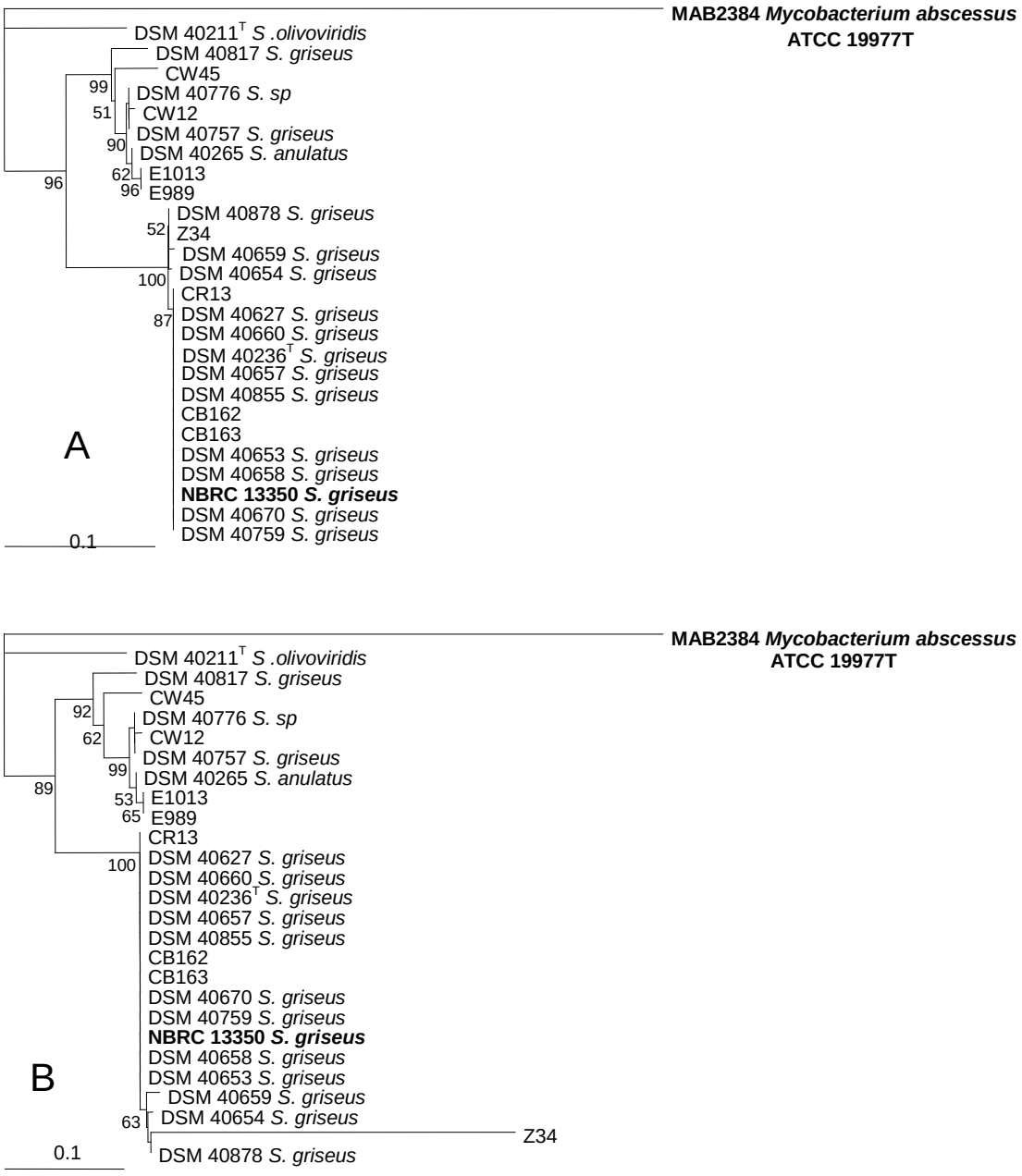


Figure 3.9. An (A) nucleotide (B) amino acid sequence phylogenetic tree of the streptomycin resistance gene *aphE*. Sequences in bold were acquired from GenBank. The tree was constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% nucleotide dissimilarity.

The great branch length for Z34 in Figure 3.9.B is due to an insertion mutation halfway through the sequence which resulted in a frameshift, indicating that the gene is non-functional in that strain. The *aphE* gene was amplified from the strains belonging to Group B1 and Group A on the *strA* phylogenetic tree (Figure 3.3) with the exception of DW15 which appears to lack that gene. *Streptomyces olivoviridis* DSM 40211^T was the only strain lacking *strA* from which an *aphE* homologue was amplified and is sensitive to streptomycin, indicating that its gene is either inactive

or too divergent to protect against the antibiotic. Pairwise analysis revealed a correlation of 0.933 for the distances between the *strA* and *aphE* nucleotide sequences, indicating that these two resistance genes have a similar evolutionary history.

3.3.9 Screening for viomycin biosynthetic and resistance genes

In order to determine whether the findings regarding the streptomycin resistance genes are unique to it or extend beyond that specific gene, it was decided to examine the distribution of a second antibiotic resistance gene and the genes associated with its biosynthesis. A list of antibiotics other than streptomycin produced by *S. griseus* was compiled, which included borrelidin, candicidin, chromomycin A3, griseorhodin, griseusin, rhodomycin, streptothricin and viomycin (Laskin & Lechevalier, 1988). From these the viomycin biosynthetic cluster was selected, as the strain *Streptomyces floridae* DSM 40938^T that had already been included in this study was known to possess it (Wink, 2009). Furthermore viomycin was available and could be used to screen the strains used in this study for resistance (Section 2.2.2) Also, the capreomycin biosynthetic gene cluster from *Saccharothrix mutabilis* was homologous to the viomycin cluster, which enabled the alignment of genes from the two clusters and the subsequent identification of conserved regions within them that were used to generate primers (Table 3.7). The primers, which were from across the entire cluster (Figure 3.10), were tested on *S. puniceus* DSM 40083^T, another strain know to possess the cluster, and the genes selected for the screening were *vioB*, *vioD* and *vph*, as their primers generated the least non-specific amplification (Figure 3.11) and were located near the two ends of the cluster which would reveal if one of the two cluster ends was absent.

Table 3.7. Primers designed to amplify genes from the viomycin biosynthetic gene cluster.

Gene	Protein Function	Primer name	Primer sequence
vioB	2,3-diaminopropionate synthase	vioB_F	TGA AGC TGG AGG GCC RCT CG
		vioB_R	GGG CAC SCC GCC GAA CAC
vioD	capreomycidine synthase	vioD_F	CCG CCT ACC ACT CGC TGT CG
		vioD_R	TCG GCG TGC TCC ACG GCG AC
vioE	permease	vioE_F	GGT TCT GGT GGG GCG GCA CC
		vioE_R	CGT CAG GGC SCG CAC CAG CG
vioJ	2,3-diaminopropionate alpha,beta-desaturase	vioJ_F	GTC SGT GCA GAC CGA CAT CG
		vioJ_R	TCG GCG TCG AGT TCG GCS AG
vioK	ornithine cyclodeaminase	vioK_F	CCC GCC GAG GGC TAY CTG C
		vioK_R	CTG CTG CCC GAG GTG TTC
vph	viomycin phosphotransferase	vph_F	GTT CCA CSA YGT GKT GAT CG
		vph_R	CCA GGT CCT CMG CCG GGT CG

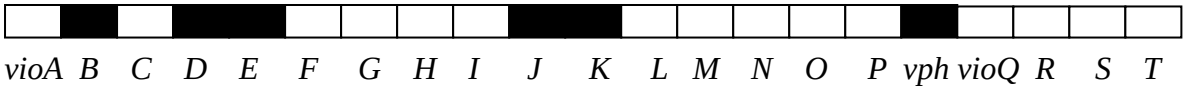


Figure 3.10. Viomycin cluster gene order. Genes for whom primers were generated are coloured black while the remaining genes are coloured white.

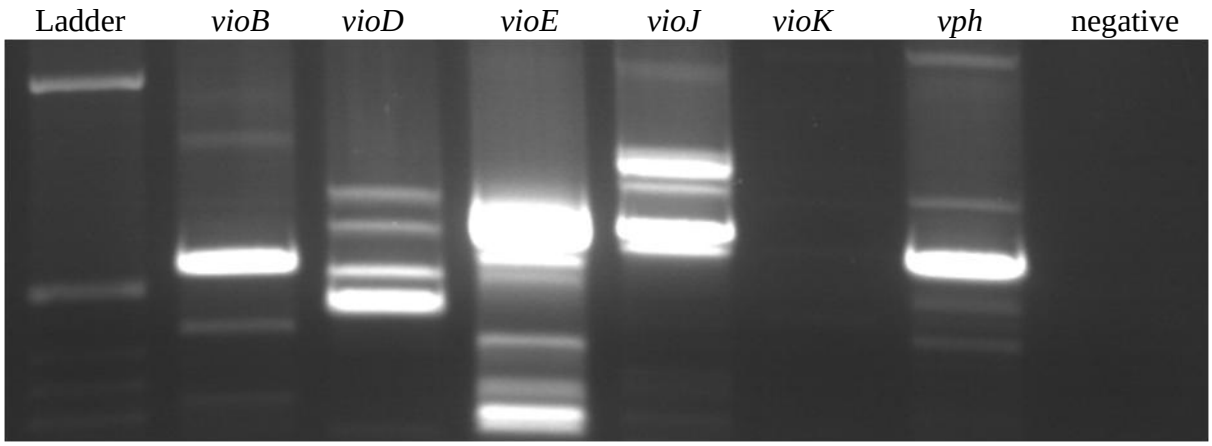


Figure 3.11. Amplification of viomycin cluster genes from *S. puniceus* DSM 40083^T to examine primer performance.

All the strains that that were screened for streptomycin genes (Table 3.5, Table 3.6) were also screened for viomycin resistance and production genes. All three genes were successfully amplified from *S. californicus* DSM 40058^T, *S. vinaceus* DSM 40257, *S. floridae* DSM 40938^T and the soil isolate CB157. These strains showed high level resistance to viomycin. The *vph* primers also generated a product from *S. bacillaris* DSM 40598^T. The blastn algorithm was used to search for homologous sequences to the *S. bacilliaris* amplicon in GenBank and revealed that the amplicon

had almost the same sequence identity to *SGR421* (80.7%), an *S. griseus* putative phosphotransferase, as to *vph* (81.1%). A new primer set was designed based on *SGR421* and its *S. bacillaris* homologue (421_F: GGT GGT STG CTT CGC CC, 421_R: CGT CCC AGT CGA TGA, 56°C annealing) and used to screen all the strain used in this study.

The blastn algorithm was used to screen the *S. griseus* NBRC 13350 genome for each individual gene from the viomycin cluster (Figure 3.10) in order to determine whether there were any additional homologues present. Other than the homology of *SGR*421 to *vph*, *vioS* had a 70% sequence identity to *strK* from the streptomycin cluster. The homology of these genes is presumably due to them being viomycin and streptomycin phosphate phosphatases respectively. The *vioR* gene had a 90% sequence identity to *SGR878* which was listed as a putative transcriptional regulator. *SGR878* was located 574 kb on the 3' direction from *SGR421*. There was no significant identity between the remaining cluster genes and the *S. griseus* chromosome.

Table 3.8. Presence of viomycin genes in streptomycete isolate groups. Screened denotes the number of strains that underwent PCR and plating on viomycin-containing plates. 0, 10 and 50 µg/ml indicates the number of strains that were able to grow in that concentration of viomycin. The gene names denote the number of strains from which they were amplified and sequenced from.

Isolates	Origin	Screened	0 µg/ml	10 µg/ml	50 µg/ml	*vph*	*SGR421*	*vioB*	*vioD*
Tolba	Various	9	5	3	1	0	1	0	0
Cayo Blanco	Cuba	36	17	8	2	1	2	1	1
E	UK	8	6	2	0	0	0	0	0
RB	UK	1	0	1	0	0	0	0	0
DSM	Various	52	28	3	21	3	17	3	3
Z	Zambia	1	0	0	1	0	1	0	0
BTG	Chemostat	8	8	0	0	0	0	0	0
Total	Various	115	64	17	25	4	21	4	4

The number of strains screened for resistance was quite small, so no firm inferences can be drawn about the resistance profile of the strains in question (Table 3.8). From the Cuban samples 25.0% of streptomycete isolates had low level resistance (10 µg/ml) and 5.56% had high level resistance (50 µg/ml). 33.3% of the UK isolates had low level and 0.0% had high level resistance. There was no significant

difference between the two soils either at the low (χ^2 p = 0.9489) or high (χ^2 p = 0.8616) levels of resistance.

Of the strains possessing *SGR421*, 9.5% demonstrated low level and 90.5% high level resistance to viomycin. This indicates that despite the divergence of *SGR421* from *vph* it is capable of providing protection from that antibiotic. Only 24.0% of the highly resistant strains lacked both *SGR421* and *vph*, though these genes were absent from 88.2% of the strains with low level resistance.

The very small branch lengths of the viomycin gene phylogenetic trees indicate that all isolated genes are very closely related (Figures Figure 3.12, Figure 3.13, Figure 3.14). The vph gene producers had 99.0% to 98.4% nucleotide and 100% to 99.4% amino acid sequence identity to one another, the *S. bacilliaris* clade 100% to 97.8% nucleotide and 100% to 97.0% amino acid sequence identity to one another and the *S. griseus* clade 100% to 98.8% nucleotide and 100% to 98.2% amino acid sequence identity to one another (Figure 10.3, Figure 10.4). There is variation in the topology of the trees; the nucleotide sequence tree of *vioB* has *S. californicus* branching out first (Figure 3.13A), however the amino acid sequence tree of *vioB* and those of *vioD* have *S. vinaceus* branching out first instead (Figure 3.13B, Figure 3.14). Conversely, the *vph* trees have CB157 in that position (Figure 3.12). These apparent contradictions in tree topology have too low bootstrap values to be viewed as being significant and pairwise analysis reveals that the distances between the genes are highly congruent (*vioB*/*vioD* r=0.738, *vioB*/*vph* r=0.825, *vioD*/*vph* r=0.747); as a result they are unlikely to be due to recombination of the genes in question.

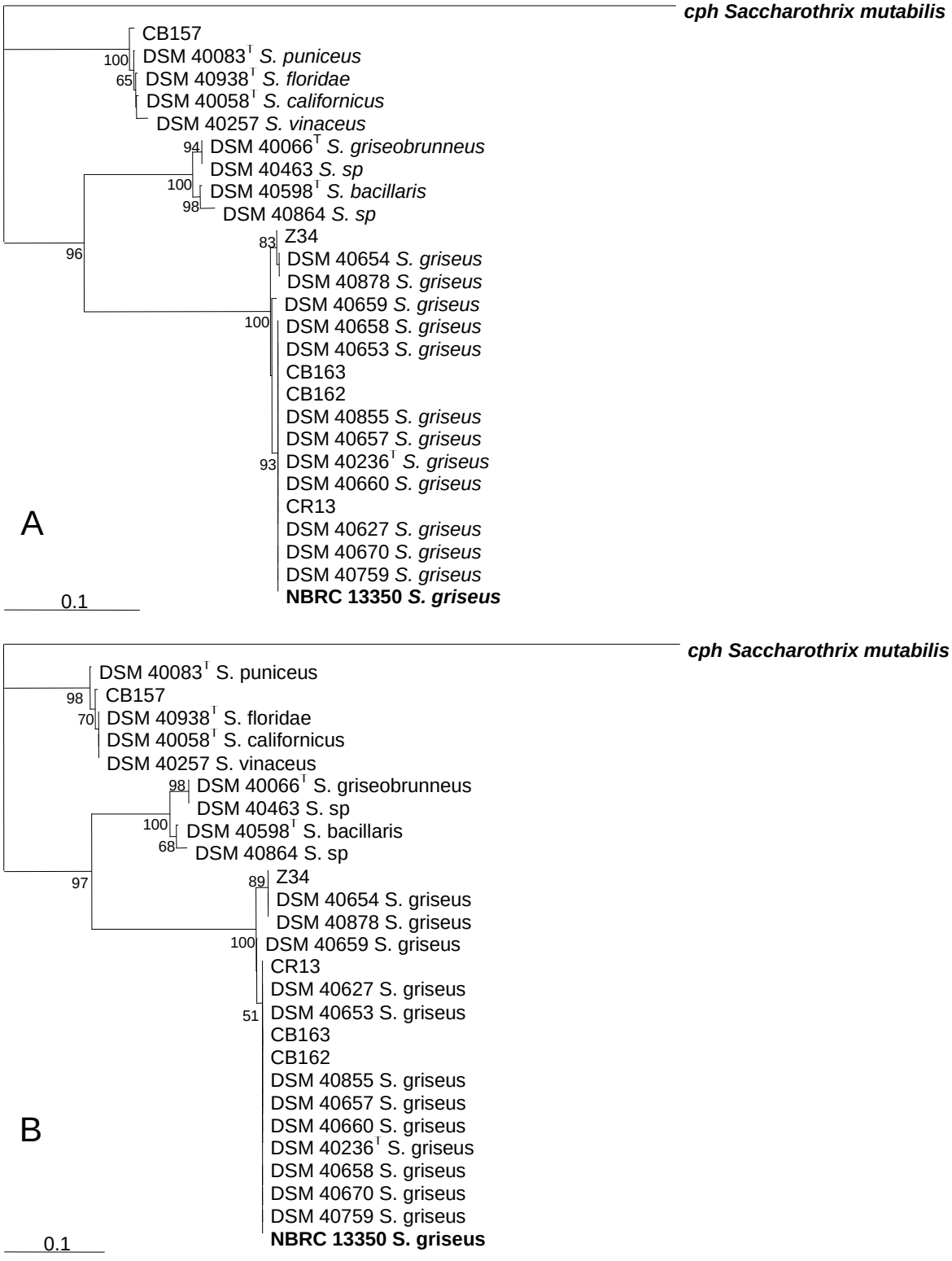


Figure 3.12. An (A) nucleotide (B) amino acid sequence phylogenetic tree of *vph*. Sequences in bold were acquired from GenBank. The tree was constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% sequence dissimilarity. The *cph* gene, from the capreomycin cluster of *Saccharothrix mutabilis*, was used as a root.

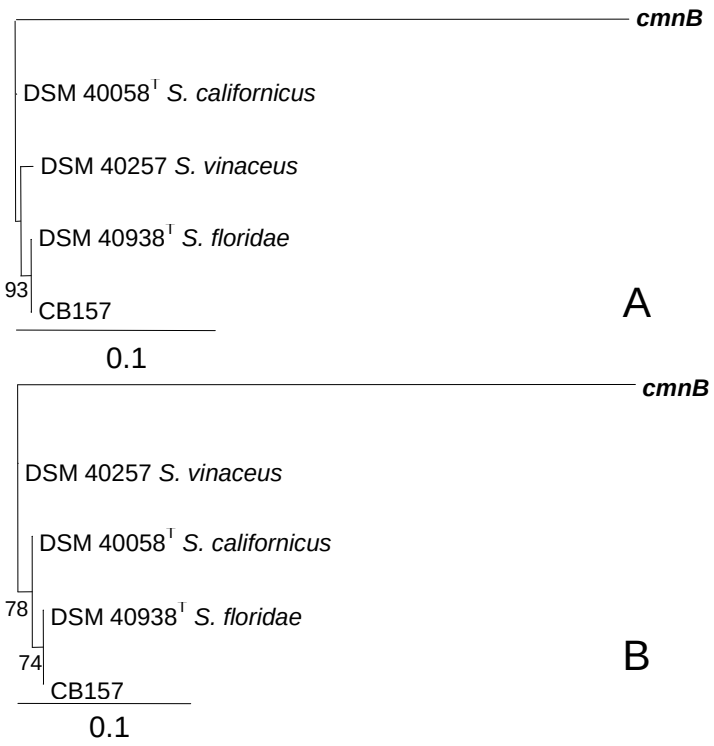


Figure 3.13. An (A) nucleotide (B) amino acid sequence phylogenetic tree of *vioB*. Sequences in bold were acquired from GenBank. The tree was constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% sequence dissimilarity. The *cmnB* gene, from the capreomycin cluster of *Saccharothrix mutabilis*, was used as a root.

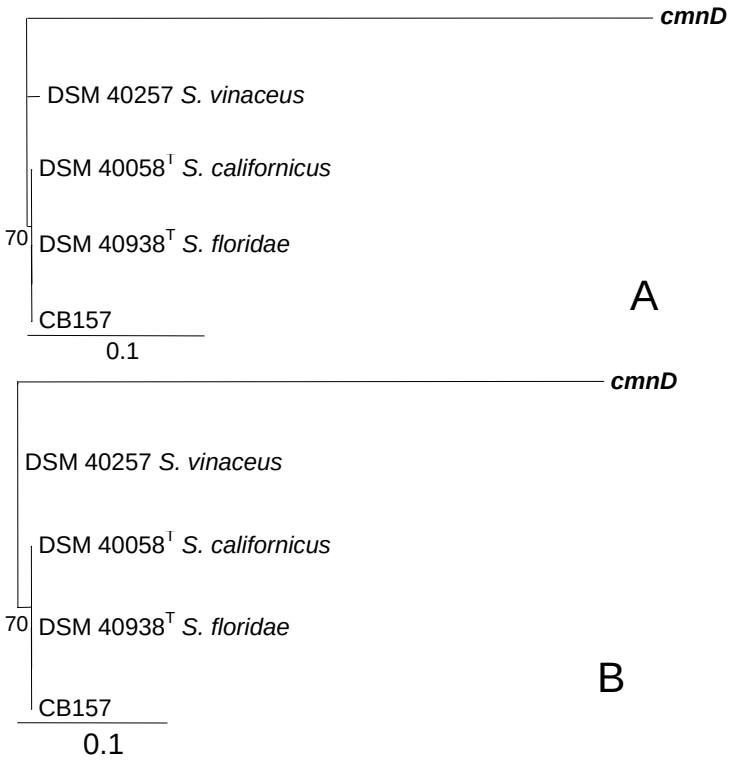


Figure 3.14. An (A) nucleotide (B) amino acid sequence phylogenetic tree of *vioD*. Sequences in bold were acquired from GenBank. The tree was constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates

10% sequence dissimilarity. The *cmnD* gene, from the capreomycin cluster of *Saccharothrix mutabilis*, was used as a root.

3.3.10 Screening for other secondary metabolite genes

3.3.10.1 Spectinomycin biosynthetic genes

The uncertain phylogenetic position of the spectinomycin and dihydrostreptomycin resistance genes raised the possibility that the Group B resistance genes are more closely related to them and may thus be flanked by biosynthetic genes from the spectinomycin or dihydrostreptomycin clusters rather than the streptomycin cluster. No sequences are available for the dihydrostreptomycin cluster and thus primers could not be designed for its biosynthetic genes, however the spectinomycin cluster sequence was available. It was unnecessary to specifically search for hydroxystreptomycin biosynthetic genes, as the *strW* primers were designed based on the homologue from that cluster in addition to *strW* from the streptomycin cluster.

Primers were designed for twelve biosynthetic genes from the spectinomycin cluster, selected for the presence of homologues in the GenBank database that could be used to determine conserved regions for primer design. Genes with no homologues were ignored. The genes selected are listed in Table 3.9 and illustrated on Figure 3.15.

Table 3.9. Primers designed to amplify biosynthetic genes from the spectinomycin cluster.

Gene	Protein Function	Primer Name	Primer Sequence
spcN	spectinomycin phosphotransferase	spcN_F	AGGCCCACSGTGTCCCAGTC
		spcN_R	GTTCGYCACCGTCTCCGACC
spcT	putative spectinomycin export protein	spcT_F	CGGCGACSGGSACCGCGAGC
		spcT_R	CCMCGCCGCGTCATGATCG
spcA	putative myo-inositol monophosphatase	spcA_F	TCGTAGTAGGCGYCGAGGCG
		spcA_R	TTCGACCGCRGGAGCGAGG
spcB	putative myo-inositol dehydrogenase	spcB_F	SAGCGAGGACCAGTACATCG
		spcB_R	GTTCCTCGCGGAACGCCTCG
spcC	putative aminotransferase	spcC_F1	CAACGCCGAGGTGTGGGACA
		spcC_F2	TACATCGACTTCTTCACCGG
		spcC_R1	GAGCCGTTCTCGATGGACTG
		spcC_R2	GTGATGACCTCGTCCAGGAT
spcD	putative epimerase	spcD_F	TTCCGCGACATGTGCGACCG
		spcD_R	GGAAGTCACCGCTGCCGCGC
spcX	homolog to a multiple sugar binding protein	spcX_F1	SCTTCCTCACCGACCACACC
		spcX_R1	GAGCGTCATGCCCGCCTCGC

		spcX_F2	ACATGCGGCTCAACGGCCCC
		spcX_R2	GTGTCGAGGTCGTAGAACTG
spcE	putative aminotransferase	spcE_F	CCCSTCATCGAGGACTGCTC
		spcE_R	CGAGCAGCGCCTGGTGGAGG
spcF	putative glycosyltransferase	spcF_F	TSGACGACGACAACGAGTGG
		spcF_R	TGTTGGAGTAGCCGCCCAGG
spcG	putative methyltransferase	spcG_F	CACGGCTCCTACCACTGGGG
		spcG_R	CGCACGTGCTGCTCCAGGCC
spcJ	putative dehydratase	spcJ_F	TCGGTGCCGCCGCCGATGTG
		spcJ_R	CGCACGTGGACCGGTCCATC
spcK	transposase homolog	spcK_F	GCTCGGCGACGGCTCGCACC
		spcK_R	AGCCGATCTTCAGGCCCTGG

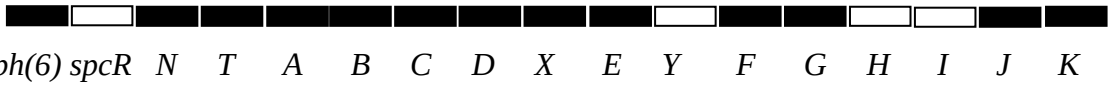


Figure 3.15. Spectinomycin cluster gene order. Genes screened for via PCR are colored black while genes not are coloured white.

A preliminary screen was performed on strains 666, DW15, DW21, CR13, CR50, AR23 along with DSM 40093 (*S. netropsis*) as a positive control. All primer sets other than spcJ_F/R either produced solely non-specific amplification or no amplification in strains other than the positive control. However, sequences homologous to *spcJ* were amplified from strains 666 and DW21 and it was thus decided to screen all the strains used in this study with the *spcJ* primer set. Homologues for the sequences generated were obtained via blastn and used in the generation of a phylogenetic tree (Figure 3.16). The sequences were translated to generate an amino acid tree for SpcJ (Figure 3.17).

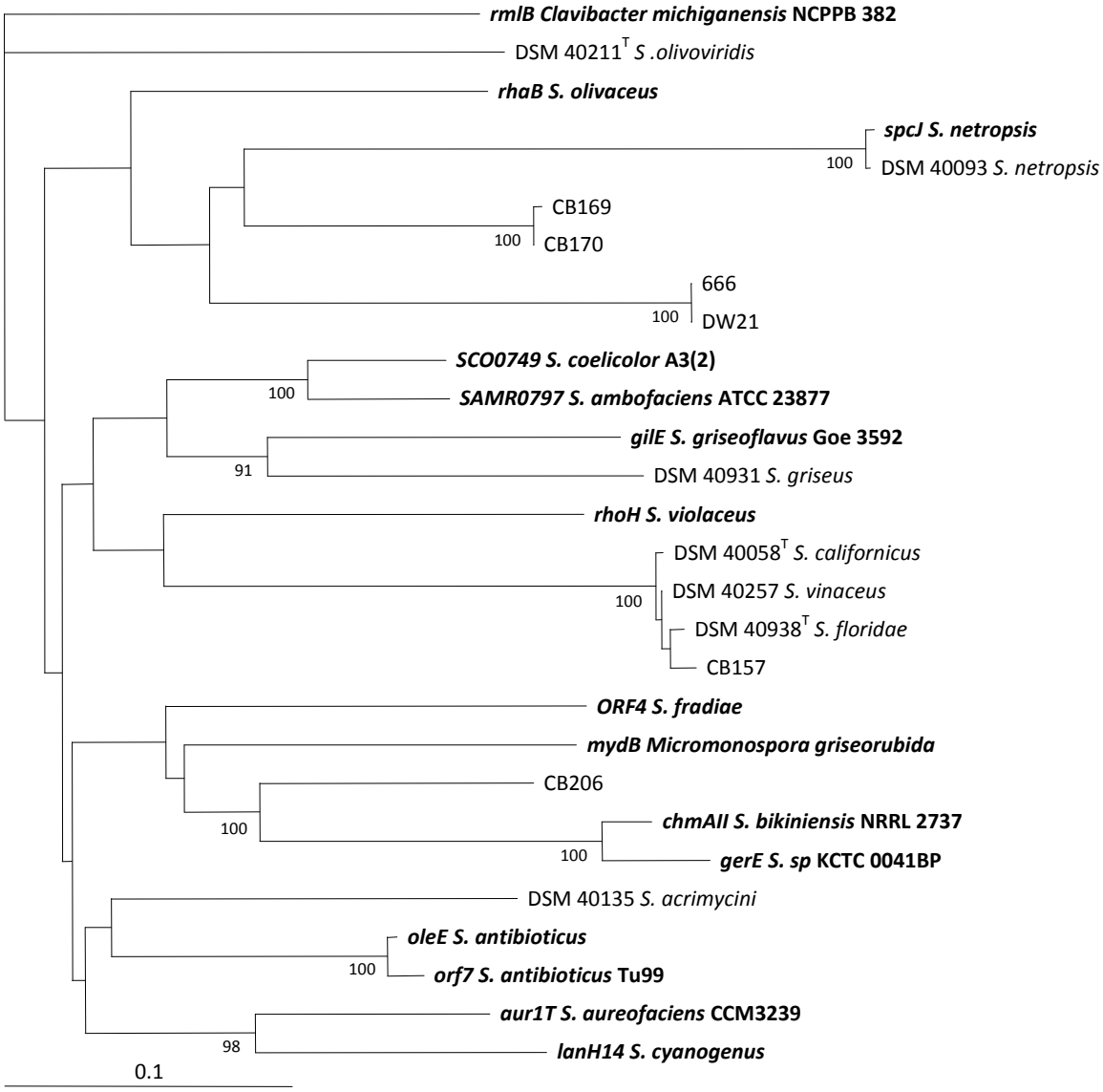


Figure 3.16. A phylogenetic tree of *spcJ* and other dehydratase-encoding genes. Sequences in bold were acquired from GenBank. The tree was constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% nucleotide sequence dissimilarity. Sequences in bold were obtained from the GenBank database.

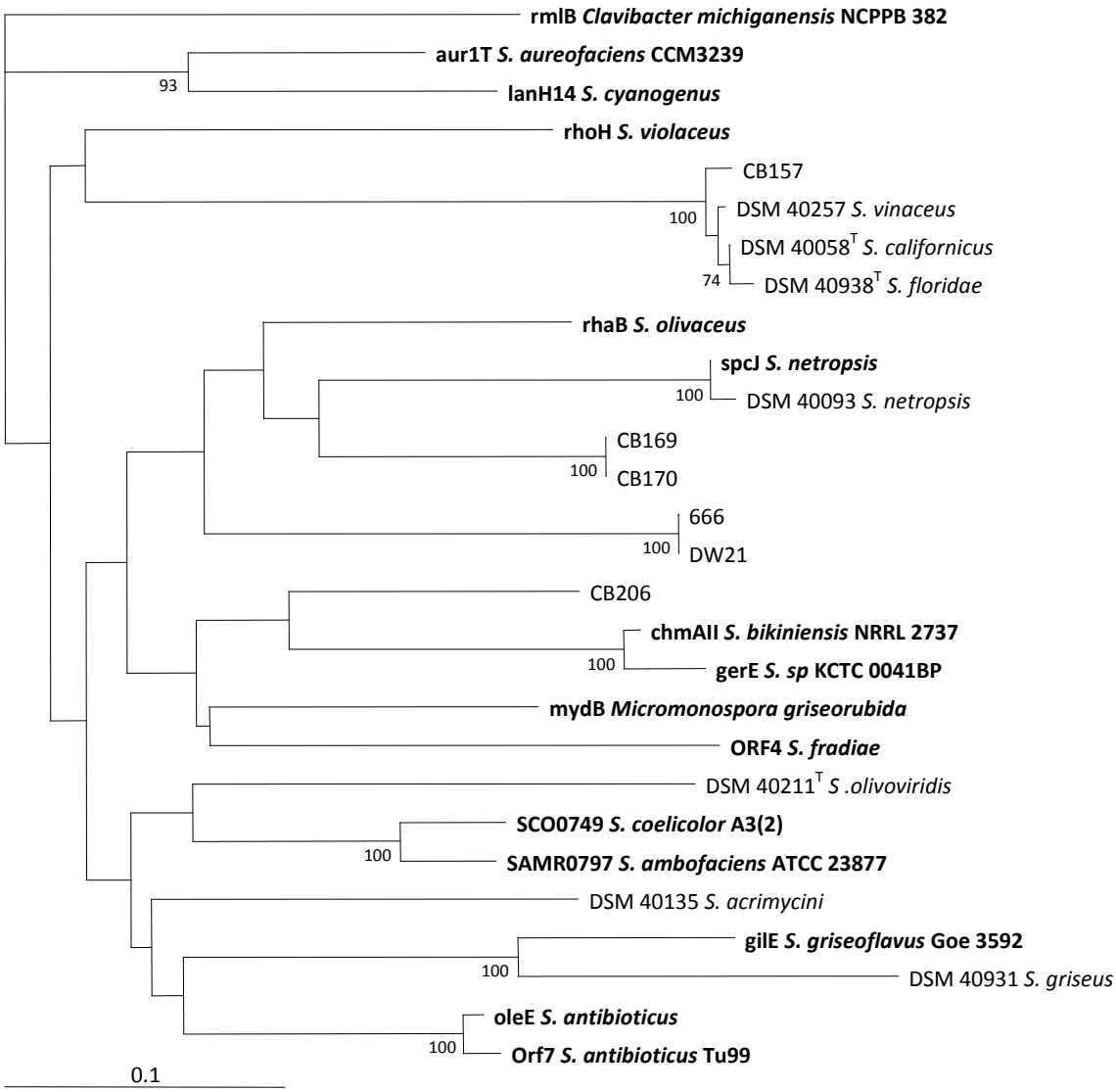


Figure 3.17. A phylogenetic tree of *spcJ* and other dehydratase-encoding genes constructed using translated protein sequences. Sequences in bold were acquired from GenBank. The tree was constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% amino acid sequence dissimilarity. Sequences in bold were obtained from the GenBank database.

The majority of sequences clustered with dehydratases other than the one from of the spectinomycin cluster. The clade that included CB157, *S. vinaceus* DSM 40257, *S. californicus* DSM 40058^T and *S. floridae* DSM 40938^T was most closely related to *rhoH*, part of the rhodomycin biosynthetic cluster. The *S. acrimycini* DSM 40135^T gene formed a clade along with *oleE*, belonging to the oleandomycin cluster, though the low bootstrap value and sequence identity (82.6%) means that they are not closely related. The CB206 sequence was related to *chmAII*, from the chalcomycin biosynthetic cluster, and *gerE*, from the dihydrochalcomyin cluster. The *S. griseus* DSM 40931 gene clustered with *gilE*, from the gilvocarcin

biosynthetic gene cluster. The *S. olivoviridis* DSM 40211^{T} sequence did not cluster with any other genes on the nucleotide tree, but appeared related to *SCO0749* and *SAMR0797*, two putative dehydratases from *S. coelicolor* A3(2) and *S. ambofaciens*, on the amino acid tree.

The sequences from 666, DW21, CB169 and CB170 fell in the same clade as *spcJ*, however the nodes linking them had very low bootstrap values and low sequence identities to *spcJ* (77.0% for CB169, 77.3% for CB170, 76.1% for 666 and 76.1% for DW21). As the sequence identities were approximately equal to that of *rhaB*, from the NDP-L-rhamnose biosynthetic cluster, to *spcJ* (77.9%), these dehydratases presumably belong to clusters that produce compounds other than spectinomycin.

3.3.10.2 Conserved secondary metabolite genes

In order to compare the distribution pattern and phylogeny of *strA* with those of other secondary metabolite genes in the same set of strains, clusters were suggested by Alice Yaxley that are widespread in the *Streptomyces* genus to increase the probability of them being present in strains possessing *strA*.

The arm regions of the chromosomes were aligned to find conserved secondary metabolite genes as they tend to be found in those areas of the streptomycete chromosome (Paradkar *et al.*, 2003). The left arm region of the *S. coelicolor* A3(2) chromosome extends from approximately 0 to 1.5 Mb and the right from 6.4 to 8.7 Mb (Bentley *et al.*, 2002). The genes located approximately on the boundaries between the arm regions and the core were SCO1407 and SCO5845. The homologues of SCO1407 were SCAB76041 on the *S. scabies* chromosome and SAV2414 on the *S. avermitilis* chromosome and those of SCO5845 were SCAB76041 and SAV6945. The three genomes were aligned using Artemis (Rutherford *et al.*, 2000) and the sequences on the 5' end of SCO1407 and its homologues as well as the 3' end of SCO5845 and its homologues were aligned to find genes conserved in all three genomes. There were two clusters and one gene found in the arm regions conserved between all three species; a hopanoid cluster (SCO6771-SCO6759, SCAB12881-SCAB13001, SAV1643-SAV1654), a carotenoid cluster (SCO0196-SCO0185, SCAB5431-SCAB5511, SAV1025-SAV1019) and a geosmin biosynthetic gene (SCO6073, SCAB21021, SAV2163).

Blastn was used to find homologues from other streptomycete species for these genes and those along with sequences from the three genomes were used to identify conserved areas from which primers were generated. PCR was then used to amplify these sequences in the strains *S. griseus* DSM 40236^T, *S. violaceoruber* DSM 40783 and *S. glaucescens* DSM 40155^T. Those that successfully amplified from all three species were then tested on *S. platensis* DSM 40041^T and *S. limosus* DSM 40131^T to confirm that these primers were sufficiently conserved to amplify from multiple *Streptomyces* species. The PCR conditions used are described in Section 2.3.1 and the annealing temperature used was 55°C.

Table 3.10. Primers designed to amplify conserved secondary metabolite genes. A (*S. griseus* DSM 40236^T), B (*S. violaceoruber* DSM 40783), C (*S. glaucescens* DSM 40155^T), D (*S. platensis* DSM 40041^T), E (*S. limosus* DSM 40131^T). A minus indicates no amplification, while the number of crosses indicates the amount of product generated during the PCR.

Gene	Primer Name	Primer Sequence	A	B	C	D	E
geoA	geoA_F1	TGC CGG MCA TCC AGT TCT GC	-	+++	+++		
	geoA_R1	TAC GGC GAC GAC TAC TAC CC					
	geoA_F2	CGC ATC TCG ATG TAC TCG AC	-	++	++		
	geoA_R2	ATG GGC ATG CTG GAG GGC TC					
crtE	crtE_F	CTC CTT CCA GGC SCT GTA CG	+++	+++	+++	-	+++
	crtE_R	CCG TCT GSG SGA AGG TRT GG					
crtI	crtI_F	CTG GGC SGA CGA CAT CGT SG	++	++	++	-	-
	crtI_R	GCA GGA AGT TGS TSA GCT GG					
crtU	crtU_F	GGC CGG TCS AGC CAS AGC CG	++	++	++	-	++
	crtU_R	STT CCG SCA GTA CTA CAA CC					
SCO0195	195_F	CCC GAG CAG GCG CAG GAC A	+	-	++		
	195_R	GTC CTC GGC GCT GCC CCA GG					
hopD	hop60_F	GCA GCA GCC GGT AGC CCT CG	+	+	+	-	++
	hop60_R	GTC GAC GAC ATC GGC GAC GG					
hopB	hop62_F1	TGT TGC CCT CCA TCT CCA GG	-	+	+		
	hop62_R1	AGA CTG TGC CCA CTG TGC CC					
	hop62_F2	GTC GCT CCA GGT CTG CTT GC	+++	+++	+++	-	+++
	hop62_R2	ACA CCG TCG CCG CCT ACC AC					
SAV1649	hop64_F	TCT GGA GCG TCC ACG ACC AC	++	+	+	-	+
	hop64_R	TCG CST GCG CGC TCG GCA TC					
SAV1648	hop65_F	GTA GCC GCA GTG CGC CAT GC	+++	++	+++	-	-
	hop65_R	CAG ATC GAC GAG ATC GTS CG					
ispG2	hop67_F	TCT CCA CGA TCT TCG ACT CG	-	-	+		
	hop67_R	GAT CTC GGT CAA GCA CAA CG					
dxs1	hop68_F	GCT CGG CGA TSC CGA CGT CC	-	++	-		
	hop68_R	TCG GCG GCC TCG CCA ACC AC					
SAV1645	hop69_F	ATC AGC TTG ATC ACC TCC AG	+	+	++	-	-
	hop69_R	GCT GAA GTT CGC CCG GTA CG					

One set of primers (Table 3.10) was selected from each cluster for further use: geoA_F2/R2 for *geoA*, crtE_F/R for *crtE*, and hop62_F2/R2 for *hopB*. The all had an optimal annealing temperature of 55°C.

These three primer sets were then used to amplify and sequence the genes from the strains 666, DW15, DW21, CR13, CR50, AR23, CW12, CW45 and *S. griseus* DSM 40236^{T}. *GeoA* failed to amplify from strains CR13 and *S. griseus* DSM 40236^{T}, c*rtE* from strains DW15 and CW12; however *hopB* was successfully amplified and sequenced from all the strains.

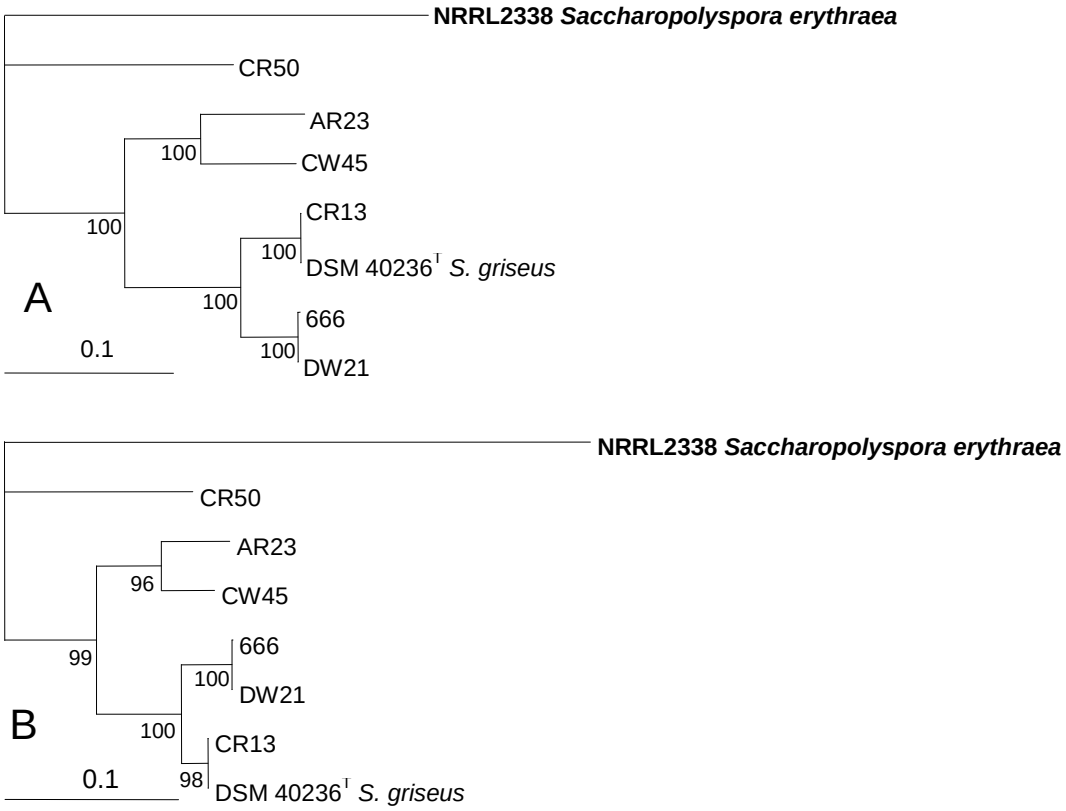
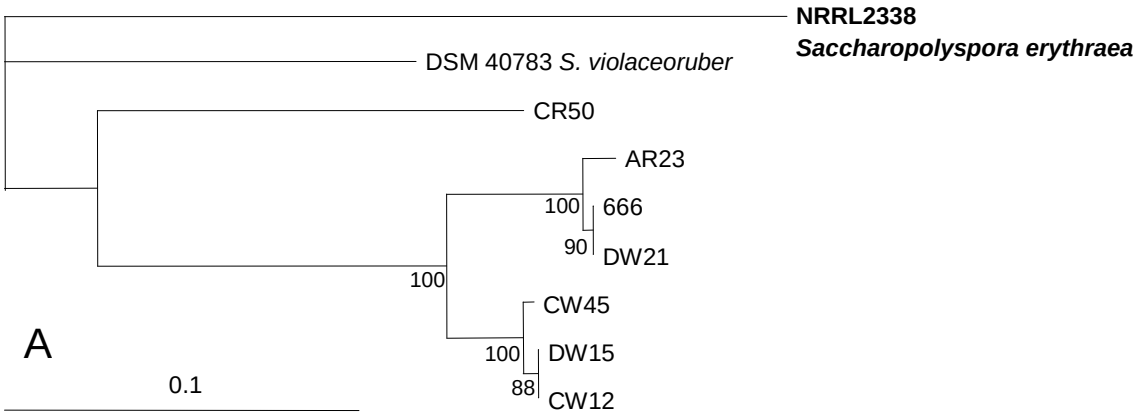


Figure 3.18. An (A) nucleotide (B) amino acid sequence phylogenetic tree of *ctrE*. Sequences in bold were acquired from GenBank. The tree was constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% sequence dissimilarity.



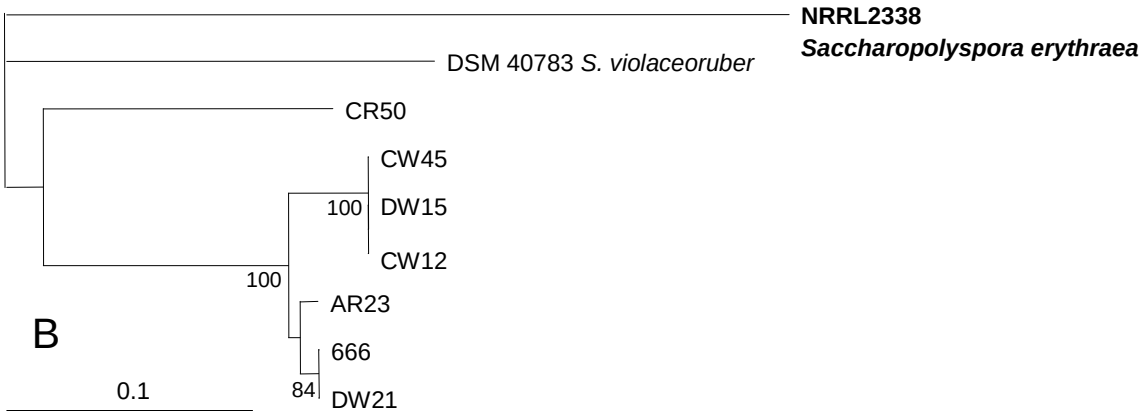


Figure 3.19. An (A) nucleotide (B) amino acid sequence phylogenetic tree of *geoA*. Sequences in bold were acquired from GenBank. The tree was constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% sequence dissimilarity.

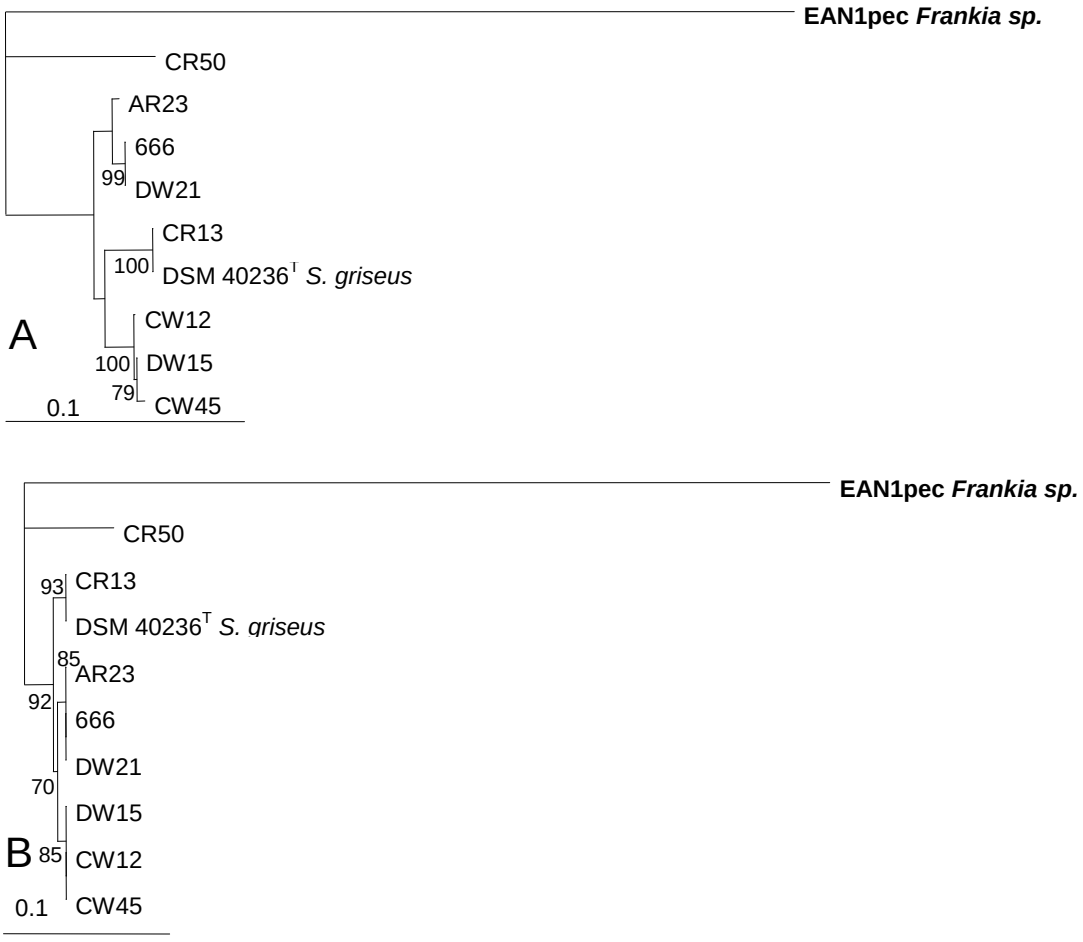


Figure 3.20. An (A) nucleotide (B) amino acid sequence phylogenetic tree of *hopB*. Sequences in bold were acquired from GenBank. The tree was constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% sequence dissimilarity.

For all three genes the nucleotide and amino acid trees had the same branching pattern (Figure 3.18, Figure 3.19, Figure 3.20), though the bootstrap values of the

amino acid trees were generally lower than those of the nucleotide trees. Disregarding the absence of some of the strains in some of the trees, the topology of *geoA* was identical to that of *hopB*. That of *crtE* differed however, as AR23 formed a clade along with CW45 and the two were in a sister clade to 666 and DW21, whereas in the others AR23 clusters with 666 and DW21. This may have been the result of recombination.

All three genes, but *hopB* in particular, had similar phylogenetic histories to the housekeeping genes (see Section 5.3.3). Because of that it was decided not to screen the remaining isolates of this study, as they would have been unlikely to produce results that would have significantly differed from those generated by the sequencing of the housekeeping genes themselves

3.3.10.3 nec1, a gene conferring a pathogenic phenotype

One secondary metabolite gene found in some streptomycetes is the *nec1* gene which encodes a secreted necrogenic protein that enables streptomycetes possessing it to necrotise and colonise plant roots (Bukhalid & Loria, 1997). It has a very low GC content (54%) compared to the average for streptomycete genomes (72%), indicating that is has been acquired recently via horizontal gene transfer (Loria *et al.*, 2006). All the strains used in this study were screened via PCR in order to compare its distribution patterns with those of *strA* and the other secondary metabolite genes. The primer pairs used were NecF1 (ACC TCG CCT GCA GAG AGG AC), NecR1 (GTT TCG TAC TCT CCG AGC GG) (Cullen & Lees, 2007) and Nf (ATG AGC GCG AAC GGA AGC CCC GGA), Nr (GCA GGT CGT CAC GAA GGA TCG). Both sets underwent PCR with a 60°C annealing temperature (Coombs & Franco, 2003).

Both primer sets failed to amplify *nec1* from the strains previously screened for *strA*, which indicates that *nec1* is not present in non-pathogenic streptomycetes. Because of that seven scab isolates, some of which had been previously screened for *nec1* (Cullen & Lees, 2007), were also incorporated in this study (Table 3.11).

Table 3.11. Scab isolates used in this study. Previous result refers to the detection of nec1 (one cross refers to detection with nested PCR, two to detection with a simple PCR) (Cullen & Lees, 2007), current result to the result produced by this study.

Strain	Species	Previous result	Current result
R2	*S. albidoflavus*	+	-
SCAB3	*S. sp*	N/A	-
SCAB4	*S. sp*	N/A	-
S46	*S. albidoflavus*	-	-
MP2	*S. rochei*	-	-
8.16	*S. diastaticus*	-	-
ASS8112	*S. diastaticus*	++	-
DSM 41668	*S. acidiscabies*	N/A	++

The absence of a positive result in R2 is to be expected (Figure 3.10), as no nested PCR was performed. This is due to the fact that a nested PCR would have produced an amplicon too small for use in the construction of reliable phylogenetic trees. No amplification was detected in ASS8112, however both primer sets were functional as they readily amplified *nec1* from a fresh culture of *S. acidiscabies* DSM 41668 a strain known to possess *nec1* (Bukhalid *et al.*, 2002). Due to a lack of sequences, attempts to compare the phylogeny of *nec1* with that of *strA* were abandoned.

3.4 Discussion

The relative frequency of streptomycin resistance in *Streptomyces* isolates from UK and Cuban soils is presumably due to native streptomycin producers. Cayo Blanco, Cuba is a pristine environment around which there has been very little human development. The same cannot be said for UK soil sites; however streptomycin is not used in large volumes in the UK (VMD, 2007) and is employed sparingly in medicine due to its toxicity (Begg & Barclay, 1995). Therefore the anthropogenic input of streptomycin in the UK environment is relatively low. This is supported by the fact that the frequency of streptomycin resistance amongst streptomycetes in UK soils (15.1%) was almost the same as that of a pristine environment (Cayo Blanco, 16.4%).

The number of isolates possessing the streptomycin cluster agrees with previous estimates of a frequency of around 10^{-2} in the environment (Baltz, 2006). Of the 207 Cayo Blanco isolates examined, two contained *strA* along with biosynthetic genes. No strains possessing the entire cluster were found amongst the UK isolates; however the number screened was only 73. Nonetheless, biosynthetic genes have

been detected in DNA extracted from UK soils, indicating that streptomycin producers are found as well (Tolba, 2004). The presence of producers explains why resistance has been selected for in other streptomycetes present in these soils. Streptomycin resistance amongst streptomycetes however does not appear to be universally distributed. A study examined 480 *Streptomyces* soil isolates from urban, agricultural and forest locations screened on selective plates with 20 μg/ml of streptomycin and no streptomycin-resistant streptomycete isolates were identified (D'Costa *et al.*, 2006). Another study which included 169 streptomycete isolates from multiple geographical locations only found streptomycin resistance (10 μg/ml) in 2.37% of the strains (Phillips *et al.*, 1994). One possible explanation for that is that there were no streptomycin producers in most of the sites sampled and as a result there was no selective pressure to evolve or retain resistance amongst the other streptomycetes in that environment.

Other groups of environmental bacteria can have a higher frequency of streptomycin resistance than streptomycetes. 90.2% of 236 Gram-negative bacteria isolated from seawater, sediment and shrimp from an industrially polluted site on south coast of Turkey were resistant to discs with 10 μg of streptomycin (Matyar *et al.*, 2008). The same applied to 88% of 100 *Enterococcus* isolates from seawater (Kimiran-Erdem *et al.*, 2007) and 96.1% of 26 *Enterococcus* isolates from diseased or dead fish (Savasan *et al.*, 2008) collected from coastal areas off Constantinople (Turkey). 100% of 103 *Salmonella* isolates form aquatic environments in Jabalpur (India) were streptomycin-resistant (Sharma & Rajput, 1996). All of these environmental sites were polluted which may have selected for other types of resistance and thus co-selected for streptomycin resistance. This is supported by the finding that *Escherichia coli* isolates from fresh faeces obtained at a Canadian swine farm had a mean streptomycin resistance of 54% ± 24% (tested on agar plates containing 64 μg/ml streptomycin) despite the fact that the only antibiotics used on the swine were oxytetracycline and penicillin G, indicating that coselection can occur (Duriez & Topp, 2007). Pollutants can have a profound effect in the levels of streptomycin resistance; 65% of 299 *Enterococcus* isolates from ponds in Thailand receiving chicken manure were resistant compared to only 29% of 111 isolates from ponds used as controls (Petersen & Dalsgaard, 2003). A comparison in Tennessee of agricultural soil, amended with manure from animals fed with antimicrobials, with

forest soil with no history of animal agriculture found that 30.4% of isolated Enterobacteriaceae were resistant to streptomycin in the former and 4.2% in the latter (Srinivasan *et al.*, 2008). Environmental *Salmonella* isolates from Portugal also had considerably lower streptomycin resistance levels compared to most of the previously described sites; only 22% of 58 isolates were resistant (Antunes *et al.*, 2006). There was significantly higher resistance in the Gram negative bacteria from Turkey, *Enterococcus* isolates from Turkey and Thailand, the *Salmonella* isolates from India compared to the *Streptomyces* soil isolates from the UK and Cuba (χ^2 $p < 0.0001$). Despite it being higher, there was no significant difference between the Portuguese *Salmonella* isolates and the *Streptomyces* isolates ($p = 0.4298, 0.3944$) however this may be due to the small sample size of the Portuguese isolates. The higher frequency of resistance may indicate that streptomycin resistance is more widespread in environmental enterococci and enterobacteria compared to streptomycetes. This may be due to a greater mobility of resistance mechanisms in these groups; there is evidence that the *strA-strB* streptomycin resistance genes have undergone multiple horizontal gene transfer events and that they are coselected with other resistance genes such as the *sulII* sulfonamide resistance gene (Sundin & Bender, 1996). Another possible reason is that these enterococci were primarily isolated from aquatic environments where antibiotics and other compounds can more readily diffuse and place selective pressure on all organisms present. Soil is considerably more heterogenous, which impedes diffusion of antibiotics especially since many of them adhere to clay or soil particles (Sarmah *et al.*, 2006).

Of the resistant streptomycete strains identified in this study, only 51% possessed *strA*. The remaining ones must have relied on a different resistance mechanism. The *strA* primers used in this study were highly conserved and therefore are unlikely to have failed to amplify *strA* if it was present in a strain, as they are capable of amplifying more diverse genes such as hydroxyurea phosphotransferases. The high level resistance of M110 and CB148 can be explained by the replacement of Lys-43 with an Asn in M110 and a Thr in CB148 and that of CB141 by the replacement of Lys-88 with an Arg, as mutations in these positions of RpsL have been previously been documented to provide resistance (Musser, 1995). The insertion of an Arg between Gly-85 and Arg-86, found in CB158, has not previously been documented. This mutation does not necessarily provide streptomycin resistance, as CB158 also

possesses *strA*. Mutations in *rpsL* accounted for the resistance of only 3% of strains screened and mutations on the 530 loop of the 16S rRNA sequence or presence of solely the *aphE* resistance gene for 0%, which left 46% of resistant strains with no known mechanism of streptomycin resistance. These strains may rely on the inactivation of the *rsmG* gene which can provide *S. coelicolor* with low level streptomycin resistance (Nishimura *et al.*, 2007) or on mutations on the 900 stem region of the 16S rRNA (Frattali *et al.*, 1990). Other possibilities include highly divergent streptomycin phosphotranspherases or adenyltransferases or completely novel resistance mechanisms.

The majority of strains with high viomycin resistance appeared to either rely on *vph* or its homologues, but only a small minority of strains with low resistance possess it. *Mycobacterium* species can gain viomycin resistance either by the G nucleotide in position 745 of the 23S rRNA being left unmethylated (Gustafsson & Persson, 1998), which can result from mutations in the *tlyA* encoding an rRNA methyltransferase (Maus *et al.*, 2005b), or from a substitution of G with A T at position 1473 of the 16S rRNA sequence (Taniguchi *et al.*, 1997). Neither the *tlyA* homologue in *Streptomyces* nor that region of the 16S gene were sequenced in this study, so it is impossible to know if the remaining strains rely on one of those mechanisms.

Previous isolation attempts have identified streptomycin producers in North and South America, Europe and Asia (Gordon & Horan, 1968; Huddleston *et al.*, 1997; Tolba *et al.*, 2002). This study identified producers from Central America and Africa, indicating that it is found in all continents but Antarctica, whose soil appears to lack streptomycin producers (Section 6.3.2). The absence of streptomycin resistance in almost all streptomycete isolates in two previous screenings (D'Costa *et al.*, 2006; Phillips *et al.*, 1994) also implies that it is not found in all soils. Nevertheless, the wide distribution and frequency (Baltz, 2008) with which they are isolated indicates that streptomycin production provides a significant evolutionary advantage to organisms possessing it.

Streptomycin is not unique in this wide distribution. Zwittermicin A, a linear aminopolyol antibiotic synthesised by *Bacillus* species commonly used in biocontrol (Emmert *et al.*, 2004), producers have been isolated from locations as diverse as the

US, Panama, Australia, the Netherlands and Honduras (Stabb *et al.*, 1994; Raffel *et al.*, 1996). 2,4-Diacetylphloroglucinol, another antibiotic used in biocontrol and synthesised by *Pseudomonas* producers have been isolated from Switzerland, the US, Italy, and Ghana (Keel *et al.*, 1996). *Myxococcus* strains producing the antibiotic have been isolated from Spain, Germany, Brazil (Gerth *et al.*, 1982) and Israel (Gaspari *et al.*, 2005). Soraphen and epothilone, two compounds with cytotoxic and antifungal activities respectively produced by *Sorangium cellulosum*, also have global distribution and a high frequency (Gerth & Muller, 2005). 1.1% to 3.6% of *Sorangium* soil isolates from Europe (616 isolates), Asia (410), Africa (277) and the US (295) produced soraphen, while 1–2.5% of them produced epothilone (Gerth *et al.*, 2003). The majority of antibiotic gene clusters however appear to have much more limited distributions (Baltz, 2005). *S. avermitilis* MA-4680, producer of the anthelmintic avermectins, has only been isolated once in Japan (Kim & Goodfellow, 2002). This could be because these compounds provide a lesser selective advantage compared to other antibiotics; however it could also be due to difficulties in isolating some of the producers. For example, more than thirty bioactive molecules including antibiotics have been detected in a relatively limited number of *Streptosporangium* isolates that was considered to be rare in the environment; however specialised isolation procedures have discovered this genus in more than half the soils screened, indicating that *Streptosporangium* may actually be relatively common (Lazzarini *et al.*, 2000).

The sequence identities to Group A *strA* sequences of the Group B *strA* sequences, the spectinomycin resistance gene *aph(6)* from *S. netropsis* and the dihydrostreptomycin resistance gene from *S. humidus* were all similar. This suggests that the Group B strains evolved from an aminoglycoside phosphotransferase other than the streptomycin resistance gene *strA*, though the fact that no biosynthetic genes belonging to the streptomycin or spectinomycin cluster could be amplified from Group B strains implies that if that is what occurred they have been lost. Another possibility however is that the loss of the streptomycin biosynthetic genes in the strains possessing Group B *strA* genes, freed the resistance gene to evolve in response to needs other than providing protection to the cell against the antibiotic its ancestor produced. In such a scenario a more diverse enzyme that could have also served additional functions or had a broader substrate specificity would be more

advantageous for the organism, thus favouring diversifying selection. However, the fact that the housekeeping genes of Group B strains have an average sequence identity of 97% and *strA* has only 75% makes it unlikely that *strA* had enough time to diverge if it originated from the streptomycin cluster. The two subgroupings B1 and B2 presumably evolved from a common ancestor, though their divergence may be due to different selective pressures operating on these two groups. Despite the divergence of these sequences, the conservation of the active sites examined compared to the remainder of the peptide indicates that *strA* is under purifying selection in all of the strains and that it therefore provides a selective advantage to the organisms possessing it. As a result it seems very likely that it still serves a function as an aminoglycoside phosphotransferase.

Like *strA*, a viomycin resistance gene homologue was detected in strains lacking viomycin biosynthetic genes. This is the first time that a *vph* homologue has been found in bacteria that were not viomycin producers. It is unknown what the function of *SGR878*, the homologue of *vioR*, is in *S. griseus*. It may regulate the *vph* homologue *SGR421*, however there is no experimental evidence supporting this. The homology of *vioS* and *strK* may be an indication of a recent common ancestor between the two genes that postdates the assembly of the remaining cluster genes in at least one of these two clusters. One possible explanation is that a viomycin cluster was acquired by *S. griseus* via conjugation and then proceeded to break up due to recombination. Most of the cluster genes were lost; however the *strK* ancestor got inserted into the streptomycin cluster, proved useful there and was retained, as was the resistance gene and its transcriptional regulator. Another possibility is that the viomycin cluster acquired *vioS* from an aminoglycoside cluster and the two other viomycin cluster gene homologues found in *S. griseus* were independently mobilised. This is supported by the finding that *strK* homologues are also present in the biosynthetic clusters of four additional aminoglycosides (hydroxystreptomycin, apramycin, paromomycin, lividomycin) while there are no other sequenced peptide antibiotic clusters possessing a homologue of *vioS*, indicating that this gene originated in aminoglycoside rather than peptide biosynthetic clusters. Furthermore, the *S. vinaceus vph* gene has 76% sequence identity to the *S. griseus SGR421* and 79% to its *S. bacillaris* homologue while their housekeeping genes demonstrate a

96% sequence identity. This great sequence divergence of the resistance gene indicates that *SGR421* originated from another peptide antibiotic gene cluster.

The short branch length and high congruence in gene distances indicates that both the viomycin resistance and biosynthetic genes found in clusters have the same phylogenetic history. This also applies to the *strA* and *strW* genes found in aminoglycoside clusters, which reinforces the idea that antibiotic biosynthesis clusters are selected as a unit in antibiotic producers.

The distribution of the geosmin biosynthetic gene and the hopanoid and carotenoid clusters was very similar to one another. The absence of *geoA* from *S. griseus* DSM 40236^{T} and CR13 indicates that it may have been lost in one of their common ancestors, as these two strains form a robust clade in the *geoA* and *hopB* phylogenetic trees. This also appears to have happened to *crtE* in DW15 and CW12, as these two strains also formed a clade in the other two trees. Another possibility is that *crtE* was lost before the divergence of CW45, but that strain then acquired it from another streptomycete via horizontal gene transfer. The fact that CW45 clusters with AR23 instead of 666 and DW21, which form a sister clade to CW45, DW15 and DW12 in the other two trees, supports this hypothesis.

The screen for *nec1* in free-living *Streptomyces* strains was undertaken to confirm that none of these soil isolates had the capacity to infect plant roots. Based on the absence of a positive result, it appears that the distribution of *nec1* is limited to pathogenic strains and is not widely distributed in streptomycetes. The same applies to production of thaxtomin A, another phytotoxin, which is limited to pathogenic streptomycete strains (Kinkel *et al.*, 1998). The failure to amplify *nec1* from ASS8112 may indicate that the gene had been deleted in the version of the strain that was available. The same applied to *strA* in the available *S. galbus* DSM 40480 sample, as the 16S rRNA sequence matched that of the stain but *strA* failed to amplify from it. As *nec1* is found on a genomic island (Loria *et al.*, 2006), the position of *strA* in *S. galbus* is unknown, it would have been simple for it to be lost and such a loss could provide a selective advantage to organisms cultured in a laboratory where there is no need to colonize plant roots or have antibiotic resistance. The fact that *nec1* was readily amplified from a fresh sample of *S.*

acidiscabies DSM 41668, and *strA* from a fresh sample of *S. galbus* DSM 40480, supports this hypothesis.

In conclusion, the results in this chapter provided support for both hypotheses. The detection of the streptomycin producers from across the world indicates that *strA* provides a major selective advantage and the amplification of a number of strains possessing *strA* indicates that the streptomycin resistance gene provides a selective advantage as well.

4 Investigation of the chromosomal location of the streptomycin cluster and *strA* gene in soil isolates

4.1 Introduction

In order to understand how streptomycin resistance evolved in soil streptomycetes and the regulation of expression, it was necessary to examine the regions of the chromosome flanking these genes.

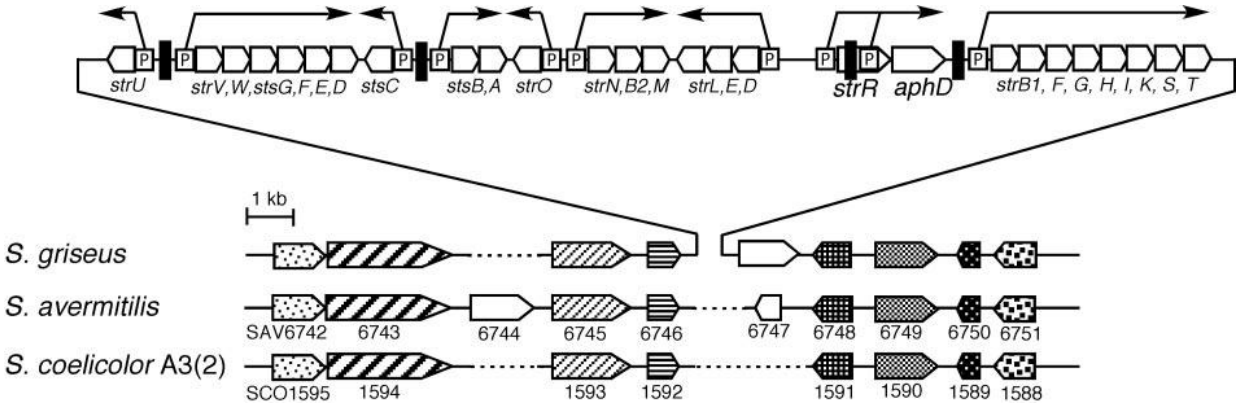


Figure 4.1. Position of the streptomycin gene cluster in S. griseus and comparison of the flanking regions with the homologous segments of the *S. avermitilis* and *S. coelicolor* chromosomes. StrR binding sequences are indicated by black bars, promoters with P and transcriptional units with arrows. After Tomono *et al.* (2005).

On the *S. griseus* genome, the streptomycin cluster includes the genes from *SGR5914* to *SGR5940* which span the region from 6,919,728 bp to 6,951,001 bp (Figure 4.1). The site between *SGR5914* to *SGR5942* where the streptomycin cluster is located may serve as a recombination hotspot, since a different gene is present between the two homologous flanking genes in *S. avermitilis* (Tomono *et al.*, 2005). The cluster is located near the 3' end of the core of the chromosome, which extends from *SGR923* to *SGR6311* in *S. griseus*, *SCO6804* to *SCO1209* in *S. coelicolor* and *SAV1625* to *SAV7128* in *S. avermitilis* (Ohnishi *et al.*, 2008).

The *strA* flanking regions of AR23 and CR50 were selected for analysis; AR23 because it is part of the resistance-only group of *strA* genes (Figure 3.3) and CR50 because it contains the cluster that is most closely related to the *S. griseus* streptomycin producers that falls outside that group.

4.2 Aims

Clone the *strA* gene and its flanking regions from representative strains. Determine if it is located in the same position as in the *S. griseus* chromosome which may indicate that the location is a recombination hotspot.

The hypothesis of this chapter is that the *strA* gene or cluster will preferentially recombine at the same location as in *S. griseus* DSM 40236^{T} since that location may act as a recombination hotspot.

4.3 Results

4.3.1 AR23 flanking regions

The regions containing the AR23 *strA* was cloned using a fosmid library (Section 2.4) and sequenced using primer walking from *strA* outwards. The sequence was confirmed via PCR and sequencing, using primers designed based on the previously obtained sequence, and annotated using Artemis.

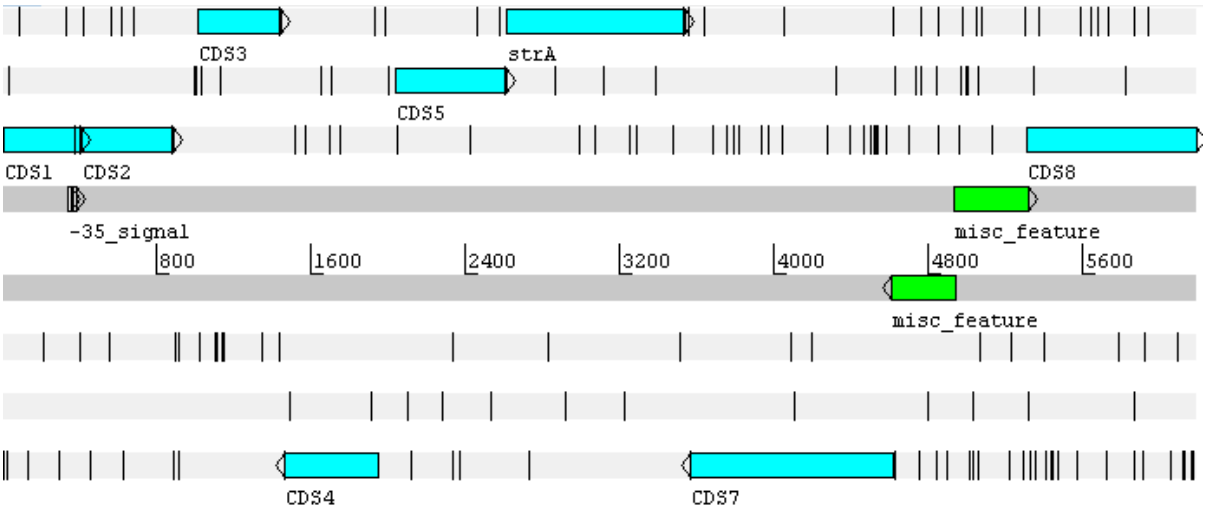


Figure 4.2. Annotated region of the chromosome containing the *strA* gene in isolate AR23 as visualised in Artemis.

Table 4.1. Putative genes located in a 6 kb area containing the *strA* gene in isolate AR23

CDS	Location	Homologue	Function
1	3..404	SGR4325	putative oxidoreductase
2	414..881	SGR4326	putative acetyltransferase
3	1015..1437	-	31% amino acid sequence identity to hypothetical protein BB14905_14455 [*Bacillus sp*. B14905]

4	1938..1465	SGR4330	putative acetyltransferase
5	2036..2608	SGR172	putative acetyltransferase
6	2611..3537	*strA*	putative streptomycin 6-phosphotransferase
7	4611.. 3568	SGR4331	conserved hypothetical protein
8	5316..6186	SGR400	putative amino acid transport integral membrane protein

Table 4.2. Features other than genes located in a 6 kb area containing the *strA* gene in isolate AR23

Location	Description
342..347	putative -35 box
366..374	putative -10 box
4934.. 4612	Homologue of intergenic region on 5' end of SGR4331
4935..5315	Homologue of intergenic region on 5' of SGR400

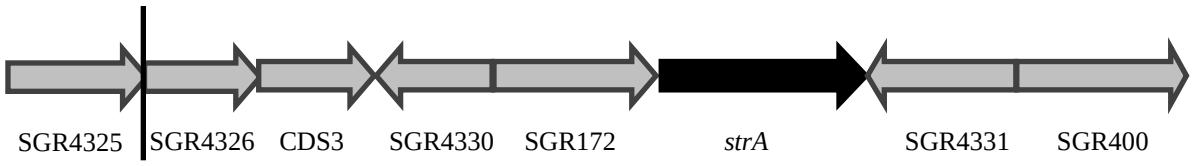


Figure 4.3. Diagram of the region of the chromosome containing the *strA* gene in isolate AR23.

The 6,186 bp fragment contained 2092 C, 2287 G, 973 T and 834 A bases. Its GC content was 70.8%. All but one (CDS3) of the genes identified in the AR23 sequence were also present in the *S. griseus* chromosome though their order differed (Figure 4.2, Figure 4.3).

The entire AR23 *strA* gene was highly divergent (75% sequence identity to Group A *strA*). The non-sequential position of neighboring ORFs SGR172 and SGR400 compared to the genome sequence of *S. griseus* indicated that there had been multiple recombination events in the flanking regions of *strA* (Table 4.1). Assuming that SGR4325, SGR4326, SGR4330 and SGR4331 have not been moved from the location they are on the *S. griseus* chromosome, the *strA* gene in AR23 was found 1.85 Mb away from the location of the streptomycin cluster.

BPROM identified three putative promoter regions, two of which were located within SGR4326 and SGR400, indicating they are unlikely to be functional; however the -35 and -10 promoter at the start of the SGR4326 homologue may be active. As the *strA* gene had no promoters at the start of its sequence, it may be

transcribed as a polycistronic mRNA molecule containing CDS2, CDS3 and CDS5 (Table 4.2). The large number of alternate sigma factors in streptomycetes limit the predictive power of bioiformatic tools such as BPROM. There may be additional unidentified promoters in that region of the chromosome (Hodgson, pers. comm.). The majority of operons in streptomycetes are up to four genes in length due to the rapid decline in expression across them, which indicates there may be another promoter before the SGR172 or *strA* genes (Hodgson, pers. comm.).

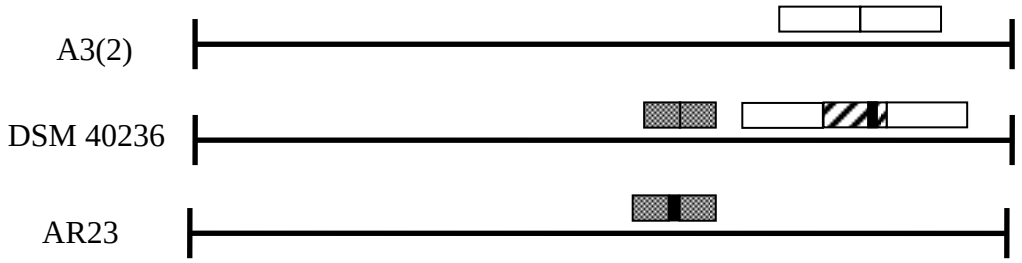


Figure 4.4. Comparison of locations of flanking regions to the streptomycin cluster and resistance gene in A3(2), DSM 40236^T and AR23. Black denotes strA, diagonal stripes the remaining streptomycin cluster and the white and dotted boxes denote the flanking regions of the AR23 strA and the DSM 40236^T cluster.

The genes flanking the AR23 *strA*, had no homologues in A3(2). *strA* and SGR172 had been inserted between SGR4330 and SGR4331 in AR23, as they flank each other in *S. griseus* DSM 40236^T (Figure 4.4). Sequencing of the flanking regions confirmed the finding from the PCR reactions that the *strA* gene is not associated with any genes of the streptomycin, or other aminoglycoside, biosynthetic cluster.

4.3.2 CR50 flanking regions

The CR50 cluster and flanking regions were reconstructed from two fosmids, one that was 42 Kb in size and included the 5' flanking region and part of the cluster until the *strA* gene and a second that was 34 Kb in size and stated in the strA gene and extended into the 3' flanking region of the cluster. The two fosmids were sequenced using shotgun sequencing (Section 2.4.3) and a small gap between the two within the *strA* sequence was filled by designing primers on the two ends and then performing a conventional PCR and sequencing reaction (Section 2.3.1, 2.3.4).

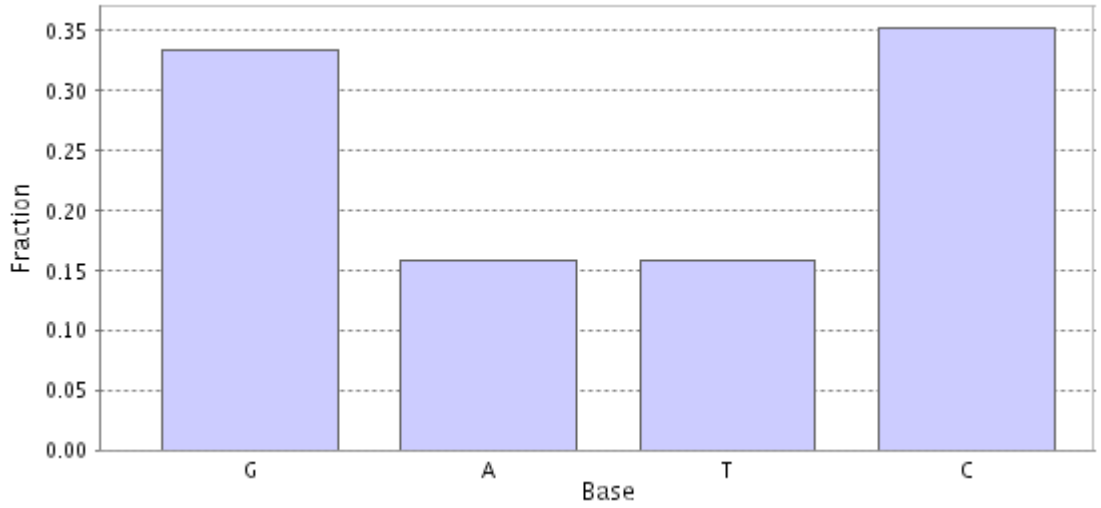


Figure 4.5. Forward strand nucleotide composition of the CR50 genome section containing the streptomycin cluster.

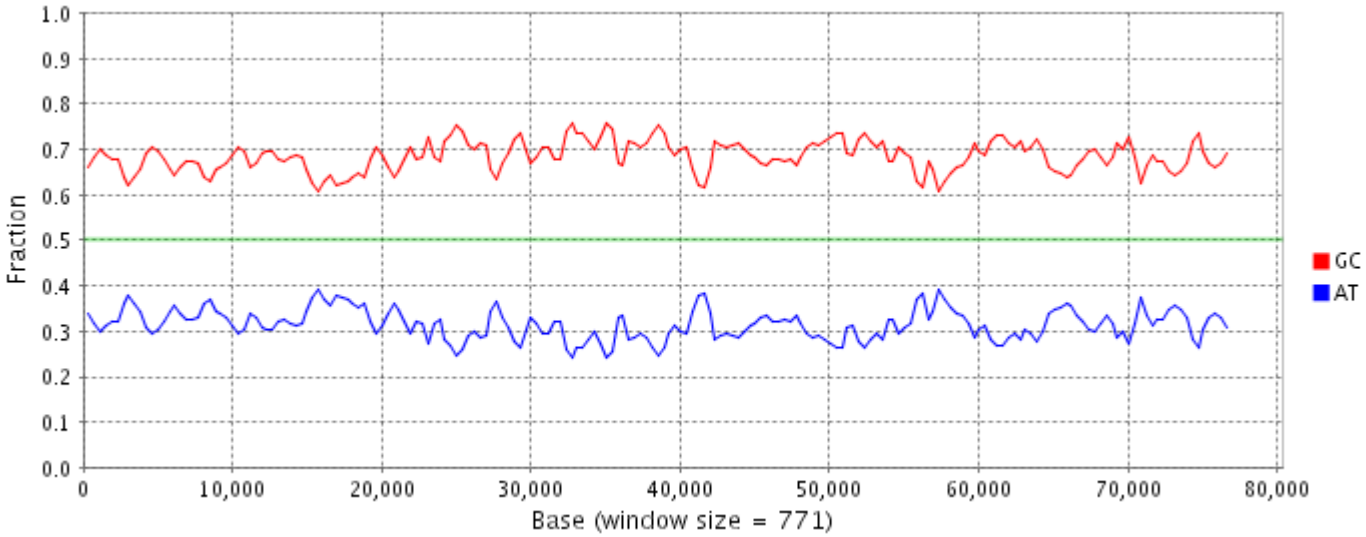


Figure 4.6. Sliding nucleotide composition of the CR50 genome section containing the streptomycin cluster.

The 77147 bp fragment contained 27164 C, 25674 G, 12173 T and 12136 A bases (Figure 4.5). Its overall GC content was 68.5%, though it ranged from 60% to 76% (Figure 4.6). The absence of any regions with highly divergent GC content indicates that there have not been any recombination events with DNA not belonging to high-GC content bacteria in this region of CR50's chromosome.

Figure 4.7. Annotated left (top) and right (bottom) flanking regions of the streptomycin biosynthetic cluster in CR50 as visualized in Artemis.

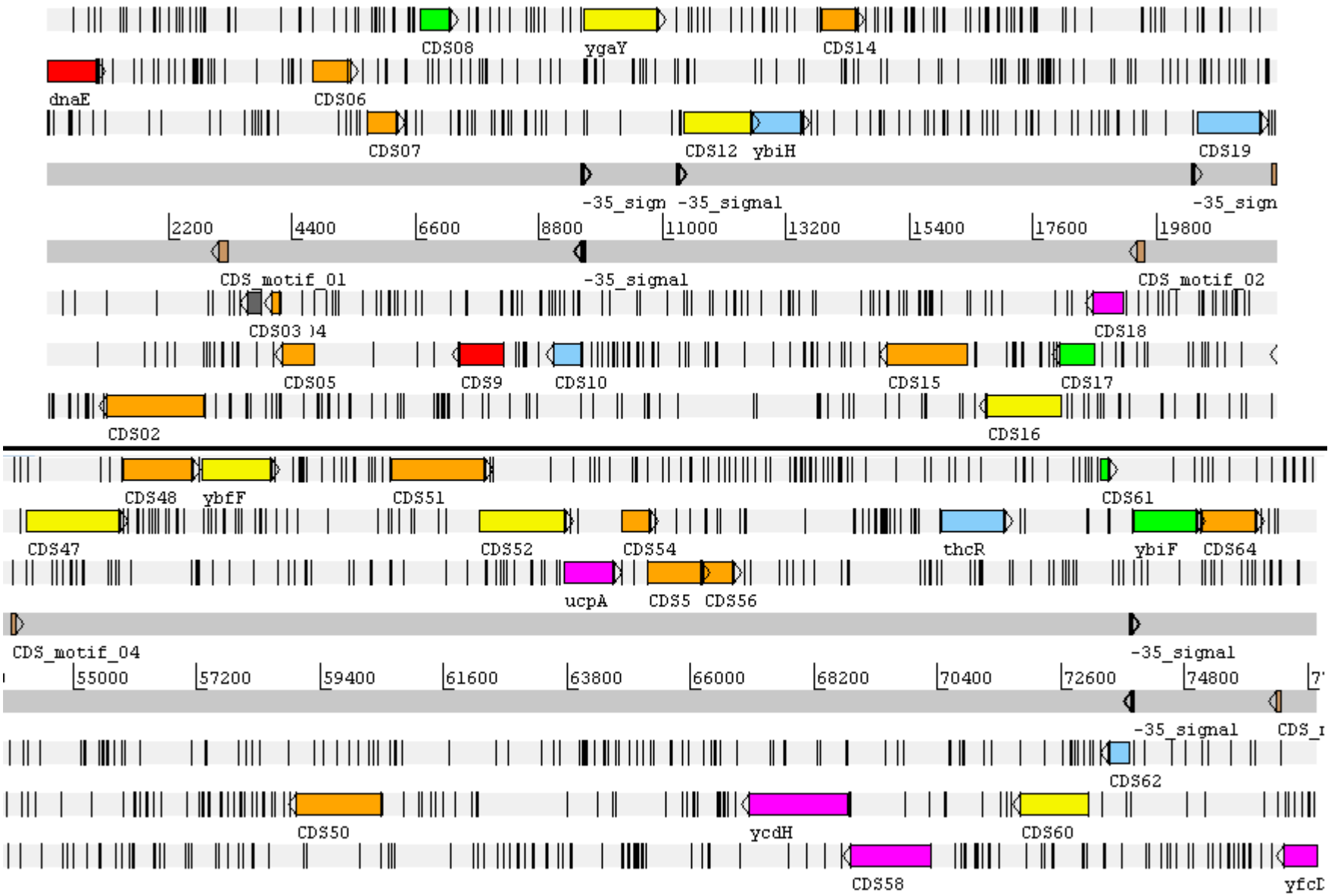

CDS08
ygaY
CDS14
dnaE
CDS06
CDS07
CDS12
ybiH
CDS19
-35_sign
-35_signal
-35_sign
2200
4400
6600
8800
11000
13200
15400
17600
19800
CDS_motif_01
-35_signal
CDS_motif_02
CDS03
CDS18
CDS05
CDS9
CDS10
CDS15
CDS17
CDS02
CDS16
CDS48
ybfF
CDS51
CDS61
CDS47
CDS52
CDS54
thcR
ybiF
CDS64
ucpA
CDS5
CDS56
CDS_motif_04
-35_signal
55000
57200
59400
61600
63800
66000
68200
70400
72600
74800
-35_signal
CDS62
CDS50
ycdH
CDS60
CDS58

Table 4.3. Putative genes located on the 5' flanking region of the CR50 streptomycin gene cluster.

CDS	Location	Homologue	Function
01	53..898	*dnaE*	DNA Polymerase III, alpha subunit
02	2786..1089	SCO_0642	Putative TPR Domain Protein
03	3816..3598	CabB	Putative calmodulin-like protein
04	4164..4030	-	Hypothetical Protein
05	4777..4220	-	Hypothetical Protein
06	4757..5401	MMAR_2851	Hypothetical Protein
07	5733..6260	nfa44150	Hypothetical Protein
08	6688..7194	SCP1.257	Putative membrane protein.
09	8146.. 7373	SLP2.30c	Putative transposase for insertion sequence element IS112
10	9532.. 9053	PFL_3943	Putative HxlR family transcriptional regulator
11	9592..10899	PFL_3944	Putative major facilitator family transporter
12	11373..12560	SACE_4644	Putative glycosyltransferase
13	12591..13460	*ovmY*	TetR-Family Transcriptional Regulator
14	13837..14442	-	Hypothetical Protein
15	14993..16411	SCO3050	Hypothetical Protein
16	18080.. 16779	ParG	Putative P450 heme thiolate protein
17	18682.. 18065	*neoT*	Putative sugar ABC transporter
18	19176.. 18679	CaciDRAFT_82810	Hypothetical esterase/lipase protein
19	20526..21641	*kanL*	Putative transcriptional regulatory protein

Table 4.4. Features other than genes located on the 5' flanking region of the CR50 streptomycin gene cluster.

Location	Description
3225.. 3089	Region homologous to segment of *S. griseus* NBRC 13350 gene SGR_277 encoding a conserved hypothetical protein
9538..9543	-35 signal
9549..9556	Promoter; rpoD17 binding site
9558..9566	-10 signal
9565.. 9557	-10 signal
9590.. 9585	-35 signal
11247..11252	-35 signal
11270..11278	-10 signal
19559.. 19466	Region homologous to segment of *S. ambofaciens* ATCC 23877 gene SAML0255 encoding a putative transposase
20435..20440	-35 signal
20451..20459	-10 signal
21858..21904	Region homologous to sequence flanking a *S. ambofaciens* ATCC 23877 putative truncated transposase

Table 4.5. Putative genes located on the 3' flanking region of the CR50 streptomycin gene cluster.

CDS	Location	Homologue	Function
47	54170..55822	SACE_6109	Putative family 39 glycosyl transferase
48	55888..57123	SGR_5941	Conserved hypothetical protein
49	57292..58509	SAV_3293	Putative dolichol-phosphate mannosyltransferase
50	60484.. 58973	NFA_43510	Hypothetical Protein
51	60667..62328	DSY4811	Hypothetical Protein
52	62240..63763	cyp0759	Putative cytochrome P450
53	63747..64619	SGR_6656	Putative oxidoreductase
54	64769..65272	SrosDRAFT_84830	Hypothetical Protein
55	65232..66194	SAV_3261	Hypothetical Protein
56	66246..66755	SSCG_00725	Hypothetical Protein
57	68770.. 67046	NFA_12490	Putative oxygenase
58	70256.. 68850	SACE_5403	Putative oxygenase
59	70484..71593	NFA_12460	Probable thc operon regulatory protein
60	73060.. 71879	SghaA1_0101 00039235	Probable acyl-CoA dehydrogenase
61	73309..73446	SCO7074	Putative membrane protein
62	73794.. 73459	FRAAL3735	Putative Transcriptional ArsR Family Regulator
63	73895..75010	SACE_1858	Putative transmembrane protein
64	75125..76060	SACE_6814	Hypothetical Protein
65	77144.. 76566	SACE_2248	Putative NUDIX hydrolase

Table 4.6. Features other than genes located on the 3' flanking region of the CR50 streptomycin gene cluster.

Location	Description
53910..53940	Region homologous to segment of *S. coelicolor* A3(2) gene SCO6391 encoding a putative IS110 transposase/integrase
73825..73830	-35_signal
73853.. 73845	-10_signal
73846..73854	-10_signal
73877.. 73872	-35_signal
76497.. 76450	Region homologous to segment of *S. griseus* NBRC 13350 gene SGR_6t encoding a hypothetical protein

There is considerable evidence for recombination in the streptomycin cluster flanking regions (Figure 4.7). CDS08 has homologues found in the *S. coelicolor* plasmid SCP1 and CDS09 encodes a homologue of a transposase found in the *S. lividans* 1326 plasmid SLP2, indicating there have been plasmid recombination events in these areas. CDS15 is a homologue of the *Mycobacterium* phage Myrna protein gp18 which suggests that there has been an insertion of viral DNA in that

location. The most relevant finding however is the presence of a region homologous to a sequence flanking a putative truncated transposase on the 5' end of the streptomycin cluster and of a region homologous a putative IS110 transposase/integrase on the 3' of the cluster (Table 4.3, Table 4.4, Table 4.5, Table 4.6). The transposase/integrase on the 3' end has been partially lost and the transposase on the 5' end has been completely lost leaving only its flanking region behind, however these remnants strongly suggest that the cluster arrived at that location via transposition.

The flanking regions contained a number of secondary metabolite genes, such as *neoT*, and genes of unknown function. That and the evidence that there have been numerous genetic transfer events in that area suggested that it is located in one of the arm regions of the CR50 chromosome. It was not possible to map the *str* cluster location on the chromosome of CR50, as these flanking regions were radically different from all sequenced genomes.

Only one of the genes flanking the streptomycin cluster in *S. griseus* had a homologue in CR50. SGR5941 is located next to *strU* in *S. griseus*, however a glycosyl transferase (CDS47) had been inserted between the *strU* and the SGR5941 homologues. SGR5941 and CDS48 had a 76% sequence homology.

4.3.3 CR50 streptomycin cluster

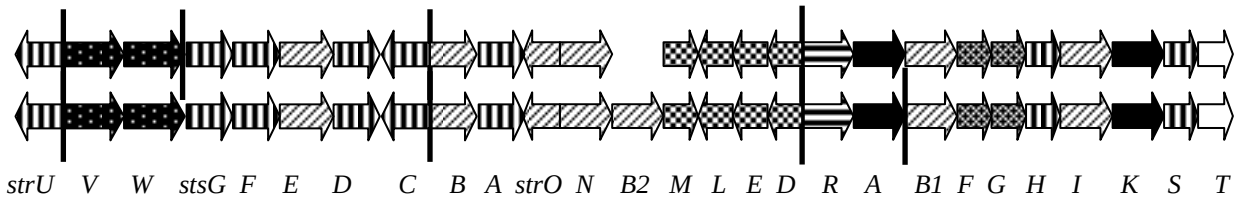


Figure 4.8. Comparison of the structure of the CR50 (top) and *S. griseus* DSM 40236^T (bottom) clusters. Black lines represent sites where promoters have been detected.

The CR50 genes were in the same order as in the *S. griseus* streptomycin cluster (Figure 4.8). The sole difference was that the *strB2* gene, encoding a scyllo-inosamine-4-phosphate amidinotransferase, had been partially deleted. Of the original 1053 bp gene, only a 204 bp fragment from the 3' end of the gene remained between *strN* and *strM*.

Table 4.7. Percentage identity between CR50 and DSM 40236^T steptomycin cluster genes. For *strB2* the homology of the 204 bp fragment present in CR50 was measured.

Gene	Identity	Gene	Identity	Gene	Identity
strT	88	*strR*	80	*stsB*	82
strS	88	*strD*	84	*stsC*	86
strK	84	*strE*	87	*stsD*	84
strI	82	*strL*	87	*stsE*	82
strH	86	*strM*	90	*stsF*	85
strG	87	*strB2* *	79	*stsG*	88
strF	91	*strN*	82	*strW*	87
strB1	89	*strO*	85	*strV*	86
strA	81	*stsA*	88	*strU*	87

The average sequence identity between the *S. griseus* and CR50 cluster genes, excluding the *strB2* pseudogene, was 86% (Table 4.7). The overall streptomycin cluster sequence identity, including the intergenic regions, was 84%.

Table 4.8. Identity of intergenic regions larger than 100 bp within the streptomycin cluster. 'CR50 positions' lists the area they occupy on the CR50 sequence and 'size' indicates the length of the intergenic regions on the CR20 and the *S.griseus* DSM 409236 streptomycin clusters.

Flanking genes	CR50 positions	CR50 size	*S. griseus* size	Identity
strU..strV	23390..23790	402	479	78%
stsC..stsB	31840..32064	225	238	N/A
strD..strR	41257..41577	321	406	N/A
strA..strB1	43756..44409	654	494	N/A
strG..strH	47019..47247	229	161	47%
strH..strI	48406..48595	190	44	N/A

Unlike the genes, the intergenic regions demonstrated very little conservation (Table 4.8). The intergenic region with the greatest sequence similarity was that between *strU* and *strV* followed by the area between *strG* and *strH*. The remaining areas were too divergent to be aligned. The little identity between the intergenic regions indicates that they are not under selective pressure to be retained, which led to their divergence. Conversely, with the exception of *strB2*, the genes that form part of the cluster have to be under positive selection as they have been retained with minor alterations.

4.4 Discussion

Examination of the flanking regions reveals that both *strA* in AR23 and the streptomycin gene cluster in CR50 are in different locations on their chromosomes

compared to the *S. griseus* DSM 40236^T streptomycin cluster. This indicates that, despite the presence of an insertion at that location in the *S. avermitilis* genome suggesting that recombination is not uncommon at that locus (Tomono *et al.*, 2005), the streptomycin cluster or resistance gene do not preferentially integrate between the homologues of *SGR5913* and *SGR5941*, leading to the rejection of this chapter's hypothesis.

Assuming that the area containing the genes *SGR4325*, *SGR4326*, *SGR4330* and SGR4331 has not undergone a major genomic rearrangement, both the AR23 *strA* gene and the *S. griseus* DSM 40236^T streptomycin cluster are located in the core region of the chromosome. The GC content of the AR23 region containing *strA* (70.8%) was very similar to the average GC content of the *S. coelicolor* (72.1%) and *S. avermitilis* (70.7%) genomes (Choulet *et al.*, 2006). This also supports the hypothesis that *strA* is located in the core of the chromosome, as chromosomal arm regions have a lower GC content that is similar to that of streptomycete plasmids due to the large number of insertion sequences present in them (Choulet *et al.*, 2006). The DSM 40236^T cluster is located 0.537 Mb away from the start of the right arm region. Assuming the same overall genomic structure, the AR23 *strA* would be 2.43 Mb from the start of the right arm region, though the genomic organisation of the area sequenced suggests that the gene order in AR23 is different from that of DSM 40236^T. The presence of a secondary metabolite gene cluster in the core area of the chromosome is not surprising. Analysis of the *S. ambofaciens*, *S. coelicolor* and *S. avermitilis* genomes demonstrated the presence of a long gradient of indel differences between the core and arm regions, indicating that there is not a simple cut-off between core and arm regions (Ventura *et al.*, 2007). The presence of the streptomycin cluster and the *strA* gene in the core region of their chromosomes may be an indication that they provide a strong selective advantage, as it is theorised that advantageous secondary metabolite genes will migrate from the unstable arm regions, where they can be easily lost, to the more stable core region (Chater & Chandra, 2006).

The streptomycin gene cluster in CR50 is likely to be located in one of the chromosomal arm regions. The area sequenced does not correspond to that of any available streptomycete genome. This may be due to CR50 being distantly related to

strains whose sequences are available, however a lack of conservation is to be expected from the arm regions as even closely related strains can have significant differences in them (Choulet *et al.*, 2006). The relatively low GC content (68.5%) in the sequenced CR50 region is similar to that of streptomycete plasmids, which can range from 68.4% for SLP2 (Huang *et al.*, 2003) to 69.2% for SAP1 (Ikeda *et al.*, 2003). The CR50 region sequenced also contains plasmid-associated genes, mobile genetic element-related genes and pseudogenes. These are found at a much higher frequency in the arm regions of the streptomycete chromosome (Choulet *et al.*, 2006), which also suggests that it is located there.

The fact that the *S. griseus* DSM 40236^{T} streptomycin cluster is located between two genes whose homologues are next to one another in *S. coelicolor* and have a different gene present between them in *S. avermitilis* strongly suggests that the streptomycin gene cluster arrived at its present location via an insertion event. The same applies to AR23, as no *strA* homolog is present between the homologues of the genes flanking it in *S. griseus*. Neither the streptomycin cluster in *S. griseus* or the *strA* gene in AR23 are flanked by sequences, such as transposase genes, translocatable elements, plasmid transfer origins or attachment sites of phage integrases (Ochman *et al.*, 2000), that might reveal how they arrived in their current location. The loss of these sequences is not surprising, as their presence is not necessary for the functioning of the genes they transferred. There are 78 predicted transposase genes in *S. coelicolor* and 99 in *S. avermitilis*, however only a small minority still retains necessary genetic elements supporting transposability (Ventura *et al.*, 2007), indicating that there is very little selective pressure to retain transposability once a genomic island has been integrated into the host's chromosome. The CR50 streptomycin cluster, unlike that of *S. griseus*, retains remnants of sequences belonging to or associated with transposases at both its ends though both of them have undergone deletions that render them non-functional. The fact that remnants of the transposases remain in CR50 however suggests that the cluster arrived in this strain more recently than in *S. griseus* and the transposases have yet to be fully lost; their presence in CR50 suggests that the aminoglycoside clusters have been transferred between streptomycetes on transposons whose transposable elements have been lost in most strains possessing them, leaving only the cluster behind. Transfer of biosynthetic gene clusters via transposons may be a

widespread phenomenon in bacteria; a gene cluster encoding enzymes involved in the biosynthesis of the LPS O antigen in *Francisella tularensis* is flanked by two functional transposases (Prior *et al.*, 2003), indicating that it recently arrived in that organism via HGT. Even if a cluster is not itself flanked by transposases, it can be mobilised as part of a larger genomic island. The thaxtomin biosynthetic cluster, encoding enzymes that produce a phytotoxin (Loria *et al.*, 2006), is flanked by a transposase pseudogene and thus can no longer be independently mobilised; however it can be mobilized as part of a 660 kb pathogenicity island (Kers *et al.*, 2005). This is unlikely to be the case for the streptomycin gene cluster however, as only the cluster itself along with *SGR5941* has been inserted in *S. griseus* DSM 40236^{T} between the two flanking genes whose positions are conserved across all sequenced streptomycete chromosomes.

It is not certain that the streptomycin cluster in CR50 is active. The great difference in conservation between the cluster's genes and intergenic regions indicates that the genes are under considerable purifying selection, which would imply that they offer a selective advantage which would necessitate them being expressed. This however is contradicted by the almost complete deletion of the *strB2* biosynthetic gene. One possible explanation is that the other genes are expressed and their enzymes function normally despite the absence of *strB2*, generating an aminoglycoside that differs somewhat from streptomycin. StrB2 is involved in the transamidination of the diaminocyclitol moiety of streptomycin. In its absence one of the NH_2 groups on the 6' end of streptomycin would not be replaced by a guanidine. However that would not necessarily result in an inactive molecule; the bluensomycin gene cluster does not include an *strB2* homologue but is able to produce an active antibiotic (Jung *et al.*, 2003). Another possibility is that an amidinotransferase whose gene is located outside the CR50 streptomycin cluster compensates for the deleted *strB2* gene. Compensation for a deleted biosynthetic cluster gene has been previously observed; the *dnfM* gene from the daunorubicin biosynthetic cluster of *S. peucetius* has a frameshift mutation that renders it inactive, however a gene located outside the cluster encodes a TDP-D-glucose 4,6-dehydratase which compensates for the loss of *dnfM* (Gallo *et al.*, 1996). Another example can be found in *Burkholderia cepacia*, where the inactivation of the *bceA* gene from the exopolysaccharide (EPS)

biosynthetic cluster results only in the reduction of EPS production, suggesting that a functional homologue compensates for the loss of *bceA* (Sousa *et al.*, 2007).

5 Diversity analysis

5.1 Introduction

In order to elucidate the evolutionary history of the secondary metabolite genes that were examined in Distribution of the streptomycin cluster, *strA* resistance gene and other secondary metabolite genes in the genus *Streptomyces*, it is necessary to compare the phylogeny of these genes with that of the strains possessing them. This will reveal the pattern in which secondary metabolism genes have been distributed, which can be used to identify whether they have spread via horizontal gene transfer (HGT) or vertical descent.

Traditionally phenotypic characters, such as mycelium colouration or spore morphology, were used to classify streptomycetes, however these methods tend not to be as reliable and different methods can produce variable results. One comparative analysis of ten independent studies involving eight different chemotaxonomic or DNA fingerprinting techniques discovered that none of the taxospecies derived by these studies were represented by all eight tests and there was variability between methods regarding which taxospecies each strains fell under (Anderson & Wellington, 2001). DNA-DNA hybridization is still used to define bacterial species, but cannot be used for classification below the species level (Staley, 2006).

It was thus decided to rely on sequence data for the classification of the streptomycete strains, both because they are more reliable and can delineate the interrelationships of different strains down to the subspecies level and because it would also enable the comparison of phylogenetic distances of housekeeping with secondary metabolism genes. The *rrn* gene, encoding the 16S ribosomal RNA, is one of the earliest (Woese & Fox, 1977) and most widely used genes (Ludwig *et al.*, 1998) for phylogenetic analysis. The 16S rRNA sequence has the advantage of being conserved across all bacteria, however due to its conservation it cannot provide sufficient resolution below the genus level (Staley, 2006). Other potential pitfalls of using 16S sequences in phylogeny include the fact that bacteria can have multiple *rrn* operons with slightly divergent sequences and that, due to the conservation of the 16S sequence, it is a likely site for homologous recombination

which can undermine phylogenetic reconstructions based on it (Acinas *et al.*, 2004). Using protein-coding genes that evolve more rapidly than 16S can provide greater phylogenetic resolution, however this does not address the issue of recombination (Feil & Spratt, 2001). This can be resolved through the use of multiple genes located across an organism's genome, as that will buffer distortions on a phylogeny generated by recombination (Hanage *et al.*, 2005). The use of seven housekeeping genes can allow the use of multilocus sequence typing, which can enable even more precise delineation of individual strains using procedures such as eBURST (Spratt *et al.*, 2004).

5.2 Aims

Determine the phylogenetic interrelationships of the strains discussed in Distribution of the streptomycin cluster, *strA* resistance gene and other secondary metabolite genes in the genus *Streptomyces* and compare the phylogenetic histories of the secondary metabolite genes analyzed therein with those of a set of housekeeping genes.

The hypothesis for this chapter is that biosynthetic clusters and resistance genes have spread principally via vertical descent but there have also been a number of HGT events within their evolutionary history.

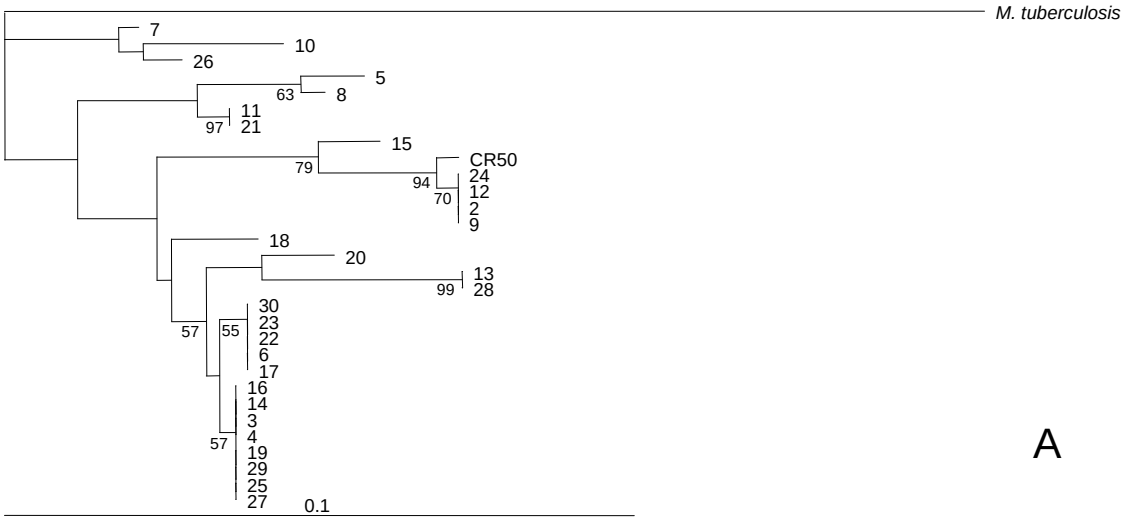
5.3 Results

5.3.1 16S rRNA Sequence Analysis

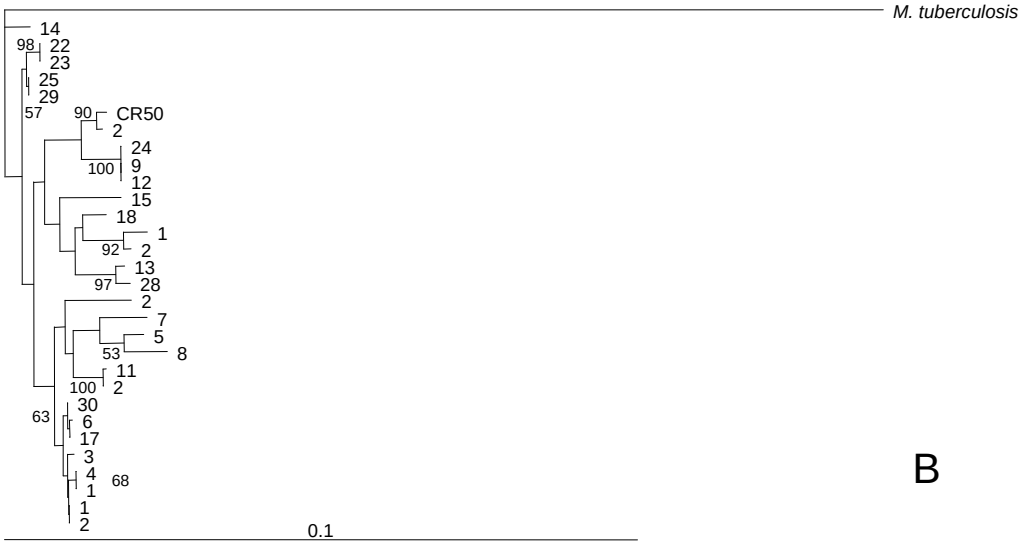
5.3.1.1 Determination of optimal 16S fragment size to use for analysis

A segment of the *rrn* gene encoding the 16S rRNA was sequenced to compare its phylogeny with those of other housekeeper genes and in order to confirm that the type strains used in this study had not been contaminated in the past with other streptomycetes. Using a partial 16S rRNA sequence would have the advantage of only requiring two rather than four sequencing reactions; a 120 bp 16S fragment has been used to classify streptomycetes in the past (Kataoka *et al.*, 1997). However using a partial 16S rRNA sequence can result in misleading phylogenetic trees; e.g. the genus *Kitasatosporia* does not tree separately from *Streptomyces* when a partial 16S rRNA sequence is used (Wellington *et al.*, 1992), but does so when the full

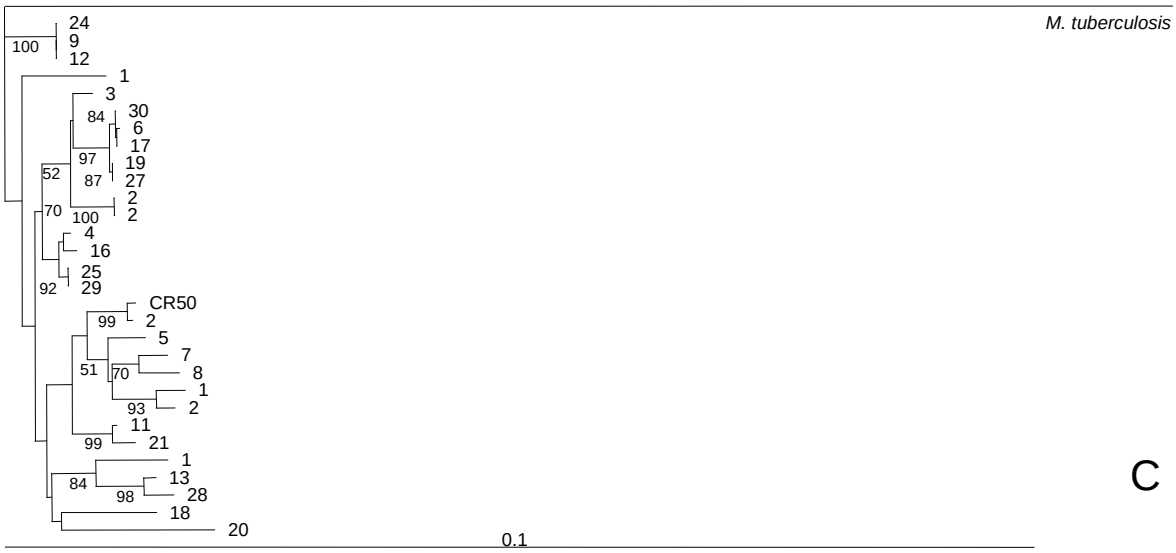
sequence is employed (Zhang *et al.*, 1997). To ensure that the fragment would be large enough to produce a reliable result the blastn algorithm was used to search GenBank for a set of 16S sequences that were closely related to that of CR50 and then trees were constructed based on a 142 bp region, containing the γ hypervariable region (Stackebrandt *et al.*, 1991), on an 800 bp region, containing the γ region and some other areas of minor variation, and on 1480 bp segment, containing almost the entire 16S rRNA sequence in order to compare the topologies and bootstrap values of the resultant trees.



A



B



C

Figure 5.1. Phylogenetic trees based on (A) an 142 bp region (B) an 800 bp region and (C) an 1480 bp region of the 16S sequence. The trees were constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% nucleotide dissimilarity.

The tree topology of the 142 bp tree is more divergent compared to that of the 800 bp and the 1480 bp trees (Fig. 5.1). This is supported by pairwise analysis of the distance matrices used to generate those trees, as the 142 bp had a correlation of r=0.863 to the 800 bp and r=0.828 to the 1480 bp distance matrix, but the 800 bp one had a correlation of r=0.993 to the full 16S sequence. There were still some topological differences between the 800 bp and the 1480 bp trees, but these were limited to nodes with low bootstrap values. Due to the very high correlation between the two distance matrices and the conservation of high bootstrap clades between them, it was deemed that this 800 bp region was sufficient to generate reliable 16S phylogenetic trees.

5.3.1.2 Generation of 16S rRNA phylogenetic tree

16S sequences were amplified with the pA/pH primer set (Edwards *et al.*, 1989) and sequenced using the pA and the novel reverse primer 852R (ACT TAA TGC GTT AGC TGC) binding 852 bp downstream of pA, which was used to generate 780 bp sequences. The procedures described in Section 2.3 were used in the amplification, purification and sequencing of all the amplicons discussed in this chapter.

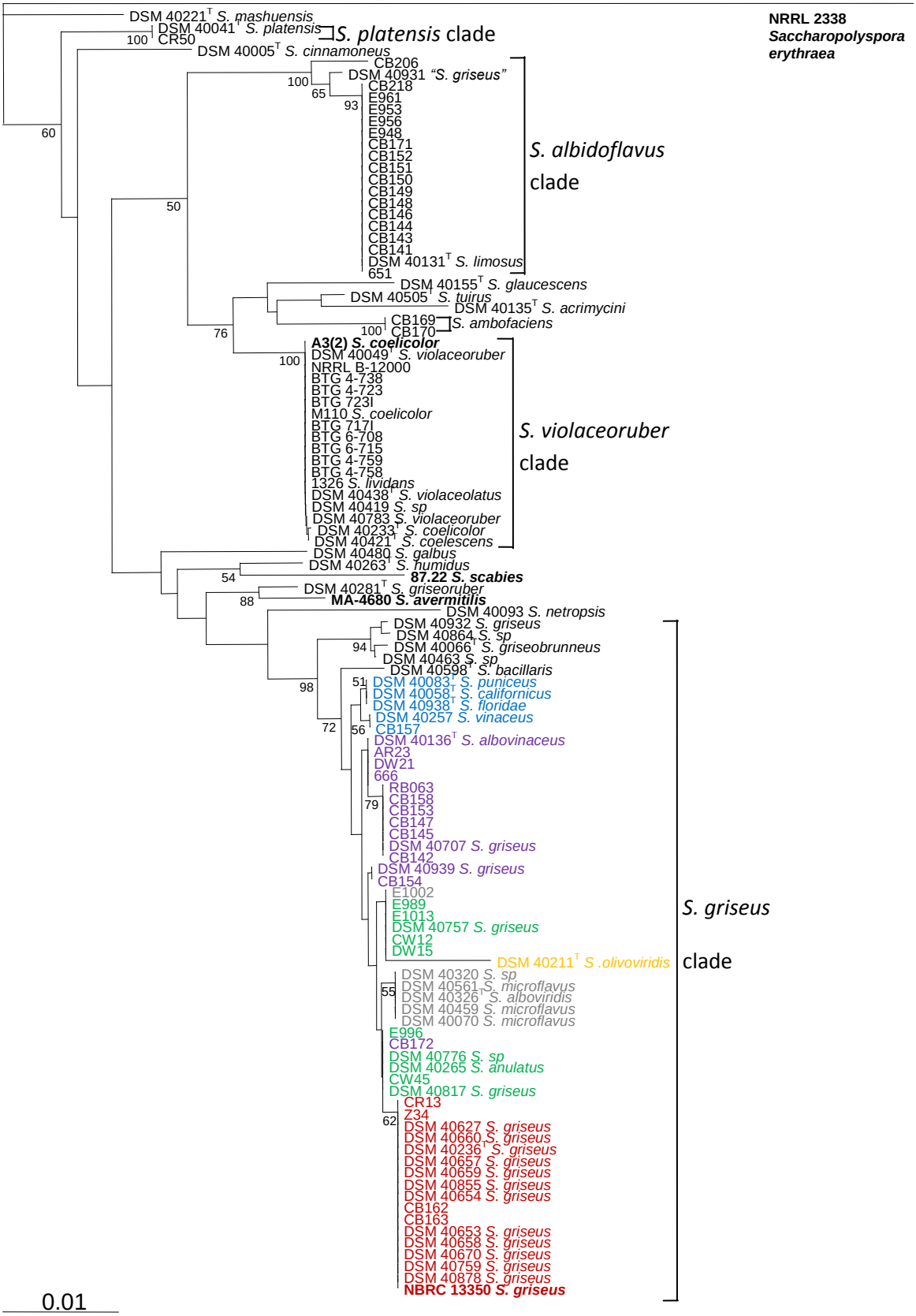


Figure 5.2. Phylogenetic trees based on a 780 bp region of the 16S rRNA gene sequence containing the γ hypervariable region. The *S. griseus* clade strains are coloured according to which clade they belong to on the concatenated housekeeper gene tree (Figure 5.4). The trees were constructed using the neighbour-

joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% nucleotide dissimilarity. ^{T} denotes a species' type strain.

5.3.2 Selection of housekeepers for MLSA

Using multiple housekeeper genes can produce trees whose branching is more statistically significant and it can also reduce problems due to long-branch attraction, as the effect of a fast-evolving gene in the phylogeny can be reduced by other more slowly evolving ones (Gontcharov *et al.*, 2004). There are a number of criteria for selecting genes to use in MLSA: (i) they must be essential for the cell's survival in its natural environment, (ii) they must be spaced far enough apart on the organism's chromosome so that they are not co-transducible (>100 kb), (iii) that sequences already exist in GenBank for as many of the target species as possible, (iv) that there be no known paralogs and (v) that the genes selected do not over-sample any one physiological process to prevent concordance in the gene trees as an artefact of co-evolution (Gontcharov *et al.*, 2004).

Table 5.1. Housekeeper genes

Gene	Protein Function	Involved in
argH	putative argininosuccinate lyase	Amino-acid biosynthesis
atpD	ATP synthase beta chain	Production of ATP from ADP
efp	elongation factor P	Peptide bond synthesis
ppc	phosphoenolpyruvate carboxylase	Citric acid cycle
rplC	ribosomal protein L3	Part of the 50S ribosomal subunit
rpoB	DNA-directed RNA polymerase beta subunit	Transcription of DNA into RNA
rpsL	ribosomal protein S12	Part of the 30S ribosomal subunit
sodF	Fe-containing superoxide dismutase	Catalysis of 2 superoxide + 2 H^+ = O_2 + H_2O_2.

The genes selected were involved in a range of cellular processes such as energy generation (*atpD*), primary metabolism (*argH*, *ppc*), transcription (*rpoB*), translation (*rplC*, *rpsL*) and antioxidant defence (*sodF*) (Table 5.1). The genes selected were distributed across the entire core region of the chromosome of all three available sequenced genomes (Figure 5.3), though some did form clusters and thus could be co-transduced. These genes were selected despite not fully fulfilling criterion (ii) due to the need for the genes to be sufficiently conserved and also because it was possible to generate efficient PCR primers that functioned across the *Streptomyces* genome from the selected strains.

A genomic rearrangements is visible in *S. avermitilis*, where the positions of the *rplC*, *rpsL* and *rpoB* genes have been switched with that of *ppc* (Figure 5.3).

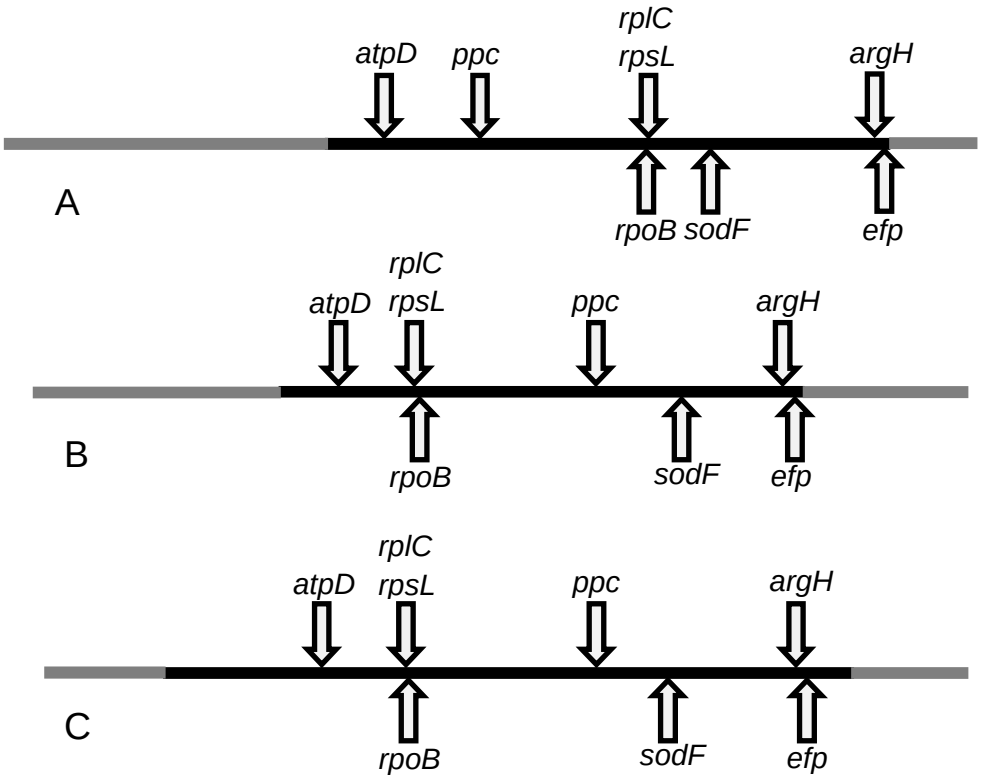


Figure 5.3. Locations of selected genes on the (A) *S. avermitilis* MA-4680, (B) *S. coelicolor* A3(2) and (C) *S. griseus* NBRC 13350 chromosomes. The core region of the chromosome is in black and the arm regions are in gray.

Primers were designed by aligning all versions of a gene available on GenBank and then basing the primers on the two most distant conserved regions on the sequence (Table 5.2).

Table 5.2. Primers for the housekeeper genes used in the multi-locus sequence analysis (MLSA)

Target Gene	Primer	Sequence	Product size (bp)	TA (°C)
argH	argH_F	CAC CTC CAG CAC GCC CAG CC	529	60
	argH_R	AGT CGA AGA CCG GCT CCT TG		
atpD	atpD_F	AAG ACC GAG ATG TTC GAG AC	466	56
	atpD_R	CCA TCT CGT CGG CCA GGT TC		
efp	efp_F	CTT CCA CGA ACG ACC TCA AG	526	56
	efp_R	GTG TCG ACC TTG ATC TTC TC		
ppc	ppc_F	AGC AGG GCG AGG TCA TCT CC	541	56
	ppc_R	ACG AGG GTG TCG ACG TAG TG		
rplC	rplC_F	AGA AGC TCG GCA TGA CGC AG	532	56
	rplC_R	TTC TCC GCG TCA ACG GCG TG		
rpoB	rpoB_F	GAG CAC GAC GAC GCC AAC CG	580	59
	rpoB_R	GGA TGA TGC CGC GCT CGT CG		
rpsL	rpsL_F	GGC AGG ACA AGG TCG AGA AG	331	56
	rpsL_R	TCC TTC TTG GCG CCG TAG C		
sodF	sodF_F	AGA TCA TCG AGC TGC ACC AC	517	56
	sodF_R	ACG TCC TGC CAG TTG ACG AC		

5.3.3 Concatenated housekeeper gene trees

There are two methods of utilizing multiple gene sequences to generate phylogenetic trees; one is to concatenate all the sequences into one and then treat it as a single gene, while the other is to generate individual trees for all the genes and to obtain a consensus tree from them. The former appears to generate more reliable trees, though concatenation has the disadvantage of potentially reinforcing biases by causing spuriously high bootstrap support for incorrect partitions (Gadagkar *et al.*, 2005).

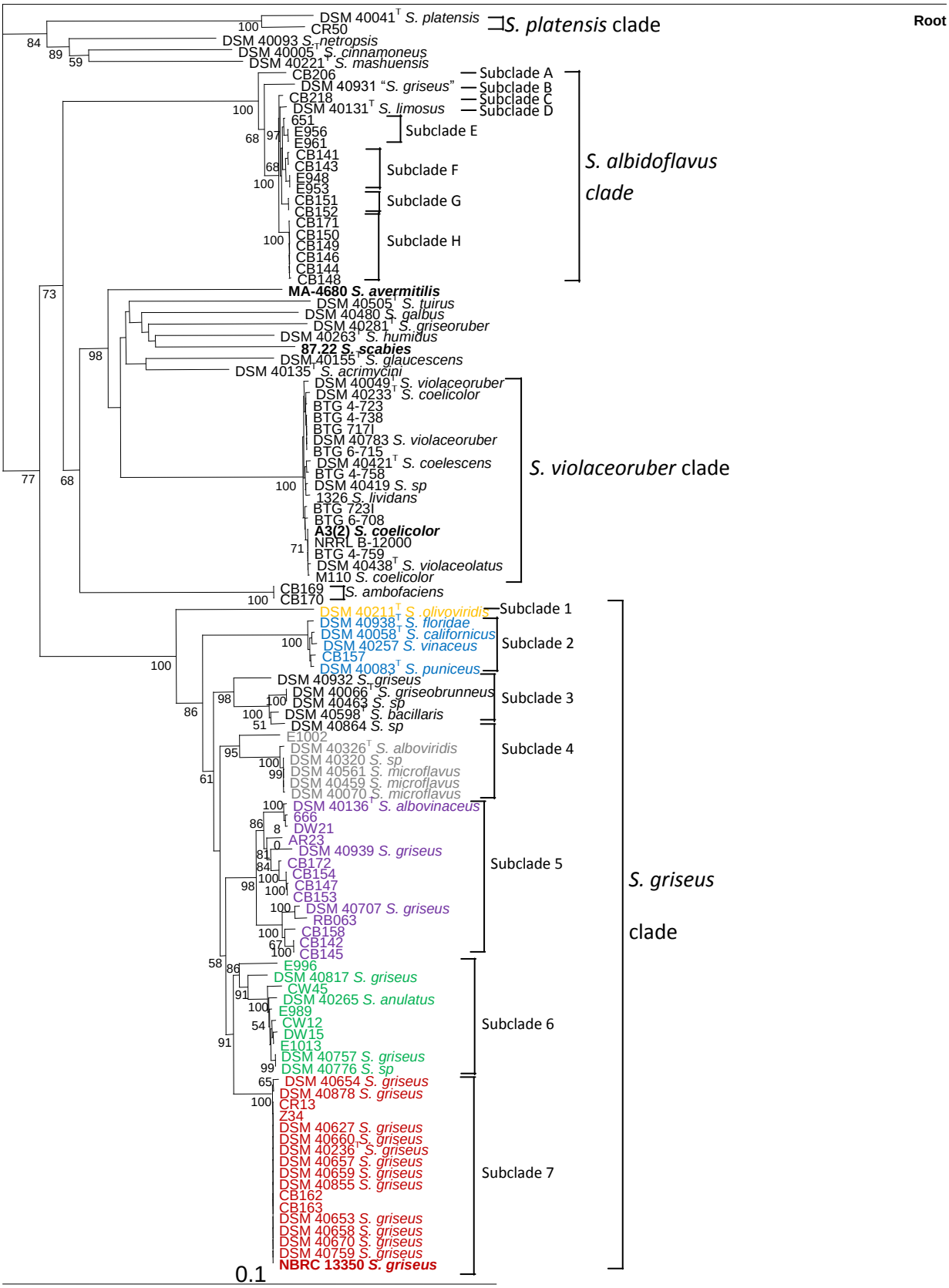


Figure 5.4. Phylogenetic trees based on the concatenated nucleotide sequences of eight protein-coding housekeeper genes. The trees were constructed using the neighbour-joining method; the numbers besides

the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% nucleotide dissimilarity. ^{T} denotes a species' type strain.

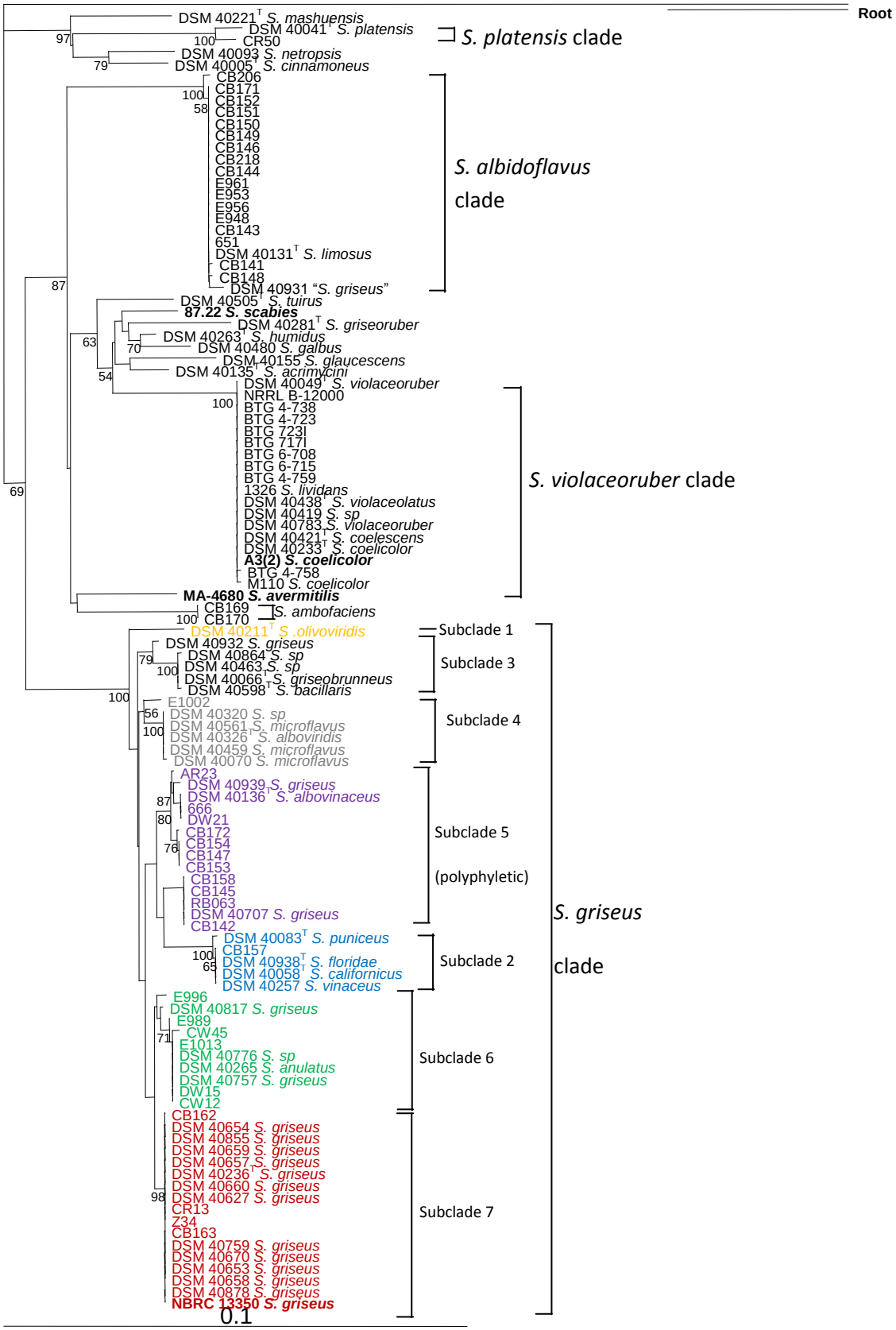


Figure 5.5. Phylogenetic trees based on the concatenated translated sequences of eight protein-coding housekeeper genes. The trees were constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% sequence dissimilarity. ^{T} denotes a species' type strain.

Both the concatenated nucleotide (Figure 5.4) and the amino acid (Figure 5.5) trees provided well-defined clades identified as species for *S. griseus*, *S. violaceoruber* and *S. albidoflavus*, as *S. limosus* is a synonym of *S. albidoflavus* (Hain *et al.*, 1997) , with improved phylogenetic resolution compared to both the 16S tree (Figure 5.2) and the individual gene trees (Appendix Figure 10.5 to Figure 10.20). It however failed resolve interspecies relationships and individual housekeeper gene phylogenies differed at the species group level. The *S. griseus* clade was more diverse in comparison to the *S. violaceoruber* and *S. albidoflavus clade*s. There was one potential instance of recombination amongst the housekeeper genes examined, where *S. griseus* DSM 40932 clustered with *S. bacillaris* DSM 40598^{T} in all trees except for *rplC*, where it was recovered with the *S. griseus* type strain (Figure 10.13).

The traditional cutoff point for the delineation of species has been a 97% 16S sequence identity , as that corresponds to a 70% result in a DNA/DNA hybridization (Stackebrandt & Goebel, 1994). More recently, it has been suggested that a cutoff point of 99% 16S sequence identity (Drancourt *et al.*, 2000). The partial 16S sequence data for novel isolates and the full 16S for characterized strains whose entire 16S sequence is available in GenBank were used to examine whether how well the clades fit the narrow (99%) and wider (97%) species definition. CR50 and DSM 40041 both belonged to the *S. platensis* species under both species definitions. *S. netropsis*, *S. cinnamoneus* and *S. mashuensis* were all different species under the narrow definition but belonged to the same species as *S.* platensis under the wider species definition. The strains in the *S. albidoflavus* clade demonstrated 99% or greater sequence identity to one another and thus qualified to be a species under both definitions, as did the *S. violaceoruber* clade. None of the strains in the sister clade to *S. violaceoruber*, containing *S. scabies*, had a sequence identity greater than 99%. All strains in that clade but *S. scabies* had a sequence identity to *S. coelicolor* above 97%, though most of them varied more than 3% from one another, indicating that these strains were different species. All the strains in the *S. griseus* clade fell in the same species under the wider species definition, however under the narrow species

definition Sublcades 4, 5, 6 and 7 form a single species, while Sublclades 1, 2 and 3 are three separate species.

Strain DSM 40931 is listed as belonging to *S. griseus*; however both the 16S rRNA sequence from GenBank and the sequence obtained in this study showed only a 95% sequence identity to that of *S. griseus* DSM 40236^{T}. It has a 99% sequence identity to *S. albus* NBRC 3711, indicating that this strain has been misclassified.

The phylogenetic trees based on nucleotide sequences had higher bootstrap values and their tree topologies were more likely to agree with that of the concatenated trees than the amino acid trees. This is to be expected due to the additional phylogenetic information from silent mutations that is lost when a sequence is translated and is thus unavailable when generating phylogenetic trees based on amino acid sequences. The increased conservation of protein sequences can assist in generating phylogenetic trees above the genus or species level, but it proves a hindrance when attempting to separate closely related strains.

The majority of soil isolates belonged to either the *S. griseus* or the *S. albidoflavus clade*s. Only CR50, which was related to *S. platensis*, and CB169 and CB170, whose closest relative based on their 16S rRNA sequence was *S. ambofaciens*, fell outside these two groups. Of the Cayo Blanco isolates, 42% (10) strains belonged to *S. griseus* and 50% (12) to *S. albidoflavus*. Of the UK isolates, 56% (5) strains belonged to *S. griseus* and 54% (4) to *S. albidoflavus*.

5.3.4 ClonalFrame analysis

In order to check for potential biases in the structure of the concatenated gene tree, a multilocus phylogenetic tree in which each gene is analyzed separately was also generated using ClonalFrame (Didelot & Falush, 2007). All eight protein-coding housekeeper genes were used in the generation of the tree.

The tree generated by ClonalFrame (Figure 5.6) is in overall agreement with the concatenated tree (Figure 5.4). All the major clades (*S. albidoflavus*, *S. violaceoruber*, *and S. griseus*) were present in both; however other less robust clades, such as the one composed of *S. platensis*, *S. netropsis*, *S. cinnamoneus* and *S. mashuensis*, were also recovered. The subclades comprising the *S. griseus* species

group are also identical, though the branching pattern was somewhat different. This is to be expected, as the positions of these subclades not strongly supported either in the concatenated or the ClonalFrame tree. This also applies to the sister clade of *S. violaceoruber*; the branching pattern of these strains is too weakly supported to provide reliable phylogenetic information.

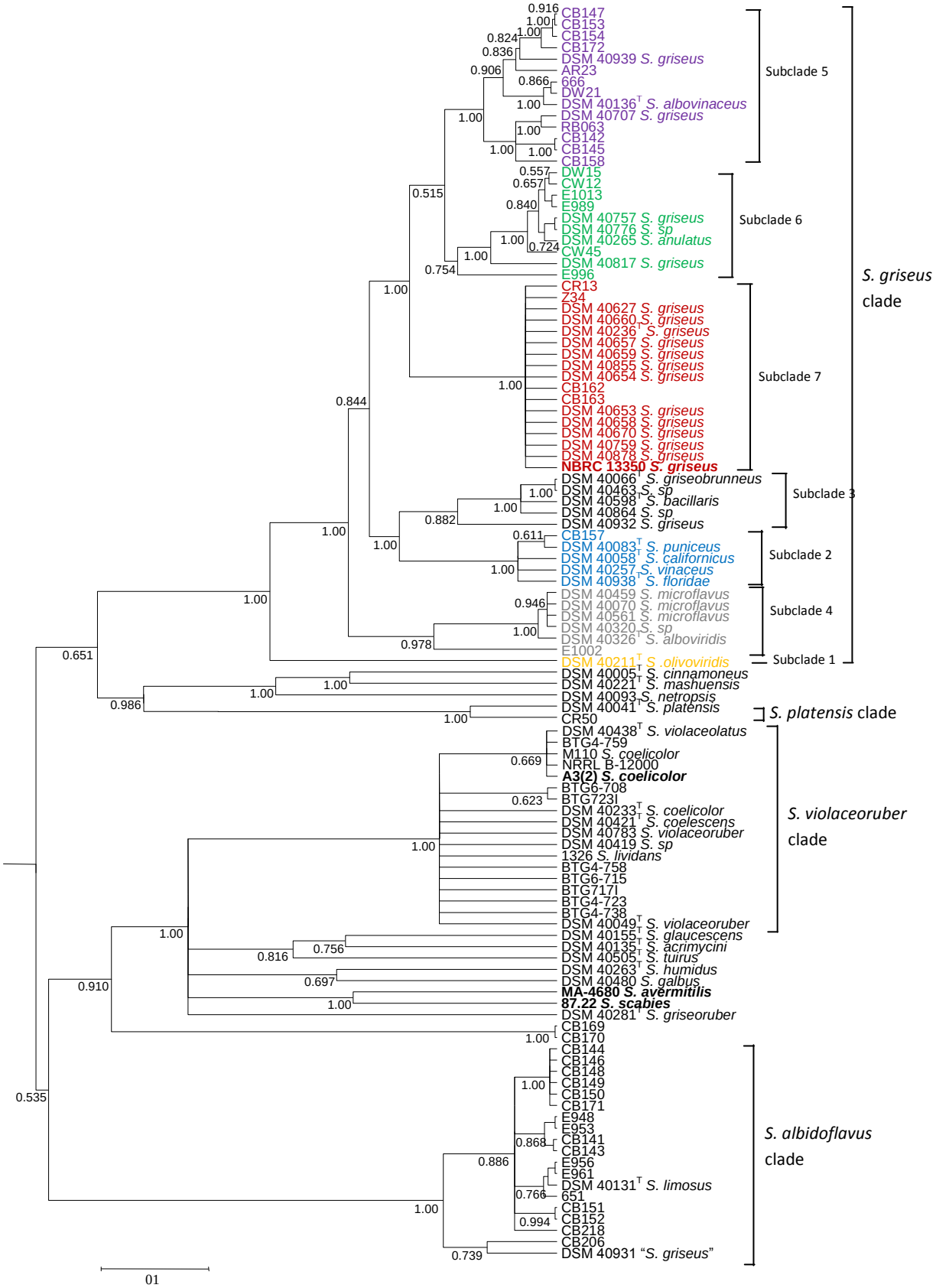


Figure 5.6. Majority-rule (0.5) consensus tree based on the posterior distribution of genealogies. The level of confidence in each node is noted next to it.

5.3.5 Detection of recombination

An attempt was made to identify recombination events both within and between housekeeper genes using the BAPS (Corander & Marttinen, 2006) and BRAT (Marttinen *et al.*, 2008) programs; however the MLST sequence types for the strain collection were too diverse for BAPS to generate realistic clusters of the strains. The only cluster of the four generated which mirrored the phylogeny of the strains examined was that containing *S. violaceoruber* isolates which were all closely related. The remaining three clusters contained a random assortment of isolates from across all species groups except *S. violaceoruber*. The GARD algorithm (Pond & Frost, 2005) was therefore used to detect recombination breakpoints within the housekeeper gene sequences instead.

Table 5.3. Location of recombination breakpoints (if present) and their level of significance. Breakpoint lists the nucleotide at which recombination has taken place.

Gene	Breakpoints	p-values
16S	184, 504	< 0.01, < 0.01
argH	N/A	N/A
atpD	222	< 0.01
efp	N/A	N/A
ppc	N/A	N/A
rplC	282	< 0.01
rpoB	300	< 0.01
rpsL	N/A	N/A
sodF	N/A	N/A

There was evidence that the genes *atpD*, *rplC* and *rpoB* have undergone recombination once and 16S twice (Table 5.3).

5.3.6 Phylogenetic network

A phylogenetic split network using the concatenated gene sequences was generated using SplitsTree (Huson & Bryant, 2006). Phylogenetic networks have the advantage that they can more easily display through their structure instances of horizontal gene transfer or recombination, making them useful if it is suspected that such events have taken place.

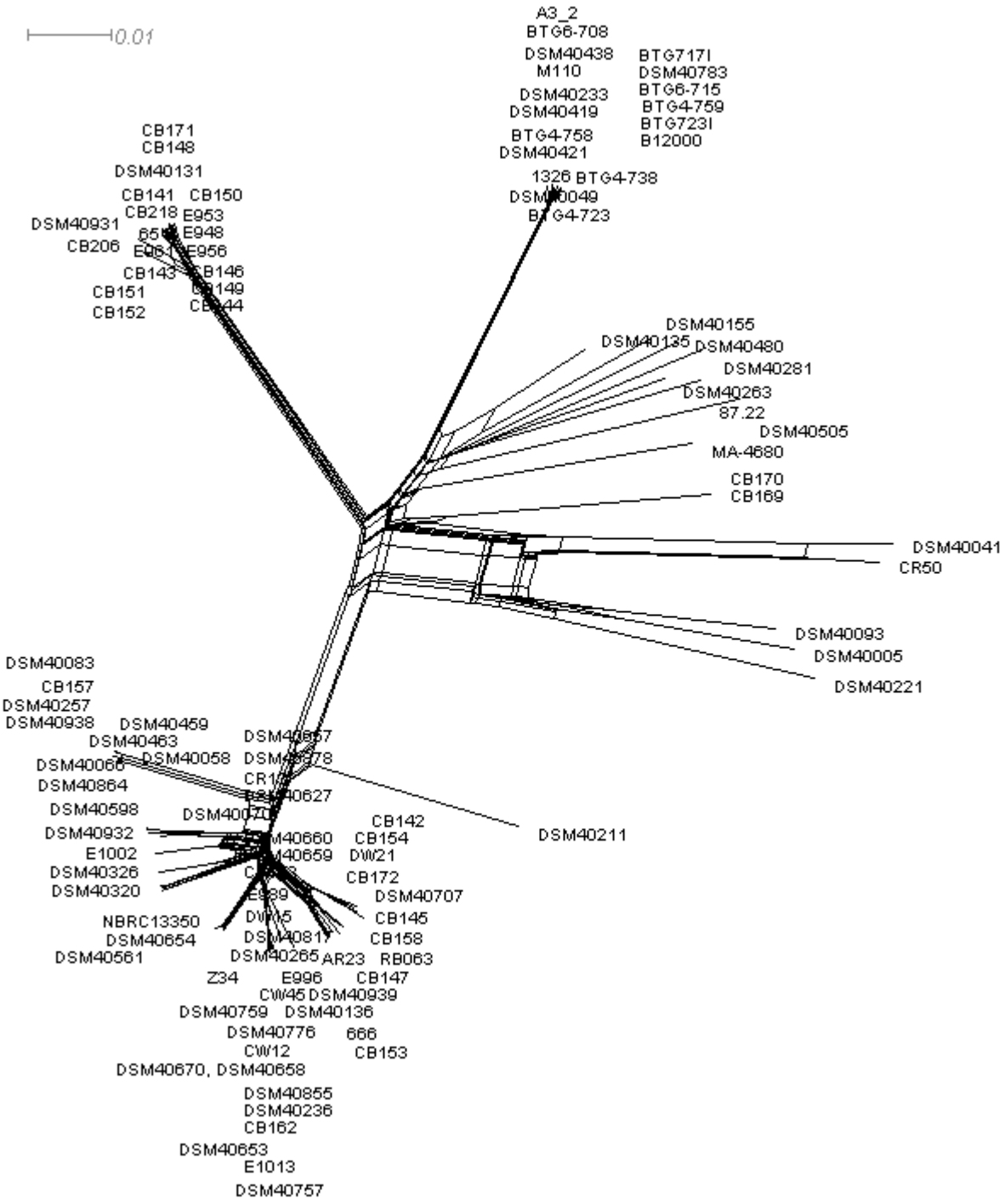


Figure 5.7. Phylogenetic split network of the concatenated housekeeping gene trees generated using the Kimura 2-parameter algorithm to calculate the distances for network creation.

The split network for the concatenated genes had the same structure as the neighbour-joining tree (Figure 5.7). *S. griseus*, S. *violaceoruber* and *S. albidoflavus* all formed discrete groupings and the looser cluster of *S. platensis*, CR50, *S. netropsis*, *S. cinnamoneus* and *S. mashuensis* was also present. The strains forming subclades within the *S. griseus* clade were also recovered as groups with the same

branching pattern on the split network, indicating that the recombination events in some of the housekeeper genes have not affected the overall structure of the concatenated tree.

5.3.7 MultiLocus Sequence Typing

There were groups of strains within the *S. griseus*, *S. albidoflavus* and *S. violaceoruber* clades that were extremely closely related to one another and whose interrelationships were impossible to resolve. Multilocus Sequence Typing (MLST) permits the separation of closely related strains, which may have only one SNP difference, into clonal complexes and can determine the putative founders of a population using algorithms such as eBURST (Spratt *et al.*, 2004). The genes used were *argH*, *atpD*, *efp*, *ppc*, *rplC*, *rpoB* and *sodF*. The *rpsL* gene was omitted, as its sequence was both the shortest and the most conserved.

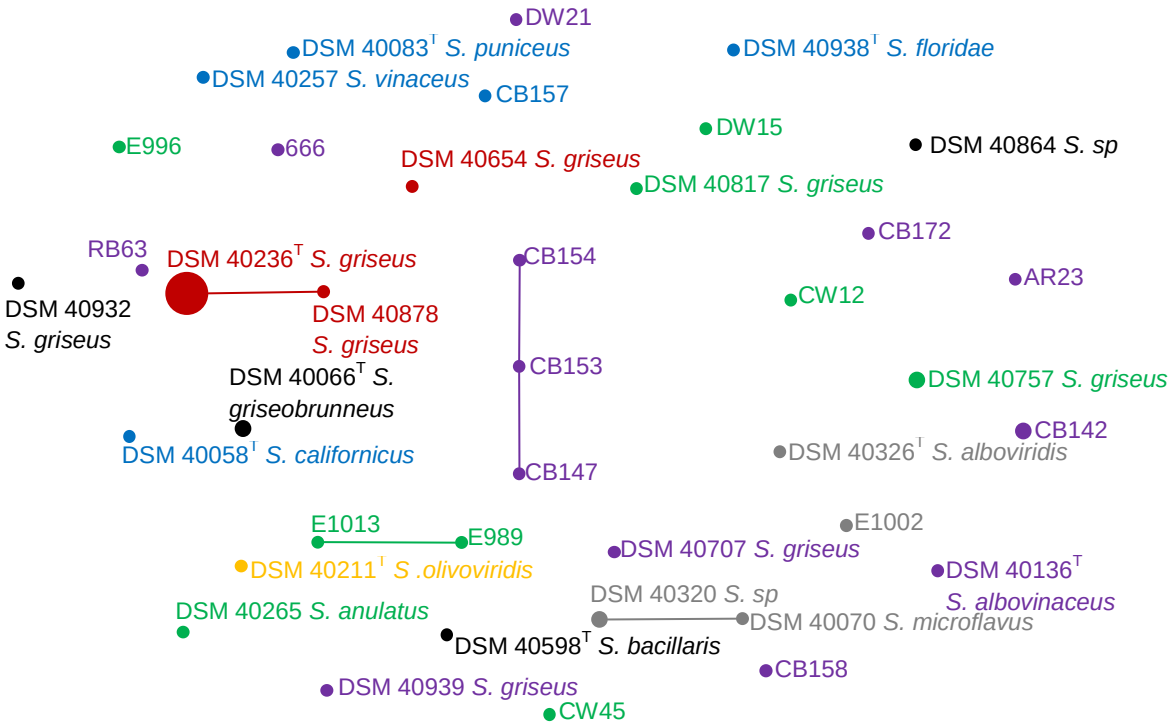


Figure 5.8. Population snapshot of the *S. griseus* clade. The sequence type of DSM 40459 was identical to that of DSM 40320, that of DSM 40463 to DSM 40066^T, those of DSM 40627, DSM 40653, DSM 40657, DSM 40658, DSM 40659, DSM 40660, DSM 40670, DSM 40759, DSM 40855, CB162, CB163, CR13, Z34 and NBRC 13350 were identical to that of DSM 40236^T, that of DSM 40776 to DSM 40757, and that of CB145 to CB142.

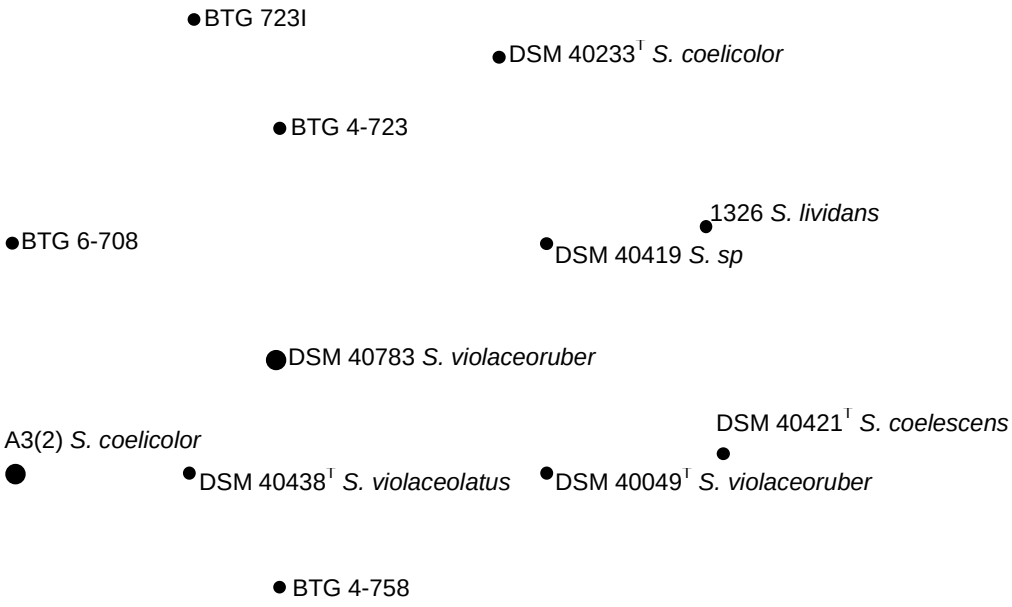


Figure 5.9. Population snapshot of the *S. violaceoruber* clade. The sequences type of NRRL B-12000, M110 and BTG4-759 were identical to that of A3(2), and those of BTG4-738, BTG6-715 and BTG717I to DSM 40783.

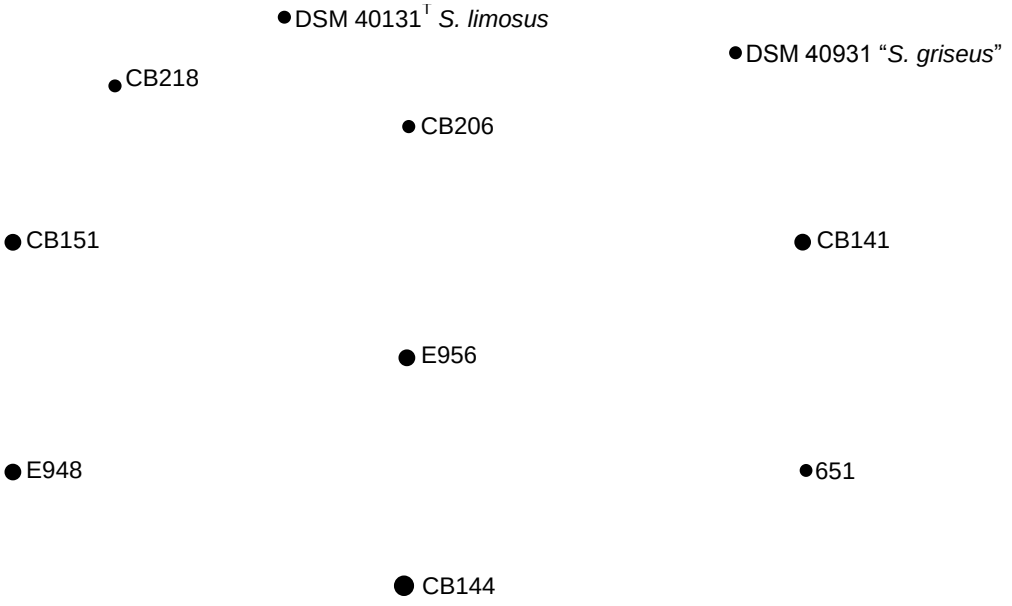


Figure 5.10. Population snapshot of the *S. albidoflavus clade*. The sequence type of CB143 was identical to that of CB141, those of CB146, CB148, CB149, CB150 and CB171 to CB144, that of CB152 to CB151, that of E953 to E948 and that of E961 to E956.

Table 5.4. Strains forming part of the *S. griseus* streptomycin producer clonal complex.

Strain	Origin
DSM 40236^T	<- E. B. Shirling, ISP <- S. A. Waksman, IMRU. [ETH 4289]. Soil.
DSM 40627	<- J. B. Routien, Chas. Pfizer & Co., FD 1076 <- S. A. Waksman
DSM 40653	<- R. E. Gordon; IMRU 3475 <- H. W. Anderson; variant 42.1 of 4. *
DSM 40657	<- R. E. Gordon; IMRU 3511 <- Parke, Davis & Co; 04506. *
DSM 40658	<- R. E. Gordon; IMRU 3523 <- Merck & Co. <- N. J. Rahway; 7R2557. *
DSM 40659	<- R. E. Gordon; IMRU 3538 <- H. Umezawa; S-3. *
DSM 40660	<- R. Gordon, IMRU <- H. Umezawa, S-4.
DSM 40670	-
DSM 40759	<- E. Merck AG, Darmstadt, M 4052 <- ATCC <- S. A. Waksman, IMRU <- Merck&Co. Inc., 1-118 MA 45
DSM 40855	<- K. Kieslich, Schering 308 <- ATCC <- S.A. Waksman, 4. Host of Phage S6 (DSM 49150).
DSM 40878	<- K. Kieslich; Schering 372 <- IFO <- K. Nakazawa; 2315.
CB162	Cayo Blanco, Cuba: soil isolate
CB163	Cayo Blanco, Cuba: soil isolate
CR13	Dossenheim, Germany: Control site soil isolate
Z34	Musonwiji , Zambia: Soil at water edge isolate
NBRC 13350	-

In all three species, the majority of strains fail to form clonal complexes. *S. griseus* had four pairs of strains with identical sequence types (STs), two clonal complexes composed of two strains, one composed of three and one composed of 16 streptomycin producers of which 15 had identical STs (Figure 5.8). The majority of streptomycin-producing *S. griseus* strains were isolated at different sites and are therefore unlikely to be the same isolate submitted under different accession numbers (Table 5.4). *S. albidoflavus* had four pairs of strains and a set of eight strains with identical STs, but no clonal complexes (Figure 5.9). *S. violaceoruber* had a set of four and a set of three strains with identical STs and no clonal complexes (Figure 5.10). A number of the isolates with identical STs from the same site (e.g. CB142 and CB145 or E956 and E961) are simply the result of multiple re-isolations of the same organism. Not all organisms with identical sequence types originated from the same site though; the BTG strains with an identical ST to DSM 40783 were isolated at different times and at different locations from the latter; the same applies for many of the strains with the same ST as DSM 40236^T.

5.3.8 Pairwise analysis of housekeeper genes

Table 5.5. Pairwise analysis of the distance matrix between housekeeper genes using the Kimura 2-parameter nucleotide substitution model. Conc is the concatenated sequence from all eight protein-coding genes.

	argH	*atpD*	*efp*	*ppc*	*rplC*	*rpoB*	*rpsL*	*sodF*	conc
16S	0.808	0.934	0.918	0.780	0.830	0.864	0.649	0.941	0.928
argH		0.823	0.805	0.789	0.894	0.872	0.853	0.867	0.929
atpD			0.932	0.734	0.857	0.872	0.708	0.912	0.939
efp				0.808	0.829	0.913	0.716	0.918	0.949
ppc					0.731	0.830	0.617	0.861	0.875
rplC						0.827	0.815	0.836	0.921
rpoB							0.792	0.905	0.951
rpsL								0.691	0.962
sodF									0.962

The distance matrices of all the genes show considerable correlation to one another (Table 5.5). The two most divergent genes are *ppc* and *rpsL*, though this is not because they have been co-selected or co-transduced as they are an average of 2.3 Mb apart on the three available genomes and have the lowest correlation value between themselves. All the housekeeping genes, including the 16S rRNA gene sequence, strongly correlate with the concatenated protein-coding housekeeper genes, indicating that they have very similar evolutionary histories. This is also supported by the fact that the composition of the species groups remains constant across all individual gene trees.

Table 5.6. Pairwise analysis of the distance matrix between translated housekeeper genes using the Jones-Taylor-Thornton amino acid substitution model. Conc is the concatenated sequence from all 8 protein-coding genes.

	AtpD	Efp	Ppc	RplC	RpoB	RpsL	SodF	conc
ArgH	0.542	0.638	0.692	0.858	0.678	0.116	0.756	0.820
AtpD		0.892	0.549	0.686	0.706	0.0395	0.854	0.854
Efp			0.680	0.760	0.853	0.0270	0.865	0.928
Ppc				0.732	0.751	0.0827	0.731	0.824
RplC					0.750	0.0962	0.814	0.903
RpoB						0.0237	0.8134	0.901
RpsL							0.0708	0.946
SodF								0.946

The amino acid distance matrices show considerably less correlation to one another than the nucleotide ones (Table 5.6). This may be because there is a loss of phylogenetic information by the elimination of synonymous mutations. For example *rpsL* displays numerous synonymous SNPs, however when translated only four

strains display a single amino acid change each. This results in its amino acid distance matrix having almost no correlation to those of other, more variable translated sequences.

5.3.9 Omega nucleotide substitution ratios for housekeeper genes

The ratio of non-synonymous to synonymous substitution ratio per nucleotide site of housekeeper genes was examined using the yn00 program from the from the PAML (Phylogenetic Analysis by Maximum Likelihood) software package (Yang, 2007). This procedure can help determine whether a gene is under purifying, neutral or diversifying selection, as mutations that change the amino acid sequence will be selected against, unlike silent mutations, in genes under purifying selection while in genes under diversifying selection non-synonymous mutations will be preferentially fixed in comparison to silent mutations (Yang *et al.*, 2000). Genes under purifying selection are expected to have ω <1, genes under no selection ω =1 and genes under diversifying selection ω >1.

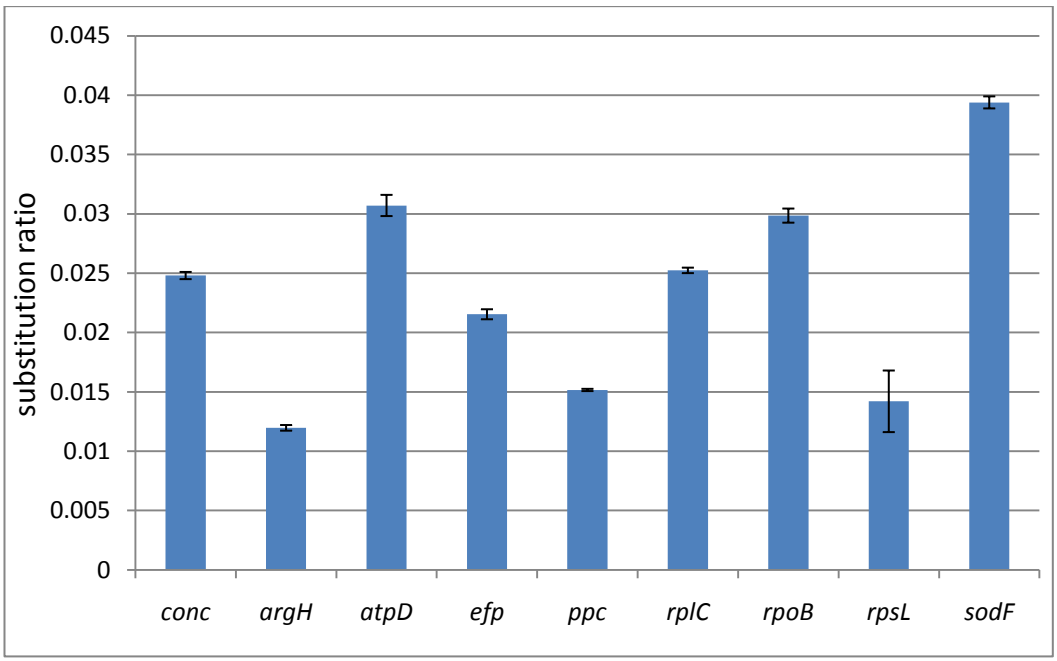


Figure 5.11. The average non-synonymous/synonymous nucleotide substitution ratio for housekeeper genes for all strains used in this study. Conc refers to the concatenation of all 8 housekeepers.

The housekeeper genes demonstrated a certain amount of variation; however their omega values ranged from 0.01 to 0.04, which are all very low and indicate that the genes are strongly conserved and under high purifying selection (Figure 5.11). *argH*, which is involved in amino-acid biosynthesis was the most conserved gene,

surpassing genes involved in transcription and translation such as *rplC*, *rpoB* and *rpsL*. *sodF* was the least conserved housekeeper, which is to be expected as its function is not as central to the cell's growth and can be performed by other enzymes as well such as nickel-containing superoxide dismutases.

5.3.10 Positive and negative selection on individual codons

Obtaining the average ω value for a gene is useful to determine under how much purifying selection it is, but positive selection tends to act in a small number of sites and can thus be obscured by the remainder of the gene. In order to examine the evolution of these genes in greater detail, individual codons were tested with the SLAC and FEL algorithms, to determine whether they were under significant positive or negative selection ($p < 0.05$). The REL algorithm was not used because the sample set was too large (Pond & Frost, 2005).

Table 5.7. Number of housekeeper gene codons under positive or negative selection ($p < 0.05$). +ve and -ve lists the number of codons under selection, % the percentage of the sequence examined the codons comprise.

Gene	SLAC		FEL		SLAC		FEL	
	+ve	%	+ve	%	-ve	%	-ve	%
argH	0	0.00	1	0.60	69	41	86	51
atpD	0	0.00	0	0.00	31	23	43	30
efp	0	0.00	1	0.60	26	16	47	28
ppc	0	0.00	0	0.00	76	44	87	49
rplC	0	0.00	0	0.00	32	19	49	29
rpoB	0	0.00	0	0.00	32	17	44	24
rpsL	0	0.00	0	0.00	14	13	18	17
sodF	1	0.60	0	0.00	58	35	64	39

SLAC produced fewer significant results than FEL because it is a more conservative algorithm (Table 5.7). There was evidence for positive selection on a total of only three codons while there were hundreds of negatively selected ones, confirming that the housekeeper genes are predominantly under purifying selection.

5.3.11 Distribution of *S. griseus* phenotype

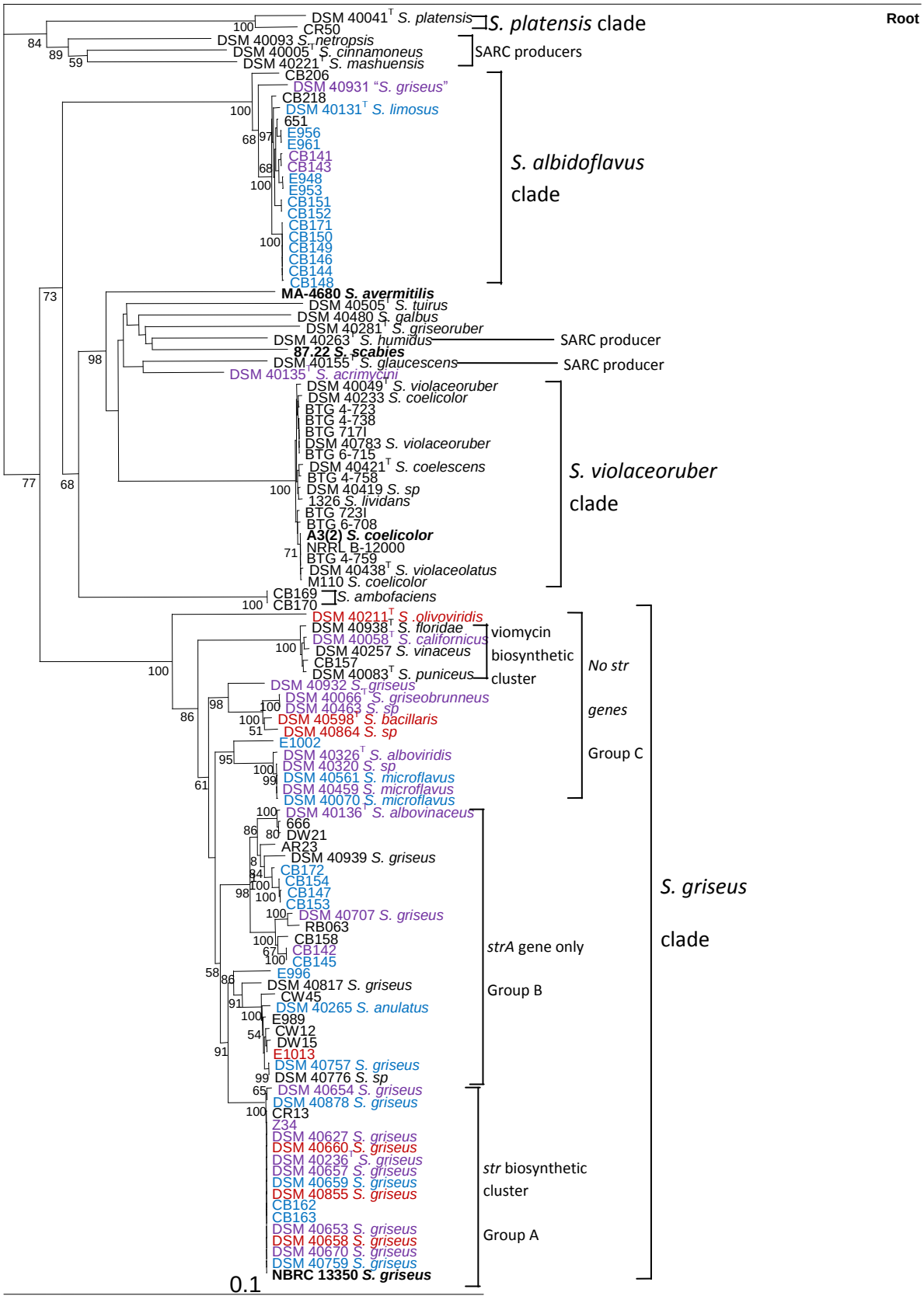


Figure 5.12. Distribution of *S. griseus* type strain phenotype. Black colouration indicates a completely different phenotype (-ve), blue means very similar spore colour (C), red means very similar colony morphology (M) and purple means very similar spore colour and colony morphology (C M).

The *S. griseus* phenotype is at least partially present in 72% of the strains belonging to the *S. griseus* clade (Figure 5.12). It is absent from the *S. violaceoruber* clade, and all other species included in this study, except for *S. acrimycini* and *S. albidoflavus*, also failed to display it. The high frequency of the *S. griseus* phenotype in *S. albidoflavus* strains (84% of strains included in this study) is one of the reasons that such a large number of isolates belonging to this clade were examined.

5.3.12 Distribution of streptomycin resistance phenotype

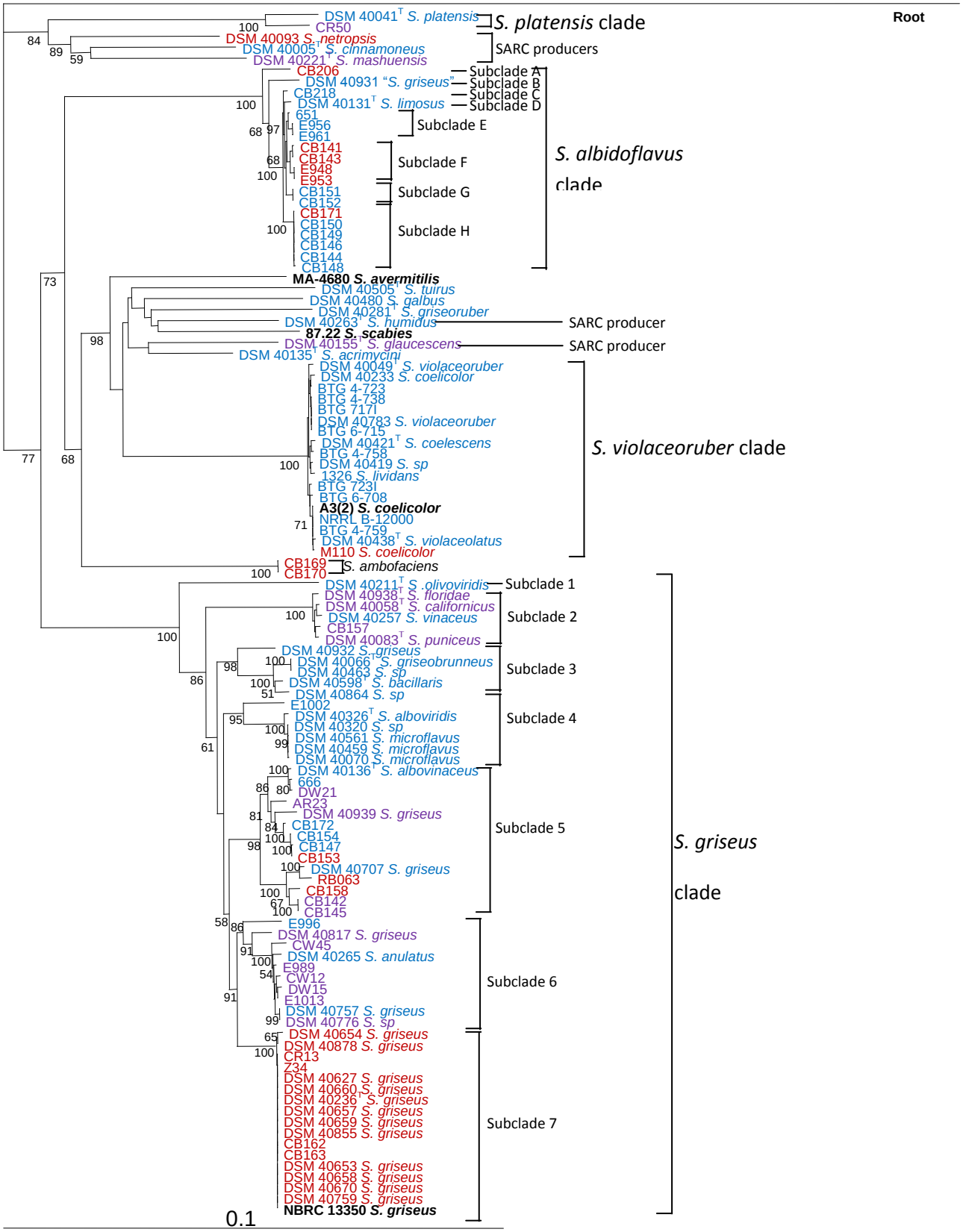


Figure 5.13. Distribution of streptomycin resistance. Blue colouration indicates sensitivity, purple means ability to grow on plates containing 10 µg ml^{-1} streptomycin and red means ability to grow on plates containing 50 µg ml^{-1}.

From the *S. violaceoruber* clade only M110, which has been engineered to have a ribosomal mutation, was resistant to streptomycin. In *S. albidoflavus*, Subclades A and F uniformly displayed high resistance. In Subclade H however, one isolate (CB171) displayed high resistance despite having an identical sequence type to the other isolates that were sensitive. In *S. griseus* Subclades 3, 4 and 7 had uniform resistance phenotypes, however Subclades 2 and 6 had a mix of sensitive and low level resistance and Subclade 5 had phenotypes ranging from sensitive to high level resistance. The two *S. ambofaciens* isolates and *S. netropsis* had high level resistance, CR50, *S. glaucescens* and *S. mashuensis* had low level resistance and all other strains were sensitive (Figure 5.13).

5.3.13 Distribution of viomycin resistance phenotype

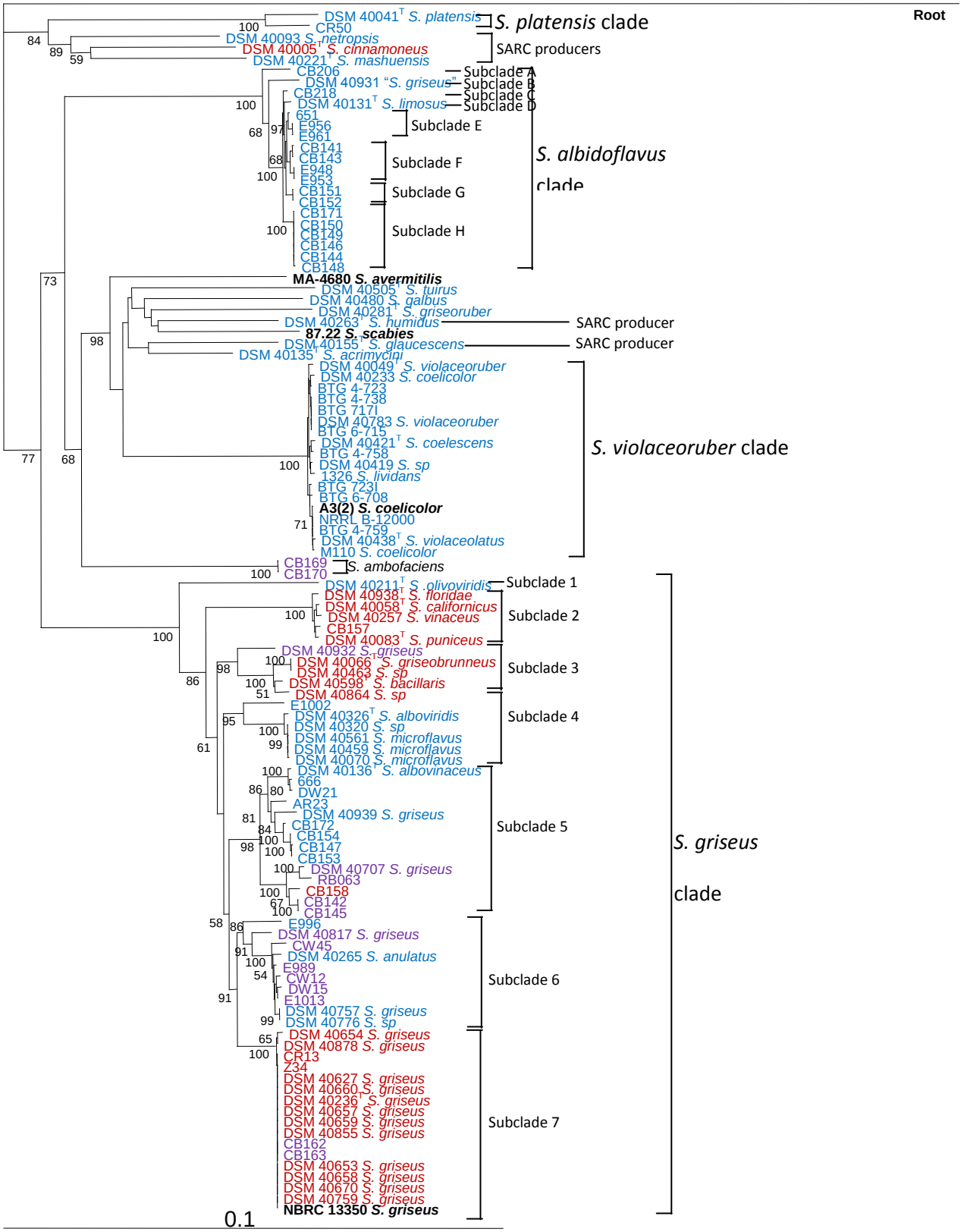


Figure 5.14. Distribution of viomycin resistance. Blue colouration indicates sensitivity, purple means ability to grow on plates containing 10 µg ml^{-1} viomycin and red means ability to grow on plates containing 50 µg ml^{-1}.

The distribution of viomycin resistance was considerably more uniform than that of streptomycin. All members of the *S. violaceoruber* and *S. albidoflavus clade*s were sensitive. From *S. griseus*, Subclades 1, 4 and one branch of clade 5 were sensitive. The other branch of Subclade 5 displayed low level resistance and had one member with high, Subclade 2 had uniformly high, Subclade 3 had high with one exception and Subclade 7 with two. The only clade that was a mix of sensitive and low level resistance strains was Subclade 6. All other strains were sensitive with the exception of the two *S. ambofaciens* isolates, which had low resistance, and *S. cinnamoneus*, which had high level resistance (Figure 5.14).

5.3.14 Distribution of *strA* gene

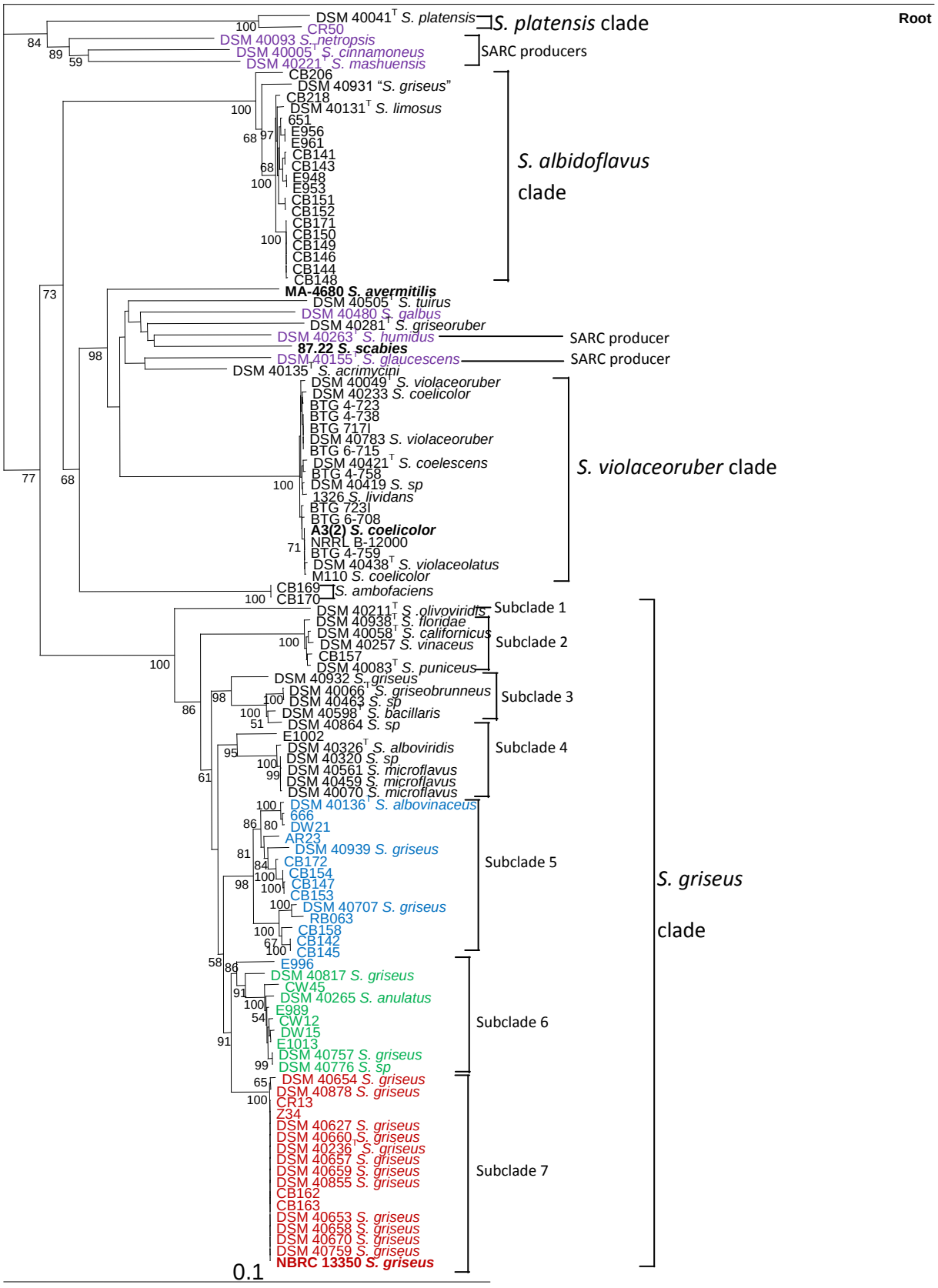


Figure 5.15. Distribution of *strA*. Red colouration indicates the presence of a Group A *strA* (found within a streptomycin cluster), green a Group B1 *strA* (not associated with a biosynthetic cluster), blue a Group B2 *strA* (not associated with a biosynthetic cluster) and purple and *strA* associated with a biosynthetic cluster other than the *S. griseus* streptomycin cluster. Black denotes an absence of the gene.

The *strA* gene belonging to Group A (Figure 3.3) was found only in Subclade 7 and it was always associated with streptomycin biosynthetic genes (Figure 5.15). Strains possessing Group B1 *strA* genes were also only found in a single subclade, with the exception of E996 belongs to Subclade 6. Group B2 *strA* genes were all found in Subclade 5, a sister clade to 6 and 7. The tree topology of the ClonalFrame tree (Figure 5.6) was more similar to that of the *strA* gene compared to the concatenated tree, as it depicted Subclade 5 and 6 as sister clades. Neither of these branching patterns however was strongly supported.

S. humidus, *S. galbus* and *S. glaucescens* all had *strA* homologues that form part of aminoglycoside clusters, however due to the poor resolution at that section of the tree it is impossible to discern whether they were more closely related to one another than to strains missing the gene. *S. netropsis*, *S. cinnamoneus* and *S. mashuensis* all form one clade. CR50 was related to them, but is unlikely to have acquired the cluster from them as its *strA* is more closely related to that of *S. galbus* and *S. griseus*.

Pairwise analysis between all the *strA* sequences and the concatenated housekeeping genes revealed an r= 0.378 for nucleotide and r= 0.320 for amino acid sequences. Performing the same analysis for Group B *strA* and housekeepers produces an r= 0.902 for nucleotide and r= 0.824 for amino acid sequences. Group A *strA* had an r= -0.0572 for nucleotide sequences, however that is due to the fact that there is only one SNP in *strA* sequences and one in the concatenated housekeeping genes, which skewers the results. The amino acid sequences produced an r= 1.00. The remaining cluster-associated *strA* genes had an r= 0.786 for nucleotide and r= 0.727 for amino acid sequences. The low overall correlation between concatenated housekeeping genes and *strA* was presumably due to HGT. The relatively high correlations within Groups A and B however indicate that there have been few to no HGT evens within these groups. The lower correlation coefficient for the non-*griseus strA* genes indicated that it is likely there have been some instances of HGT. Using the GARD algorithm revealed that there has been a recombination event at position 331 ($p < 0.01$).

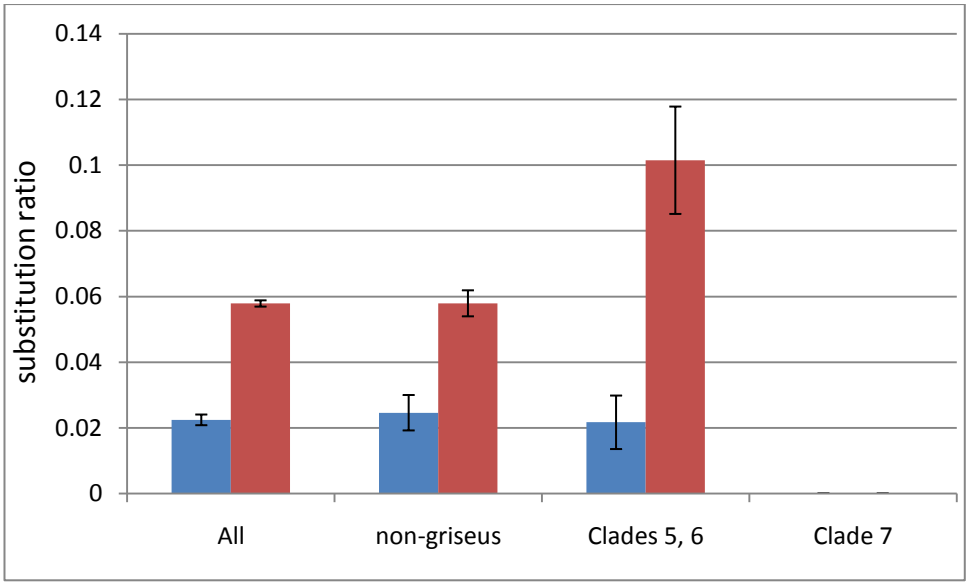


Figure 5.16. Average non-synonymous/synonymous nucleotide substitution ratio for the concatenated housekeeping genes (blue) and the *strA* gene (red). All represents the omega values for every strain possessing *strA*, non-*griseus* includes all strA genes found outside the S. griseus clade while Subclade 5, 6 and Subclade 7 includes the genes found in the respective *S. griseus* subclades.

None of the Subclade 7 *strA* or housekeeping genes had any non-synonymous substitutions and as a result it is impossible to draw inferences about the rate of evolution of the genes in that clade as its strains have not had enough time to diverge. The Group B *strA* strains (Subclades 5, 6) was the most diverged, indicating that they are under less purifying selection. The *strA* homologues from non-*griseus* strains were more conserved, which is to be expected as they form part of biosynthetic clusters and may play a role in prevention of cellular suicide (Figure 5.16).

Using SLAC, FEL and REL to detect selection on the codon level of all the *strA* sequences reveals no codons under positive selection and 84 codons under negative selection ($p < 0.05$). When examining solely the Group B *strA* genes, SLAC identifies codon 140 as being under positive selection and all three algorithms identify a total of 33 sites as being under negative selection ($p < 0.05$).

5.3.15 Distribution of *strW* gene

In *S. griseus*, *strW* was found exclusively in Subclade 7 (Figure 5.15). Outside that species it was present in *S. galbus*, *S. glaucescens*, *S. mashuensis* and CR50. Pairwise analysis between all the *strW* and the concatenated housekeeping genes produces an $r = 0.946$ for nucleotide and $r = 0.936$ for amino acid sequences,

indicating that the distances of the two gene sets are very congruent. The GARD algorithm failed to identify any instances of recombination within the amplified sequence.

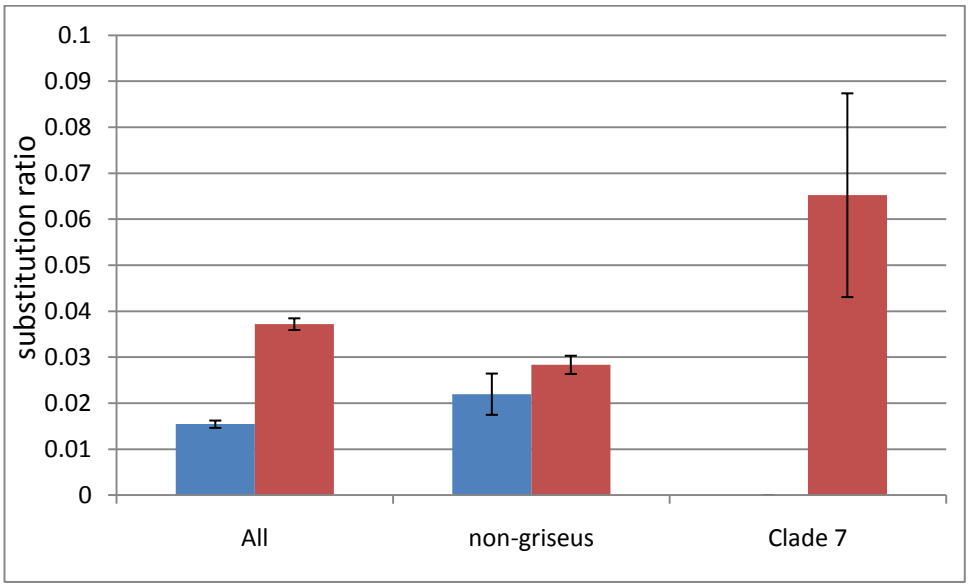


Figure 5.17. Average non-synonymous/synonymous nucleotide substitution ratio for the concatenated housekeeping genes (blue) and the *strW* gene (red). All represents the omega values for every strain possessing *strW*, non-*griseus* includes all *strW* genes found outside the *S. griseus* clade while Subclade 7 includes the genes found in that *S. griseus* subclade.

The *strW* gene is under less purifying selection than the housekeeping genes (Figure 5.17). This is especially evident from Subclade 7, as the housekeeping genes in that clade have no non-synonymous substitutions but the *strW* gene have four. Use of SLAC, FEL and REL detected two codons (at positions 53 and 157) as being under positive selection and 38 under negative selection ($p < 0.05$).

5.3.16 Distribution of *aphE* gene

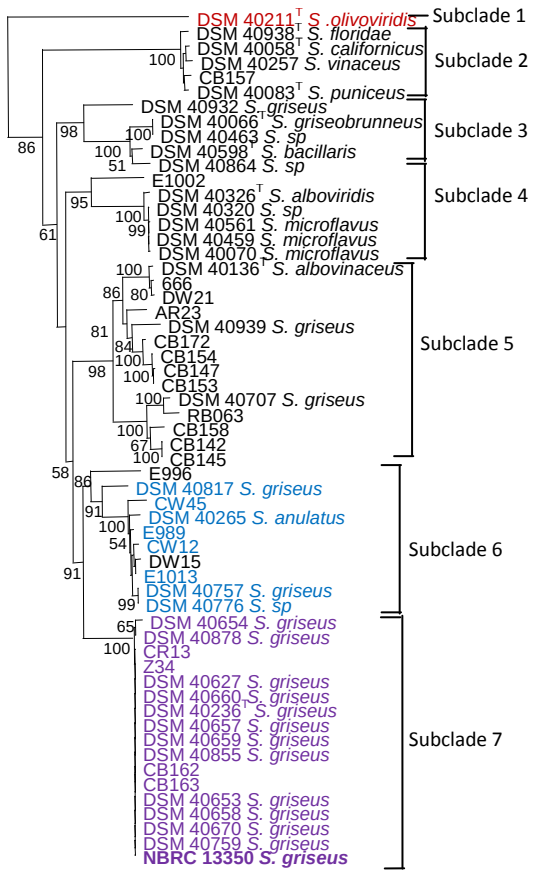


Figure 5.18. Distribution of *aphE*. Purple colouration indicates the presence of a Group A *aphE*, blue a Group B *aphE* (not associated with a biosynthetic cluster) and red a Group C *aphE*. Black denotes an absence of the gene. Clades other than that of *S. griseus* are omitted because the gene was absent in them.

The *aphE* gene was amplified from three subclades which mirrored the three clades of *aphE*'s phylogenetic tree (**Error! Reference source not found.**, Figure 5.18). Correlating the nucleotide distance matrices of the concatenated housekeeping genes and *aphE* produces an r= 0.923 for nucleotide sequences, though removing *S. olivoviridis* (Subclade 1) from the analysis results in r= 0.990. This indicates that the evolutionary history of housekeeping genes is almost identical with that of *aphE* and that it spread via vertical descent from the last ancestor of the Subclade 6 and Subclade 7 strains, though isolate DW15 appears to have lost *aphE*. The exception to that must be *S. olivoviridis*, as its housekeeping genes are too diverged in comparison to *aphE*. Presumably it arrived in that strain via HGT. The GARD algorithm failed to identify any instances of recombination within the amplified sequence.

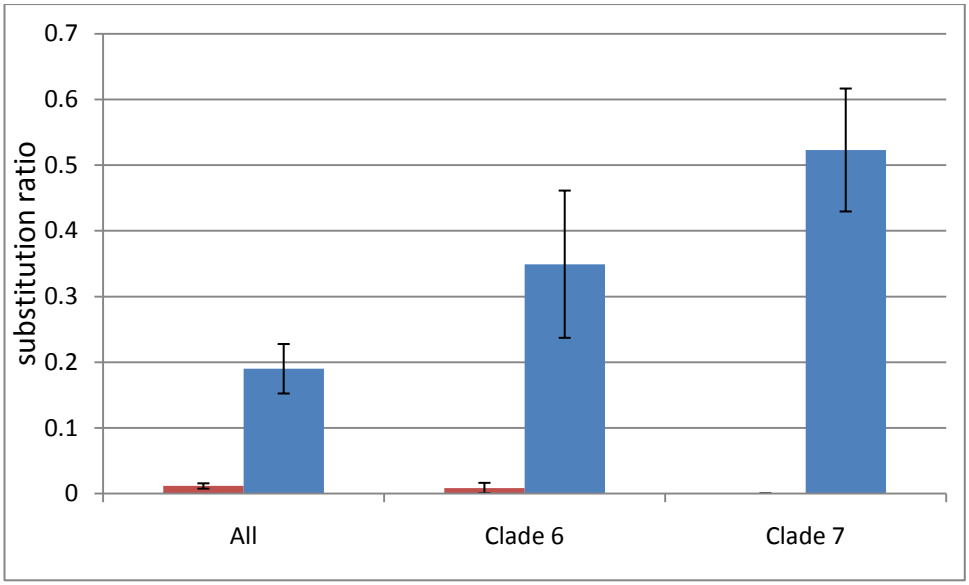


Figure 5.19. Average non-synonymous/synonymous nucleotide substitution ratio for the concatenated housekeeping genes (red) and the *aphE* gene (blue). All represents the omega values for every strain possessing *aphE* while Subclade 6 and Subclade 7 is the ratio for strains only each respective clade.

In all the strains the *aphE* non-synonymous/synonymous substitution ratio was significantly higher compared to the housekeeping genes (Figure 5.19). The reason that the ratio was lower when all the strains were included than either Subclade 6 or Subclade 7 is because the *S. olivoviridis* gene had a large number of synonymous mutations. The fact that there are approximately as half as many non-synonymous SNPs as there are synonymous ones in Subclade 7 indicates that *aphE* in these strains is under very little purifying selection. The genes in Subclade 6 are under slightly more purifying selection, which suggests that they are more important for those organisms.

5.3.17 Distribution of *vph*/*SGR421* gene

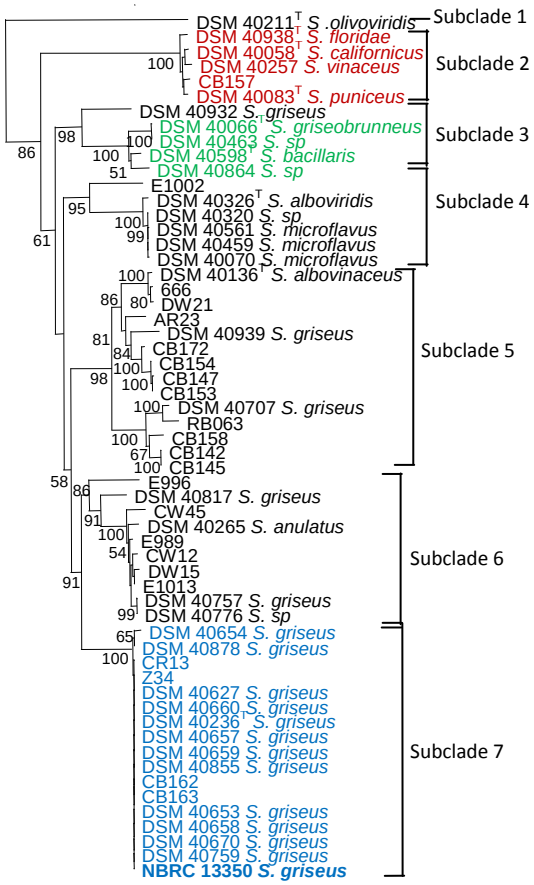


Figure 5.20. Distribution of *vph* and *SGR421*. Red colouration indicates the presence of *vph* (as well as *vioB* and *vioD*), green a Group B *SGR421* (not associated with a biosynthetic cluster) and blue a Group A *SGR421* (not associated with a biosynthetic cluster). Black denotes an absence of either gene. Clades other than that of *S. griseus* are omitted because the gene was absent from them.

The subclades containing strains possessing the *vph* and *SGR421* genes on the concatenated housekeeping gene tree (Figure 5.20) also mirror that of the *vph* tree. The *vph* gene, along with the biosynthetic *vioB* and *vioD*, were found exclusively in Subclade 2. The distance matrices of *vph/SGR412* had a correlation of *r*= 0.991 for the nucleotide and *r*= 0.987 for the amino acid sequences and the GARD algorithm failed to identify any instances of recombination, which indicates that there have been no recombination or HGT events in this gene's history. Comparison the distances of the nucleotide sequences of *vioB* and *vph* produced a correlation of *r*= 0.740, *vioD* and *vph* *r*= 0.652, *vioB* and *vioD* *r*= 0.738. These relatively low values are due to the very small total number of SNPs, which results in deceptively low

correlation coefficients. No recombination events were detected in either *vioB* or *vioD*.

All strains possessing the viomycin cluster, that includes the *vph* gene, had high level resistance to viomycin. The same applied to strains possessing the *SGR421* gene with the exception of CB162 and CB163, which only had low level resistance. Conversely, only two strains lacking these genes (CB158 and *S. cinnamoneus*) had high level resistance which indicates that SGR421 is the gene responsible for providing viomycin resistance to the strains possessing it.

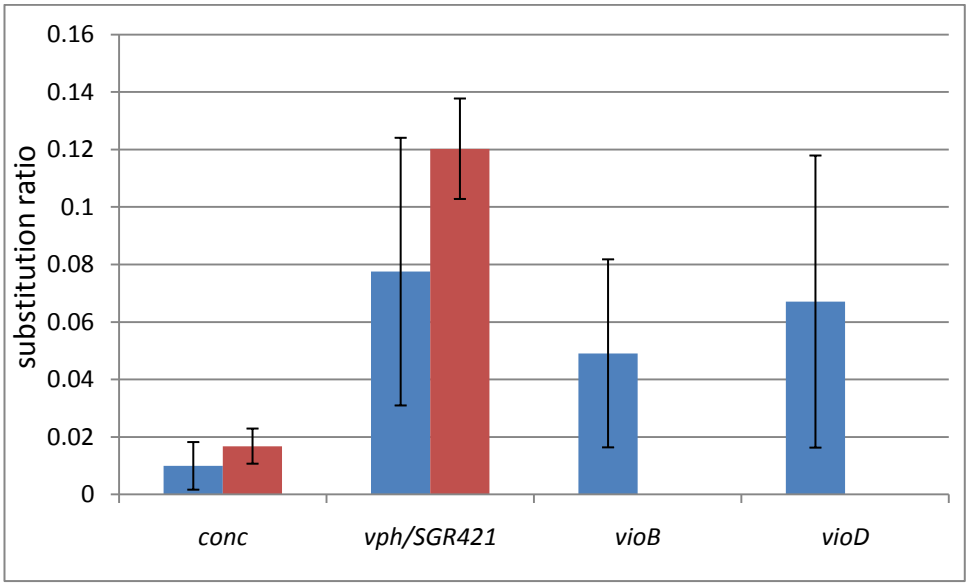


Figure 5.21. Average non-synonymous/synonymous nucleotide substitution ratio for the concatenated housekeeping genes, the *vph/SGR421* resistance genes and two viomycin biosynthetic genes in viomycin producers (blue) and resistance-only strains (red).

The *vph* and *SGR421* genes were significantly less conserved than the housekeepers (Figure 5.21). The biosynthetic genes from the producers (resistance-only strains lacked them) were also more diverged than the housekeeping genes but not to a significant extent. The concatenated housekeeping genes appear slightly less conserved in resistance-only strains because they are more distantly related to one another than the producers. The *SGR421* gene appears more diverged than *vph* as well; however their difference is not significant.

Using SLAC, FEL and REL to detect selection on the codon level the *vph*, *vioB* and *vioD* sequences reveals no codons under positive selection and 28 codons for *vph*, 6 for *vioB* and 5 for vioD under negative selection ($p < 0.05$).

5.3.18 Distribution of other secondary metabolite genes

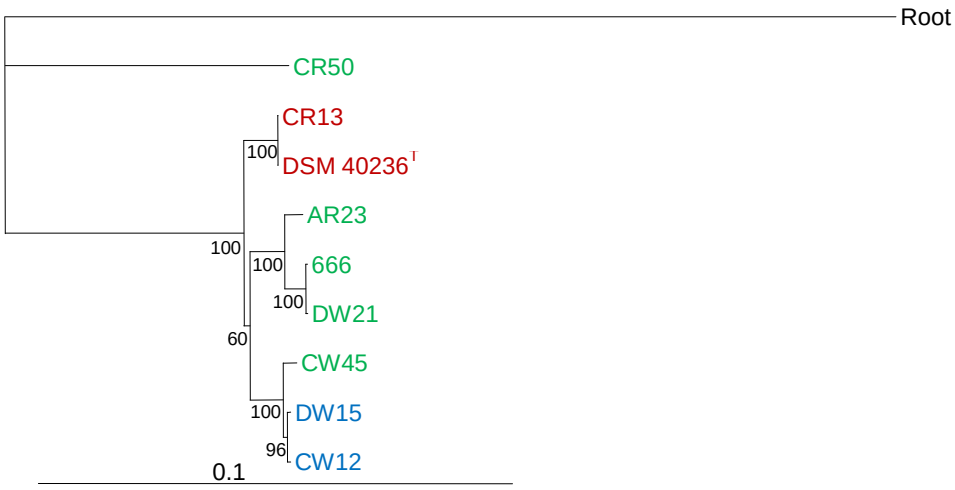


Figure 5.22. Neighbour-joining phylogenetic tree of the strains screened for the conserved secondary metabolite genes (See section 3.3.10.2). Strains possessing all three genes are in green, those lacking *crtE* are in blue and those lacking *geoA* are in red.

The fact that the *crtE* and *geoA* are absent in one subclade each but present in all others indicates that their absence is due to gene loss (Figure 5.22). No recombination events were detected in any of the genes. The correlation between *crtE* and the housekeepers was *r*= 0.801 for nucleotide and *r*= 0.822 for amino acid sequences, that between *geoA* and the housekeeping genes was *r*= 0.990 for nucleotide and *r*= 0.998 for amino acid sequences and that between *hopB* and the housekeepers was *r*= 0.980 for nucleotide and *r*= 0.991 for amino acid sequences. The evolutionary histories of *hopB* and *geoA* closely mirrored those of the housekeeping genes; however that of *crtE* was more diverged which may indicate an HGT event. No codons were detected to be under positive selection in any of the genes; however 59 sites were under negative selection in *crtE*, 71 in *geoA* and 6 in *hopB*.

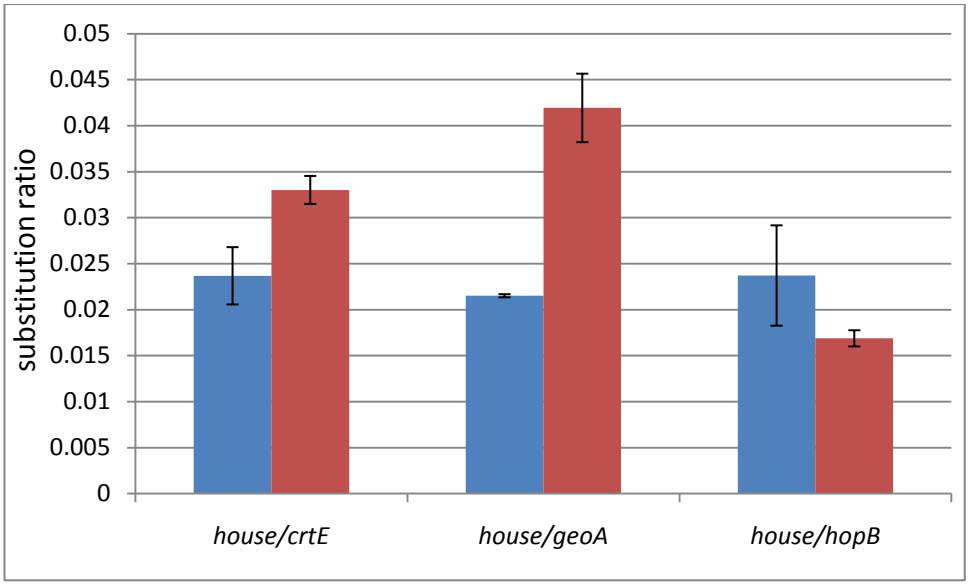


Figure 5.23. Average non-synonymous/synonymous nucleotide substitution ratio for the concatenated housekeeping genes (blue) and the secondary metabolite genes (red).

The *crtE* and *geoA* secondary metabolite genes are under less purifying selection than housekeeping genes (Figure 5.23), however *hopB* is under less than the average for the housekeeping genes used in this study.

5.4 Discussion

Using eight concatenated housekeeping genes produced a more robust phylogenetic tree with higher bootstrap values than the partial 16S tree or the individual protein-coding housekeeping genes. This agrees with previous observations that multiple housekeeping genes are more reliable for determining the phylogeny of a group of organisms than individual genes (Hanage *et al.*, 2005; Guo *et al.*, 2008), as they can represent a much larger area of the chromosome and will only be partially affected by recombination events. Recombination can not only result in incorrect topologies, but also in the overestimation of the number of mutations and apparent substitution rate heterogeneity among sites which will effect attempts to calculate the ω value (Schierup & Hein, 2000). The correlation of the genetic distances between the concatenated housekeeping genes and the partial 16S rRNA sequence was fairly high (r= 0.928) however the difference between the two indicates that the use of multiple housekeeping genes will generate more accurate phylogenetic trees. The reliability of the concatenated housekeeper neighbour-joining tree is also supported

by the agreement between trees derived from nucleotide and amino acid sequences. The topology of the maximum likelihood tree generated from nucleotide sequences matched those of the neighbour-joining trees; however that of amino acid tree was divergent (results not shown). The majority-rule consensus tree generated by ClonalFrame also agreed with that of the neighbour-joining tree in all nodes having a confidence value greater than 0.95. The same structure is also apparent in the split network, indicating that the neighbour-joining algorithm did not produce an incorrect phylogenetic tree due to systemic biases (Huson & Bryant, 2006).

Using spore colour and colony morphology for the phenotypic identification of *S. griseus* proved unreliable. A number of unrelated streptomycetes, belonging to *S. albidoflavus* in particular but to other species as well, were misidentified as *S. griseus*. However the majority of strains belonging to the *S. griseus* clade matched that phenotype, so phenotypic characteristics can be useful for a preliminary identification of strains provided that it is followed up with a reliable taxonomic identification method. Nevertheless, alternative phenotypic markers, such as the utilization of light microscopy to observe the spore chain type (*S. griseus* DSM 40236^T has straight spore chains) could have provided a more reliable phenotypic marker than spore and mycelium colour (Dietz, 1986).

A comparison of the MLSA results with the numerical classification (Figure 1.8, Figure 5.4) (Williams *et al.*, 1983) reveals areas of agreement and of disagreement. The *S. netropsis* (cluster 56), *S. cinnamoneus* (55), *S. mashuensis* (55) clade is represented, as the strains belong to the same or to neighbouring taxonomic clusters. *S. platensis* (29) is next to this clade in the concatenated tree, however cluster 29 is a sister clade to the clusters 1-28 in the numerical taxonomy study instead. *S. coelescens* (21), *S. lividans* (21) and *S. violaceolatus* (21) all were in the same cluster;the *S. coelicolor* (1A) strain used in that study differed from the "*S. coelicolor*" A3(2) which belongs to the *S. violaceoruber* species. The species *S. tuirus* (21), *S. galbus* (24), *S. griseoruber* (21), *S. humidus* (19), *S. scabies* (3) and *S. glaucescens* (28) formed a sister clade to *S. violaceoruber* in the MLSA analysis; however none of them were situated within cluster 1A along with *S. coelicolor* in the numerical taxonomy. *S. ambofaciens* (23) was in the same clade as *S. coelicolor* (1A) in the MLSA instead of with *S. microflavus* (23). *S. griseobrunneus* (1B), *S.*

alboviridis (1B), *S. albovinaceus* (1B), *S. anulatus* (1B) all belonged to the same cluster as *S. griseus* (1B) in the phenotypic classification, however *S. microflavus* (23), *S. californicus* (9), *S. puniceus* (9) and *S. olivovidiris* (3) did not. *S. microflavus* (23) in particular was almost genetically identical to *S. alboviridis* (1B) but they were nevertheless situated in very different clusters. This finding is unsurprising, since classifications based on 16S rRNA and the ribosomal AT-I30 protein have also demonstrated relatively poor correlation with the phenetic data (Ochi, 1995). Phenotypic characters can therefore produce useful taxonomic information but such data will often fail to correspond to the genetic interrelations of the studied microorganisms.

The discovery of strains with identical housekeeping genes that have different resistance profiles to streptomycin or viomycin is not surprising. In certain groups of bacteria, such as *Serratia*, there is correspondence between antibiotic resistance profiles and phylogeny (Berg, 2000); however this does not appear to be the case in *Pseudomonas* (Lottmann & Berg, 2001) or *Streptomyces* (Davelos Baines *et al.*, 2007). Other phenotypes, such as temperature sensitivity, can also vary between actinobacterial strains with identical 16S sequences (Hahn & Pockl, 2005). These differences in correlation between genotype and phenotype may depend on the frequency of HGT and recombination which can result in the transfer of resistance determinants. Recombination occurs with high frequency in both *Pseudomonas* (Lugtenberg *et al.*, 2001) and *Streptomyces* (Hopwood, 2006). The high instability of the *Streptomyces* genome, which can result in the deletion of large sections of the chromosome (Redenbach *et al.*, 1993), can also account for the phenotypic variation in the genus as resistance genes can be easily lost, rendering a member of a resistant group sensitive. The greater heterogeneity in the resistance phenotype of streptomycin compared to viomycin might indicate that streptomycin resistance is more easily acquired or lost.

The fact that more strains were resistant to streptomycin than to viomycin might indicate that, since streptomycin production is one of the most frequently isolated antibiotic production phenotypes (Baltz, 2006), streptomycin resistance determinants are also more widespread. A study of the resistance patterns of streptomycetes from prairie soil suggested that genetically similar isolates from

different environments may have different resistance profiles (Davelos Baines *et al.*, 2007). Some of the strains with identical housekeeping genes but varying resistance levels were isolated from different locations (e.g. CB162 and Z34 for viomycin or RB063 and DSM 40707 for streptomycin), however there were also isolates with identical housekeepers and varying levels of streptomycin resistance that had been isolated from the same site (e.g. CB171 and CB150, CB147 and CB153). This implies that the site is not the sole factor influencing resistance of closely related organisms, though it should be noted that, due to the heterogeneity of soil, organisms isolated from one site may have inhabited distinct microenvironments with different selective pressures (Ettema & Wardle, 2002).

The fact that four out of nine housekeeping genes appeared to have undergone recombination is unsurprising, as the streptomycete genome is highly unstable (Chen *et al.*, 2002). The most recombinogenic sequence was that of the 16S gene which contained two recombination breakpoints, while the protein-coding genes had one at most, though this might to an extent be due to the longer length of the 16S rRNA sequence studied. This finding supports the idea that 16S is highly recombinogenic and can therefore potentially generate misleading phylogenetic trees (Acinas *et al.*, 2004). HGT of housekeeping genes is generally rare (Chen *et al.*, 2002). The one instance of potential housekeeping gene HGT that was observed (*rplC*) indicated that even core genes can undergo recombination with a diverged homologue as noted in other studies (Daubin & Lerat, 2006). Based on the soil isolates used in this study, it can be concluded that strains in the *Streptomyces* genus form clusters with high bootstrap values composed of strains with less than 97% 16S rRNA sequence diversity, that match the definition of a species cluster (Hanage *et al.*, 2005) rather than a continuous spectrum of genotypic variation (Gevers *et al.*, 2005).

The uncertain status of interspecies relationships found in this study was also evident in other MLSA studies of streptomycetes. One contained a six-gene MLSA that included *S. griseus* strains, *S. coelicolor* A3(2) and *S. avermitilis* MA-4680 whose nodes had bootstrap values below 50 (Antony-Babu *et al.*, 2008) while another with a seven-gene MLSA produced the same result for *S. griseus* and *S. flavidofuscus* (Guo *et al.*, 2008). The same pattern can be seen in *Streptomyces*

phylogenies based on the full 16S rRNA sequence (Rintala *et al.*, 2002). The low phylogenetic resolution between strains with a 16S rRNA sequence identity below 97% may be due to HGT or recombination between the ancestors of the species groups that may have resulted in the phylogenetic signal being lost. This runs counter with the findings from the pairwise analysis, which indicates that all the housekeeping genes have very similar phylogenetic histories. Also the interspecies clades were poorly supported in the trees composed of housekeeping genes where no recombination was detected. Another possible explanation is that the ancestors of these species groups diverged in a very brief period and thus a phylogenetic tree cannot resolve the exact pattern of divergence.

All the genes included in this study were overall under purifying selection, as they had an $\omega < 1$. That is to be expected as diversifying selection is seen very rarely, e.g. in the V3 region of the HIV-1 *env* gene (Yang *et al.*, 2000). Non-viral genes will not display such levels of fixation of non-synonymous substitutions. A few genes contained codons that were under positive selection, but these codons were a tiny minority indicating that the genes examined are mainly under negative selection. The *strA* resistance gene overall had significantly higher ω values compared to the housekeeping genes. However, the ω values of the *strA* genes associated with an aminoglycoside cluster were significantly lower than those of resistance-only genes, indicating that the *strA* in antibiotic producers are under more purifying selection. The same pattern can be seen in the *vph* gene which is under significantly less purifying selection than the housekeeping genes and whose resistance-only homologue is under less purifying selection than *vph*, though the difference between the latter is not statistically significant due to the small sample size of viomycin producers. The biosynthetic genes of these two clusters were also under significantly less selective pressure than housekeeping genes. *strW* was more highly conserved than *strA* in non-*S. griseus* strains though the opposite was observed in *S. griseus* streptomycin producers. *vioB* and *vioD* appeared to be more conserved than *vph*, however the differences were not significant. *aphE* was the least conserved gene in this study, having an ω value 19 times larger than the housekeeping genes. The very low purifying selection indicates that *aphE* plays a minor role to the strains possessing it. As with the other secondary metabolite genes discussed so far, *crtE* and *geoA* were less conserved than the housekeeping genes. However *hopB* was

significantly under more purifying selection than housekeeping genes, which indicates that hopanoids play a vital role in the growth of the organism.

A calculation of ω for polyketide synthases from lichenized fungi, which also produce secondary metabolites, found that their values averaging between 0.0051 and 0.0325 (Muggia *et al.*, 2008). These ratios are lower than those of the viomycin biosynthetic genes, but their values are in the same range as those of *strW*, *crtE*, *geoA* and *hopB*. The testing of individual codons of *Streptococcus pneumoniae* penicillin-binding proteins 1a, 2b and 2x identified 46 codons as being under significant positive selection in *pbp2x*, 13 in *pbp2b* and 22 in *pbp2a* (Stanhope *et al.*, 2008). In *strA* there was only a single codon in a subset of *strA* sequences that appeared to be under positive selection and none in *vph*. The *strA* codon 140 does not form part of any functional domain so the advantage that could be conferred by its diversification is unknown. It therefore appears that, while antibiotic resistance genes can be subjected to strong positive selection, this does not seem to apply to *strA* and *vph*. One possible explanation for this discrepancy is that since *Streptococcus pneumonia* is a pathogenic organism that is subjected to very high doses of antibiotic it will be under tremendous selective pressure to acquire resistance which can result in diversifying selection. Soil-inhabiting organisms are subjected to only periodic exposure of considerably lower concentrations of antibiotics and as a result are under less selective pressure to diversify.

The high correlation of evolutionary distances between housekeeping genes, as well as the high level of purifying selection which they are under, provides clear evidence of the conservation and stability of the housekeeping genes found within the core chromosome as defined by Bentley (Bentley *et al.*, 2002). However the fact that *hopB* is more conserved than the majority of housekeeping core genes despite being located in the arm region of the chromosome demonstrates that despite having mostly non-essential genes the arm regions can contain genes that are vital for the cell's survival.

The *strA*, *aphE* and *vph* resistance genes all have a discontinuous distribution amongst the strains examined. From the phylogenetic trees of *strA* and the concatenated housekeeping genes and the very low correlation between *strA* and housekeeping gene distances it can be inferred that there have been at least three

HGT events for *strA* to arrive to the strains possessing it. One for CR50, one for *S. netropsis*, *S. cinnamoneus* and *S. mashuensis*, one for Subclades 5, 6 and 7 of *S. griseus* and one for *S. humidus*, *S. glaucescens* and *S. galbus*. The last two groups may have had multiple HGT events, but the phylogenetic resolution on their branching is too low to be certain. The ancestral streptomycin-like gene cluster must have spread and diversified early on in its evolution, as evidenced by the presence of related clusters in distantly related species such as *S. glaucescens* and *S. netropsis*. There are two possible explanations for the origin of the resistance gene in the non-streptomycin producing *S. griseus* strains. One is that the common ancestor of the producers and resistance-only *S. griseus* strains lacked a biosynthetic cluster and the non-producers acquired their resistance gene from another species with a related cluster via HGT and the common ancestor of the producers similarly acquired the entire streptomycin cluster independently. This transfer could have consisted solely of the resistance gene or of an entire or partial cluster that decayed, leaving only *strA* present. The independent acquisition of *strA* by the resistance-only strains is supported by the higher sequence identity of the resistance-only *strA* to the *aph(6)* of the spectinomycin cluster (79-83%) and the *sph* gene of the hydroxystreptomycin cluster (81-83%) compared to the *strA* of the streptomycin cluster (73-77%). The *strA* phylogenetic tree also places *aph(6)* and *sph* in a single clade alongside the resistance-only *strA* sequences, however their bootstrap values are low, which indicates that this may not be a real clade.

The alternative and possibly more parsimonious hypothesis for the origin of these genes is that the common ancestor of the resistant strains and the streptomycin producers possessed a streptomycin cluster but that the biosynthetic and regulatory genes were lost in the resistant strains and *strA* arrived at a different location in the chromosome via recombination. The advantage of this hypothesis is that it only needs to posit one HGT event, that of the original cluster to the ancestor of all the *S. griseus* strains possessing *strA*, rather than one for the biosynthetic cluster and one or two such events for the resistance gene. Its disadvantage is that *strA* is highly divergent between Group A and Group B strains compared to the housekeeping genes, which makes it very unlikely to have diversified this rapidly, especially considering the fact that its ω value is not very high. The two streptomycin resistance-only clades (Subclades 5 and 6) successively branch out from the clade

containing the producers, however since the relevant bootstrap values are low they may form part of a single clade that underwent a HGT event to acquire *strA* and then rapidly diverged into two subgroups. This latter model is supported by the ClonalFrame phylogenetic tree.

The *aphE* gene requires at least two transfer events to explain its current distribution; one for *S. olivoviridis* and one for Subclades 6 and 7. Its function in *S. olivoviridis* is unknown, as that strain is sensitive to streptomycin. It may therefore be inactive.

The *vph* and its homologues are found in the *S. griseus* Subclades 2, 3 and 7. The high correlation of phylogenetic distances between the concatenated housekeeping and *vph* genes (which is higher than that of most individual housekeeping genes) suggest that there have not been HGT events in its history, however this would require the *vph* gene to have been lost four times. The alternative explanation is that there have been three incidents of HGT for it to arrive in its current distribution. If the last common ancestor of clades 2 to 7 had the gene, then it must also have possessed the remaining viomycin cluster as *vph* is normally contained within it as in the Subclade 2 strains. In that case, the biosynthetic genes were lost in clades 3 to 7 and the *vph* gene itself was lost in DSM 40932 and Subclades 4 to 6. As with *strA*, the disadvantage of this hypothesis is that *vph* is highly divergent between Group A and Group B strains compared to the housekeeping genes, which makes it very unlikely for it to have diversified this rapidly, especially considering the fact that its ω value is not very high. The alternative hypothesis is that there have been three separate HGT events to the Subclades 2, 3 and 7. Such a transfer could have included an entire or partial peptide cluster or *vph* by itself; however the *vph* homologue (*SGR421*) in *S. griseus* NBRC 13350 is not flanked by sequences homologous to the viomycin or other peptide clusters, indicating that if they were transferred alongside the resistance gene these genes have since been lost with the possible exception of *vioR* (Section 3.3.9).

From the biosynthetic secondary metabolite genes, *strW* appears to have undergone fewer HGT events than *strA* due to its higher correlation with housekeeping genes. This presumably is because it has a more limited distribution than *strA*. The phylogenetic tree structure of *crtE* is different to that of the housekeeping genes and

has it has a lower correlation to housekeepers compared to *geoA* and *hopB*. Because of that it appears likely that it has undergone HGT or recombination. The correlation coefficients of *geoA* and *hopB* to the concatenated housekeeping sequence was very high, suggests that have not undergone HGT within the set of strains examined despite being located in the arm regions of the chromosome. The same applies to *chiF*, a gene encoding a chitinase (Ul-Hassan, 2007).

Both streptomycin producers and resistance-only strains are found in similar locations worldwide. All the strains possessing a streptomycin biosynthetic cluster fell under Subclade 7. Of the 17 strains, 16 strains formed part of a clonal group and 15 of those had identical sequence types. Not all of these strains were genetically identical though; DSM 40659 and DSM 40654 had SNPs on their *aphE* gene and *Z34* on its *aphE*, *SGR421* and *strA* genes that were absent from other strains with identical housekeeping genes, indicating that they have undergone some divergence. The fact that all the streptomycin producers are so closely related demonstrates that the streptomycin production phenotype is highly advantageous, as it has enabled the global radiation of this clonal group in a very short period of time, marking them as highly successful soil colonists. Isolates belonging to producer clade (Group A) have been isolated from Europe, Asia, Africa and America. Numerous more isolates with an identical phenotype have also been found in additional screens (Gordon & Horan, 1968). The viomycin producers and their cluster are more diverse and have a more limited distribution, indicating that they have spread more gradually and that viomycin provides a smaller selective advantage than streptomycin. The streptomycin resistance-only group is also widespread; however it is more diverse than producers, being split into two subgroups which in turn are more distantly related to one another. The resistance gene in streptomycin producers is more conserved, as it is necessary for the cell to avoid suicide. In non-producers it is more divergent, as a high level of streptomycin resistance is not essential for these organisms, enabling them to tolerate a decrease in the efficiency of the enzyme as streptomycin levels in soil are probably low. The diversification of the resistance gene may also be driven by the organism's exposure to other related aminoglycoside antibiotics, which have selected for an enzyme that can provide resistance at a lower efficiency but against a wider spectrum of antibiotics. The level of diversity of the *vph* resistance-only gene is much more similar to that of viomycin producers

compared to *strA*. This is reflected in the fact that strains possessing it almost universally have high-level viomycin resistance. Presumably there is a selective advantage for these strains to maintain an enzyme that provides high resistance.

6 Streptomycin resistance gene distribution in soil

6.1 Introduction

Obtaining gene sequences from organisms grown in pure culture is more straightforward than from environmental samples and ensures that all the genes amplified and sequenced from a sample originated from the same organism, enabling the identification of the bacterium possessing a secondary metabolism gene and phylogenetic comparisons between genes. The disadvantage of a culture-based approach is that it requires the isolation of microorganisms from the environment, which will generate a considerable bias. It is roughly estimated that only approximately 1% of soil bacteria can be grown on plates (Davis *et al.*, 2005). Isolation therefore can only be used to detect the presence of genes in a small minority of bacteria present in the environment. Furthermore, in this study only *Streptomyces* isolates were selected for screening, which would prevent the detection of *strA* or *vph* genes present in organisms belonging to this genus. Extracting total community DNA (TCDNA) directly from soil samples has the advantage that it can bypass the bias generated by isolation procedures and allows the detection of at least 200 times higher diversity than that of bacterial isolates from the same soil (Torsvik *et al.*, 1998). Direct DNA extraction from soil however can also generate biases. A protocol that is too gentle may fail to lyse Gram positive cells and thus preferentially extract DNA from Gram negative bacteria, whereas one that is too harsh may shear DNA, thus resulting in a loss of DNA and the formation of chimeric molecules during PCR (von Wintzingerode *et al.*, 1997). The PCR amplification step can also result in bias. Coextracted humic acids can interfere with PCR and there can be differential amplification of target genes due to different affinities of primers to templates, different copy numbers of target genes, hybridisation efficiency and primer specificity as well as preferential amplification for sequences with lower GC content due to their faster denaturation during PCR (Kirk *et al.*, 2004). Utilising PCR and sequencing on DNA from soil isolates and on DNA extracted from soil can help compensate for these biases.

There has been one previous attempt to specifically amplify the *strA* resistance gene from environmental DNA. The *strA* gene, as well as the *strB1* and *stsC* biosynthetic genes, was amplified from the Dossenheim AR and CR soils. The PCR produced

faint DNA bands on a gel electrophoresis for all three genes from both sites, but when they underwent southern hybridisation to the *S. griseus* DSM 40632 versions of the genes they hybridised poorly (Tolba, 2004). This, as well as the fact that these amplicons were never sequenced, leaves some uncertainty about whether these gel bands were genuine copies of streptomycin gene cluster genes or false positives.

6.2 Aims

To determine the frequency and diversity of *strA* in the two Dossenheim sites and in a number of additional soils. Identify the composition of streptomycete populations in the soils examined via the amplification of a housekeeping gene.

This chapter's hypotheses are that (i) both the *strA* gene and the streptomycin producers should be found in a high density across a variety of different soils as the *S. griseus* streptomycin producer is a successful soil colonist and therefore widespread in soil as indicated by its successful isolation from several different continents (Table 5.4) (ii) That the application of streptomycin in soil has an effect on the microbial soil flora..

6.3 Results

6.3.1 Screening of soils for *strA* and streptomycete diversity

Table 6.1. Primers used to screen DNA extracted from soils

Gene	Primer	Sequence	Annealing	Size
strA	strA_F	GCG GCT GCT CGA CCA CGA C	63°C	570
	strA_R	CCG TCC TCG ATG TCC CAC AGG G		
	strA_F2	AGG CCT CCC TCG TGS TGC	60°C	615
	strA_R3	SGT CAG CAG GTC GAA GCG		
atpD	atpD_F	AAG ACC GAG ATG TTC GAG AC	56°C	466
	atpD_R	CCA TCT CGT CGG CCA GGT TC		

Two primer sets were used to detect for *strA* (Table 6.1); the F2/R3 pair (designed by Dr. Leo Calvo-Bado) bound to relatively variable sites and was thus specific to the *strA* gene associated with the biosynthetic clusters of *S. griseus* DSM 40236^{T} and CR50 while F/R pair (Tolba, 2004) bound to conserved sites and was able to amplify from more diverse homologues to streptomycin 6-phosphotransferases genes such as hydroxyurea phosphotransferases.

The blastn algorithm was used to compare the housekeeping gene primers to all available streptomycete sequences in GenBank to confirm that they would bind to all known versions of the genes present in the soil. They were also compared with other actinobacterial sequences in order to ensure that they would be distinct enough not to amplify non-streptomycete strains. Some, such as the *rpoB* primers, were too conserved while others, such as the *sodF* primers, were too specific. The primer set selected was that of *atpD*, as its binding sites were both conserved in streptomyetes and varied in other bacteria. Furthermore, a large number of streptomycete *atpD* sequences were available in GenBank which could be used in the construction of phylogenetic trees.

Table 6.2. Soil sites screened for the *strA* and *atpD* genes (Tolba *et al.*, 2002; Garcia & Hernandez, 2004)

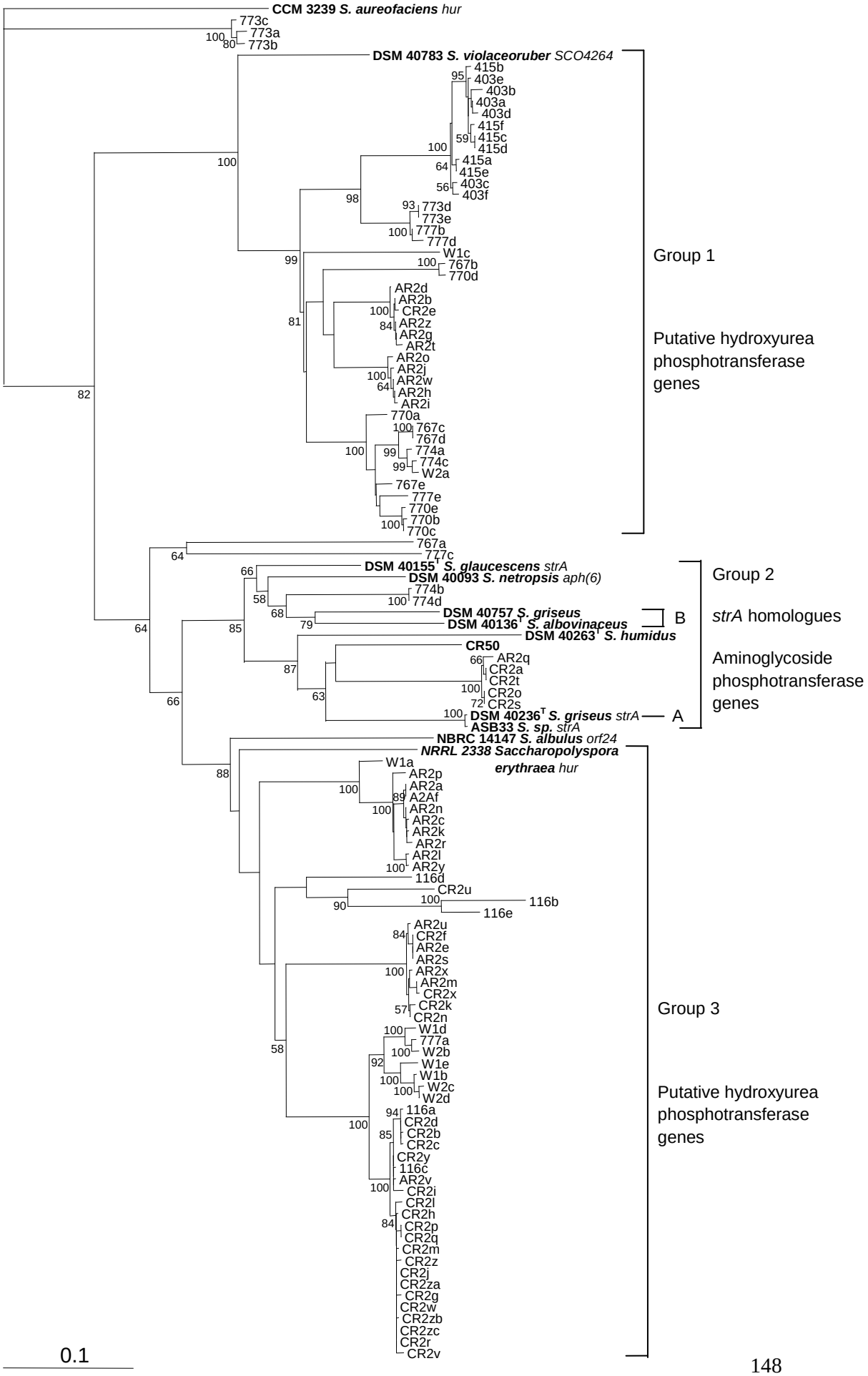
Location	Soil information	Abbreviation
Skopelos, Greece	Agricultural site	116
	Pine forest	Skop1
Cayo Blanco, Cuba	Scrubland	403
	Fir forest	415
Bacilicata, Italy	Conventional agricultural site	602
El Aguilucho, Spain	Terracing plus pines with mycorrhiza	728
Santomeras, Spain	Forest	773
	Bare land	774
	Scrubland	777
Dossenheim, Germany	Apple orchard where plantamycin was regularly applied	AR
	Control site with no plantamycin application	CR
Sourhope, Scotland	Limed soil	5A
	Control site with no liming	5B
Cotswalds, England	Well drained alkaline soil	C
Cryfield, England	Grassland	W

The soils selected for this study (Table 6.2) either contained large amounts of actinobacteria (Skopelos, Cayo Blanco, Bacilicata, Santomeras, Sourhope) or had streptomycetes possessing *strA* previously isolated from them and they were selected to determine how common these genes were in them (Cayo Blanco, Dossenheim, Cryfield).

DNA was extracted (Section 2.6.4) and the *atpD* gene successfully amplified from all the soils (Section 2.6.5), indicating that streptomycetes were present at detectable levels. However both sets of *strA* primers failed to produce any amplicons, demonstrating that *strA* was present below the detection threshold of the PCR used.

Soil microcosms in which streptomycetes were selected for via the addition of chitin and starch were created from soil samples 116, 403, 415, 767, 770, 773, 774, 777, AR, CR and W in order to increase the probability of detecting *strA* (Section 2.6.2). A set of microcosms from all the soils used in this study seeded with *S. griseus* DSM 40236^{T} (Section 2.6.3) were also prepared in order to confirm that the absence of results was not due to inhibition by humic acids or other co-extracted inhibitors. The *strA* gene was successfully amplified from all the positive controls, indicating that the absence of *strA* amplification was not due to inhibition. No amplicons were generated from the microcosm DNA using the F2/F3 primers, which were specific to the *strA* associated with the biosynthetic cluster. The more conserved strA_F/R primers amplified a product of the correct size, as did the atpD_F/R primers. The *strA* and *atpD* PCR products from all the microcosm DNA extracts were cloned and a selection of the clones sequenced. Particular emphasis was placed on the AR soil, as the use of plantamycin at that site may have had selected for resistance and CR because it was the control site for AR and because the strain CR50 was isolated from it. The blastn algorithm was used to identify the closest homologues to all the *strA* and *atpD* sequences and these were included in the phylogenetic trees.

Figure 6.1. A phylogenetic tree of cloned *strA* amplicons from soil. Sequences in bold were from known isolates. The tree was constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% nucleotide dissimilarity.


CCM 3239 S. aureofaciens hur
773c
773a
773b
DSM 40783 S. violaceoruber SCO4264
415b
403e
403b
403a
403d
415f
415c
415d
415a
415e
403c
403f
773d
773e
777b
777d
W1c
767b
770d
AR2d
AR2b
CR2e
AR2z
AR2g
AR2t
AR2o
AR2j
AR2w
AR2h
AR2i
770a
767c
767d
774a
774c
W2a
767e
777e
770e
770b
770c
767a
777c
DSM 40155T S. glaucescens strA
DSM 40093 S. netropsis aph(6)
774b
774d
DSM 40757 S. griseus
DSM 40136T S. albovinaceus
DSM 40263T S. humidus
CR50
AR2q
CR2a
CR2t
CR2o
CR2s
DSM 40236T S. griseus strA
ASB33 S. sp. strA
NBRC 14147 S. albulus orf24
NRRL 2338 Saccharopolyspora erythraea hur
W1a
AR2p
AR2a
A2Af
AR2n
AR2c
AR2k
AR2r
AR2l
AR2y
116d
CR2u
116b
116e
AR2u
CR2f
AR2e
AR2s
AR2x
AR2m
CR2x
CR2k
CR2n
W1d
777a
W2b
W1e
W1b
W2c
W2d
116a
CR2d
CR2b
CR2c
CR2y
116c
AR2v
CR2i
CR2l
CR2h
CR2p
CR2q
CR2m
CR2z
CR2j
CR2za
CR2g
CR2w
CR2zb
CR2zc
CR2r
CR2v
Group 1
Putative hydroxyurea phosphotransferase genes
Group 2
B
strA homologues
A
Aminoglycoside phosphotransferase genes
Group 3
Putative hydroxyurea phosphotransferase genes
0.1

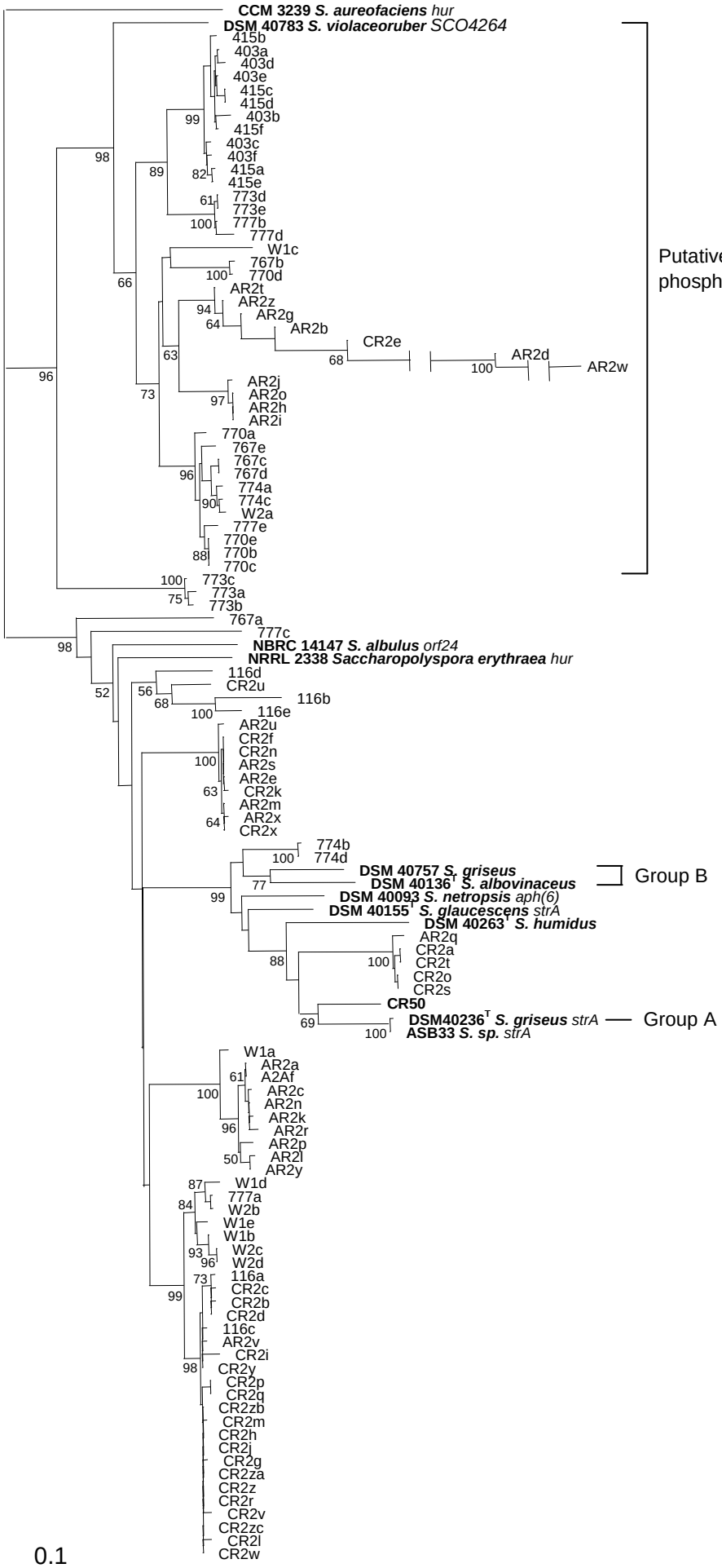

CCM 3239 S. aureofaciens hur
DSM 40783 S. violaceoruber SCO4264
415b
403a
403d
403e
415c
415d
403b
415f
403c
403f
415a
415e
773d
773e
777b
777d
W1c
767b
770d
AR2t
AR2z
AR2g
AR2b
CR2e
AR2d
AR2w
AR2j
AR2o
AR2h
AR2i
770a
767e
767c
767d
774a
774c
W2a
777e
770e
770b
770c
773c
773a
773b
Putative hydroxyurea phosphotransferase genes
767a
777c
NBRC 14147 S. albulus orf24
NRRL 2338 Saccharopolyspora erythraea hur
116d
CR2u
116b
116e
AR2u
CR2f
CR2n
AR2s
AR2e
CR2k
AR2m
AR2x
CR2x
774b
774d
DSM 40757 S. griseus
DSM 40136T S. albovinaceus
Group B
DSM 40093 S. netropsis aph(6)
DSM 40155T S. glaucescens strA
DSM 40263T S. humidus
AR2q
CR2a
CR2t
CR2o
CR2s
CR50
DSM40236T S. griseus strA
ASB33 S. sp. strA
Group A
W1a
AR2a
A2Af
AR2c
AR2n
AR2k
AR2r
AR2p
AR2l
AR2y
W1d
777a
W2b
W1e
W1b
W2c
W2d
116a
CR2c
CR2b
CR2d
116c
AR2v
CR2i
CR2y
CR2p
CR2q
CR2zb
CR2m
CR2h
CR2j
CR2g
CR2za
CR2z
CR2r
CR2v
CR2zc
CR2l
CR2w
0.1

Figure 6.2. A phylogenetic tree of the translated cloned *strA* amplicons from soil. Sequences in bold were from known isolates. The tree was constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% amino acid dissimilarity. AR2d and AR2w had frameshift mutations which resulted in their very high branch lengths. The branches have been partially removed as they were too long to display.

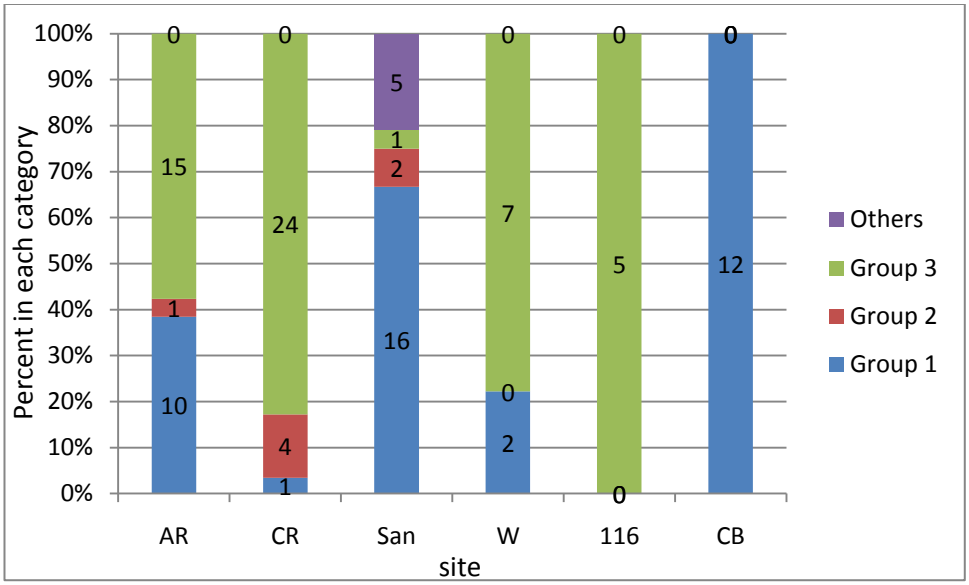


Figure 6.3. Distribution of the *strA* homologues isolated from each soil site on the three main groups identified on the nucleotide phylogenetic tree (Fig 6.1). 'San' indicates all the Santomera strains (767, 770, 773, 774, 777). 'Others' refers to the two minor clades that do not form part of the three main groupings.

Table 6.3. Level of significance, calculated using the χ^2 test, of the differences in number of *strA* sequences belonging to each group between soil sites. Significant values ($p < 0.05$) are highlighted in grey. The upper right corner has the p values for Group 1 and the lower left those for Group 2 from Figure 6.1.

	AR	CR	San	W	116	CB
AR		0.0037	0.0871	0.6332	0.2450	0.0013
CR	0.4172		< 0.0001	0.2639	0.3118	< 0.0001
San	0.9430	0.8501		0.0586	0.0256	0.0654
W	0.5729	0.5781	0.9406		0.7327	0.0011
116	0.3492	0.8945	0.7634	1		0.0004
CB	0.6880	0.4378	0.7970	1	1	

Table 6.4. Level of significance, calculated using the χ^2 test, of the differences in number of *strA* sequences belonging to each group between soil sites. Significant values ($p < 0.05$) are highlighted in grey. The upper right corner has the p values for Group 3 and the lower left those for the Others from Figure 6.1.

	AR	CR	San	W	116	CB
AR		0.0808	0.0002	0.4999	0.1934	0.0025
CR	1		< 0.0001	0.8765	0.7477	< 0.0001
San	0.0475	0.0348		0.0001	< 0.0001	0.7199
W	1	1	0.3465		1	< 0.0001
116	1	1	0.6375	1		0.0303
CB	1	1	0.2330	1	1	

All *strA* homologues that are part of an antibiotic gene cluster belonged to a single clade (Group 2) with a bootstrap value of 85 (Figure 6.1) or 99 (Figure 6.2). Only 7 out of 105 sequenced amplicons (6.7%) belonged to this clade. Two of the Santomeras bare land amplicons formed a sister clade to the Group B *strA* strain and four of the Dossenheim control site clones as well as one clone from the Dossenheim apple orchard formed a sister clade to the Group A *strA* sequence. The position of CR50 in relation to them was phylogenetically uncertain as it grouped both with them (Figure 6.1) and with the Group A sequence (Figure 6.2). On the nucleotide tree, 40 of the sequences (38%) formed a sister clade to *SCO4264* (Group 1) and 51 (49%) formed a sister clade to the hydroxyurea phosphotransferase gene *hur* (Group 3). In addition there were two clades composed of three (2.9%) and two (1.9%) sequences that fell outside those main groupings. On the amino acid tree, the aminoglycoside phosphotransferase clade fell within the sister clade of *hur*.

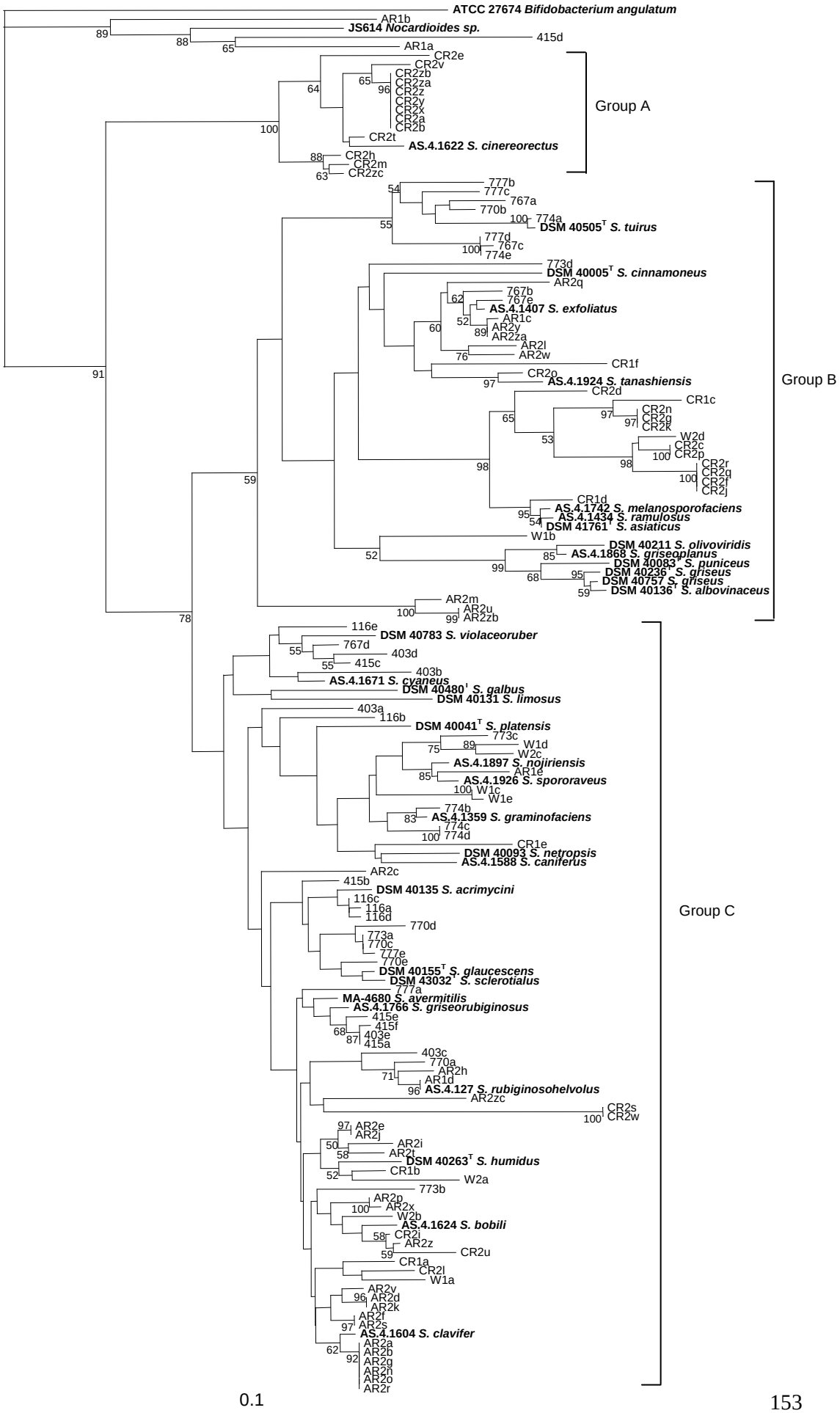
The sequences within the clade containing aminoglycoside producer resistance genes are likely to be phosphorylate aminoglycosides, however the function of the remaining genes is less certain. *SCO4264* is listed as a probable aminoglycoside phosphotransferase on GenBank, however its sequence is most similar to hydroxyurea phosphotransferase genes than to APH genes. The product of *orf24* is listed as being a streptomycin 6-phosphotransferase; however it too demonstrates greater similarity to the *hur* gene than to *strA*. It therefore seems probable that the majority of the isolated sequences do not encode aminoglycoside phosphotrasferases and may encode enzymes serving other functions such as hydroxyurea phosphotransferases.

The sequences from each area examined formed distinct subclades in all three groups with only a few cases of mixed clades (Figure 6.3). In Group 1 a Cryfield sequence formed part of a clade containing Santomeras sequences, while in Group 3 two Skopelos sequences fell into a clade containing Dossenheim *strA* homologues and one Santomeras sequence appeared in a clade containing Cryfield sequences. There appeared to be little difference between different soil sites within an area. Fir forest and scrubland samples from Cayo Blanco were intermingled, as were forest, bare land and scrubland samples from Santomeras. There was mixing between AR and CR sequences from Dossenheim and no significant differences in the number of

sequences from each site belonging to Group 2 or 3; however there was a significant difference between the two for Group 1 (Table 6.3, Table 6.4). This indicates that the application of plantamycin may have caused a significant change in the diversity of the *strA* homologues present in these soils.

In addition there were significant differences in the ratios of *strA* homologue groups between Santomeras and the Dossenheim, Cryfield and Skopelos soil sites, as well as between Cayo Blanco and Dossenheim, Cryfield and Skopelos. This demonstrates that despite the fact that the majority of areas examined contained sequences from all three groups, with the exception of Skopelos and Cayo Blanco where only members of a single group were sequenced, the frequency with which these homologues are found in soil varies.

Figure 6.4. A phylogenetic tree of cloned *atpD* amplicons from soil. Sequences in bold were from known isolates. The tree was constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% nucleotide dissimilarity.


ATCC 27674 Bifidobacterium angulatum
AR1b
JS614 Nocardioides sp.
415d
AR1a
CR2e
CR2v
CR2zb
CR2za
CR2z
CR2y
CR2x
CR2a
CR2b
CR2t
AS.4.1622 S. cinereorectus
CR2h
CR2m
CR2zc
Group A
777b
777c
767a
770b
774a
DSM 40505T S. tuirus
777d
767c
774e
773d
DSM 40005T S. cinnamoneus
AR2q
767b
767e
AS.4.1407 S. exfoliatus
AR1c
AR2y
AR2za
AR2l
AR2w
CR1f
CR2o
AS.4.1924 S. tanashiensis
CR2d
CR1c
CR2n
CR2g
CR2k
W2d
CR2c
CR2p
CR2r
CR2q
CR2f
CR2j
CR1d
AS.4.1742 S. melanosporofaciens
AS.4.1434 S. ramulosus
DSM 41761T S. asiaticus
W1b
DSM 40211 S. olivoviridis
AS.4.1868 S. griseoplanus
DSM 40083T S. puniceus
DSM 40236T S. griseus
DSM 40757 S. griseus
DSM 40136T S. albovinaceus
Group B
AR2m
AR2u
AR2zb
116e
DSM 40783 S. violaceoruber
767d
403d
415c
403b
AS.4.1671 S. cyaneus
DSM 40480T S. galbus
DSM 40131 S. limosus
403a
116b
DSM 40041T S. platensis
773c
W1d
W2c
AS.4.1897 S. nojiriensis
AR1e
AS.4.1926 S. spororaveus
W1c
W1e
774b
AS.4.1359 S. graminofaciens
774c
774d
CR1e
DSM 40093 S. netropsis
AS.4.1588 S. caniferus
AR2c
415b
DSM 40135 S. acrimycini
116c
116a
116d
770d
773a
770c
777e
770e
DSM 40155T S. glaucescens
DSM 43032T S. sclerotialus
777a
MA-4680 S. avermitilis
AS.4.1766 S. griseorubiginosus
415e
415f
403e
415a
403c
770a
AR2h
AR1d
AS.4.127 S. rubiginosohelvolus
AR2zc
CR2s
CR2w
AR2e
AR2j
AR2i
AR2t
DSM 40263T S. humidus
CR1b
W2a
773b
AR2p
AR2x
W2b
AS.4.1624 S. bobili
CR2i
AR2z
CR2u
CR1a
CR2l
W1a
AR2v
AR2d
AR2k
AR2f
AR2s
AS.4.1604 S. clavifer
AR2a
AR2b
AR2g
AR2n
AR2o
AR2r
Group C
0.1

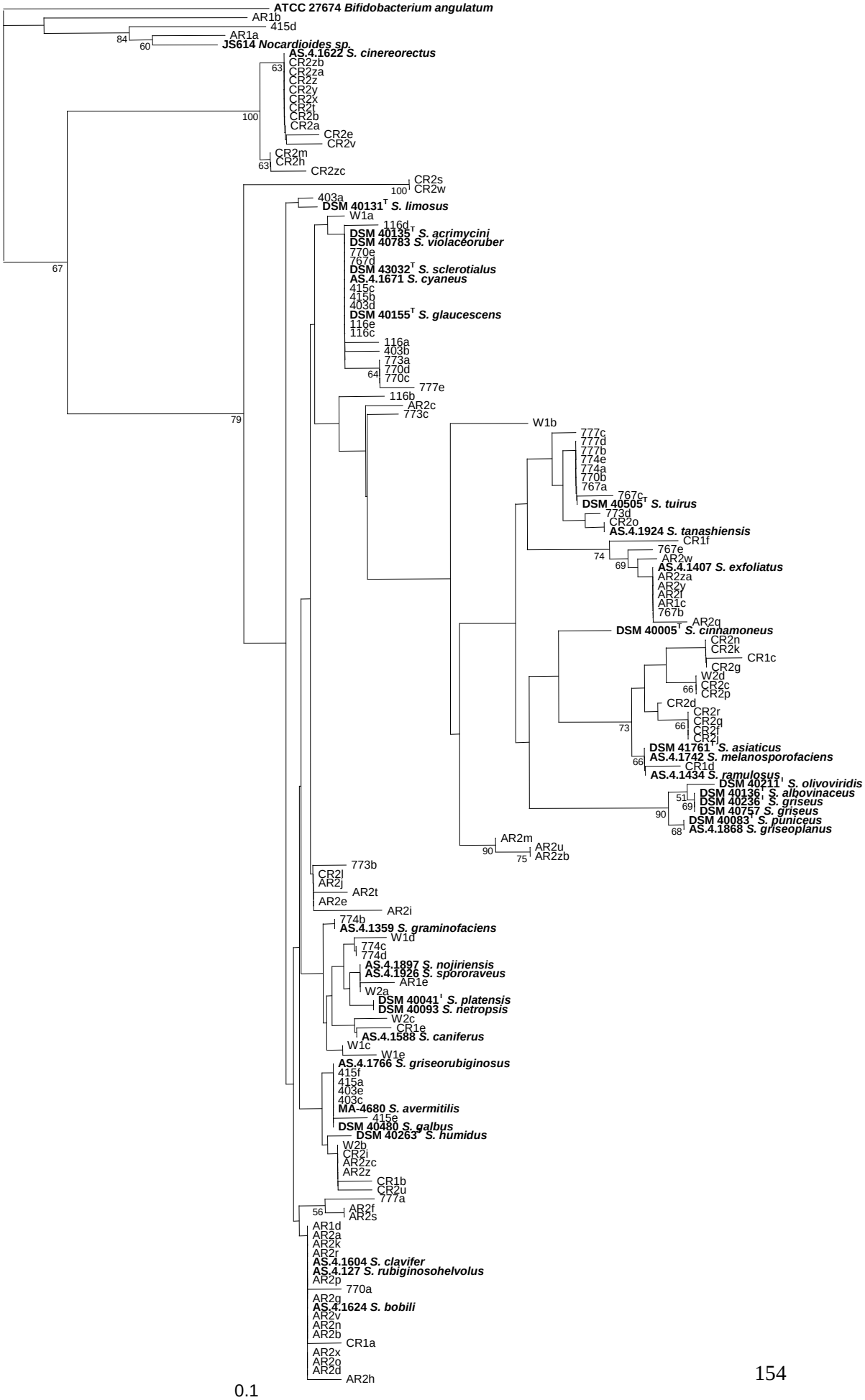

ATCC 27674 Bifidobacterium angulatum
AR1b
415d
AR1a
JS614 Nocardioides sp.
AS.4.1622 S. cinereorectus
CR2zb
CR2za
CR2z
CR2y
CR2x
CR2t
CR2b
CR2a
CR2e
CR2v
CR2m
CR2h
CR2zc
CR2s
CR2w
403a
DSM 40131T S. limosus
W1a
116d
DSM 40135T S. acrimycini
DSM 40783 S. violaceoruber
770e
767d
DSM 43032T S. sclerotialus
AS.4.1671 S. cyaneus
415c
415b
403d
DSM 40155T S. glaucescens
116e
116c
116a
403b
773a
770d
770c
777e
116b
AR2c
773c
W1b
777c
777d
777b
774e
774a
770b
767a
767c
DSM 40505T S. tuirus
773d
CR2o
AS.4.1924 S. tanashiensis
CR1f
767e
AR2w
AS.4.1407 S. exfoliatus
AR2za
AR2y
AR2l
AR1c
767b
AR2q
DSM 40005T S. cinnamoneus
CR2n
CR2k
CR1c
CR2g
W2d
CR2c
CR2p
CR2d
CR2r
CR2q
CR2f
CR2j
DSM 41761T S. asiaticus
AS.4.1742 S. melanosporofaciens
CR1d
AS.4.1434 S. ramulosus
DSM 40211T S. olivoviridis
DSM 40136T S. albovinaceus
DSM 40236T S. griseus
DSM 40757 S. griseus
DSM 40083T S. puniceus
AS.4.1868 S. griseoplanus
AR2m
AR2u
AR2zb
773b
CR2l
AR2j
AR2t
AR2e
AR2i
774b
AS.4.1359 S. graminofaciens
W1d
774c
774d
AS.4.1897 S. nojiriensis
AS.4.1926 S. spororaveus
AR1e
W2a
DSM 40041T S. platensis
DSM 40093 S. netropsis
W2c
CR1e
AS.4.1588 S. caniferus
W1c
W1e
AS.4.1766 S. griseorubiginosus
415f
415a
403e
403c
MA-4680 S. avermitilis
415e
DSM 40480 S. galbus
DSM 40263T S. humidus
W2b
CR2j
AR2zc
AR2z
CR1b
CR2u
777a
AR2f
AR2s
AR1d
AR2a
AR2k
AR2r
AS.4.1604 S. clavifer
AS.4.127 S. rubiginosohelvolus
AR2p
770a
AR2g
AS.4.1624 S. bobili
AR2v
AR2n
AR2b
CR1a
AR2x
AR2o
AR2d
AR2h
0.1

Figure 6.5. A phylogenetic tree of the translated *atpD* amplicons from soil. Sequences in bold were from known isolates. The tree was constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% amino acid dissimilarity.

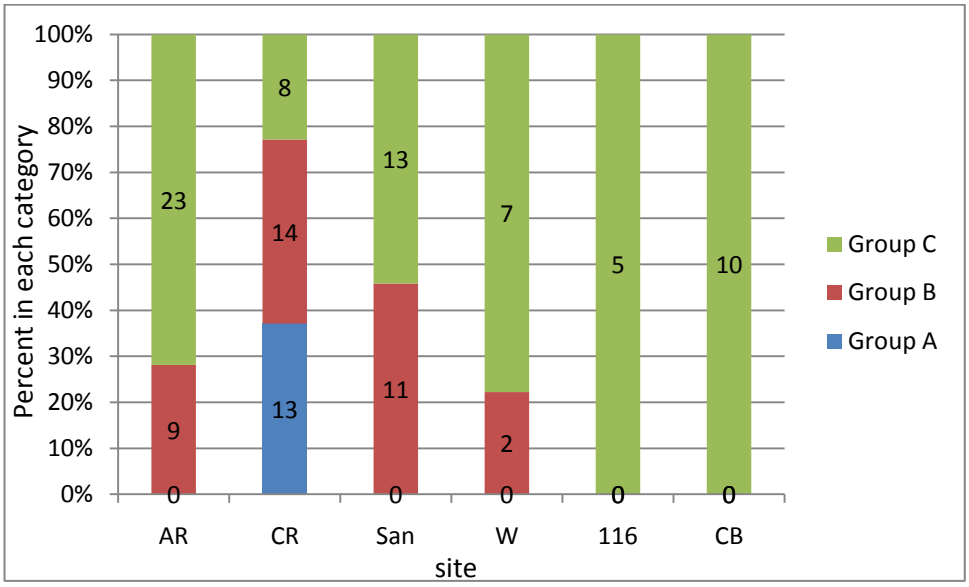


Figure 6.6. Distribution of the *atpD* sequences isolated from each soil site on the three main groups identified on the nucleotide phylogenetic tree (Figure 6.4). 'San' indicates all the Santomera strains (767, 770, 773, 774, 777). The *Nocardioides* sequences are not included.

Table 6.5. Level of significance, calculated using the χ^2 test, of the differences in number of sequences belonging to each group between soil sites. Significant values ($p < 0.05$) are highlighted in grey. The upper right corner has the p values for Group A and the lower left those for Group B of Figure 6.4.

	AR	CR	San	W	116	CB
AR		0.0004	1	1	1	1
CR	0.4443		0.0022	0.0770	0.2508	0.0588
San	0.2771	0.8593		1	1	1
W	0.9420	0.5483	0.4030		1	1
116	0.4221	0.2102	0.1571	0.7327		1
CB	0.1469	0.0431	0.0278	0.4080	1	

Table 6.6. Level of significance, calculated using the χ^2 test, of the differences in number of sequences belonging to each group between soil sites. Significant values ($p < 0.05$) are highlighted in grey. The upper right corner has the p values for Group C of Figure 6.4.

	AR	CR	San	W	116	CB
AR		0.0002	0.2771	0.942	0.4221	0.1469
CR			0.0285	0.0068	0.0033	0.0001
San				0.4030	0.1571	0.0278
W					0.0257	0.0024
116						1
CB						

The *atpD* primers demonstrated high specificity to the *Streptomyces* genus; only three out of 107 (2.8%) sequenced amplicons fell outside it, which belonged to the actinobacterial genus *Nocardioides*. After excluding the *Nocardioides* sequences, the amplicons can be divided into three major groups (Group A, B, C) according to the nucleotide tree (Figure 6.4). In the amino acid tree Group B appears as a sublcade to group C (Figure 6.5). The low bootstrap values at the roots of Group B and C make it difficult to be certain that they are genuine clades, but this division was used in the χ^2 analysis and useful for the determination of population ratios between different areas based on *atpD*.

As with *strA*, *atpD* sequences from the same area tended to cluster together (Figure 6.6). All the Group A sequences were from CR. Group B had clades composed of sequences from a single sample site, such as the clade containing *Streptomyces tuirus*, but also included instances of strains from multiple areas in one clade; the *Streptomyces exfoliatus* clade contained both AR and Santomeras sequences and the *Streptomyces asiaticus* clade had a Cryfield isolate along with eleven CR sequences. There was considerable intermingling in Group C where some subclades, such as the sister clade to *S. platensis* or the clade containing *S. violaceoruber*, contained sequences from up to three areas.

There were significant differences in the frequencies of *atpD* sequences between AR and CR sample sites. The most obvious example were the Group A strains that comprised 37% of the sequenced amplicons but were entirely absent from AR and the other soil sites. The number of Group C sequences was also significantly different between the two though there was no significant difference for Group B. In Group B there were also significant differences between Cayo Blanco and CR as well as Cayo Blanco and Santomera (Table 6.5). In Group C there were also significant differences between CR and Santomera; Cryfield, Skopelos and Cayo Blanco; Santomera and Cayo Blanco; Cryfield and Skopelos as well as Cryfield and Cayo Blanco (Table 6.6). This demonstrates that the population structure varies between different regions though again there were no clear distinctions between different sample sites within a given area with the exception of AR and CR soils.

None of the sequenced amplicons fell within the *S. griseus* or the *S. albidoflavus clade*s. Of the three species groups examined in Chapter 5, only *S. violaceoruber*

was represented by amplicons from TCDNA. The remaining *atpD* sequences either clustered with other species or failed to cluster with known sequences at all.

6.3.2 Screening of Antarctic soil clone library for antibiotic resistance and streptomycete diversity

Antarctica represents a relatively pristine environment, as there has been very little human activity on that continent. It is therefore ideal to investigate the presence of antibiotic resistance genes in a soil site where there has been no anthropogenic input of antibiotics. Soil samples from Mars Oasis, located on the south eastern coast of Alexander Island on the western Antarctic Peninsula (71° 52'42"S, 68° 15'00"W) (Bridge & Newsham, 2009) and TCDNA was extracted from them and cloned into fosmids. A library consisting of 260 samples each containing 96 fosmids with 30-40 Kb inserts, for a total amount of sequence of approximately 874 Mb (constructed by Dr. Leo Calvo-Bado and Dr. Kevin Hughes), were screened for the presence of the *strA* gene utilising the strA_F/R and strA_F2/R3 primer sets, *vph* utilising the vph_F/R primer set and *atpD* utilising the atpD_F/R primer set (Table 6.1, Table 6.7) (Section 2.4.4).

Table 6.7. Additional primers used to screen Antarctic isolates

Gene	Primer	Sequence	Annealing	Size
vph	vph_F	GTT CCA CSA YGT GKT GAT CG	57°C	590
	vph_R	CCA GGT CCT CMG CCG GGT CG		
NRPS	NRPS_ FP1	CGC TGA CCC CCA Acg gna arb tny a	50°C	180
	NRPS_ FP2	CGC GCG CAT GTA CTG Gac ngg nga yyt	50°C	420
	NRPS_ RP1	GGA GTG GCC GCC Car nyb raa raa		

None of these primer sets generated sequences homologous to the target genes, instead either producing non-specific amplicons or no amplification, indicating that streptomycetes are present in very low numbers, if at all, in that Antarctic soil site. The absence of *strA* and *vph* indicates that there are no organisms, streptomycete or not, possessing these genes at detectable levels.

6.3.3 Screening of Antarctic soil clone library for NRPS genes

To confirm that secondary metabolite genes could be detected in this library, it was screened for the presence of nonribosomal peptide-synthetase (NRPS) genes using

the NRPS_FP1/RP1 and NRPS_FP2/RP1 primer pairs (designed by Dr. Martin Krsek). NRPS genes are involved in the biosynthesis of a number of antibiotics, such as viomycin and CDA 46 (calcium-dependent antibiotic) in *S. coelicolor* A3(2) (Donadio *et al.*, 2007), as well as immunosuppressive agents and toxins (Hill, 2006).

The FP1/RP1 primer set generated two positive results from the Antarctic library (Figure 6.7) whereas the FP2/RP1 generated 13 (Figure 6.8). That, in addition to the larger amplicon generated by the latter primer set, indicate that FP2/RP1 is a better primer set for amplifying NRPS gene sequences. However the two FP1/RP1 sequences, along with the NFA50330 and ECA1487, formed a sister clade with a bootstrap value of 57 to all the FP2/RP1 sequences (results not shown) indicating that the two primer sets amplified different groups of NRPS genes.

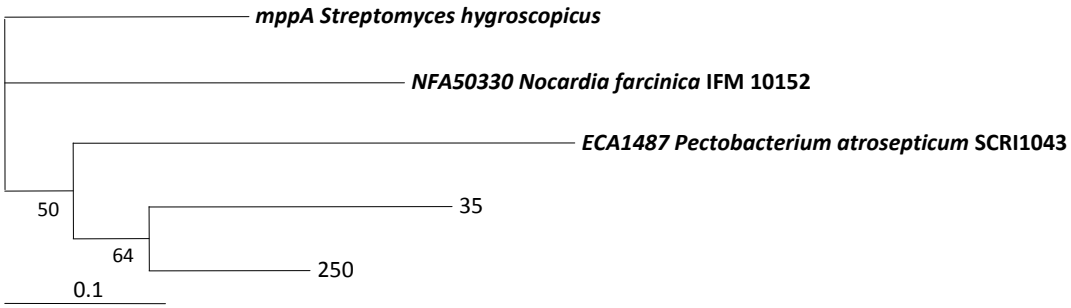


Figure 6.7. A phylogenetic tree of the NRPS sequences amplified using the FP1/ RP1 primers on the Antarctic fosmid library. Sequences in bold were taken from the GenBank database. The tree was constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% nucleotide dissimilarity.

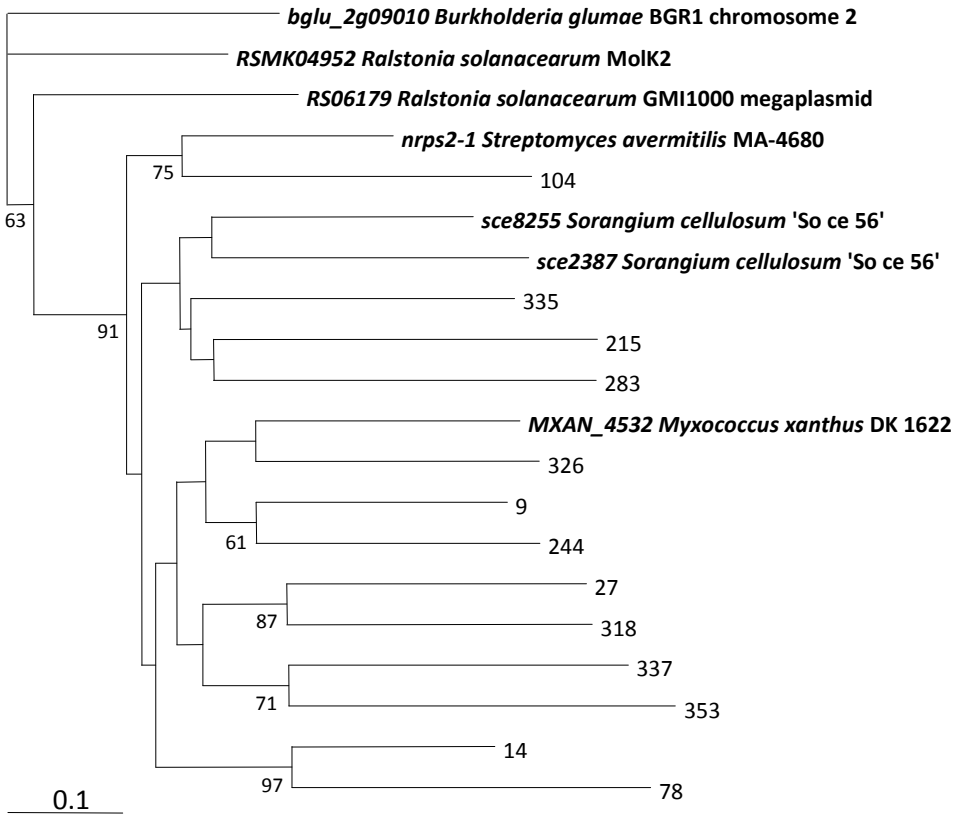


Figure 6.8. A phylogenetic tree of the NRPS sequences amplified using the FP2/ RP1 primers on the Antarctic fosmid library. Sequences in bold were taken from the GenBank database. The tree was constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% nucleotide dissimilarity.

Most of the interrelationships of the isolated sequences were uncertain, as the majority of bootstrap values of the nucleotide tree were below 50 (Figure 6.8). The topology of the NRPS tree is therefore highly unlikely to reflect the true phylogeny of these genes. 60% of the Antarctic NRPS sequences had the *Myxococcus xanthus MXAN_4532* gene as their closest homologue, 20% the *Sorangium cellulosum* sce2387 gene, 13% the *Pectobacterium atrosepticum* ECA1487 gene and 7% the *Streptomyces avermitilis nrps2-1* gene. None of these NRPS genes have an assigned product; the *ECA1487* gene is listed as having a 43.13% sequence identity to the *Pseudomonas syringae* syringomycin synthetase *SyrE*, however the low sequence identity indicates that it is highly unlikely they synthesise the same product. Both *Sorangium* and *Myxococcus* belong to the order Myxococcales, *Pectobacterium* belongs to the Enterobacteriales and *Streptomyces* to the Actinobacteria. This suggests that 80% of the sequenced NRPS genes are found in myxobacteria, though HGT renders the identity of the organisms possessing them uncertain.

6.3.4 Comparison of Antarctic NRPS sequences with UK sequences

A tree was constructed incorporating NRPS sequences amplified by Dr. Martin Krsek from BAC libraries containing DNA extracted from a sand dune in Druridge Bay and from an untreated agricultural hay meadow in Cockle Park Plot 6, both situated in Northumberland, UK.

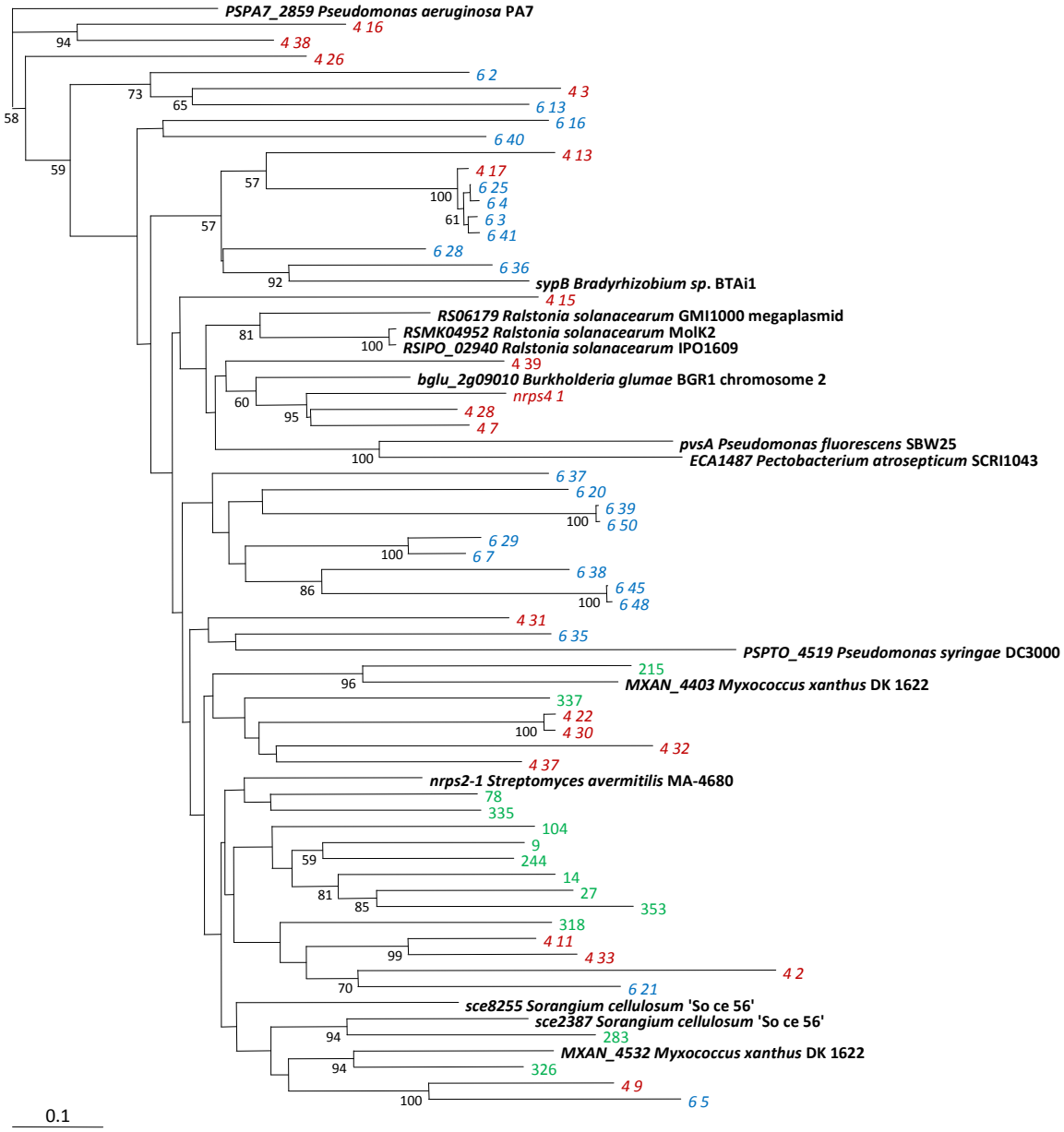


Figure 6.9. A phylogenetic tree of the NRPS sequences amplified using the FP2/ RP1 primers on fosmid libraries from the Antarctic (green), Druridge Bay, UK (red) and Cockle Park, UK (blue). Sequences in bold were taken from the GenBank database. The tree was constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% nucleotide dissimilarity.

The topology of the tree that includes the UK NRPS genes (Figure 6.9) differs from that containing solely the Antarctic sequences (Figure 6.8). The majority of the sequences were found in a clade containing the *Streptomyces avermitilis* NRPS and only two were found in a clade along with *Myxococcus xanthus* and *Sorangium cellulosum* genes. However the fact that another *Myxococcus xanthus* NRPS gene formed a sister clade to the aforementioned sequences suggests that the *Streptomyces avermitilis* NRPS arrived in that organism from a myxobacterium via HGT.

The Druridge Bay and Cockle Park sequences did not form distinct clades, indicating that the bacterial populations in these two sites are not isolated. The Antarctic sequences all fell into a single distinct clade which included some Northumberland amplicons. The latter contained considerably greater diversity as in addition to the sequences that clustered with the Myxobacteria, which belong to the δ proteobacteria, there were also amplicons that fell in clades containing genes from the α (*Bradyrhizobium*), β (*Ralstonia*) and γ (*Pseudomonas*) proteobacteria. This demonstrates that the UK soils have a greater diversity of NRPS genes than the Antarctic soils.

6.4 Discussion

Amongst randomly screened soil actinobacteria 1% can synthesise streptomycin, which is very high compared to other antibiotics (Baltz, 2006). However despite the high frequency with which streptomycin producers are isolated from soil, this study failed to amplify any copies of *strA* that belong to the streptomycin gene cluster using either the *S. griseus* producer-specific primers or the more conserved *strA* primers. Therefore, despite its wide distribution, the cluster is not common in soil streptomycetes. Only a small number of sequences fell within the branch containing phosphotransferase genes found within aminoglycoside biosynthetic clusters and the *S. griseus* resistance-only *strA* genes (Subclade 2). The remaining sequences formed two clades (Subclade 1 and 3) along with more distant *strA* homologues whose function is uncertain. Streptomycin-resistant streptomycetes have been isolated from the Dossenheim soils using selective plates containing streptomycin (Tolba, 2004). Despite the much higher numbers of Subclade 1 and 3 genes in these environments, none of the resistant isolates possessed *strA* homologues belonging to these clades,

indicating that these genes cannot provide streptomycin resistance. They presumably serve another function, such as the phosphorylation of hydroxyurea or another related compound. The detection of a small number of closely related homologues to *strA* indicates that these genes, which may form part of an aminoglycoside biosynthetic cluster or be independent, are more common than *strA* belonging to the streptomycin cluster. Nonetheless, they comprise only a small minority of the genes encoding aminoglycoside/hydroxyurea antibiotic resistance kinases from the APH_6_hur superfamily. This indicates that streptomycin resistance genes are found in low numbers in soil.

There have been a number of studies on the distribution of aminoglycoside or specifically streptomycin biosynthetic genes in the environment which can serve as an indicator of the frequency at which the streptomycin biosynthetic cluster or other aminoglycoside clusters are found in other soils. Two studies screened DNA extracted from Jordanian soil samples for the presence of the streptomycin biosynthetic gene *strB1* via PCR, however these studies only examined the amplicons via gel electrophoresis instead of sequencing (Gharaibeh *et al.*, 2003; Saadoun & Gharaibeh, 2008). Their results suggest that streptomycin producers were detected in 37 different sites in Jordan without any enrichment for streptomycetes, however it is more probable that this result was the product of mispriming or the amplification of a homologous gene with a different function.

Two other studies used PCR and cloning to detect the presence of the 2-deoxy-scyllo-inososе gene in DNA extracted from Japanese soils from the Shinjuku Central Park (from which 45 separate amplicons were sequenced), from Shinjuku Gyoen National Garden (6 clones), from Mount Fuji (3 clones) and from Meiji Jigu Gaiden Park (0 clones). None of these sequences however were closely related to the *stsC* gene from the streptomycin cluster, instead forming clades with genes from other biosynthetic clusters such as kanamycin or gentamicin (Tamegai *et al.*, 2006; Nagaya *et al.*, 2005). DNA samples extracted from marine sediments collected from the Hiroo Valley in the north-western Pacific Ocean were also screened for the 2-deoxy-scyllo-inososе gene. A total of 34 distinct genes were identified, however none were closely related to *stsC* (Aoki *et al.*, 2008). It therefore appears that while streptomycin producers can be readily isolated from a variety of soils from across

the world (Section 3.4), the population size of producers is very low in all soils examined.

The low density of streptomycin producers is also supported by the data generated for *atpD*. The most common producer of streptomycin is *S. griseus* DSM 40236^T and its close relatives, as it is the most frequently isolated strain. However none of the cloned *atpD* genes matched or were related to any strains that fell in the *S. griseus* clade. Other studies also support this finding. An investigation of 16S sequences from soil crusts in the Colorado Plateau failed to uncover any *S. griseus*, though only 13 *Streptomyces* sequences were isolated which is too low to allow any firm inferences to be drawn about the streptomycete population in that site (Gundlapally & Garcia-Pichel, 2006). In the 16S sequences of 5000 actinobacteria isolated from rainforests in Singapore, there were 68 *Streptomyces* isolates of which the isolates closest to *S. griseus* were two that were closely related to *S. purpureus*; a species that has 98% sequence identity to the 16S of *S. griseus* (Wang *et al.*, 1999b). The 16S sequences of 156 *Streptomyces* isolates from prairie soil in the Cedar Creek Natural History Area were classified into species based on them having ≥99% 16S rRNA sequence identity. None of these strains belonged to *S. griseus* (Davelos *et al.*, 2004). A study that used streptomycete-specific 16S primers on DNA extracted from temperate forest soils, one near Siena, Italy one in Gerenzano, Italy, produced a total of 22 sequences. Of these, one from each site had a very great similarity to the *S. griseus* 16S sequence, indicating that *S. griseus* is relatively common in these sites (Monciardini *et al.*, 2002). The findings of this project along with the results available in the literature indicate that despite the frequency with which *S. griseus* is isolated in antibiotic screens, it is a relatively rare organism that is infrequently detected in the environment.

There have been a few attempts to measure the amount of APH genes in the environment which can both provide information on the frequency of resistant strains as well as assist in assessing the frequency of aminoglycoside producers. In an agricultural field site from Costa Rica 13 from a set of 69 actinobacterial isolates (27%) possessed an *strA* homologue while in a coastal salt marsh from the USA the figure was 22 out of 48 isolates (32%). However these included hydroxyurea phosphotransferases and other distantly related homologues; inclusion of only the

genes that fell in the same clade as resistance genes from aminoglycoside clusters reduced the number of homologues to 2 (2.9%) from Costa Rika, that were more closely related to the *S. glaucescens strA*, and 0 from the USA (Anderson *et al.*, 2002). The low percentage of close *strA* homologues is analogous to this study's findings amongst the soils examined. DNA extracted from three Swedish wastewater-associated environments (soil from an overland flow area treating landfill leachates, biofilm from a municipal wastewater treatment plant and sludge from a hospital wastewater pipeline) were screened for the presence of the aminoglycoside resistance genes *aac(6')-Ie* and *aph(2")* using Real-Time PCR. The genes were detected at all the sites, though the soil had very low levels of the genes compared to the other samples that were close to the detection limit (Borjesson *et al.*, 2009). A BAC library containing 5.6 Gb of DNA extracted from soil contained nine isolates of which six were AAC(6') and none were APH genes (Riesenfeld *et al.*, 2004). Assuming that an average bacterial chromosome is 4.5 Mb (Casjens, 1998), it can be estimated that only 0.08% of soil bacteria have resistance to aminoglycosides These findings agree with the results of this study and indicate that aminoglycoside resistance genes are generally rare in soil.

The average genome size of free-living bacteria is 3-6 Mb (Casjens, 1998) and it is estimated there are approximately 10^4 to 10^6 bacterial species in a 10 g soil sample (Janssen, 2006). Assuming that due to the inhospitable environment in Antarctic soil the species diversity will be at the lower end of that range and that the library contained DNA from approximately 100 to 200 distinct genomes, the sequence coverage of the Antarctic site from the fosmid library can be estimated to be a maximum of 1-2% of the total genetic diversity present in that soil. The hypothesis that Antarctic soils are less diverse than other soils is supported by the calculation of soil diversity indices (Smith *et al.*, 2006) as well as by the comparison of the NRPS sequences amplified from the Antarctic library when compared to the two sites in Northumberland, UK. The fact that none of the NRPS genes identified in this study were sequenced twice indicates that they all formed a small fraction of the total DNA present, also supporting the hypothesis that this library has a limited sequence coverage.

It is estimated that there is a total of 10^8 to 10^9 bacteria g^{-1} of soil of which 10^7 to 10^8 are actinobacteria (Sylvia *et al.*, 2005). Siberian permafrost is estimated to contain 10^7 - 10^9 cells g^{-1}; however that number is likely to be lower in Antarctic permafrost due to the lower temperatures and considerably less organic carbon present (Steven *et al.*, 2006). A screening of a diverse number of Antarctic soil sites revealed that an average of 11% of the 16S sequences present belonged to Actinobacteria, which matches the commonly accepted actinobacterial population density (Yergeau *et al.*, 2007). A similar percentage was found in a screen of 126 16S sequences amplified from DNA of agricultural soil from Menangle in south western Sydney, Australia where 9.5% belonged to the order Actinobacteria (Ge *et al.*, 2009), in 324 clones from subtropical Australian exotic pine plantations, where the value was 14.7% (Stahley & Strobel, 2006) and in 300 soil DNA clones from soil from Vadodara in India where actinobacteria comprised of 9.5% of 16S sequences isolated (Desai *et al.*, 2009). This differed in other sites, such the tundra soils from Samoylov Island in Siberia where 16.8% of 430 16S sequences belonged to the Actinobacteria (Liebner *et al.*, 2008), a recently cleared forest area near Skulow Lake in Canada where they comprised 30% of 2063 sequences (Hartmann *et al.*, 2009) and in the highly saline and alkaline soils from the former lake Texcoco where that value was 21 (4.2%) out of a total of 500 clones (Valenzuela-Encinas *et al.*, 2009), indicating that the percentage of actinobacteria can vary considerably. A comparison of five 16S libraries derived from agricultural, undisturbed tall grass prairie and forest soils where Actinobacteria ranged from being the 4^{th} to the 10^{th} most diverse bacterial order also supports this hypothesis (Youssef & Elshahed, 2009).

Actinobacteria tend to be predominant in extremely dry, low nutrient soils such as those found in deserts or volcanic debris (Costello *et al.*, 2009). The rarity of streptomycetes in Antarctic soils is supported by the fact that only one out of 28 (3.6%) amplified 16S sequences from DNA extracted from soil from the McMurdo Dry Valleys in Eastern Antarctica belonged to the genus *Streptomyces* (Babalola *et al.*, 2009) and in another Dry Valley soil study none of the 14 actinobacterial sequences isolated belonged to *Streptomyces* (Smith *et al.*, 2006). This paucity of streptomycete sequences amongst 16S actinobacterial sequences obtained via culture-independent methods extends beyond the Antarctic; in the former lake Texcoco, only two out of 21 (9.5%) of the actinobacterial 16S sequences belonged

to the *Streptomyces* genus (Valenzuela-Encinas *et al.*, 2009). In a 16S clone library from soil of the Sturt National Park in New South Wales, Australia, nine clones out of 38 (26%) belonged to the order Actinobacteria but none of them fell in the genus *Streptomyces* (Holmes *et al.*, 2000). Out of 55 actinobacterial 16S clones obtained from soil of Alston and the University of Kent campus in the UK as well as marine sediments from a Norwegian fjord, no isolates belonged to the genus *Streptomyces*, though one campus sequence formed a sister clade to the genus (Stach *et al.*, 2003). From 30 actinobacterial 16S sequences amplified from the desert of Tataouine in south Tunisia only one was a streptomycete (Chanal *et al.*, 2006).

Despite this apparent rarity, members of the *Streptomyces* genus are the Actinobacteria which are most commonly isolated from soil. When an attempt was made to culture actinobacteria from the McMurdo Dry Valleys soil, more than 80% of the isolates were streptomycetes; a marked difference compared to the 3.6% value derived from the direct amplification of 16S from environmental DNA (Babalola *et al.*, 2009). This is similar to the results from diverse environments such as the Yunnan province in China, where soil samples taken from tropical, subtropical plateau, cool temperate mountain, and snowy mountain sites, which had an average of 90% of the actinobacterial diversity belonging to the *Streptomyces* genus (Xu *et al.*, 1996) and from the hyper-arid soils from the Atacama desert in Chile, where 75% of the actinobacterial isolates were streptomycetes (Okoro *et al.*, 2009). This indicates that the apparent high density of streptomycetes compared to other Actinobacteria may be the result of a bias generated from cultivation-based approaches in determining diversity. The *Streptomyces* species isolates may exist in very low population numbers or only as spores in soil but appear as the dominant genus in isolation attempts due to the ease with which they are cultured compared to other Actinobacteria. Another possibility is that there is a bias against streptomycetes during extraction and amplification of environmental DNA. This seems improbable though; while some Actinobacteria have thinner cell walls which might make it easier for DNA to be extracted, this alone cannot account for the great change in frequency. Furthermore, other Actinobacteria still have a high GC content which will affect PCR amplification (Kirk *et al.*, 2004). *Streptomyces* species have a GC content of around 70 to 73% which is somewhat higher than that the majority of other actinobacterial genera with the exception of *Frankia* (Ventura *et al.*, 2007)

which could result in a slight bias. However as 99% of environmental bacteria cannot be grown on plates (Davis *et al.*, 2005), it appears more probable that biases in isolation protocols are principally responsible for the different results obtained by culture-dependent and culture-independent procedures. Therefore, the direct amplification of 16S from environmental DNA is more likely to present a more accurate picture of streptomycete population sizes compared to isolation attempts.

It can be concluded that this chapter's first hypothesis is incorrect and that despite the global distribution of *S. griseus* and the streptomycin cluster (Section 3.4), *S. griseus* is not the dominant streptomycete group in any of the soils examined in this study. In addition, the Antarctic metagenomic library, as well other actinobacterial libraries, suggest that streptomycetes are not the dominant genus of Actinobacteria in a number of soil sites despite being the dominant group amongst actinobacterial isolates. Furthermore, the *strA* gene was below the PCR detection limit in all the soils examined when *Streptomyces* species were not selected for via the addition of starch and chitin, indicating that both streptomycin producing and *strA*-containing strains are uncommon in soil and are likely to be present in only a very small minority of soil bacteria. There was support for the second hypothesis however, as the TCDNA PCR results indicated there were significant differences in the streptomycete populations of the AR and CR soils despite the fact that all other soil sites from the same region had streptomycete populations that could not be differentiated from one another.

7 Regulation of the streptomycin and viomycin resistance genes

7.1 Introduction

Streptomycin production in *S. griseus* NBRC 13350 is controlled by the A-factor regulatory cascade (Horinouchi, 2007). AfsA synthesizes A-factor (2-isocapryloyl-3R-hydroxymethyl-γ-butyrolactone) in a growth-dependent manner (Horinouchi, 2002). When the concentration of A-factor passes a threshold, at or near the middle of exponential growth, it binds to ArpA (A-factor receptor protein) and causes it to dissociate from the promoter of *adpA* (Ohnishi *et al.*, 1999). The removal of the repressor allows the transcription and translation of AdpA which in turn activates a large number of genes involved in secondary metabolism and differentiation (Horinouchi, 2007). One of the sites it binds to is on the streptomycin gene cluster, where it induces the expression of the streptomycin regulatory gene *strR* (Vujaklija *et al.*, 1993) as well as of the resistance gene *strA* due to read-through (Tomono *et al.*, 2005). StrR in turn further induces its own expression in addition to the expression of the biosynthetic and resistance genes in the streptomycin cluster (Retzlaff & Distler, 1995). In addition to the A-factor cascade, the transcriptional activator AtrA-g, which binds upstream of the *strR* gene, appears to act as a tuner for the expression of the streptomycin gene cluster. Δ*artA-g S. griseus* mutants produce slightly smaller amounts of streptomycin than the wild-type under certain growth conditions (Hirano *et al.*, 2008).

The *strA* gene in strain AR23 is located in a different chromosomal position compared to that of *S. griseus* NBRC 13350 (Section 4.3.1), which is likely to apply to other Group B (resistance-only) strains as well. Furthermore, the absence of the *strR* regulatory gene in these strains indicates that *strA* is regulated in a manner that differs from that of the producers. It is not known whether the streptomycin cluster in CR50 is expressed, as *S. platensis* is not closely related to *S. griseus*. The genetic background of CR50 is therefore likely to differ significantly from that of *S. griseus* NBRC 13350 and CR50 may lack the proteins that regulate the streptomycin cluster in *S. griseus* NBRC 13350.

A previous study examined the regulation of AR23 *strA* in comparison to *S. griseus* DSM 40236^{T} *strA* using reverse transcription and conventional PCR followed by gel

electrophoresis. Quantification was performed by measuring band intensity in comparison to standards and hybridisation with an *strA* probe from *S. griseus* DSM 40236^{T} was used to confirm the identity of the PCR product (Tolba, 2004). It found that *strA* was constitutively expressed in AR23, however the quantification method was inexact, did not use sample replicates and did not examine other genes of interest.

The regulatory mechanism of the viomycin resistance gene (*SGR421*) is also likely to differ from that of viomycin producers, as *SGR421* in *S. griseus* NBRC 13350 is not flanked by viomycin cluster genes.

7.2 Aims

In this part of the project the mechanism of expression for *strA* in non-producing isolates AR23 and E1013 was investigated and a possible role for the flanking regions defined in Chapter 5 established. The first objective was to determine whether the A-factor cascade is present in CR50 or strains possessing solely the resistance gene and then to examine whether the *strA* and *aphE* resistance genes are expressed in representatives of Group B. A further objective was to discover whether the *strA, strR* and *strW* genes are expressed in CR50 and compare the expression pattern of those genes with that of a *S. griseus* streptomycin producer to determine whether the reduced resistance level in the resistance-only strains is due to alterations in the StrA protein sequence or due to lower expression levels. The final objective was to examine the expression pattern of the putative viomycin resistance gene *SGR421* and a putative regulator *SGR878* in an *S. griseus* streptomycin producer.

The hypotheses of this chapter were that (i) *strA* is constitutively expressed at a lower level in Group B strains compared to streptomycin producers, (ii) the expression pattern in the CR50 streptomycin cluster differs from that of *S. griseus* strains and (iii) the viomycin resistance gene *SGR421* is constitutively expressed at a high level in *S. griseus* streptomycin producers.

7.3 Results

7.3.1 Test for antibiotic production

The supernatant from TSB cultures growing *S. griseus* DSM 40236^T, AR23, CR50, *S. coelicolor* M145 taken at certain time points was tested for the production of zones of inhibition on plates seeded with *Bacillus subtilis* DSM 347 (Section 2.5). This test cannot determine whether the compound causing the zones of inhibition is streptomycin or another antibiotic, however the absence of zones of inhibition indicates either an absence of streptomycin production or production at subinhibitory concentrations.

Table 7.1. Size (in mm) of zones of clearing generated by paper discs containing culture supernatant from selected strains or certain amounts of streptomycin

	24 h	48 h	72 h
40236^T	0	0.5	0
AR23	0	0	0
CR50	0	0	0
M145	0	0	0
0 µg	0	0	0
2.5 µg	2.3	2.0	1.8
50 µg	6.7	7.7	7.2

The only strain that exhibited antibiotic production was *S. griseus* DSM 40236^T and it only demonstrated that phenotype on the 48 h time point (Table 7.1). It therefore appears that streptomycin is not produced in large amounts by that strain in liquid culture. None of the other organisms appeared to produce any antibiotics at concentrations high enough to inhibit growth of *B. subtilis*. An attempt was made to detect streptomycin production with the isogenic pair of *E. coli* ATCC 29842 (streptomycin-resistant strain) and ATCC 29839 (sensitive strain), however the isogenic sensitive *E. coli* strain proved to be less sensitive to streptomycin than *B. subtilis* and no zones of clearing were detected.

7.3.2 Screen for A-factor cascade genes

Homologues of *afsA* and *adpA* from streptomycetes other than *S. griseus* NBRC 13350 were obtained from GenBank and used to design consensus primers. There were no homologous sequences of *arpA* and primers had to be designed for that gene using only the *S. griseus* sequence (Table 7.2).

Table 7.2. Primers used to detect A-factor cascade genes.

Gene	Primer	Sequence	Annealing
adpA	adpA_F	TCG CGG TGC TGC TGT TC	52
	adpA_R	GAC CTC GTC CAC CGA GTA G	
afsA	afsa_F	CAC GAT CCG CTG CTG	50
	afsa_R	ATC CAG CAC GGG CTG	
arpA	arpA_F	TTC GCG GAG GCT TTC GAC	53
	arpA_R	CTG GTC CTT CTC GCT GCC	

Table 7.3. Percentage identity of A-factor cascade genes homologues in selected strains to the *S. griseus* type strain.

Gene	CR50	AR23	A3(2)	DSM 40236^T
adpA	???	97%	86%	100%
afsA			74%	100%
arpA				100%

The presence of none of the A-factor cascade genes could be confirmed in CR50. A faint band of the correct size was visible after gel electrophoresis for *adpA*, however attempts to sequence it directly and after cloning it both failed to yield a sequence homologous to *adpA*. AR23 had a gene that was almost identical, 97% nucleotide and 100% amino acid sequence homology, to *adpA*, however none of the other A-factor cascade could be detected. *S. coelicolor* A3(2) lacks an A-factor cascade but nonetheless possessed an *adpA* gene as well as an *afsA* homologue called *scbA* (Table 7.3).

It therefore appeared that neither CR50 nor AR23 possessed a functional A-factor cascade.

7.3.3 Screen for the *atrA-g* transcriptional modulator

The AtrA protein, a TetR-family transcriptional regulator, binds upstream of *strR* in than *S. griseus* NBRC 13350 and modulates expression of *strR* and by extension *strA* (Hirano *et al.*, 2008). The *atrA* gene could therefore play a role in the regulation of the streptomycin cluster or *strA* gene even in the absence of the A-factor cascade. The real time *atrA* primer set (Table 7.5) was used to screen a set of strains.

Table 7.4. Percentage identity of *atrA* genes homologues in selected strains to the *S. griseus* type strain.

Gene	CR50	AR23	E1013	DSM 40236^T
atrA	89%	93%	97%	100%

All the strains examined possessed a homologue to *atrA* (Table 7.4). The AR23 and E1013 genes demonstrated a highest sequence identity to the *S. griseus* NBRC 13350 gene, respectively having an 85% and 86% sequence identity to the *S. coelicolor* A3(2) *SCO4118* gene and an 82% and 84% sequence identity to the *S. avermitilis* MA-4680 *SAV4110* gene. The CR50 gene however had an 89% sequence identity to *SCO4118* and 91% to *SAV4110*, indicating that it was more closely related to the *S. avermitilis* rather than to the *S. griseus* gene. The presence of homologues in all three streptomycete genomes stored on GenBank suggests that these genes play roles other than regulation of the streptomycin cluster and therefore their presence in CR50, AR23 and E1013 does not mean that they have the same regulatory functions as in *S. griseus* NBRC 13350.

7.3.4 Viomycin resistance gene expression

Using the blastn algorithm the *S. griseus* NBRC 13350 genome was screened for the presence of homologues to *vioR* and *vioT*, two putative transcriptional regulators from the viomycin gene cluster. For *vioT*, the closest homologue was a 70% match to a 149 bp region from the gene *SGR4105* encoding a putative two-component system response regulator. The fact that such a small section of the genes matched indicates that the region is a conserved domain belonging to a protein that serves a different function. A *vioR* homologue, *SGR878*, displaying 90% sequence identity was however present in the genome. The gene was listed as being a putative transcriptional regulator and was located 574 kb on the 3' direction from *SGR421*. The fact that the *SGR*878 is no longer close to *SGR421* dies not eliminate the possibility that it regulates the resistance gene. It was therefore decided to examine the expression levels of the *SGR421* putative resistance gene and the *SGR878* putative transcriptional regulator in *S. griseus* DSM 40236^{T}.

7.3.5 Real time PCR

The strains selected for study were *S. griseus* DSM 40236^{T}, due to it being the type strain and its regulation having been well-studied, CR50, as it possessed a variant streptomycin cluster whose level of expression was unknown, E1013 and AR23, as representatives of Group B1 and B2 resistance-only *strA* strains respectively. The

real-time PCR was carried out using SYBR Green (Section 2.7). The primers used (Table 7.5) were designed as discussed in section 2.7.5.

Table 7.5. Primers used to measure gene expression via RT real time PCR. All primers were used with an annealing temperature of 60° C.

Gene	Primer name	Sequence	Amplicon Size
hrdB	hrdB_F	CGC GAG CTG GAG ATC ATC	156
	hrdB_R	GCG GAT CAG ACC GAG GTT	
Group A *strA*	strA_A_F	GGC AGA CCC ATC CCA TGT	96
	strA_A_R	ACC GCT TCT GCT GTA CGA	
Group B1 *strA*	strA_B1_F	CCG CAC ACT GGC TTC GA	139
	strA_B1_R	GAC CTG CGC CAG CAT CT	
Group B2 *strA*	strA_B2_F	GTG GCG ATG GGC ATC CT	110
	strA_B2_R	GGG ACC TGC TCC AGC AT	
CR50 *strA*	strA_50_F	GGT CGT ACA CGA GGA ATG G	123
	strA_50_R	CCT CAC CGG AGC TGA GA	
strR	*strR*_F	CGT CAT CGA CGG CAT GCA	118
	*strR*_R	GTT GGA YTT CAC CGC GAA GA	
strW	strW_F	TGC GCA GCT CCT GGT A	161
	strW_R	GTC GTC CAG CGG GAA GA	
adpA	adpA_F	ACC CGC TGG GTG ATC AG	191
	adpA_R	AGG TCT TTA CCG GAG GAG ATC	
SGR172	172_F	CGA CAC BCT GGA CCT GA	147
	172_R	ACC CAG GAG TCG GAS AC	
atrA	atrA_F	GTC TCC TCC TCG GCT ATC C	129
	atrA_R	GTT CGG GGA GCT GGG TTA	
Group A *aphE*	aphE_A_F	GGT GCA TCC GGA GTT CCT	143
	aphE_A_R	TGG GCA GGC AGA GAT CAC	
Group B1 *aphE*	aphE_B1_F	CGG GTC CTC GAC TGG TAT G	106
	aphE_B1_R	AGG AAG CGC GAA GAT CCT	
SGR878	878_F	CCA GCC GGA GAT CTC GAA	118
	878_R	CGA GTG GGG CTT CGA ACT	
SGR421	421_F	GAA CCG GCC TAC CTG GTA	112
	421_R	AGA GGA GTC CGG CGT ACT	

Table 7.6. Genes examined in each strain and the R2 values of their standard curves.

	AR23	CR50	E1013	DSM 40236^T
hrdB	0.999255	0.993854	0.999257	0.995659
strA	0.998374	0.99757	0.998094	0.994964
strR		0.996448		0.996516
strW		0.997657		0.995333
adpA	0.994708	0.965059	0.996441	0.994765
SGR172	0.997474			0.994358
atrA	0.996029	0.970846	0.996687	0.990033
aphE			0.997344	0.971992
SGR878				0.994926
SGR421				0.992101

The correlation coefficient (R2) for the DNA standard curve was greater than 0.99 for all samples examined with the exception of the CR50 *adpA* and *atrA*, which were above 0.96 (Table 7.6). This indicates that the standards used produced reliable quantification results for all samples.

7.3.6 Strain growth curves

Table 7.7. Optical density readings at 600 nm of 1 in 10 dilutions of liquid cultures. The mean and standard deviations are derived from the three biological replicates of each strain. The vertical column represents the hour after seeding that samples were taken.

	AR23		CR50		E1013		DSM 40236^T	
h	mean	std dev	mean	std dev	mean	std dev	mean	std dev
10	0.027333	0.000577	0.023333	0.000577	0.021667	0.000577	0.025000	0.001000
12	0.048000	0.002000	0.045000	0.002646	0.048667	0.001528	0.069333	0.002309
14	0.114333	0.007371	0.089667	0.008083	0.135333	0.003055	0.198333	0.003512
16	0.266333	0.008963	0.203333	0.002082	0.318333	0.029263	0.385000	0.029052
18	0.526000	0.018520	0.413000	0.028827	0.602667	0.024132	0.688333	0.059969
20	0.809333	0.100977	0.736333	0.025697	0.778333	0.032130	0.780333	0.075798
22	0.816667	0.044004	0.947333	0.018148	0.812667	0.027592	0.846667	0.07654
24	0.970333	0.104026	0.841000	0.121766	0.951000	0.010536	0.938333	0.118035
36	1.221333	0.079103	0.909667	0.008145	1.088000	0.00755	0.911000	0.008888

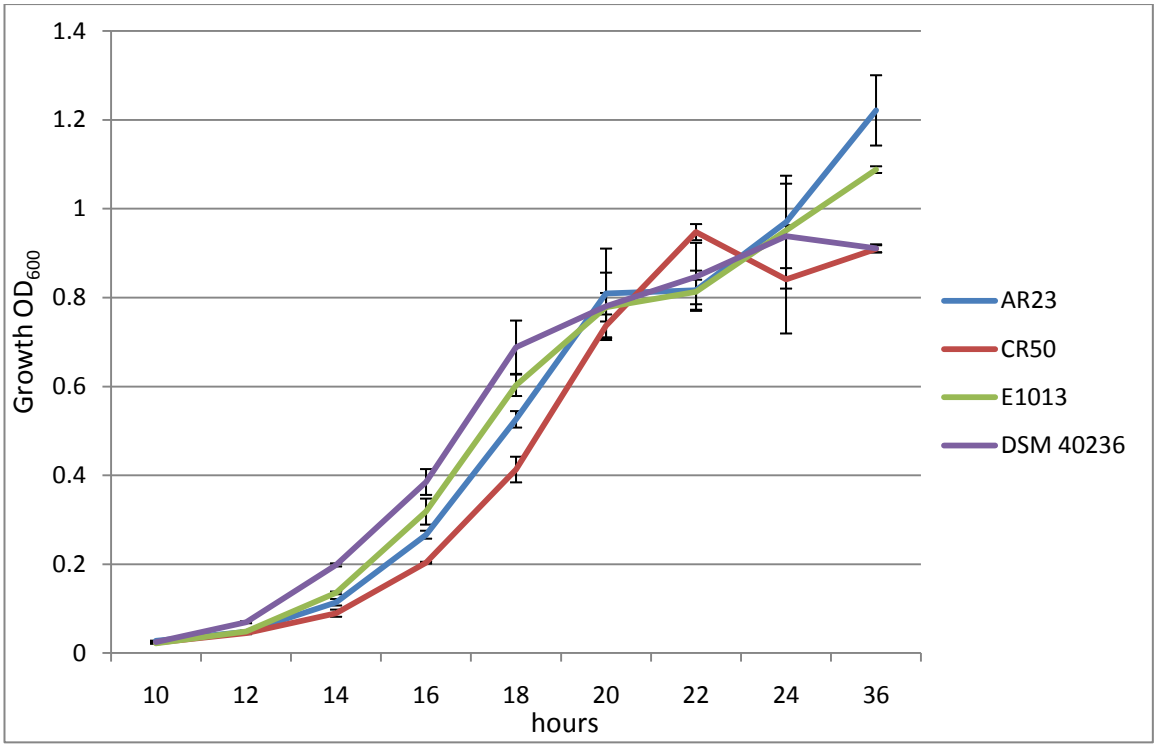


Figure 7.1. Mean values (with standard deviations represented as error bars) of OD_{600} readings from three biological replicates of the four strains utilized in this experiment. X-axis is the OD values and y-axis the hours at which the sampling took place.

The strains were grown and sampled as described in Section 2.7.2. All four strains appeared to be in the exponential phase with their OD_{600} readings doubling approximately every two hours from the 10 h time point onwards (Figure 7.1, Table 7.7). There was retardation in the growth of DSM 40236^T and E1013 past 18 h and of AR23 and CR50 past 20 h. There was relatively little increase in biomass over the 16 hours past the 20 h time point; the relative increase in AR23 and E1013 is likely to have been due to extracellular material and/or pigment synthesis by those strains.

7.3.7 *S. griseus* AR23

Table 7.8. The mean values of the gene/*hrdB* expression ratio from three biological replicates and their standard deviations. The vertical column represents the hour after seeding that samples were taken.

	strA		*SGR172*		*adpA*		*atrA*	
h	mean	std dev	mean	std dev	mean	std dev	mean	std dev
12	0.016630	0.002283	0.073513	0.015940	1.410551	0.447420	0.286746	0.020266
16	0.005538	0.001427	0.018459	0.001094	2.777808	0.411525	0.325964	0.047633
20	0.023112	0.003247	0.205841	0.047361	2.692806	0.571930	6.960597	1.385433
24	0.049829	0.027320	0.196282	0.096704	5.300103	1.207296	6.620566	4.340161
36	0.154589	0.137666	0.366638	0.186436	2.424762	0.190363	9.244915	1.583064

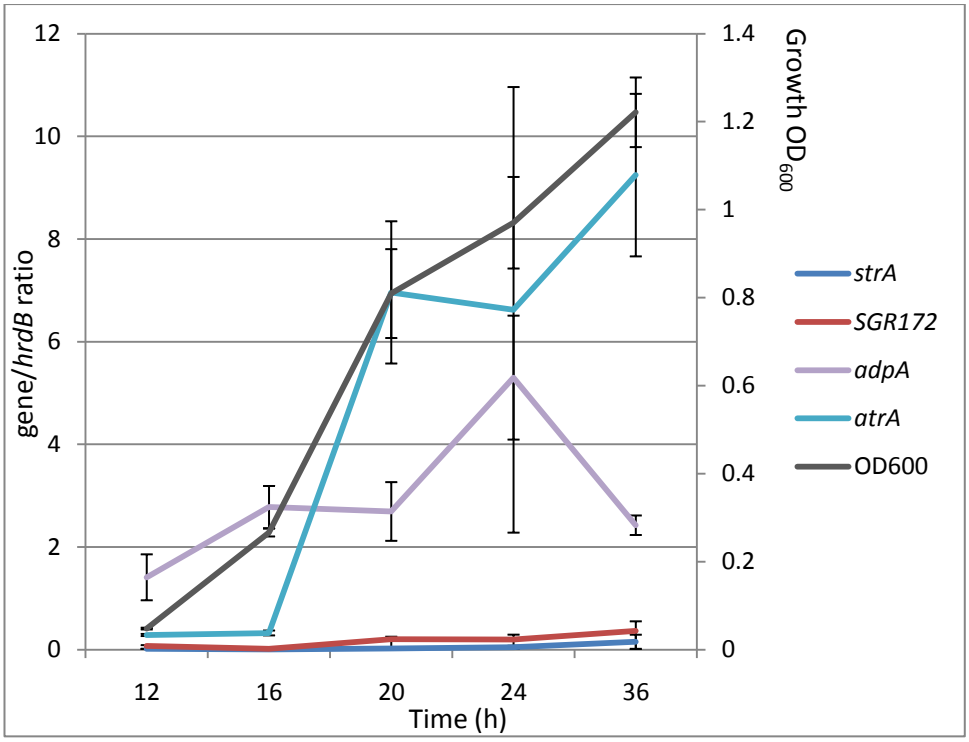


Figure 7.2. Mean values (with standard deviations represented as error bars) of expression levels of genes and optical density of culture from three biological replicates of strain *S. griseus* AR23. X-axis is the hours at which the culture was sampled; left y-axis is the gene/*hrdB* ratio for the genes and right y-axis is optical density for the OD_{600} line.

The *SGR172* and *strA* genes had an almost identical expression pattern, with *SGR172* being slightly more highly expressed. The increase on the 36 h time point in the expression of *SGR172* and *strA* was due to a single of the three biological replicates. The other two biological replicates were as expressed at 36 h as they were at 24, indicating that these two genes have a low level constitutive expression. The *adpA* gene appeared to be highly expressed from the 12 h time point; however it reached the peak of its expression at 24 h. *atrA* demonstrated low expression at 12 h and 16 h, but was highly expressed from 20 h onwards. (Figure 7.2, Table 7.8).

7.3.8 *S. platensis* CR50

Table 7.9.The mean values of the gene/*hrdB* expression ratio from three biological replicates of strain *S. platensis* CR50 and their standard deviations. The vertical column represents the hour after seeding that samples were taken.

	strA		*strR*		*strW*	
h	mean	std dev	mean	std dev	mean	std dev
12	0.001288	0.000307	0.000283	6.45E-05	9.98E-05	2E-05
16	0.003873	0.001085	0.001523	0.000586	0.000362	0.000119
20	0.074524	0.048681	0.108271	0.055646	0.001568	0.000748
24	0.013563	0.005787	0.028398	0.005890	0.000475	6.26E-05
36	0.075542	0.011098	0.072000	0.005360	0.001062	0.000142
	adpA		*atrA*			
h	mean	std dev	mean	std dev		
12	0.025441	0.018839	0.024210	0.002083		
16	0.105665	0.049007	0.083000	0.010286		
20	4.268442	2.904946	0.722126	0.485402		
24	1.293714	0.112910	1.274205	0.208963		
36	0.026821	0.021014	0.036825	0.020779		

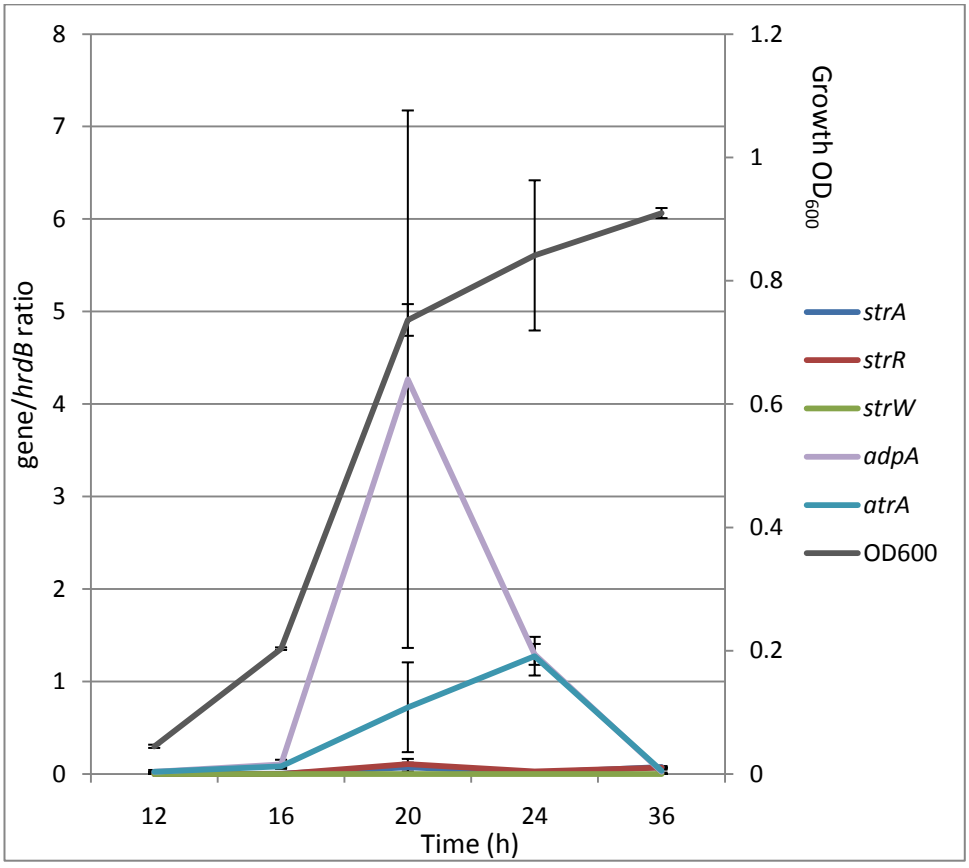


Figure 7.3. Mean values (with standard deviations represented as error bars) of expression levels of genes and optical density of culture from three biological replicates of strain *S. platensis* CR50. X-axis is the hours at which the culture was sampled; left y-axis is the gene/*hrdB* ratio for the genes and right y-axis is optical density for the OD_{600} line.

The expression of the *strA*, *strR* and *strW* genes had low levels of transcription at 12 h and 16 h, were more strongly induced at 20 h and their transcription fell at 24 h. Their overall level of expression was very low. The *adpA* gene was very strongly expressed at 20 h and then declined after that. The high error at that time point is because one of the three biological replicates had an expression that was twice as high as that of the other two. While it was not possible to sequence *adpA* from genomic DNA, its dissociation curves indicated the presence of a single product that had a very similar melting temperature to the *adpA* of the other strains suggesting that this was the correct product. The *atrA* expression increased at 20 h but peaked at 24 h (Figure 7.3, Table 7.9). The fact that *strA*, *strR*, *strW* and *adpA* all reached their maximum expression level at 20 h implies that the transition phase in CR50 occurred at that time.

7.3.9 *S. griseus* E1013

Table 7.10. The mean values of the gene/*hrdB* expression ratio from three biological replicates of strain *S. griseus* E1013 and their standard deviations. The vertical column represents the hour after seeding that samples were taken.

	strA		*adpA*		*atrA*		*aphE*	
h	mean	std dev	mean	std dev	mean	std dev	mean	std dev
12	0.013006	0.002128	0.80556	0.086084	0.056843	0.011596	0.020942	0.004228
16	0.007724	0.001056	0.684352	0.062725	0.086165	0.008023	0.021162	0.002018
20	0.035885	0.002829	2.86566	0.067733	2.905058	0.337093	0.042316	0.005289
24	0.027333	0.001418	3.115386	0.426352	0.855797	0.214077	0.065673	0.019484
36	0.042363	0.00678	2.593366	0.392764	4.538233	1.041583	0.193449	0.060761

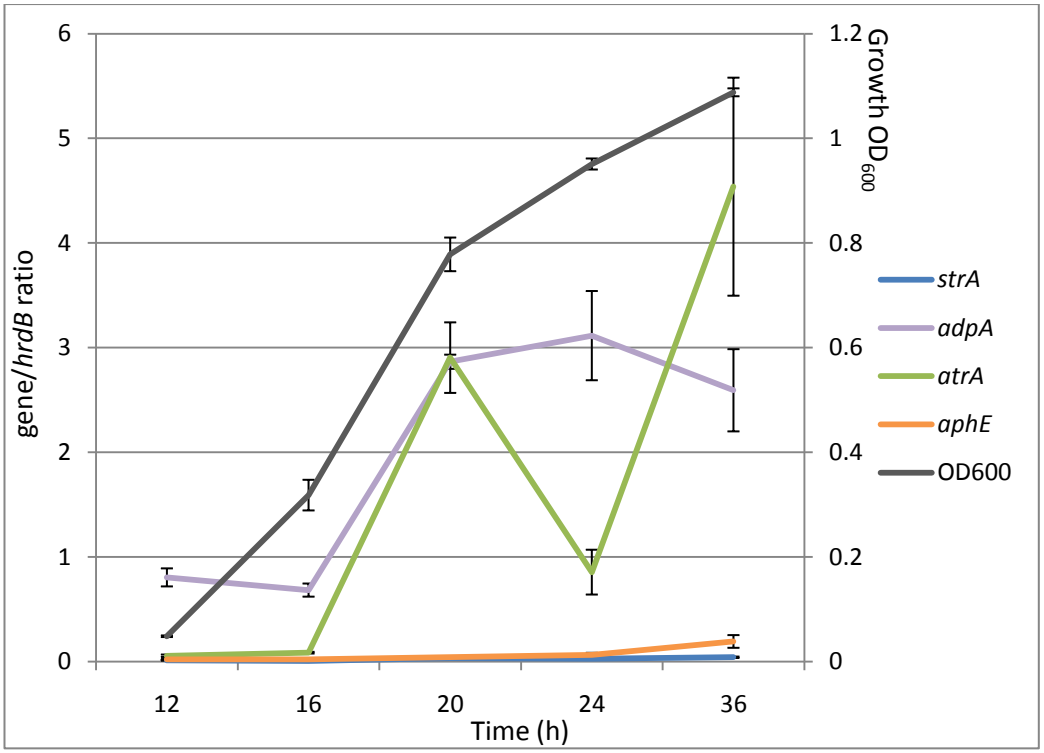


Figure 7.4. Mean values (with standard deviations represented as error bars) of expression levels of genes and optical density of culture from three biological replicates of strain *S. griseus* E1013. X-axis is the hours at which the culture was sampled; left y-axis is the gene/*hrdB* ratio for the genes and right y-axis is optical density for the OD_{600} line.

The *adpA* gene was upregulated after the 16 h time point and its expression remained relatively stable after 20 h. *atrA* and *strA* were also upregulated after 16 h but they had a temporary drop in expression at the 24 h time point. The expression of *aphE* was more stable and appeared to gradually increase over time (Figure 7.4, Table 7.10).

7.3.10 *S. griseus* DSM 40236^{T}

Table 7.11. The mean values of the gene/*hrdB* expression ratio from three biological replicates and their standard deviations. The vertical column represents the hour after seeding that samples were taken.

	strA		*strR*		*strW*	
h	mean	std dev	mean	std dev	mean	std dev
12	0.080953	0.002241	0.013991	0.001606	0.006793	0.000778
16	0.494067	0.013553	0.2234	0.015554	0.008101	0.000158
20	0.62524	0.304344	0.400743	0.18748	0.00862	0.004483
24	1.332843	0.195086	0.918156	0.104328	0.046393	0.011045
36	0.426726	0.017451	0.59546	0.087584	0.008666	0.001805
	adpA		*SGR172*		*atrA*	
h	mean	std dev	mean	std dev	mean	std dev
12	0.58652	0.197735	0.013141	0.002763	0.129779	0.054503
16	0.945798	0.075121	0.018003	0.001387	0.351252	0.047745
20	1.010265	0.112661	0.06925	0.031449	0.27314	0.090578
24	1.684085	0.178224	0.072019	0.016817	0.358474	0.090076
36	0.822416	0.40226	0.088077	0.061782	0.324481	0.246502
	aphE		*SGR878*		*SGR421*	
h	mean	std dev	mean	std dev	mean	std dev
12	0.006587	0.00219	0.045979	0.004466	0.012886	0.002104
16	0.008151	0.000895	0.133663	0.031924	0.024169	0.002841
20	0.00782	0.003174	0.098256	0.052642	0.035715	0.006741
24	0.016006	0.012124	0.127305	0.033011	0.041256	0.028325
36	0.01853	0.002364	0.175399	0.037805	0.052154	0.015065

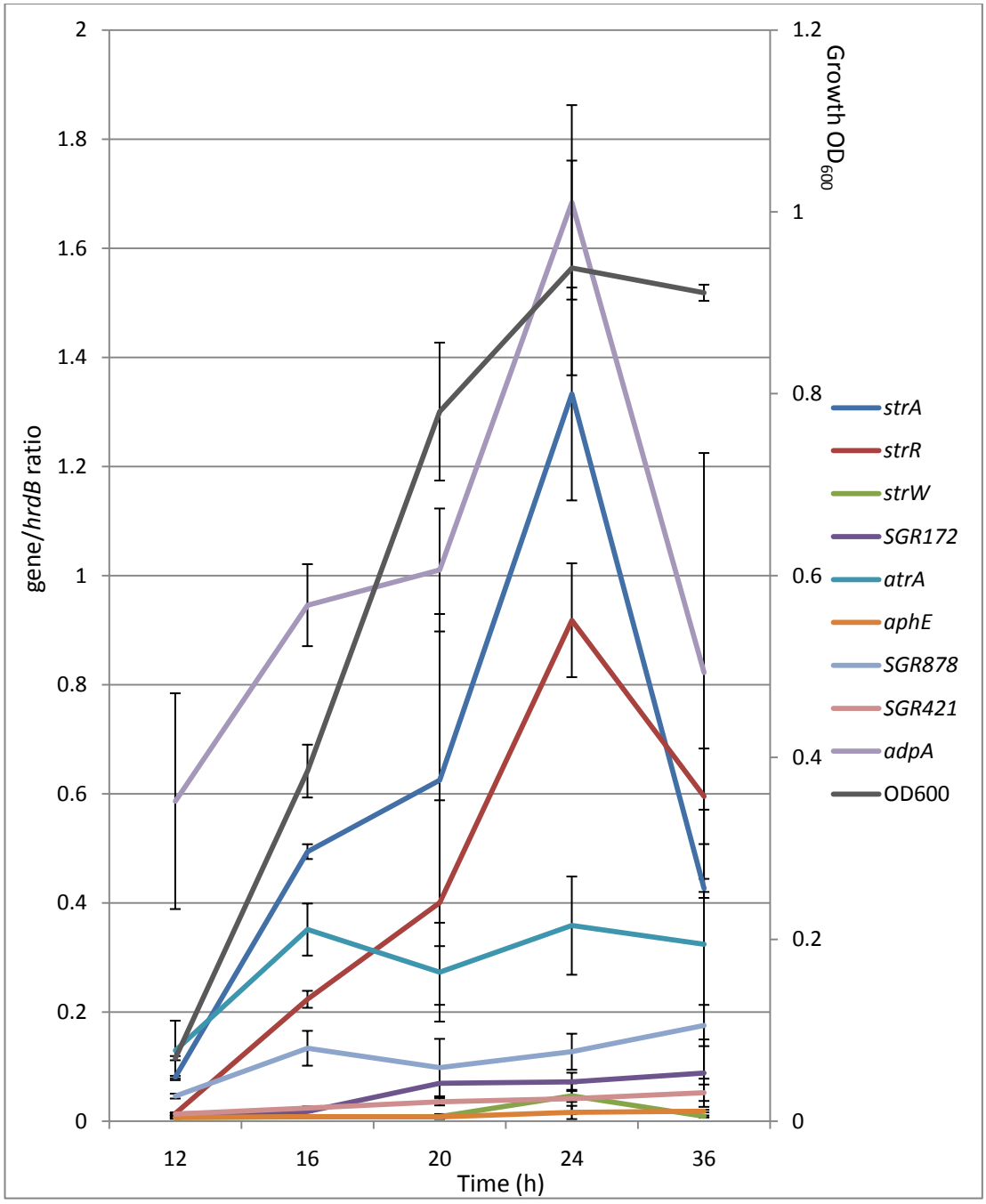


Figure 7.5. Mean values (with standard deviations represented as error bars) of expression levels of genes and optical density of culture from three biological replicates of strain *S. griseus* DSM 40236^{T}. X-axis is the hours at which the culture was sampled; left y-axis is the gene/*hrdB* ratio for the genes and right y-axis is optical density for the OD_{600} line.

The *strA*, *strR* and *adpA* genes had low expression initially, but demonstrated rapid increase from 16 h onwards and peaked after 24 h before declining at 36 h. The expression of *strW* only started to increase from 20 h onwards and also peaked at 24 h, though its peak expression level was 29-fold lower than that of *strA*. *SGR172*, *atrA*, *aphE*, *SGR878* and *SGR421* all appeared to have an expression profile that was close to stationary, at most displaying a slight increase in expression over time

compared to the 12 h time point. The transcriptional regulators *atrA* and *SGR878* had a moderate expression level; however the remaining genes were very poorly expressed (Figure 7.5, Table 7.11).

7.3.11 Comparison of individual gene expression across strains

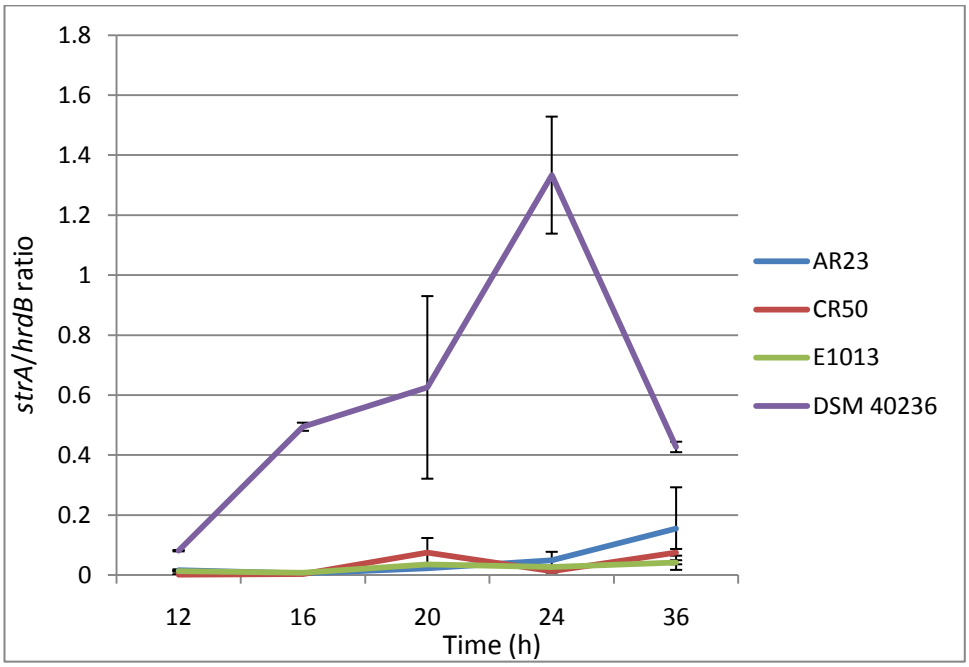


Figure 7.6. Mean values (with standard deviations represented as error bars) of *strA* expression levels. X-axis is the hours at which the cultures were sampled; y-axis is the *strA*/*hrdB* ratio.

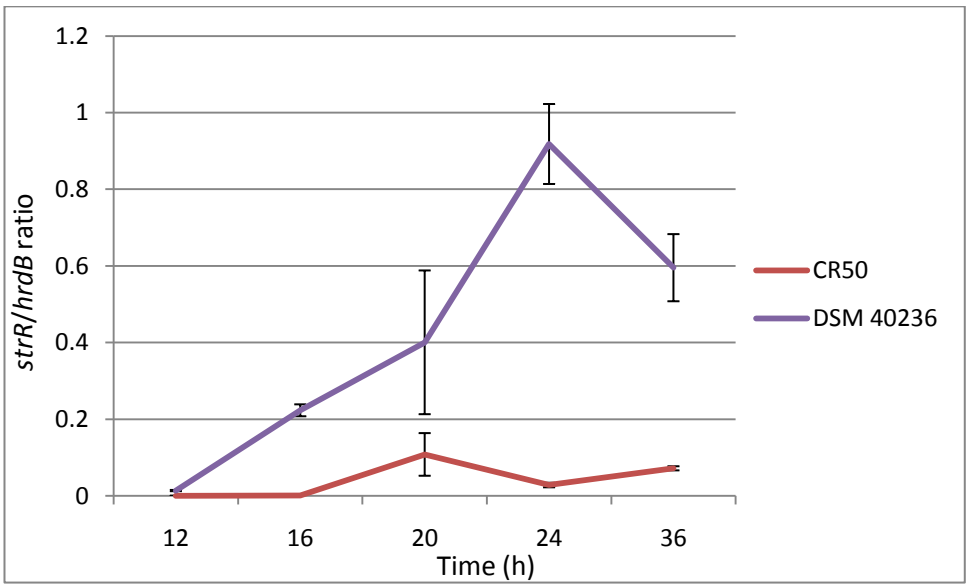


Figure 7.7. Mean values (with standard deviations represented as error bars) of *strR* expression levels. X-axis is the hours at which the cultures were sampled; y-axis is the *strR*/*hrdB* ratio.

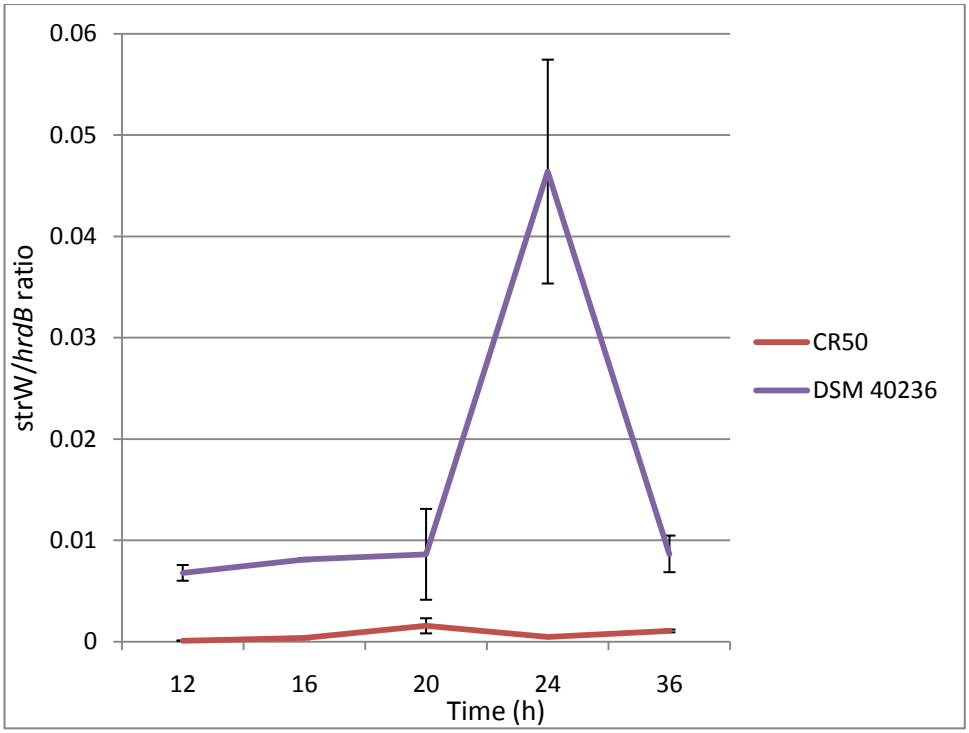


Figure 7.8. Mean values (with standard deviations represented as error bars) of *strW* expression levels. X-axis is the hours at which the cultures were sampled; y-axis is the *strW*/*hrdB* ratio.

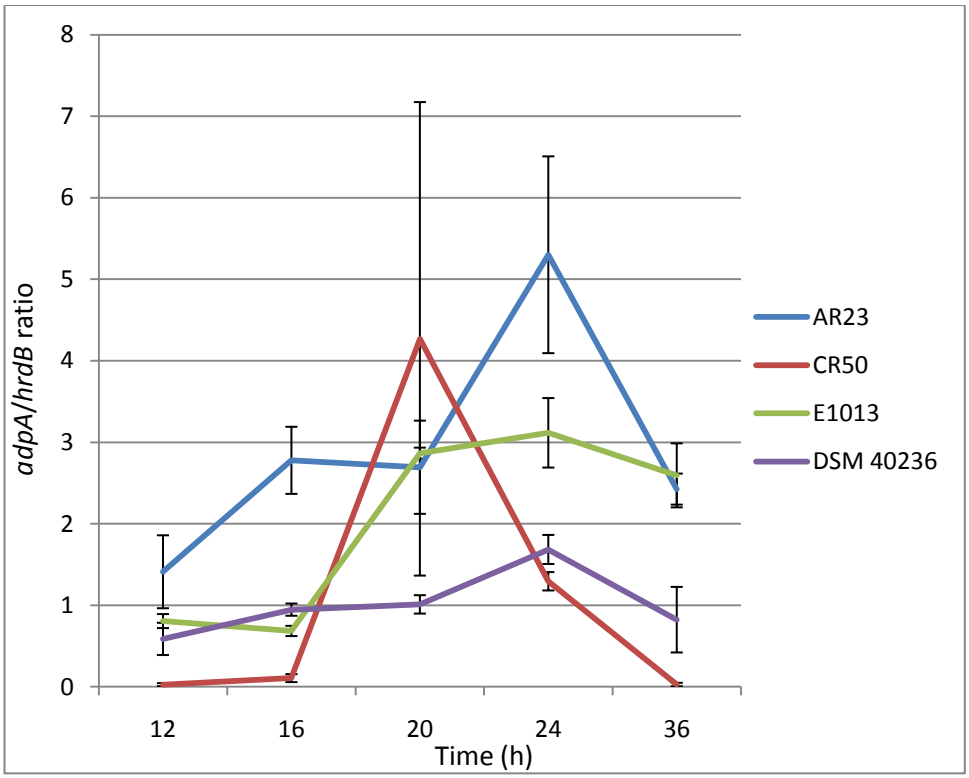


Figure 7.9. Mean values (with standard deviations represented as error bars) of *adpA* expression levels. X-axis is the hours at which the cultures were sampled; y-axis is the *adpA*/*hrdB* ratio.

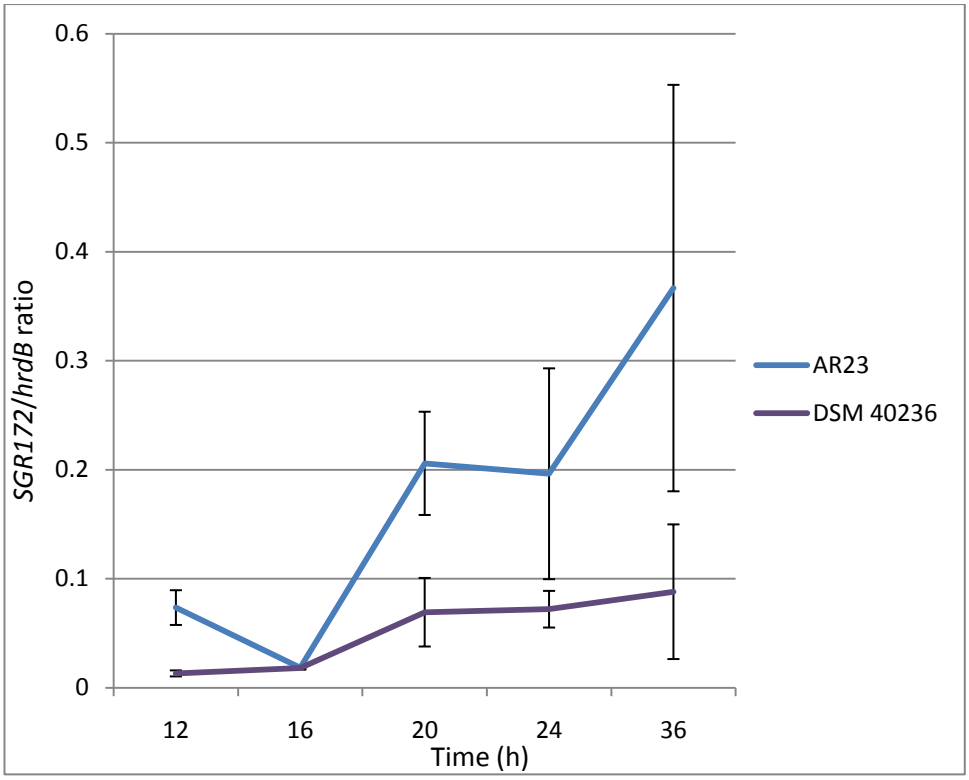


Figure 7.10. Mean values (with standard deviations represented as error bars) of *SGR172* expression levels. X-axis is the hours at which the cultures were sampled; y-axis is the *SGR172*/*hrdB* ratio.

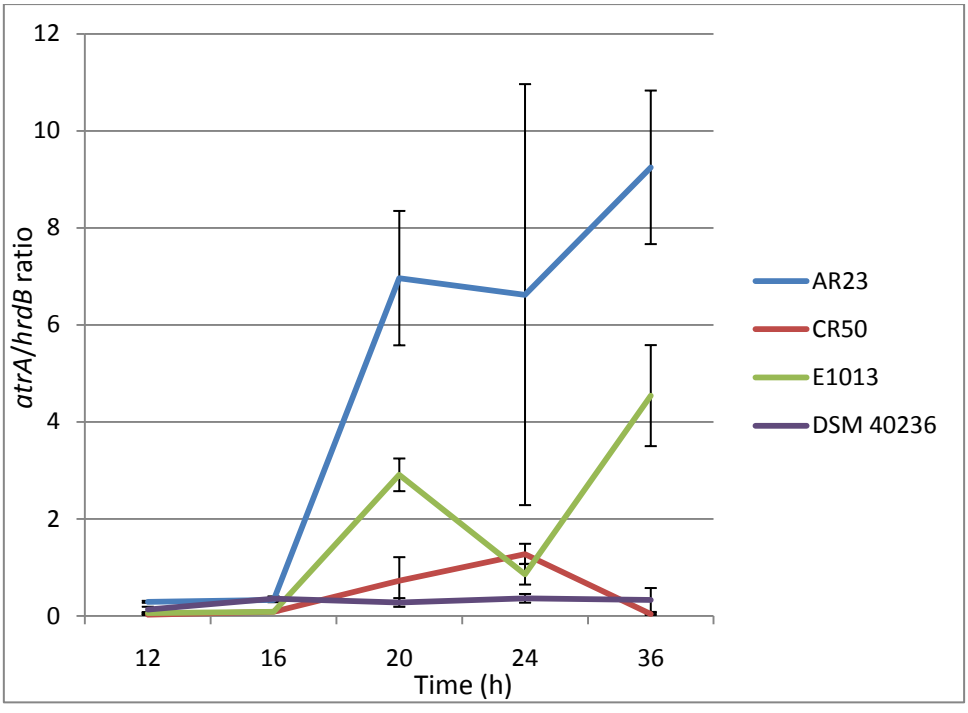


Figure 7.11. Mean values (with standard deviations represented as error bars) of *atrA* expression levels. X-axis is the hours at which the cultures were sampled; y-axis is the *atrA*/*hrdB* ratio.

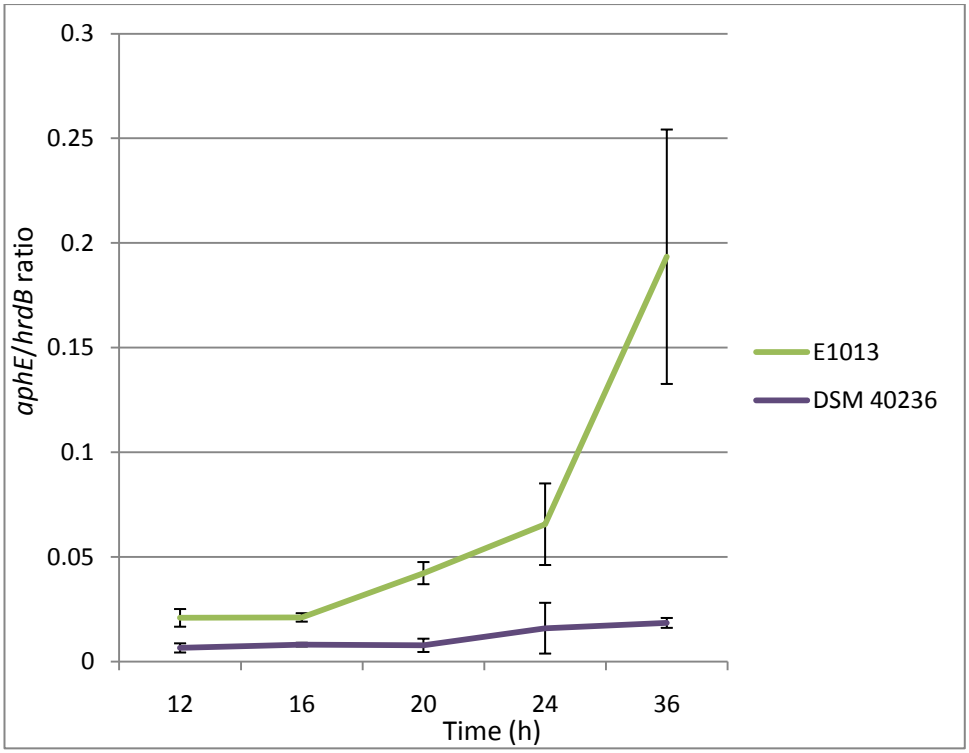


Figure 7.12. Mean values (with standard deviations represented as error bars) of *aphE* expression levels. X-axis is the hours at which the cultures were sampled; y-axis is the *aphE*/*hrdB* ratio.

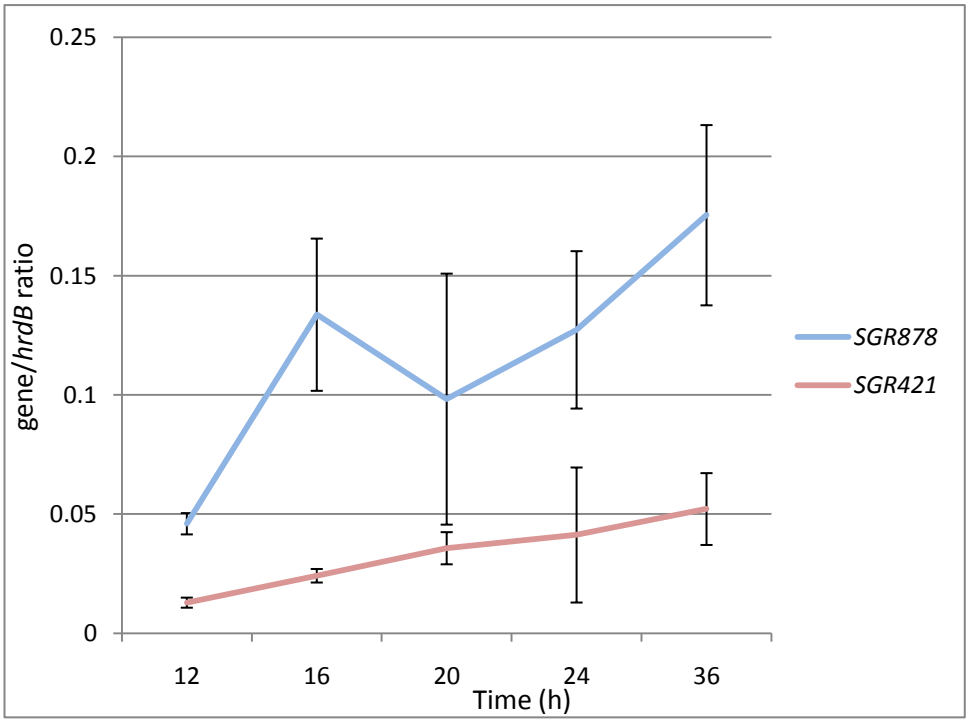


Figure 7.13. Mean values (with standard deviations represented as error bars) of *SGR878* and *SGR421* expression levels from *S. griseus* DSM 40236^{T}. X-axis is the hours at which the cultures were sampled; y-axis is the gene/*hrdB* ratio.

The *strA* expression in DSM 40236^T was upregulated before the start of antibiotic biosynthesis and reached its peak at 24 h. CR20 expression was most highly expressed at 20 h but overall its expression was at the same level as the strains lacking a streptomycin cluster. AR23 and E1013 showed very little change in their *strA* expression levels over time, indicating that *strA* expression in these strains is constitutive (Figure 7.6).

The expression of *strR* and *strW* was also significantly higher in DSM 40236^T compared to CR50. As with *strA*, their expression peaked at 24 h for DSM 40236^T and at 20 h for CR50 (Figure 7.7, Figure 7.8).

The *adpA* gene was upregulated after 16 h in all the strains, though in AR23 it was already expressed at a high level. Its expression peaked at 24 h for all strains except for CR20 where it peaked at 20 h (Figure 7.9).

The expression profiles of *SGR172* in DSM 40236^T and AR23 were significantly different. In AR23 its expression level was higher and demonstrated a dip at 16 h. Conversely in DSM 40236^T the expression only showed a very slight increase over time (Figure 7.10).

There was considerable variation in the expression profiles of *atrA*. In AR23 and E1013 the gene was upregulated after 16 h, had a drop in expression at 24 h and reached its peak at 36 h. In CR50 it was upregulated after 16 h, reached its peak at 24 h and had a decrease in expression at 36 h. In DSM 40236^T the gene's expression was almost stable from the start until the end of the experiment (Figure 7.11).

The *aphE* expression appeared stable in DSM 40236^T; however in E1013 it increased at 24 h and peaked at 36 h (Figure 7.12).

The expression patterns of *SGR878* and *SGR421* in DSM 40236^T appeared somewhat different; however the standard deviations of the *SGR878* data points were too high to draw any firm conclusions about whether their expression truly varied (Figure 7.13). The putative viomycin resistance gene *SGR421* appeared to be constitutively expressed.

7.4 Discussion

The patterns of expression observed for DSM 40236^T genes agree with the findings of previous studies on the expression of the streptomycin gene cluster. As previously reported, the *adpA* gene is expressed first, the *strR* and *strA* genes second, *strW* third (Tomono *et al.*, 2005) and *atrA* appears to be constitutively active (Hirano *et al.*, 2008). The different expression pattern of *SGR172* compared to *strA* indicates that the regulation of the two genes is completely different in DSM 40236^T. The low level of constitutive expression of the 3" phosphoryltransferase gene *aphE* might serve to provide resistance to streptomycin when the *strA* gene is not highly expressed. The presence of an additional resistance determinant is not surprising. In *S. hygroscopicus* a knockout mutant in which the *hyg21* gene is inactivated, which is situated in the hygromycin A biosynthetic cluster and encodes an O-phosphotransferase providing self-resistance, retains significant resistance to hygromycin A, indicating that that producer has additional active resistance mechanisms as well (Dhote *et al.*, 2008).

The expression pattern of *adpA* and streptomycin cluster genes in CR50 was very similar to that of DSM 40236^T, though their expression peaked at 20 h, indicating that the transition phase in CR50 may occur earlier than in DSM 40236^T. The expression levels of all the cluster genes were extremely low, which suggests that if there is antibiotic production from this cluster it is likely to be at a very low level. This is also supported by the complete absence of zones of clearing on *Bacillus subtilis* by CR50 culture supernatant, though the minimal zones of clearing induced by DSM 40236^T indicate that this test is not very sensitive. The *adpA* gene is present and active in *S. coelicolor* A3(2) (Nguyen *et al.*, 2003), which lacks the A-factor cascade, and it can therefore potentially act as a regulator for the streptomycin gene cluster even if it is the only part of the A-factor cascade that is present in CR50.

The very similar expression profiles of *SGR172* and *strA* in AR23 lend credence to the hypothesis that these two genes are both transcribed as a polycistronic mRNA molecule. The higher level of *SGR172* compared to *strA* may be due to the rapid decline in expression along streptomycete polycistronic mRNA molecules (Hodgson, pers. comm.). Another possibility is that the mRNA molecules are degraded by 3' exonucleases (Kennell, 2002), which would result in the *strA*

sequence being lost first resulting in the presence of a comparatively larger number of *SGR172* mRNA transcripts. Finally this could be an artifact of the experiment; the random hexamer primers are more likely to reverse transcribe sequences at the 5' end of an RNA molecule, which would bias the results in favor of *STR172* which is located upstream of *strA*. The expression profile of *strA* in E1013 is also very similar to that of AR23 which suggests that it is regulated in the same manner. Since the expression of *strA* appears to be constitutive in these two strains and *adpA* is strongly induced after the 16 h time point, it is unlikely that *adpA* plays a role in the regulation of *strA* in AR23 and E1013 as this upregulation is not reflected in the expression of the resistance gene. The *aphE* expression in E1013 appears to gradually increase over time, though its levels remain approximately the same as those of *strA*. AphE is likely to play a supplementary role to StrA in the inactivation of streptomycin in E1013.

The results from this real time PCR study conflicted with those done by the measurement of DNA band intensities on a gel after the completion of a conventional PCR (Tolba, 2004). In this study the expression of *strA* at 24 h in DSM 40236^T was 27 times higher compared to *strA* in AR23 at the same time point, while in the previous study that number was 1.5 (Tolba, 2004). This can be explained by the fact that a PCR reaches a saturation point after a certain number of cycles after which little to no new product is generated. As a result, even if the amount of cDNA at the start of two reactions was considerably different both reactions would be likely to stop amplifying after reaching a similar saturation point. Conversely, real time PCR measures the amount of product during the exponential phase of the amplification and can thus provide reliable information about the amount of product initially present. Therefore the difference in the two results is likely due to the limitations of the procedure used in the previous study.

The expression levels of the putative viomycin resistance gene *SGR421* in DSM 40236^T are approximately the same as those of *strA* in AR23, CR50 and E1013. Despite the low expression level of *SGR421*, DSM 40236^T demonstrates high level resistance to viomycin (Section 5.3.13). The discrepancy between the resistance levels of DSM 40236^T to viomycin and AR23, CR50 and E1013 to streptomycin in spite of the fact that their resistance genes are equally expressed might be due to the

presence of additional genes that also provide viomycin resistance or due to the SGR421 enzyme being a significantly more efficient antibiotic inactivator than the StrA found in resistance-only strains.

Despite the fact that these genes are transcribed, it cannot be certain that they are also translated, as gene expression can be subject to posttranslational control (Nogueira & Springer, 2000). In order to confirm that the genes examined serve the functions that have been described, it would be necessary to perform gene knockouts and then examine the phenotype of the resultant mutants. Nevertheless, the fact that *strA* is highly expressed in a strain displaying high resistance and has low expression in strains with low streptomycin resistance is suggestive of a link between the level of *strA* transcription and the level of streptomycin resistance.

In *S. griseus* DSM 40236^{T}, the expression of *strA* is integrated into the organism's metabolic network, as it is part of the streptomycin gene cluster. The A-factor cascade's activation of streptomycin production is part of a wider set of changes in the metabolism of *S. griseus* that also includes the biosynthesis of a polyketide (Yamazaki *et al.*, 2004) and of grixazone, a diffusible yellow pigment (Higashi *et al.*, 2007), the regulation of two extracellular proteases (Hirano *et al.*, 2006) as well as the formation of aerial hyphae which are necessary for streptomycete sporulation (Horinouchi, 2002). γ-Butyrolactones are found across the *Streptomyces* genus and in some other acinobacteria, indicating that this is a general regulatory system for antibiotic production (Takano, 2006). They regulate the production of a number of antibiotics other than streptomycin such as lankacidin and lankamycin production in *S. rochei* (Yamamoto *et al.*, 2008), virginiamycin in *S. virginiae* (Kawachi *et al.*, 2000), showdomycin and minimycin in *S. lavendulae* (Kitani *et al.*, 2001) and actinorhodin and undecylprodigiosin in *S. coelicolor* A3(2) (Takano *et al.*, 2001). Since *adpA* is likely to be present in CR50, it is possible that its cluster is regulated via a γ-butyrolactone as well. However the possibilities that a gene other than *adpA* regulates the cluster or that *adpA* is regulated differently in CR50 cannot be excluded.

The aminoglycoside resistance genes not associated with a biosynthetic cluster however tend not to be integrated into the organism's metabolic network or modulated by the cell and are instead expressed constitutively (Shaw *et al.*, 1993;

Kim *et al.*, 2006). The same applies to other independent resistance genes such as β-lactamases (Ogawara *et al.*, 1999). This is not surprising for genes that can undergo HGT, as the organism they find themselves in may not contain the same regulatory networks as their original host and they will need an alternative method of expression (Martinez *et al.*, 2009a). Constitutive expression will be favored because the organism may be exposed to an antibiotic without warning and thus it will need to be resistant at all times (Kim *et al.*, 2006). There are however some exceptions to that; *Staphylococcus aureus* KU5801 possesses a gene encoding an aminoglycoside-6′-N-acetyltransferase and 2"-O-phosphotransferase which is expressed in the presence of β-lactam antibiotics. This is due to the *aac(6')-Ie-aph(2")-Ia* gene having integrated within a *blaZ* gene, which causes it to be expressed by the pathway regulating penicillinase production (Ida *et al.*, 2002). The -35 promoter of the *aac(6')-Ic* gene, found on the *Serratia marcescens* chromosome and encoding an aminoglycoside 6' acetyltransferase (Snelling *et al.*, 1993), overlaps with a large palindromic sequence which may function as an operator where an unlinked repressor binds to (Shaw *et al.*, 1992). The *aac(6')-Ij* gene, encoding a 6'-N-acetyltransferase, from *Acinetobacter sp.* 13 BM2689 can be expressed at various levels indicating some form of regulation (Lambert *et al.*, 1994; Depardieu *et al.*, 2007). The expression of the *aac(2')-Ia* gene, encoding a gentamicin 2'-N-acetyltransferase, is controlled by at least two regulators in *Providencia stuartii* (Rather *et al.*, 1993; Rather & Orosz, 1994). However this regulation is due to the gene's primary function in the acetylation of peptidoglycans and it only acetylates aminoglycosides by chance (Payie & Clarke, 1997). Unlike the aforementioned genes however, the *strA* gene in AR23 and E1013 and the *SGR421* and *aphE* genes in DSM 40236^{T} all appear to be constitutively expressed. This is unsurprising though, as this is the expression pattern displayed by the majority of aminoglycoside phosphotransferases (Shaw *et al.*, 1993).

It can therefore be concluded that hypothesis (i) is correct and *strA* is expressed constitutively at significantly lower amounts in resistance only strain, hypothesis (ii) is partially correct, as the CR50 follows a similar pattern of expression, though its expression level is considerably lower and peaks earlier than in DSM 40236^{T} and hypothesis (iii) is incorrect as the *SGR421* gene is expressed at a level similar to that of *strA* in the resistance-only strains.

8 General discussion

8.1 Role of antibiotics in the environment

Despite the widespread clinical use of antibiotics, there is considerable controversy over their function in the environment. The first theory to be developed was that their purpose is to inhibit the growth of microbial competitors (Waksman & Woodruff, 1940). In streptomycetes, the initiation of secondary metabolism, including antibiotics, coincides with the formation of the aerial mycelium and sporulation. In some species, such as *S. coelicolor*, the processes of antibiotic production and morphological differentiation are controlled separately and are merely initiated simultaneously while in others, such as *S. griseus*, the two processes are regulated by the same transcription factor (Ohnishi *et al.*, 2005; Flardh & Buttner, 2009). As sporulation is prompted by nutrient limitations in the organism's environment, the substrate mycelium lyses releasing stored glycogen and protein (Sprusansky *et al.*, 2005) in order to provide the developing aerial mycelia with a nutrient source (Chater *et al.*, 1989). It has therefore been postulated that antibiotics are produced at that stage in the organism's life cycle in order to eliminate other microorganisms that might attempt to compete with the aerial hyphae for the nutrients released by the lysed mycelium (Hodgson, 2000). Another proposed hypothesis is that the synthesised antibiotics are packaged in the spores and are meant to provide a competitive advantage to the germlings that emerge from them by inhibiting the growth of other organisms on the site where they germinate (Tate, 2000).

However, the limited success in detecting naturally produced antibiotics within soil (Morningstar *et al.*, 2006) has resulted in considerable doubt about whether antibiotics can accumulate in concentrations high enough to inhibit the growth of microorganisms in natural environments (Aminov, 2009). It has been suggested that the reason bacteria produce antibiotics in quantities sufficient to inhibit other microorganisms under laboratory conditions is that they stop being under the starvation diet they encounter in their natural environment and are instead grown in an extremely nutrient-rich liquid monoculture with a constant temperature to which their metabolism is not adapted. Under these artificial conditions the cell's metabolism can become disturbed, resulting in a drastic overproduction of small

molecules that would never occur in nature (Davies, 2006a). In addition, most streptomycetes appear to possess the genetic capacity to produce more than 25 different small molecules, the majority of which are not detected in traditional antimicrobial screens. This suggests that they are not produced in concentrations sufficient to inhibit growth or that they do not act as antibiotics (Yim *et al.*, 2006).

One proposed function of secondary metabolites, including antibiotics, is that they act as form of metabolic waste (Chadwick & Whelan, 1992). This hypothesis fails to explain why there is such diversity in secondary metabolism if all the compounds are simply waste products. Furthermore, some secondary metabolites accumulate in substantial amounts in the producing organism and are often stored at a separate location from the site of synthesis (Williams *et al.*, 1989).

It has also been suggested that secondary metabolites are the product of overflow metabolism in streptomycetes. This hypothesis states that under conditions where a major nutrient is unavailable, secondary metabolism could be induced to keep the major primary metabolic pathways operating until the restoration of the missing metabolite (Hodgson, 2000). Soil is an environment with an abundance of carbon sources but a limited supply of nitrogen sources. This is reflected in the carbon catabolic pathways, which are tightly regulated in order to allow selection of the optimal carbon sources for growth. Conversely, many of the nucleotide and amino acid catabolic pathways are constitutively expressed at a low level, which may be a reflection of the limited amount of nitrogen sources, necessitating them to be constantly active so that they can utilise all available nitrogen sources (Hodgson, 2000). The absence of nitrogen sources, including amino acids, in soil also explains why most of the amino acid biosynthesis pathways are also constitutive at a low level, unlike the majority of bacteria where they are regulated. If streptomycetes were placed in a situation where there was an overabundance of nitrogen sources, they would have no way of down regulating the biosynthesis of amino acids. In such a scenario secondary metabolism could be activated to utilise the excess amino acids to prevent them from reaching excessively high levels (Hodgson, 2000). One piece of evidence supporting this view is that *S. coelicolor* A3(2) mutants with a reduced ability to degrade proline overproduced the antibiotic Red (undecylprodigiosin), which suggests that the Red biosynthetic pathway may act as a sink for excess

proline (Hood *et al.*, 1992). Like the metabolic waste hypothesis, this hypothesis fails to explain why there is such great diversity in secondary metabolism if the final product is unimportant (Williams *et al.*, 1989). Furthermore, secondary metabolism is co-ordinately regulated along with sporulation in a number of streptomycetes, such as *S. griseus*. In order for this theory to apply to them these streptomycetes would have to only sporulate when under conditions with highly imbalanced nutrient availability. Also, the formation of an aerial mycelium utilises many of the same primary metabolic pathways that are involved in the growth of the vegetative mycelium. Therefore, when streptomycetes activate their secondary metabolism they need to possess the substrates necessary for primary metabolic pathways in order to grow aerial mycelia (Maplestone *et al.*, 1992). This hypothesis may apply to some pathways, but it fails to account for the diversity and expression patterns of most secondary metabolites.

Another suggestion is that secondary metabolite, including antibiotic, biosynthesis pathways act as a reserve pool for the development of new pathways that might prove useful for the cell in the future (Chadwick & Whelan, 1992). This would imply that secondary metabolite clusters are retained by bacteria in order to provide a future benefit to the organism, which is contrary to what is known about the operation of natural selection (Williams *et al.*, 1989). The high cost that carrying secondary metabolite genes on their genome must entail for streptomycetes, the arm regions of the *S. coelicolor* genome are 3.8 Mb and mostly code nonessential genes (Bentley *et al.*, 2002), and the complex and energetically expensive biosyntheitc pathways necessary to produce them suggest that secondary metabolites must provide an immediate selective advantage to their host, though this does not preclude the possibility of genes involved in secondary metabolism interacting in new ways and giving rise to novel compounds.

A more recently proposed alternative function for antibiotics is that they act as signalling molecules when present at subinhibitory concentrations (Davies, 2006a). It is estimated that there are 10^{11} to 10^{13} microorganisms in one kilogram of soil based on microscopic counts (Yim *et al.*, 2007) and approximately 10^{4} to 10^{6} bacterial species in a 10 g soil sample (Janssen, 2006). The large number and diversity of microbes in soil may therefore favour the evolution and maintenance of

methods of intermicrobial communication. Subinhibitory concentrations of antibiotics appear to effect the expression of genes not directly linked to stress responses. In *Bacillus subtilis* chloramphenicol up-regulated transport/binding, ribosomal and carbohydrate metabolism proteins and down-regulated amino acid metabolism, gentamicin down regulates transport/binding proteins and erythromycin down-regulated amino acid and carbohydrate metabolism and purine and pyrimidine biosynthetic enzymes, and had a mixed effect on transport/binding and ribosomal proteins (Davies, 2006a). Despite the fact that all these antibiotic compounds are protein synthesis inhibitors, they triggered different alterations in the cell's expression profiles. The protein synthesis inhibitors puromycin, tetracycline, chloramphenicol and erythromycin all up-regulated purine nucleotide biosynthesis in *Streptococcus pneumoniae*, while the protein synthesis inhibitors kasugamycin and puromycin down-regulated the *rpoS* regulon, which is a central regulator of the general stress response, in *E. coli* (Davies *et al.*, 2006). Furthermore, in experiments where *Salmonella typhimurium* was subjected to subinhibitory amounts of erythromycin and rifampicin, approximately 5% of its genes were modulated even when genes involved in the SOS, the heat shock or the universal stress response were knocked out (Goh *et al.*, 2002). These findings suggest that antibiotics can have an effect on bacterial gene expression beyond the simple triggering of a stress response.

Isolates with mutations to the target sites of erythromycin, rifampicin and trimethoprim (ribosome, RNA polymerase, and dihydrofolate reductase respectively) exhibited severely reduced or undetectable transcription responses to those antibiotics. The same applied to *E. coli* isolates with mutations in their ribosomes or RNA polymerase rendering them resistant to rifampicin, spectinomycin, streptomycin, and erythromycin (Yim *et al.*, 2006). This suggests that interaction with the normal target of an antibiotic plays an important role in the alteration of gene expression. However there are exceptions to this; e.g. the aminoglycoside tobramycin interacts with an aminoglycoside response regulator (*arr*) in *Pseudomonas aeruginosa* to induce biofilm formation rather than with the ribosome (Hoffman *et al.*, 2005). According to this theory therefore, antibiotics are hormetic compounds; i.e. they serve a beneficial function at low concentrations but are detrimental at higher concentrations (Calabrese, 2005). Some compounds that

were initially recognised as signalling molecules also have antimicrobial activity at higher concentrations. One example of that is 4-hydroxy-2-heptylquinoline, which is used as a signalling molecule by *P. aeruginosa* (Deziel *et al.*, 2004) but can also inhibit microbial growth (Lacey *et al.*, 1995).

Despite these alternative hypotheses concerning the function of antibiotics, there is still considerable evidence that antibiotics act antagonistically towards other microorganisms in the environment. One suggested explanation for the distribution of secondary metabolites, they are common in organisms lacking an immune system and rare in organisms possessing one, is that their production increases the organism's fitness for survival by acting as an alternative defence mechanism (Maplestone *et al.*, 1992). The failure to detect antibiotics in soil is due both to nutrient limitations placing constraints on bacterial growth and secondary metabolite production (Anukool *et al.*, 2004) and because many antibiotics strongly adsorb onto soil or clay particles making them difficult to extract and detect (Sarmah *et al.*, 2006). A study following the fate of tetracycline and tylosin added to two soil types proved that these antibiotics remain biologically active even when tightly bound to soil particles (Chander *et al.*, 2005). Antibiotic compounds are unlikely to diffuse across the soil, which can give rise to a microenvironment around the producing microbe where the concentration of the antibiotic is sufficient to inhibit growth even if the overall concentration in soil is extremely low. Antibiotics can therefore retain their bactericidal properties whether free or adsorbed and are thus biologically relevant in soil antagonism even when they cannot be detected.

A number of the transcriptional effects triggered by subinhibitory concentrations of antibiotics can be interpreted as an attempt by the microorganisms to ready themselves for a higher and potentially lethal amount of antibiotic. In *B. subtilis* sublethal concentrations of chloramphenicol and erythromycin (inhibitors of translation elongation) induced the synthesis of the stringently controlled ribosomal proteins and elongation factors while gentamicin (which interferes with ribosomal translation accuracy and does not affect elongation) induced the expression of ribosomal proteins and not elongation factors (Lin *et al.*, 2005). A number of antibiotics have been documented to induce the expression of heat shock proteins, such as gentamicin and other aminoglycosides in *B. subtilis* (Lin *et al.*, 2005;

Bandow *et al.*, 2003), streptomycin and puromycin in *S. pneumoniae* (Ng *et al.*, 2003), and kanamycin and 4-azaleucine in *E. coli* (Sabina *et al.*, 2003). Other antibiotics, such as erythromycin and chloramphenicol induce cold shock and repress heat shock genes in *E. coli* (Bianchi & Baneyx, 1999). The general stress sigma factor, *sigB*, is upregulated by bacitracin in *B. subtilis* (Mascher *et al.*, 2003) and β-lactamase and peptidoglycan synthesis genes are upregulated by the cell wall inhibitor imipenem in *P. aeruginosa* (Bagge *et al.*, 2004). Chloramphenicol induced the expression of the *yheIH* operon, encoding ABC transporter-like proteins, while chloramphenicol, erythromycin and gentamicin induced the expression of the *ysbAB* operon, encoding membrane proteins that inhibit the formation of murein hydrolase transport channels in the bacterial membrane of *B. subtilis*. Both these operons are hypothesised to provide antibiotic tolerance (Lin *et al.*, 2005).

The formation of biofilms in response to subinhibitory concentrations of antibiotics may also be a defence mechanism, as biofilms display a 100-1000-fold increase in the MIC for all antimicrobials tested in all species examined (Lewis, 2005). A biofilm's exctracellular matrix (ECM) limits the access of antibiotics due to them either binding to the ECM, e.g. aminoglycosides, or due to the retardation of diffusion combined with active degradation, e.g. β-lactams. Some bacterial cells enter a persister state, in which the level of their metabolic activity decreases significantly, rendering the cells more tolerant to antibiotic compounds (Lewis, 2005). The fact that subinhibitory concentrations of a wide variety of antibiotics are capable of triggering such a number of stress responses across both Gram positive and negative bacteria suggests that they have evolved to prepare the organisms in case they encounter lethal concentrations of antibiotic compounds. If inhibitory concentrations of antibiotics were not present in the natural environment then all those unrelated microorganisms would not have had a reason to evolve a response for such an eventuality. Nonetheless, this does not mean that antibiotics cannot play a dual role for both communication and antibiosis for their producers. The N-(3-oxododecanoyl)-L-homoserine lactone, which belongs to the family of N-acylhomoserine lactones, is a molecule with such a dual function. This compound functions as an autoinducer used in quorum sensing by Gram negative bacteria; however it also displays bactericidal activity against Gram positive bacteria (Kaufmann *et al.*, 2005).

Antibiotic production by biocontrol bacteria has long been implicated as an important characteristic in the control of soil-borne plant pathogens. Haas and Keel (2003) reviewed the evidence for production in situ based on chemical extraction and expression of antibiotic biosynthesis genes. There was convincing evidence for the production by pseudomonads of antifungal antibiotics including phenazines, 2,4-diacetylphloroglucinol and pyoluteorin and antibiotic-negative mutants gave reduced protection against fungi (Haas & Keel, 2003). A number of *Streptomyces* species are also used for biocontrol (Raaijmakers *et al.*, 2002) and can reduce the extent of infection caused by the potato pathogen *S. scabies.* This is due at least in part to the antibiotics produced by the suppressive strains, since *S. scabies* mutants resistant to some of these antibiotics are more able to infect potato tubers (Neeno-Eckwall *et al.*, 2001). The rhizosphere therefore is a niche for which there is convincing evidence of antibiotics acting as inhibitors against other competitors. Another niche where antibiotics have been confirmed to be present in lethal concentrations is on the bodies of fungus-farming insects. These insects create fungus gardens which they feed upon and can combat infections by parasitic fungal species in their gardens through the use of antimicrobial compounds derived from bacteria grown on specialised regions of their own bodies (Mueller & Gerardo, 2002).

However, there is also evidence that antibiotics can inhibit microbial growth in soil. The introduction of antibiotic-producing streptomycetes into agricultural soil suppressed the pathogen-related population by 85- 93% in the first and 36-44% in the second successive year (Bowers *et al.*, 1996). The fact that antibiosis plays a role in population reduction can be deduced by the fact that there was no reduction to the severity of disease in experiments where pathogens with mutations granting antibiotic resistance were used and because the population density in soil was lower when producers and pathogens were co-inoculated compared to when only disease-suppressing isolates were inoculated (Neeno-Eckwall *et al.*, 2001). Furthermore, the lack of correspondence between 16S rRNA sequence and antibiotic resistance profile suggests that the location where a streptomycete is isolated from determines which compounds it is resistant to (Davelos Baines *et al.*, 2007). This in turn implies that selective pressure applied by local antibiotics can rapidly produce antibiotic

resistance in streptomycetes, indicating that antibiotics can apply strong selective pressure to the organisms in the surrounding soil environment.

The distribution of resistance phenotypes, and by extension of antibiotic resistance genes, in natural environments may therefore be able to inform on the role of antibiotics in nature. A screen of 480 *Streptomyces* isolates (collected from various environments) against 21 antibiotics showed that all had resistance to one or more antibiotics (D'Costa *et al.*, 2006). The organisms were on average resistant to seven to eight antibiotics and resistance to every compound used in the trial was demonstrated. It was recently reported that soil bacteria were able to subsist on antibiotics and strains had levels of resistance ranging from 20-1000 mg l^{-1} of antibiotic (Dantas *et al.*, 2008). These studies prove that soil bacteria can exhibit a resistance phenotype exceeding the levels of antibiotics found in both serum of antibiotic-treated patients (MacGowan & Wise, 2001) and in antibiotic-polluted soils (Sarmah *et al.*, 2006). Large numbers of bacteria resistant to gentamicin, kanamycin, streptomycin, chloramphenicol or tetracycline have been isolated from multiple Siberian permafrost sediments dating back three million years (Mindlin *et al.*, 2008), long before humans started releasing antibiotics in the environment, therefore environmental strains displaying resistance cannot be dismissed as an anthropogenic phenomenon.

It has been suggested that the function of antibiotic resistance may be in attenuating signal intensity similar to quorum quenching in the quorum sensing communication (Aminov, 2009) and that they thus modulate the signalling activity of small molecules in nature (Yim *et al.*, 2007). However the hypothesis that antibiotic resistance genes are meant to modulate the effects of subinhibitory concentrations of antibiotics fails to account for the fact that genes providing resistance to extremely high levels of antibiotic are recovered from natural isolates. The hypothesis that only subinhibitory concentrations of antibiotics are common in soil does not explain why all the streptomycetes tested by D'Costa et al. (2006) demonstrated high level resistance to multiple antibiotics. Furthermore, high level resistance is not found solely in producers which would need it due to the fact that they are by necessity exposing themselves to high concentrations of antibiotic; but also in non-producers as the current study found two clades within the *S. griseus* branch that demonstrated

high level viomycin resistance despite lacking the genes necessary to manufacture the compound. The existence of antibiotic resistance genes that can provide high level resistance in non-producing soil bacteria therefore indicates that there is selective pressure in the environment for the development of resistance to inhibitory concentrations of antibiotics. This in turn implies that high levels of antibiotic can be found in certain microenvironments and that antibiotics do play a role as inhibitors of microbial growth in the environment.

The low level resistance to streptomycin of the majority of the group B strains noted in the current study may also indicate that they are generally exposed to low levels of antibiotic in their environment and that the *strA* gene might serve as a modulator. One problem with this hypothesis however is that resistance genes not associated with a biosynthetic cluster from both this and other studies (Kim *et al.*, 2006) are constitutively expressed, which could reduce their utility as modulators. The low level resistance might not be the result of streptomycin being present solely at low levels in the environment; it could also be the result of the streptomycetes possessing *strA* resistance genes reaching an optimal cost/benefit ratio. Expressing higher amounts of the *strA* gene may be more costly for the overall growth the bacterium in the environment compared to occasionally encountering sites with high levels of streptomycin and dying as a result. The results of Section 6.3.1 suggest that streptomycin producers are not in high abundance in soil, so other bacteria are less likely to encounter them. A low level streptomycin resistance phenotype may therefore be a compromise between the metabolic cost of resistance and the need for it in the environment.

8.2 Origin of antibiotic resistance genes and biosynthetic clusters

Another important topic with medical implications is the question of where antibiotic resistance genes originated. Antibiotic resistance genes are ancient; it is estimated that β-lactamases originated more than 2 billion years ago and some serine β-lactamases have been present on plasmids for millions of years (Aminov, 2009). It has been suggested that antibiotics acted as co-factors or effectors in condensation reactions, such as transcription and translation, before polypeptides assumed that role and that they can interfere with these processes today because the sites they bind to have been conserved from that period (Davies, 1990). This hypothesis may

explain why antibiotic resistance appeared this early in evolutionary history. As mentioned in the introduction, resistance can arise either due to mutations or due to the acquisition of genes from HGT events. The former has been observed in *Mycobacterium*; the latter however is much more common. The two main sources for antibiotic resistance genes are housekeeping genes and self-protection genes from antibiotic-producing microbes (Mazel & Davies, 1999). An enzyme that targets a metabolite with a structure similar to a given antibiotic might affect the antibiotic itself and confer resistance. An example of that is the *aac(2')-Ia* gene, encoding a gentamicin 2'-N-acetyltransferase, from *Providencia stuartii* (Rather *et al.*, 1993). The enzyme's normal function is to modify peptidoglycan, however it can inactivate gentamicin due to structural similarities between the two compounds (Payie & Clarke, 1997). Multidrug efflux pumps are another example, as they are highly conserved, present in all organisms and their primary functions are detoxification, virulence, homeostasis and intercellular signal trafficking (Martinez *et al.*, 2009b). Metabolic genes providing resistance can easily undergo gene duplication (Sandegren & Andersson, 2009) and some of these duplicates could well become specialised in targeting antibiotics. An example of this phenomenon are β-lactamases, whose structure suggests they evolved from penicillin-binding proteins whose function is the synthesis of peptidoglycan (Massova & Mobashery, 1998).

One well-studied example of HGT of resistance genes from producers to pathogens is the *vanHAX* operon, which encodes enzymes that synthesise peptidoglycan terminating in D-alanyl-D-lactate instead of D-alanyl-D-alanine. This results in a 1000-fold reduction in glycopeptide binding, rendering the organisms possessing it resistant to vancomycin and teicoplanin (Leclercq & Courvalin, 1997). The *vanHAX* operon present in *Enterococcus* species has homologues in the vancomycin producer *Amycolatopsis orientalis* and the A47934 producer *Streptomyces toyocaensis*. The enterococci and the producers demonstrate 54 to 61% amino acid sequence identity for VanH, 59 to 63% for VanA, and 61 to 64% for VanX and the orientations of the genes on the *vanHAX* operon are identical in both the enterococci and the producers (Marshall *et al.*, 1998). Another clear example is that homologues to the *otrA* and *otrB* tetracycline resistance genes present in tetracycline producer *S. rimosus* have been found in *Mycobacterium fortuitum* (Pang *et al.*, 1994). OtrA confers resistance by ribosomal protection and OtrB by reduced accumulation (Schnappinger & Hillen,

1996). Since streptomycete plasmids can be acquired by *Mycobacterium* species via natural transformation (Bhatt *et al.*, 2002), it is likely that they were transferred via this route.

The fact that other organisms can acquire resistance genes from producers in turn raises the question of where the resistance genes present in antibiotic biosynthetic clusters and those clusters originated from. The hypothesis that antibiotics had a prebiotic function might explain why certain cellular processes are susceptible to antibiotics; however it does not explain how antibiotic biosynthesis evolved, as the mechanisms involved in their production today require polypeptides and thus could have only evolved after antibiotics had stopped serving their suggested archaic functions as effector molecules (Stone & Williams, 1992). Secondary metabolic pathways, including those responsible for antibiotic biosynthesis, are thought to have arisen from primary metabolism.

One hypothesis for their origin is that they arose in a stepwise manner from pre-existing primary metabolic pathways. Secondary pathways could have begun as the result of duplication and subsequent mutation of a primary metabolic enzyme that generated an abnormal product. If not immediately beneficial, this product might have become advantageous after modification by other enzymes with low substrate specificity. This primitive pathway could have undergone gradual improvement to further increase the efficiency of the end product and thus ensured maintenance of the pathway (Vining, 1992). Possible examples of that include the pathways involved in polyketide and β-lactam biosynthesis, which show similarities to the general pathway for fatty acid biosynthesis, and the early steps in chloramphenicol synthesis, which is similar to the pathways for the biosynthesis of a number of amino acids (Ridley *et al.*, 2008; Vining, 1992). The efflux systems that secrete antibiotics in the environment are also derived from primary metabolism, such as ABC transporters (Martin *et al.*, 2005). It has been suggested that secondary metabolism genes formed clusters because coordinated regulation of biosynthetic genes could be achieved by organising the genes into large operons controlled by a single promoter (Brakhage *et al.*, 2005). The generation of novel and diverse compounds must have been assisted by the fact that enzymes involved in secondary metabolism tend to have wide substrate specificity, unlike primary metabolism

enzymes (Firn & Jones, 2000). Polyketide synthases for example have low substrate specificities, which allows them both to accept a variety of substrates and to tolerate the omission or substitution of previous enzymatic modification steps, allowing the generation of a considerable diversity of metabolites (Byford *et al.*, 1997). Some of the enzymes involved in astaxanthin biosynthesis in *Agrobacterium aurantiacum* also have low substrate specificity and are capable of generating a diverse assortment of carotenoids (Misawa *et al.*, 1995).

HGT has played a very important role in the evolution of secondary metabolism. Biosynthetic gene clusters are transmitted by selfish genetic elements such as pathogenicity islands and plasmids. Restrictions on HGT can be imposed by the host range of the vectors carrying the cluster and the expression of some clusters may be impossible if the new host lacks the metabolites necessary for the pathway; e.g. propionate is necessary for the biosynthesis of polyketides such as erythromycin (Fischbach *et al.*, 2008). Clusters can be transported between very distantly related groups; the polyketide synthases found in ascomycete fungi arrived in that phylum via HGT from actinobacteria (Schmitt & Lumbsch, 2009). Clusters synthesising β-lactams are found in fungi, Gram negative and Gram positive bacteria and it is thought that they all evolved from one ancestral cluster which either moved to fungi from bacteria once 370 million years ago or multiple times more recently. The bacterial origin of the clusters is supported by the absence of introns in the majority of the biosynthetic genes in fungi (Liras & Martin, 2006). Another example is the antibiotic andrimid which has been isolated from a marine *Vibrio cholera* strain, a sponge-associated marine *Pseudomonas fluorescens* strain, an endosymbiotic *Enterobacter* strain and a free-living terrestrial *Pantoea agglomerans* strain. The *P. agglomerans* cluster is flanked by a pseudogene resembling a transposase, indicating that it has been mobilised via a transposon (Fischbach *et al.*, 2008). HGT has not only allowed the spread of clusters, but also their diversification as highly diverse clusters were able to arrive in the same host and fuse or exchange genes, thus allowing a great increase in the diversity of antibiotics and secondary metabolites in general.

Antibiotic clusters can evolve in a number of ways. One is for them to merge and form superclusters. These clusters can be co-ordinately regulated or even merge

their biosynthetic metabolites to produce novel compounds (Fischbach, 2009). An example of that is simocyclinone from *S. antibioticus*, which is composed of an aminocoumarin moiety conjugated to an anthracycline moiety. The enzymes involved in aminocoumarin biosynthesis show homology to the enzymes that synthesise coumermycin in *S. rishiriensis* and novobiocin in *S. spheroides* (Trefzer *et al.*, 2002). The genes involved in the biosynthesis of the anthracycline moiety are homologous to those of the landomycin, from *S. cyanogenus*, and urdamycin, from *S. fradiae*, biosynthetic gene clusters (Galm *et al.*, 2002). Albomycin from *S. subtropicus* is another example, as its structure is that of a hydroxamate siderophore conjugated to a nucleoside antibiotic. This structure allows the antibiotic to readily enter bacterial cells via active outer membrane transporters that bind to siderophores (Ferguson *et al.*, 2000).

A second way for a biosynthetic cluster to evolve is via gene duplication, the insertion of new genes or the replacement of previous genes (Fischbach *et al.*, 2008). The duplication of two genes in a lantibiotic gene cluster from *Bacillus subtilis* A1/3 has resulted in the synthesis of two distinct but structurally similar antibiotics (Stein *et al.*, 2002). Polyketide biosynthetic pathways can exchange or add novel enzymes which enables the generation of new compounds (Ridley *et al.*, 2008). A study on the composition of glycopeptide clusters from four actinobacterial families found they had a mosaic structure, indicating that these clusters consist of distinct gene cassettes specifying enzymes that participate in subpathways which can be replaced by alternate gene cassettes (Donadio *et al.*, 2005). The diversity of both the teicoplanin, a family of non-ribosomal peptide synthetase compounds, and the aminoglycoside antibiotics can be attributed to tailoring enzymes, enzymes that modify the basic scaffold of the compound. In both instances, the presence of different tailoring enzymes in a cluster can result in the production of variant compounds (Fischbach *et al.*, 2008).

Finally, point mutations can also play an important role in the evolution of antibiotic production; e.g. single nucleotide polymorphisms in the active sites of terpene cyclases or polyketide synthases can result in the generation of different end products (Fischbach *et al.*, 2008).

Like other antibiotic clusters, the streptomycin gene cluster is ancient, having originated at least 610 million years ago (Wright, 2007). The synthesis of streptomycin requires *myo*-Inositol 1-phosphate (MIP), which is also used for the generation of compounds involved in signal transduction pathways, stress response, and cell wall biogenesis (Flatt & Mahmud, 2007). The streptomycin biosynthetic pathway is therefore likely to have originated from enzymes involved in the modification of MIP. The fact that there are three subpathways involved in the biosynthesis of streptomycin suggests that the cluster originated from the fusion of three smaller clusters into one which led to the formation of a composite molecule. The distribution of streptomycin and related aminoglycoside clusters on the concatenated housekeeping phylogenenetic tree in this study strongly indicates that the cluster has undergone HGT, as the alternative would require the independent loss of the cluster in a large number of clades. Furthermore, the fact that the cluster from *S. platensis* CR50 is almost identical in structure to the *S. griseus* DSM 40236^{T} cluster when strains more closely related to *S. griseus*, such as *S. humidus* or *S. glaucescens*, are considerably more divergent is also indicative of HGT. The discovery of transposase remnants flanking both sides of the CR50 cluster is strong evidence that the streptomycin gene cluster has spread via a transposon, as have other clusters such as the one involved in andrimid biosynthesis. The absence of transposon remnants in *S. griseus* is probably indicative of a more ancient transposition event, which has allowed the transposon sequences to be completely eliminated. The fact that the streptomycin cluster in *S. griseus* is located between two genes that neighbour one another in the *S. coelicolor* A3(2) genome is strongly indicative of an insertion event (Tomono *et al.*, 2005). The deletion of *strB2* in CR50 is an example of the streptomycin gene cluster evolving via the loss of one of its tailoring enzymes.

The *strA* resistance-only gene could have theoretically either evolved from a primary metabolism gene, as the structure of phosphotransferases is highly conserved (Figure 8.1), or originated from an aminoglycoside cluster. The Group B resistance-only genes all cluster in the same clade as *strA* genes from aminoglycoside clusters when placed in a phylogenetic tree containing more distantly related sequences (Figure 6.1). The most plausible explanation for their origin is that they are derived from a resistance gene that was involved in self-protection rather than from a

phosphotranspherase that used to serve a different purpose but also was capable of antibiotic inactivation in a manner similar to the *aac(2')-Ia* gene whose primary purpose is the modification of peptidoglycan. This raises the question of which aminoglycoside cluster the Group B *strA* sequences originated from. The gene is unlikely to have originated from the *strA* of a streptomycin gene cluster present in the last common ancestor of the producers and resistance-only strains that lost its biosynthetic genes. This is due to both its high sequence divergence, the housekeeping genes have a 97% sequence identity and *strA* has 75%, and its different position on the chromosome. While one could posit that the *strA* has diversified more rapidly and also undergone recombination to arrive at a different location, the high sequence divergence suggests that the *strA* originated from elsewhere and was retained by the close relatives of streptomycin producers because it granted them a selective advantage. The fact that *strA* is at a different chromosomal position is not a very strong piece of evidence, as the GARD algorithm reveals that there has been a recombination event within the sequence of *strA*, indicating both that *strA* has a mosaic structure in some of the examined strains and that it can undergo recombination which would explain a different chromosomal position. It is impossible to know whether the resistance-only *strA* was transferred to the Group B *S. griseus* strains alone, or together with a cluster that decayed after the transfer or alone but the latter seems more parsimonious. The *SGR172* gene, a homologue of which flanks *strA* in the AR23 Group B strain, is classified as a putative acetyltransferase, indicating that it might also be involved in resistance. If that is the case, then *SGR172* and *strA* could have been transported together as a resistance operon.

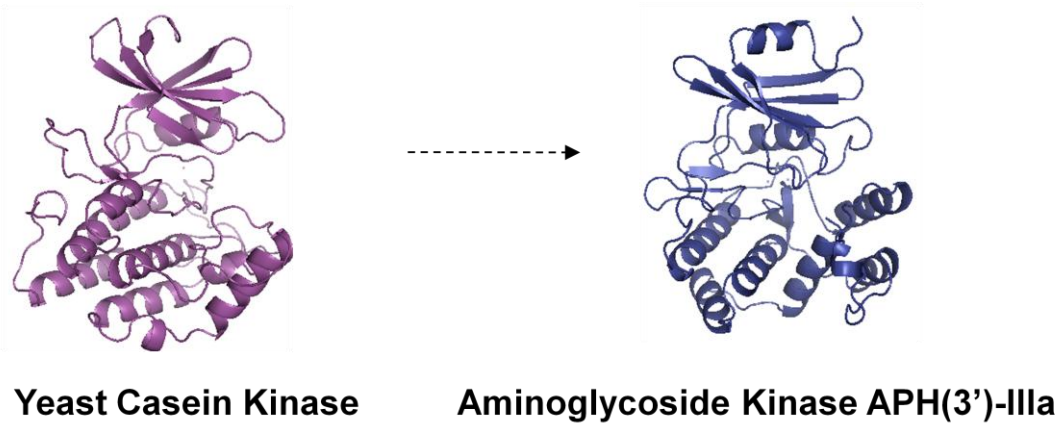


Figure 8.1. Structural similarity between a eukaryotic kinase and an aminoglycoside phosphotransferase (Wright, pers. comm.)

The low level of expression of *strA* in resistance only strains suggests that a reduced amount of enzyme is responsible for the low resistance of these strains. This is supported by the fact that the active sites in resistance-only StrA are all conserved. It should be noted though that the *strA* homologue from the *S. glaucescens* hydroxystreptomycin cluster also provides low streptomycin resistance and has conserved active sites despite the fact that it must presumably be highly expressed in order to protect *S. glaucescens* from the hydroxystreptomycin that it produces. Therefore alterations in locations other than the active sites might also influence the specificity of the enzyme. This would not be surprising, as single amino acid changes can suffice to cause large variations in the substrate specificity of aminoglycoside acetyltransferases (Rather *et al.*, 1992).

The viomycin gene cluster has spread vertically within the clade containing it and no other branches possess a homologue, indicating that if it has undergone HGT to the other strains examined it was not retained by them. The presence of the *vioS* gene in the viomycin cluster, which was found in this study to be homologous to *strK* from the streptomycin and four other aminoglycoside clusters, indicates that this biosynthetic gene was transferred from an aminoglycoside cluster into a peptide cluster. This is an example of the creation of novel compounds via the incorporation of genes from different clusters.

The *S. vinaceus vph* gene has 76% sequence identity to the *S. griseus SGR421* and 79% to its *S. bacillaris* homologue. The housekeeping genes demonstrate a 96% sequence homology between *S. vinaceus* and *S. griseus* or *S. bacillaris*. The high

divergence between sequence homologies of *vph* and the housekeepers might be due to the fact that the latter are highly conserved, but as with *strA* it seems more probable that *SGR421* in *S. griseus* and *S. bacillaris* originated from another unknown peptide cluster. The fact that *S. griseus* DSM 40236^{T} is highly resistant to viomycin despite the fact that *SGR421* is expressed at a low level might indicate that SGR421 is a more efficient enzyme than StrA or that *S. griseus* relies on multiple resistance mechanisms.

8.3 Future work

Chapter 3: No attempts were made to isolate strains resistant to viomycin from soil. As a result, this study did not provide as much information about the frequency and distribution of viomycin resistance genes and biosynthetic cluster as it did for streptomycin. This could be rectified with the screening of soil isolates on viomycin-selective plates. The *rsmG* and full 16S genes could be amplified and sequenced in order to determine whether they have mutations providing resistance to some of the isolates.

Chapter 4: The flanking regions of representatives of the Group B1 *strA* genes, such as the one found in strain E1013, were not examined. It would be interesting to examine whether Group B1 and B2 *strA* genes are located on the same chromosomal position and flanked by the same genes. The flanking regions of *strA* in strains sensitive to streptomycin would also be interesting to examine in order to determine whether there are alterations in the promoter regions. The genomic area containing *SGR421* in *S. bacillaris* would also be worth exploring to determine whether it is located in the same position as in *S. griseus* and how it may be regulated.

Chapter 5: Any new viomycin-resistant soil isolates would need to have their 16S and the eight other housekeeping genes amplified and sequenced to determine their phylogenetic interrelationships with other studied strains.

Chapter 6: The number of *strA* and *atpD* gene sequences examined from DNA extracted from soil was relatively small due to the cost of Sanger sequencing. As a result it probably failed to capture the full diversity of these genes in the soils that were studied. Ideally the PCR products from *strA*, *atpD* and also *vph* should be subjected to 454 pyrosequencing. This would generate thousands of reads from each

site and thus provide much better resolution of the phylogenetic diversity at these sites.

Chapter 7: An attempt to detect the A-factor cascade genes using Southern hybridisation would be useful in order to confirm the results obtained via PCR. It would also be of interest to examine the expression profiles of *strA* from some of the streptomycin-sensitive strains in order to determine whether it is silent in those strains. The expression of the *vph* resistance and the viomycin biosynthetic genes from *S. vinaceus* could also be examined in order to determine whether it has a similar expression pattern to the streptomycin cluster genes. The expression of *SGR421* from *S. bacillaris* could also be explored in order to see whether it was constitutive as in *S. griseus*. Furthermore, the expression of all these genes could be studied in soil microcosms in order to determine whether it was the same as in liquid culture. Finally, knocking out *strA* in Group B strains and *SGR421* in *S. griseus* and *S. bacillaris* would be a way of confirming that these genes are responsible for the resistance phenotype.

8.4 Conclusion

It was traditionally thought that if two species occupied the same niche then the inferior competitor would be excluded. Though there are doubts about how applicable the competitive exclusion principle is due to the complexity of natural environments and organismal interactions, it has been demonstrated that closely related species are more likely to share the same niche than unrelated species (Denboer, 1986), though this is not always the case (Losos, 2008). Examples of competitive exclusion between closely related species have been identified in a number of groups of organisms, including tracheophytes (Verdu & Pausas, 2007; Davies, 2006b; Ackerly *et al.*, 2006), mammals (Cooper *et al.*, 2008), birds (Lovette & Hochachka, 2006) and fish (Helmus *et al.*, 2007). Since closely related organisms tend to have similar morphology and genetics, it follows that the same type of niche will be optimal for all of them.

As the biosynthesis of secondary metabolites carries a metabolic cost, it can be assumed that they grant a selective advantage to the organisms producing them. The use of antibiotics to eliminate niche competitors would not be surprising, as

interference competition, which involves direct negative interactions between organisms, is utilised by 38% of competing species examined and is more common in microhabitats (Schoener, 1983). An example of the elimination of niche competitors by bacteria is *Streptococcus pneumonia*, which releases hydrogen peroxide to activate the SOS response in neighbouring *Staphylococcus aureus* bacteria. The SOS response causes the activation of prophages, which the majority of *S. aureus* strains possess, causing the competitors of *S. pneumoniae* to lyse (Selva *et al.*, 2009). However, the assumption that closely related organisms will exploit the same niche does not always apply, especially with bacteria which can easily exploit new niches by acquiring novel genetic material. *Bacillus anthracis* for example is extremely similar in both physiology and genetics to *B. cereus*; however the possession of two virulence plasmids and a mutation in their regulator allows *B. anthracis* to exploit a niche that is unavailable to *B. cereus* (Kolsto *et al.*, 2009). *Streptomyces* species can easily acquire novel secondary metabolite clusters that can allow them to exploit novel niches, one example being the pathogenicity island that allows *S. turgidiscabies* to colonise plant roots (Loria *et al.*, 2006). As a result, while it is probable that two closely related bacterial strains will share the same niche it is not a certainty.

In *S. griseus* the streptomycin resistance-only strains are closely related to producers, demonstrating a 97% sequence identity in their housekeeping genes and forming sister clades. The viomycin resistance-only strains were slightly more distantly related but still demonstrated a sequence identity of 96% and belonged in sister clades. Furthermore, the number of resistant strains was significantly higher in the resistance-only clades compared to other more distantly related strains. The suggestion that the evolution of resistance is due to competition for the same niche agrees with the previous observations in other organisms where closely related organisms also compete. The fact that not all these closely related strains are resistant may be attributed to them having adapted to a slightly different niche or to the heterogeneous nature of soil, as soil contains a large number of distinct microhabitats rather than forming a continuous environment (Ettema & Wardle, 2002). The division of a habitat into numerous subsections can reduce competitive inhibition (Ettema & Wardle, 2002; Wang *et al.*, 2000), which might explain why not all the close relatives of antibiotic producers are under selection for retention of

antibiotic resistance. Nonetheless, a significant number of them have resistance, which supports the hypothesis that antibiotics are used in interference competition and thus closely related organisms have coevolved resistance mechanisms to these compounds.

9 Bibliography

Acinas, S.G., Marcelino, L.A., Klepac-Ceraj, V. & Polz, M.F. (2004) Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. *J Bacteriol* **186**(9), 2629-35.

Ackerly, D.D., Schwilk, D.W. & Webb, C.O. (2006) Niche evolution and adaptive radiation: testing the order of trait divergence. *Ecology* **87**(7 Suppl), S50-61.

Alliance, T. (2008) Streptomycin. *Tuberculosis (Edinb)* **88**(2), 162-3.

Aminov, R.I. (2009) The role of antibiotics and antibiotic resistance in nature. *Environ Microbiol* **11**(12), 2970-88.

Anderson, A.S., Clark, D.J., Gibbons, P.H. & Sigmund, J.M. (2002) The detection of diverse aminoglycoside phosphotransferases within natural populations of actinomycetes. *J Ind Microbiol Biotechnol* **29**(2), 60-9.

Anderson, A.S. & Wellington, E.M. (2001) The taxonomy of *Streptomyces* and related genera. *Int J Syst Evol Microbiol* **51**(Pt 3), 797-814.

Antony-Babu, S., Stach, J.E. & Goodfellow, M. (2008) Genetic and phenotypic evidence for *Streptomyces griseus* ecovars isolated from a beach and dune sand system. *Antonie van Leeuwenhoek* **94**(1), 63-74.

Antunes, P., Machado, J. & Peixe, L. (2006) Characterization of antimicrobial resistance and class 1 and 2 integrons in *Salmonella enterica* isolates from different sources in Portugal. *J Antimicrob Chemother* **58**(2), 297-304.

Anukool, U., Gaze, W.H. & Wellington, E.M. (2004) *In situ* monitoring of streptothricin production by *Streptomyces rochei* F20 in soil and rhizosphere. *Appl Environ Microbiol* **70**(9), 5222-8.

Aoki, R., Nagaya, A., Arakawa, S., Kato, C. & Tamegai, H. (2008) Identification and diversity of putative aminoglycoside-biosynthetic aminotransferase genes from deep-sea environmental DNA. *Biosci Biotechnol Biochem* **72**(5), 1388-93.

Aumen, N.G. (1980) Microbial succession on a chitinous substrate in a woodland stream. *Microbial Ecology* **6**(4), 317-327.

Babalola, O.O., Kirby, B.M., Le Roes-Hill, M., Cook, A.E., Cary, S.C., Burton, S.G. & Cowan, D.A. (2009) Phylogenetic analysis of actinobacterial populations associated with Antarctic Dry Valley mineral soils. *Environ Microbiol* **11**(3), 566-76.

Bagge, N., Schuster, M., Hentzer, M., Ciofu, O., Givskov, M., Greenberg, E.P. & Hoiby, N. (2004) *Pseudomonas aeruginosa* biofilms exposed to imipenem exhibit changes in global gene expression and beta-lactamase and alginate production. *Antimicrob Agents Chemother* **48**(4), 1175-87.

Baltz, R.H. (2005) Antibiotic discovery from actinomycetes: will a renaissance follow the decline and fall. *SIM News* **55**, 186-196.

Baltz, R.H. (2006) Marcel Faber Roundtable: is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? *J Ind Microbiol Biotechnol* **33**(7), 507-13.

Baltz, R.H. (2008) Renaissance in antibacterial discovery from actinomycetes. *Curr Opin Pharmacol* **8**(5), 557-63.

Bandow, J.E., Brotz, H., Leichert, L.I., Labischinski, H. & Hecker, M. (2003) Proteomic approach to understanding antibiotic action. *Antimicrob Agents Chemother* **47**(3), 948-55.

Barkei, J.J., Kevany, B.M., Felnagle, E.A. & Thomas, M.G. (2009) Investigations into viomycin biosynthesis by using heterologous production in *Streptomyces lividans*. *Chembiochem* **10**(2), 366-76.

Begg, E.J. & Barclay, M.L. (1995) Aminoglycosides--50 years on. *Br J Clin Pharmacol* **39**(6), 597-603.

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., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A.,

Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J. & Hopwood, D.A. (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**(6885), 141-7.

Benveniste, R. & Davies, J. (1973) Aminoglycoside antibiotic-inactivating enzymes in actinomycetes similar to those present in clinical isolates of antibiotic-resistant bacteria. *Proc Natl Acad Sci U S A* **70**(8), 2276-80.

Berg, G. (2000) Diversity of antifungal and plant-associated *Serratia plymuthica* strains. *J Appl Microbiol* **88**(6), 952-60.

Bhatt, A., Kieser, H.M., Melton, R.E. & Kieser, T. (2002) Plasmid transfer from *Streptomyces* to *Mycobacterium smegmatis* by spontaneous transformation. *Mol Microbiol* **43**(1), 135-46.

Bianchi, A.A. & Baneyx, F. (1999) Stress responses as a tool to detect and characterize the mode of action of antibacterial agents. *Appl Environ Microbiol* **65**(11), 5023-7.

Borjesson, S., Dienues, O., Jarnheimer, P.A., Olsen, B., Matussek, A. & Lindgren, P.E. (2009) Quantification of genes encoding resistance to aminoglycosides, beta-lactams and tetracyclines in wastewater environments by real-time PCR. *Int J Environ Health Res*, 1-12.

Bowers, J.H., Kinkel, L.L. & Jones, R.K. (1996) Influence of disease-suppressive strains of *Streptomyces* on the native *Streptomyces* community in soil as determined by the analysis of cellular fatty acids. *Can J Microbiol* **42**(1), 27-37.

Brakhage, A.A., Al-Abdallah, Q., Tuncher, A. & Sprote, P. (2005) Evolution of beta-lactam biosynthesis genes and recruitment of trans-acting factors. *Phytochemistry* **66**(11), 1200-10.

Bridge, P.D. & Newsham, K.K. (2009) Soil fungal community composition at Mars Oasis, a southern maritime Antarctic site, assessed by PCR amplification and cloning. *Fungal Ecology* **2**, 66-74.

Bujnicki, J.M., Blumenthal, R.M. & Rychlewski, L. (2002) Sequence analysis and structure prediction of 23S rRNA:m1G methyltransferases reveals a conserved core augmented with a putative Zn-binding domain in the N-terminus and family-specific elaborations in the C-terminus. *J Mol Microbiol Biotechnol* **4**(1), 93-9.

Bukhalid, R.A. & Loria, R. (1997) Cloning and expression of a gene from *Streptomyces scabies* encoding a putative pathogenicity factor. *J Bacteriol* **179**(24), 7776-83.

Bukhalid, R.A., Takeuchi, T., Labeda, D. & Loria, R. (2002) Horizontal transfer of the plant virulence gene, nec1, and flanking sequences among genetically distinct Streptomyces strains in the *Diastatochromogenes* cluster. *Appl Environ Microbiol* **68**(2), 738-44.

Busse, H.J., Wostmann, C. & Bakker, E.P. (1992) The bactericidal action of streptomycin: membrane permeabilization caused by the insertion of mistranslated proteins into the cytoplasmic membrane of *Escherichia coli* and subsequent caging of the antibiotic inside the cells due to degradation of these proteins. *J Gen Microbiol* **138**(3), 551-61.

Byford, M.F., Baldwin, J.E., Shiau, C.Y. & Schofield, C.J. (1997) The mechanism of ACV synthetase. *Chem Rev* **97**(7), 2631-2650.

Calabrese, E.J. (2005) Paradigm lost, paradigm found: the re-emergence of hormesis as a fundamental dose response model in the toxicological sciences. *Environ Pollut* **138**(3), 379-411.

Carter, A.P., Clemons, W.M., Brodersen, D.E., Morgan-Warren, R.J., Wimberly, B.T. & Ramakrishnan, V. (2000) Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* **407**(6802), 340-8.

Casjens, S. (1998) The diverse and dynamic structure of bacterial genomes. *Annu Rev Genet* **32**, 339-77.

Chadwick, D. & Whelan, J. (Eds.) (1992) *Secondary metabolites: Their function and evolution.* . Chichester: John Wiley & Sons.

Challis, G.L. & Hopwood, D.A. (2003) Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species. *Proc Natl Acad Sci U S A* **100 Suppl 2**, 14555-61.

Chanal, A., Chapon, V., Benzerara, K., Barakat, M., Christen, R., Achouak, W., Barras, F. & Heulin, T. (2006) The desert of Tataouine: an extreme environment that hosts a wide diversity of microorganisms and radiotolerant bacteria. *Environ Microbiol* **8**(3), 514-25.

Chander, Y., Kumar, K., Goyal, S.M. & Gupta, S.C. (2005) Antibacterial activity of soil-bound antibiotics. *J Environ Qual* **34**(6), 1952-7.

Chater, K.F., Bruton, C.J., Plaskitt, K.A., Buttner, M.J., Mendez, C. & Helmann, J.D. (1989) The developmental fate of *S. coelicolor* hyphae depends upon a gene product homologous with the motility sigma factor of *B. subtilis*. *Cell* **59**(1), 133-43.

Chater, K.F. & Chandra, G. (2006) The evolution of development in *Streptomyces* analysed by genome comparisons. *FEMS Microbiol Rev* **30**(5), 651-72.

Chen, C.W., Huang, C.H., Lee, H.H., Tsai, H.H. & Kirby, R. (2002) Once the circle has been broken: dynamics and evolution of *Streptomyces* chromosomes. *Trends Genet* **18**(10), 522-9.

Chinault, A.C., Blakesley, V.A., Roessler, E., Willis, D.G., Smith, C.A., Cook, R.G. & Fenwick, R.G., Jr. (1986) Characterization of transferable plasmids from *Shigella flexneri* 2a that confer resistance to trimethoprim, streptomycin, and sulfonamides. *Plasmid* **15**(2), 119-31.

Choi, E.C., Misumi, M., Nishimura, T., Tanaka, N., Nomoto, S., Teshima, T. & Shiba, T. (1979) Viomycin resistance: alterations of either ribosomal subunit affect the binding of the antibiotic to the pair subunit and the entire ribosome becomes resistant to the drug. *Biochem Biophys Res Commun* **87**(3), 904-10.

Choulet, F., Gallois, A., Aigle, B., Mangenot, S., Gerbaud, C., Truong, C., Francou, F.X., Borges, F., Fourrier, C., Guerineau, M., Decaris, B., Barbe, V., Pernodet, J.L. & Leblond, P. (2006) Intraspecific variability of the terminal inverted repeats of the linear chromosome of *Streptomyces ambofaciens*. *J Bacteriol* **188**(18), 6599-610.

Coombs, J.T. & Franco, C.M. (2003) Isolation and identification of actinobacteria from surface-sterilized wheat roots. *Appl Environ Microbiol* **69**(9), 5603-8.

Cooper, N., Rodriguez, J. & Purvis, A. (2008) A common tendency for phylogenetic overdispersion in mammalian assemblages. *Proc Biol Sci* **275**(1646), 2031-7.

Corander, J. & Marttinen, P. (2006) Bayesian identification of admixture events using multilocus molecular markers. *Mol Ecol* **15**(10), 2833-43.

Costello, E.K., Halloy, S.R., Reed, S.C., Sowell, P. & Schmidt, S.K. (2009) Fumarole-supported islands of biodiversity within a hyperarid, high-elevation landscape on Socompa Volcano, Puna de Atacama, Andes. *Appl Environ Microbiol* **75**(3), 735-47.

Courvalin, P. & Fiandt, M. (1980) Aminoglycoside-modifying enzymes of *Staphylococcus aureus*; expression in *Escherichia coli*. *Gene* **9**(3-4), 247-69.

Cullen, D.W. & Lees, A.K. (2007) Detection of the *nec1* virulence gene and its correlation with pathogenicity in *Streptomyces* species on potato tubers and in soil using conventional and real-time PCR. *J Appl Microbiol* **102**(4), 1082-94.

Cundliffe, E. (1984) Self defence in antibiotic-producing organisms. *Br Med Bull* **40**(1), 61-7.

Cundliffe, E. (1989) How antibiotic-producing organisms avoid suicide. *Annu Rev Microbiol* **43**, 207-33.

D'Costa, V.M., McGrann, K.M., Hughes, D.W. & Wright, G.D. (2006) Sampling the antibiotic resistome. *Science* **311**(5759), 374-7.

Danilchanka, O., Pavlenok, M. & Niederweis, M. (2008) Role of porins for uptake of antibiotics by *Mycobacterium smegmatis*. *Antimicrob Agents Chemother* **52**(9), 3127-34.

Dantas, G., Sommer, M.O., Oluwasegun, R.D. & Church, G.M. (2008) Bacteria subsisting on antibiotics. *Science* **320**(5872), 100-3.

Daubin, V. & Lerat, E. (2006) Evolution of the core of genes. In *SGM Symposium 66* Edited by N.A. Logan, H.M. Lappin-Scott & P.C.F. Oyston. Cambridge: Cambridge University Press.

Davelos, A.L., Xiao, K., Samac, D.A., Martin, A.P. & Kinkel, L.L. (2004) Spatial variation in *Streptomyces* genetic composition and diversity in a prairie soil. *Microb Ecol* **48**(4), 601-12.

Davelos Baines, A.L., Xiao, K. & Kinkel, L.L. (2007) Lack of correspondence between genetic and phenotypic groups amongst soil-borne streptomycetes. *FEMS Microbiol Ecol* **59**(3), 564-75.

Davies, J. (1990) What are antibiotics? Archaic functions for modern activities. *Mol Microbiol* **4**(8), 1227-32.

Davies, J. (2006a) Are antibiotics naturally antibiotics? *J Ind Microbiol Biotechnol* **33**(7), 496-9.

Davies, J., Spiegelman, G.B. & Yim, G. (2006) The world of subinhibitory antibiotic concentrations. *Curr Opin Microbiol* **9**(5), 445-53.

Davies, J. & Wright, G.D. (1997) Bacterial resistance to aminoglycoside antibiotics. *Trends Microbiol* **5**(6), 234-40.

Davies, T.J. (2006b) Evolutionary ecology: when relatives cannot live together. *Curr Biol* **16**(16), R645-7.

Davis, K.E., Joseph, S.J. & Janssen, P.H. (2005) Effects of growth medium, inoculum size, and incubation time on culturability and isolation of soil bacteria. *Appl Environ Microbiol* **71**(2), 826-34.

Denboer, P.J. (1986) The present status of the competitive exclusion principle. *Trends in Ecology & Evolution* **1**(1), 25-28.

Depardieu, F., Podglajen, I., Leclercq, R., Collatz, E. & Courvalin, P. (2007) Modes and modulations of antibiotic resistance gene expression. *Clin Microbiol Rev* **20**(1), 79-114.

Desai, C., Parikh, R.Y., Vaishnav, T., Shouche, Y.S. & Madamwar, D. (2009) Tracking the influence of long-term chromium pollution on soil bacterial community structures by comparative analyses of 16S rRNA gene phylotypes. *Res Microbiol* **160**(1), 1-9.

Deziel, E., Lepine, F., Milot, S., He, J., Mindrinos, M.N., Tompkins, R.G. & Rahme, L.G. (2004) Analysis of *Pseudomonas aeruginosa* 4-hydroxy-2-alkylquinolines (HAQs) reveals a role for 4-hydroxy-2-heptylquinoline in cell-to-cell communication. *Proc Natl Acad Sci U S A* **101**(5), 1339-44.

Dhote, V., Gupta, S. & Reynolds, K.A. (2008) An O-phosphotransferase catalyzes phosphorylation of hygromycin A in the antibiotic-producing organism *Streptomyces hygroscopicus*. *Antimicrob Agents Chemother* **52**(10), 3580-8.

Diaz, I., Ehrenberg, M. & Kurland, C.G. (1986) How do combinations of rpsL- and miaA- generate streptomycin dependence? *Mol Gen Genet* **202**(2), 207-11.

Didelot, X. & Falush, D. (2007) Inference of bacterial microevolution using multilocus sequence data. *Genetics* **175**(3), 1251-66.

Dietz, A. (1986) Structure and Taxonomy of *Streptomyces*. In *The Bacteria: A Treatise on Structure and Function, Volume IX Antibiotic-Producing Streptomyces*Edited by S.W. Queener & L.E. Day. Orlando: Academic Press, Inc.

Distler, J., Braun, C., Ebert, A. & Piepersberg, W. (1987) Gene cluster for streptomycin biosynthesis in *Streptomyces griseus*: analysis of a central region including the major resistance gene. *Mol Gen Genet* **208**(1-2), 204-10.

Distler, J., Mansouri, K., Mayer, G., Stockmann, M. & Piepersberg, W. (1992) Streptomycin biosynthesis and its regulation in streptomycetes. *Gene* **115**(1-2), 105-11.

Donadio, S., Monciardini, P. & Sosio, M. (2007) Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. *Nat Prod Rep* **24**(5), 1073-109.

Donadio, S., Sosio, M., Stegmann, E., Weber, T. & Wohlleben, W. (2005) Comparative analysis and insights into the evolution of gene clusters for glycopeptide antibiotic biosynthesis. *Mol Genet Genomics* **274**(1), 40-50.

Dorak, M.T. (Ed.) (2006) *Real-time PCR.* New York: Taylor & Francis.

Drancourt, M., Bollet, C., Carlioz, A., Martelin, R., Gayral, J.P. & Raoult, D. (2000) 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. *J Clin Microbiol* **38**(10), 3623-30.

DSMZ (2009) Bacterial Nomenclature Up-to-Date, Genus: Streptomyces.

Duriez, P. & Topp, E. (2007) Temporal dynamics and impact of manure storage on antibiotic resistance patterns and population structure of *Escherichia coli* isolates from a commercial swine farm. *Appl Environ Microbiol* **73**(17), 5486-93.

Edwards, U., Rogall, T., Blocker, H., Emde, M. & Bottger, E.C. (1989) Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Res* **17**(19), 7843-53.

Emmert, E.A., Klimowicz, A.K., Thomas, M.G. & Handelsman, J. (2004) Genetics of zwittermicin a production by *Bacillus cereus*. *Appl Environ Microbiol* **70**(1), 104-13.

Ermolenko, D.N., Spiegel, P.C., Majumdar, Z.K., Hickerson, R.P., Clegg, R.M. & Noller, H.F. (2007) The antibiotic viomycin traps the ribosome in an intermediate state of translocation. *Nat Struct Mol Biol* **14**(6), 493-7.

Ettema, C.H. & Wardle, D.A. (2002) Spatial soil ecology. *Trends Ecol Evol* **17**(4), 177-183.

Fei, X., Yin, X., Zhang, L. & Zabriskie, T.M. (2007) Roles of VioG and VioQ in the incorporation and modification of the Capreomycidine residue in the peptide antibiotic viomycin. *J Nat Prod* **70**(4), 618-22.

Feil, E.J. & Spratt, B.G. (2001) Recombination and the population structures of bacterial pathogens. *Annu Rev Microbiol* **55**, 561-90.

Ferguson, A.D., Braun, V., Fiedler, H.P., Coulton, J.W., Diederichs, K. & Welte, W. (2000) Crystal structure of the antibiotic albomycin in complex with the outer membrane transporter FhuA. *Protein Sci* **9**(5), 956-63.

Firn, R.D. & Jones, C.G. (2000) The evolution of secondary metabolism - a unifying model. *Mol Microbiol* **37**(5), 989-94.

Fischbach, M.A. (2009) Antibiotics from microbes: converging to kill. *Curr Opin Microbiol* **12**(5), 520-7.

Fischbach, M.A., Walsh, C.T. & Clardy, J. (2008) The evolution of gene collectives: How natural selection drives chemical innovation. *Proc Natl Acad Sci U S A* **105**(12), 4601-8.

Flardh, K. (2003) Growth polarity and cell division in *Streptomyces*. *Curr Opin Microbiol* **6**(6), 564-71.

Flardh, K. & Buttner, M.J. (2009) *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat Rev Microbiol* **7**(1), 36-49.

Flatt, P.M. & Mahmud, T. (2007) Biosynthesis of aminocyclitol-aminoglycoside antibiotics and related compounds. *Nat Prod Rep* **24**(2), 358-92.

Fling, M.E., Kopf, J. & Richards, C. (1985) Nucleotide sequence of the transposon Tn7 gene encoding an aminoglycoside-modifying enzyme, 3"(9)-O-nucleotidyltransferase. *Nucleic Acids Res* **13**(19), 7095-106.

Frattali, A.L., Flynn, M.K., De Stasio, E.A. & Dahlberg, A.E. (1990) Effects of mutagenesis of C912 in the streptomycin binding region of *Escherichia coli* 16S ribosomal RNA. *Biochim Biophys Acta* **1050**(1-3), 27-33.

Gadagkar, S.R., Rosenberg, M.S. & Kumar, S. (2005) Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. *J Exp Zoolog B Mol Dev Evol* **304**(1), 64-74.

Gallo, M.A., Ward, J. & Hutchinson, C.R. (1996) The dnrM gene in *Streptomyces peucetius* contains a naturally occurring frameshift mutation that is suppressed by another locus outside of the daunorubicin-production gene cluster. *Microbiology* **142 (Pt 2)**, 269-75.

Galm, U., Schimana, J., Fiedler, H.P., Schmidt, J., Li, S.M. & Heide, L. (2002) Cloning and analysis of the simocyclinone biosynthetic gene cluster of Streptomyces antibioticus Tu 6040. *Arch Microbiol* **178**(2), 102-14.

Gao, B. & Gupta, R.S. (2005) Conserved indels in protein sequences that are characteristic of the phylum Actinobacteria. *Int J Syst Evol Microbiol* **55**(Pt 6), 2401-12.

Gao, B., Paramanathan, R. & Gupta, R.S. (2006) Signature proteins that are distinctive characteristics of Actinobacteria and their subgroups. *Antonie van Leeuwenhoek* **90**(1), 69-91.

Garcia, C. & Hernandez, T. (2004) Site Description. Available from: http://www.soil-index.com/english/documents/dl/index_site_description.pdf

Gaspari, F., Paitan, Y., Mainini, M., Losi, D., Ron, E.Z. & Marinelli, F. (2005) Myxobacteria isolated in Israel as potential source of new anti-infectives. *J Appl Microbiol* **98**(2), 429-39.

Ge, Y., Chen, C., Xu, Z., Eldridge, S.M., Chan, K.Y., He, Y. & He, J.Z. (2009) Carbon/nitrogen ratio as a major factor for predicting the effects of organic wastes on soil bacterial communities assessed by DNA-based molecular techniques. *Environ Sci Pollut Res Int*.

Gerth, K., Irschik, H., Reichenbach, H. & Trowitzsch, W. (1982) The myxovirescins, a family of antibiotics from *Myxococcus virescens* (Myxobacterales). *J Antibiot (Tokyo)* **35**(11), 1454-9.

Gerth, K. & Muller, R. (2005) Moderately thermophilic myxobacteria: novel potential for the production of natural products isolation and characterization. *Environ Microbiol* **7**(6), 874-80.

Gerth, K., Pradella, S., Perlova, O., Beyer, S. & Muller, R. (2003) Myxobacteria: proficient producers of novel natural products with various biological activities--past and future biotechnological aspects with the focus on the genus *Sorangium*. *J Biotechnol* **106**(2-3), 233-53.

Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., Stackebrandt, E., Van de Peer, Y., Vandamme, P., Thompson, F.L. & Swings, J. (2005) Opinion: Re-evaluating prokaryotic species. *Nat Rev Microbiol* **3**(9), 733-9.

Gharaibeh, R., Saadoun, I. & Mahasneh, A. (2003) Evaluation of combined 16S rDNA and strb1 gene targeted PCR to identify and detect streptomycin-producing *Streptomyces*. *J Basic Microbiol* **43**(4), 301-11.

Goh, E.B., Yim, G., Tsui, W., McClure, J., Surette, M.G. & Davies, J. (2002) Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics. *Proc Natl Acad Sci U S A* **99**(26), 17025-30.

Gontcharov, A.A., Marin, B. & Melkonian, M. (2004) Are combined analyses better than single gene phylogenies? A case study using SSU rDNA and rbcL sequence comparisons in the Zygnematophyceae (Streptophyta). *Mol Biol Evol* **21**(3), 612-24.

Goodfellow, M., Kumar, Y., Labeda, D.P. & Sembiring, L. (2007) The *Streptomyces violaceusniger* clade: a home for streptomycetes with rugose ornamented spores. *Antonie van Leeuwenhoek* **92**(2), 173-99.

Goodfellow, M. & Williams, S.T. (1983) Ecology of actinomycetes. *Annu Rev Microbiol* **37**, 189-216.

Gordon, R.E. & Horan, A.C. (1968) A piecemeal description of *Streptomyces griseus* (Krainsky) Waksman and Henrici. *J Gen Microbiol* **50**(2), 223-33.

Gregory, S.T. & Dahlberg, A.E. (2009) Genetic and structural analysis of base substitutions in the central pseudoknot of *Thermus thermophilus* 16S ribosomal RNA. *Rna* **15**(2), 215-23.

Gregory, S.T., Demirci, H., Belardinelli, R., Monshupanee, T., Gualerzi, C., Dahlberg, A.E. & Jogl, G. (2009) Structural and functional studies of the *Thermus thermophilus* 16S rRNA methyltransferase RsmG. *Rna* **15**(9), 1693-704.

Gundlapally, S.R. & Garcia-Pichel, F. (2006) The community and phylogenetic diversity of biological soil crusts in the Colorado Plateau studied by molecular fingerprinting and intensive cultivation. *Microb Ecol* **52**(2), 345-57.

Guo, Y., Zheng, W., Rong, X. & Huang, Y. (2008) A multilocus phylogeny of the *Streptomyces griseus* 16S rRNA gene clade: use of multilocus sequence analysis for streptomycete systematics. *Int J Syst Evol Microbiol* **58**(Pt 1), 149-59.

Gustafsson, C. & Persson, B.C. (1998) Identification of the *rrmA* gene encoding the 23S rRNA m1G745 methyltransferase in *Escherichia coli* and characterization of an m1G745-deficient mutant. *J Bacteriol* **180**(2), 359-65.

Haas, D. & Keel, C. (2003) Regulation of antibiotic production in root-colonizing *Peudomonas spp.* and relevance for biological control of plant disease. *Annu Rev Phytopathol* **41**, 117-53.

Hahn, M.W. & Pockl, M. (2005) Ecotypes of planktonic actinobacteria with identical 16S rRNA genes adapted to thermal niches in temperate, subtropical, and tropical freshwater habitats. *Appl Environ Microbiol* **71**(2), 766-73.

Hain, T., Ward-Rainey, N., Kroppenstedt, R.M., Stackebrandt, E. & Rainey, F.A. (1997) Discrimination of *Streptomyces albidoflavus* strains based on the size and number of 16S-23S ribosomal DNA intergenic spacers. *Int J Syst Bacteriol* **47**(1), 202-6.

Hanage, W.P., Fraser, C. & Spratt, B.G. (2005) Fuzzy species among recombinogenic bacteria. *BMC Biol* **3**, 6.

Hancock, R.E. (1984) Alterations in outer membrane permeability. *Annu Rev Microbiol* **38**, 237-64.

Hartmann, M., Lee, S., Hallam, S.J. & Mohn, W.W. (2009) Bacterial, archaeal and eukaryal community structures throughout soil horizons of harvested and naturally disturbed forest stands. *Environ Microbiol* **11**(12), 3045-62.

Hawkey, P.M. (2008) Molecular epidemiology of clinically significant antibiotic resistance genes. *Br J Pharmacol* **153 Suppl 1**, S406-13.

Heinzel, P., Werbitzky, O., Distler, J. & Piepersberg, W. (1988) A second streptomycin resistance gene from *Streptomyces griseus* codes for streptomycin-3"-phosphotransferase. Relationships between antibiotic and protein kinases. *Arch Microbiol* **150**(2), 184-92.

Helmus, M.R., Savage, K., Diebel, M.W., Maxted, J.T. & Ives, A.R. (2007) Separating the determinants of phylogenetic community structure. *Ecol Lett* **10**(10), 917-25.

Hertweck, M., Hiller, R. & Mueller, M.W. (2002) Inhibition of nuclear pre-mRNA splicing by antibiotics in vitro. *Eur J Biochem* **269**(1), 175-83.

Higashi, T., Iwasaki, Y., Ohnishi, Y. & Horinouchi, S. (2007) A-factor and phosphate depletion signals are transmitted to the grixazone biosynthesis genes via the pathway-specific transcriptional activator GriR. *J Bacteriol* **189**(9), 3515-24.

Hill, A.M. (2006) The biosynthesis, molecular genetics and enzymology of the polyketide-derived metabolites. *Nat Prod Rep* **23**(2), 256-320.

Hirano, S., Kato, J.Y., Ohnishi, Y. & Horinouchi, S. (2006) Control of the *Streptomyces* subtilisin inhibitor gene by AdpA in the A-factor regulatory cascade in *Streptomyces griseus*. *J Bacteriol* **188**(17), 6207-16.

Hirano, S., Tanaka, K., Ohnishi, Y. & Horinouchi, S. (2008) Conditionally positive effect of the TetR-family transcriptional regulator AtrA on streptomycin production by *Streptomyces griseus*. *Microbiology* **154**(Pt 3), 905-14.

Hodgson, D.A. (2000) Primary metabolism and its control in streptomycetes: a most unusual group of bacteria. *Adv Microb Physiol* **42**, 47-238.

Hoffman, L.R., D'Argenio, D.A., MacCoss, M.J., Zhang, Z., Jones, R.A. & Miller, S.I. (2005) Aminoglycoside antibiotics induce bacterial biofilm formation. *Nature* **436**(7054), 1171-5.

Holmes, A.J., Bowyer, J., Holley, M.P., O'Donoghue, M., Montgomery, M. & Gillings, M.R. (2000) Diverse, yet-to-be-cultured members of the Rubrobacter subdivision of the Actinobacteria are widespread in Australian arid soils. *FEMS Microbiol Ecol* **33**(2), 111-120.

Hong, H.J., Hutchings, M.I., Neu, J.M., Wright, G.D., Paget, M.S. & Buttner, M.J. (2004) Characterization of an inducible vancomycin resistance system in *Streptomyces coelicolor* reveals a novel gene (*vanK*) required for drug resistance. *Mol Microbiol* **52**(4), 1107-21.

Honore, N. & Cole, S.T. (1994) Streptomycin resistance in mycobacteria. *Antimicrob Agents Chemother* **38**(2), 238-42.

Honore, N., Marchal, G. & Cole, S.T. (1995) Novel mutation in 16S rRNA associated with streptomycin dependence in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **39**(3), 769-70.

Hood, D.W., Heidstra, R., Swoboda, U.K. & Hodgson, D.A. (1992) Molecular genetic analysis of proline and tryptophan biosynthesis in *Streptomyces coelicolor* A3(2): interaction between primary and secondary metabolism--a review. *Gene* **115**(1-2), 5-12.

Hopwood, D.A. (2003) The *Streptomyces* genome--be prepared! *Nat Biotechnol* **21**(5), 505-6.

Hopwood, D.A. (2006) Soil to genomics: the *Streptomyces* chromosome. *Annu Rev Genet* **40**, 1-23.

Horinouchi, S. (2002) A microbial hormone, A-factor, as a master switch for morphological differentiation and secondary metabolism in *Streptomyces griseus*. *Front Biosci* **7**, d2045-57.

Horinouchi, S. (2007) Mining and polishing of the treasure trove in the bacterial genus streptomyces. *Biosci Biotechnol Biochem* **71**(2), 283-99.

Huang, C.H., Chen, C.Y., Tsai, H.H., Chen, C., Lin, Y.S. & Chen, C.W. (2003) Linear plasmid SLP2 of *Streptomyces lividans* is a composite replicon. *Mol Microbiol* **47**(6), 1563-76.

Huddleston, A.S., Cresswell, N., Neves, M.C., Beringer, J.E., Baumberg, S., Thomas, D.I. & Wellington, E.M. (1997) Molecular detection of streptomycin-producing streptomycetes in Brazilian soils. *Appl Environ Microbiol* **63**(4), 1288-97.

Huson, D.H. & Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**(2), 254-67.

Ida, T., Okamoto, R., Nonoyama, M., Irinoda, K., Kurazono, M. & Inoue, M. (2002) Antagonism between aminoglycosides and beta-lactams in a methicillin-resistant *Staphylococcus aureus* isolate involves induction of an aminoglycoside-modifying enzyme. *Antimicrob Agents Chemother* **46**(5), 1516-21.

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**(5), 526-31.

Ito, T. & Wittmann, H.G. (1973) Amino acid replacements in proteins S5 and S12 of two *Escherichia coli* revertants from streptomycin dependence to independence. *Mol Gen Genet* **127**(1), 19-32.

Janssen, P.H. (2006) Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl Environ Microbiol* **72**(3), 1719-28.

Jenne, A., Hartig, J.S., Piganeau, N., Tauer, A., Samarsky, D.A., Green, M.R., Davies, J. & Famulok, M. (2001) Rapid identification and characterization of hammerhead-ribozyme inhibitors using fluorescence-based technology. *Nat Biotechnol* **19**(1), 56-61.

Jerinic, O. & Joseph, S. (2000) Conformational changes in the ribosome induced by translational miscoding agents. *J Mol Biol* **304**(5), 707-13.

Jiang, C. & Xu, L. (1996) Diversity of aquatic actinomycetes in lakes of the middle plateau, Yunnan, China. *Appl Environ Microbiol* **62**(1), 249-253.

Johansen, S.K., Maus, C.E., Plikaytis, B.B. & Douthwaite, S. (2006) Capreomycin binds across the ribosomal subunit interface using tlyA-encoded 2'-O-methylations in 16S and 23S rRNAs. *Mol Cell* **23**(2), 173-82.

Ju, J., Ozanick, S.G., Shen, B. & Thomas, M.G. (2004) Conversion of (2S)-arginine to (2S,3R)-capreomycidine by VioC and VioD from the viomycin biosynthetic pathway of *Streptomyces sp*. strain ATCC11861. *Chembiochem* **5**(9), 1281-5.

Jung, Y.G., Kang, S.H., Hyun, C.G., Yang, Y.Y., Kang, C.M. & Suh, J.W. (2003) Isolation and characterization of bluensomycin biosynthetic genes from *Streptomyces bluensis*. *FEMS Microbiol Lett* **219**(2), 285-9.

Kamal, A., Azeeza, S., Malik, M.S., Shaik, A.A. & Rao, M.V. (2008) Efforts towards the development of new antitubercular agents: potential for thiolactomycin based compounds. *J Pharm Pharm Sci* **11**(2), 56s-80s.

Karimi, R. & Ehrenberg, M. (1994) Dissociation rate of cognate peptidyl-tRNA from the A-site of hyper-accurate and error-prone ribosomes. *Eur J Biochem* **226**(2), 355-60.

Karimi, R. & Ehrenberg, M. (1996) Dissociation rates of peptidyl-tRNA from the P-site of *E.coli* ribosomes. *Embo J* **15**(5), 1149-54.

Karlowsky, J.A., Hoban, D.J., Zelenitsky, S.A. & Zhanel, G.G. (1997) Altered *denA* and *anr* gene expression in aminoglycoside adaptive resistance in *Pseudomonas aeruginosa*. *J Antimicrob Chemother* **40**(3), 371-6.

Kashiwagi, K., Tsuhako, M.H., Sakata, K., Saisho, T., Igarashi, A., da Costa, S.O. & Igarashi, K. (1998) Relationship between spontaneous aminoglycoside resistance in *Escherichia coli* and a decrease in oligopeptide binding protein. *J Bacteriol* **180**(20), 5484-8.

Kataoka, M., Ueda, K., Kudo, T., Seki, T. & Yoshida, T. (1997) Application of the variable region in 16S rDNA to create an index for rapid species identification in the genus *Streptomyces*. *FEMS Microbiol Lett* **151**(2), 249-55.

Kaufmann, G.F., Sartorio, R., Lee, S.H., Rogers, C.J., Meijler, M.M., Moss, J.A., Clapham, B., Brogan, A.P., Dickerson, T.J. & Janda, K.D. (2005) Revisiting quorum sensing: Discovery of additional chemical and biological functions for 3-oxo-N-acylhomoserine lactones. *Proc Natl Acad Sci U S A* **102**(2), 309-14.

Kawachi, R., Akashi, T., Kamitani, Y., Sy, A., Wangchaisoonthorn, U., Nihira, T. & Yamada, Y. (2000) Identification of an AfsA homologue (BarX) from *Streptomyces virginiae* as a pleiotropic regulator controlling autoregulator biosynthesis, virginiamycin biosynthesis and virginiamycin M1 resistance. *Mol Microbiol* **36**(2), 302-13.

Keel, C., Weller, D.M., Natsch, A., Defago, G., Cook, R.J. & Thomashow, L.S. (1996) Conservation of the 2,4-diacetylphloroglucinol biosynthesis locus among fluorescent *Pseudomonas* strains from diverse geographic locations. *Appl Environ Microbiol* **62**(2), 552-63.

Kennell, D. (2002) Processing endoribonucleases and mRNA degradation in bacteria. *J Bacteriol* **184**(17), 4645-57; discussion 4665.

Kers, J.A., Cameron, K.D., Joshi, M.V., Bukhalid, R.A., Morello, J.E., Wach, M.J., Gibson, D.M. & Loria, R. (2005) A large, mobile pathogenicity island confers plant pathogenicity on *Streptomyces* species. *Mol Microbiol* **55**(4), 1025-33.

Khachatourians, G.G. (1998) Agricultural use of antibiotics and the evolution and transfer of antibiotic-resistant bacteria. *Can Med Assoc J* **159**(9), 1129-1136.

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 beta-subunit gene (rpoB) sequences. *Int J Syst Evol Microbiol* **54**(Pt 2), 593-8.

Kim, C., Cha, J.Y., Yan, H., Vakulenko, S.B. & Mobashery, S. (2006) Hydrolysis of ATP by aminoglycoside 3'-phosphotransferases: an unexpected cost to bacteria for harboring an antibiotic resistance enzyme. *J Biol Chem* **281**(11), 6964-9.

Kim, S.B. & Goodfellow, M. (2002) *Streptomyces avermitilis* sp. nov., nom. rev., a taxonomic home for the avermectin-producing streptomycetes. *Int J Syst Evol Microbiol* **52**(Pt 6), 2011-4.

Kimiran-Erdem, A., Arslan, E.O., Sanli Yurudu, N.O., Zeybek, Z., Dogruoz, N. & Cotuk, A. (2007) Isolation and identification of enterococci from seawater samples: assessment of their resistance to antibiotics and heavy metals. *Environ Monit Assess* **125**(1-3), 219-28.

Kinkel, L.L., Bowers, J.H., Shimizu, K., Neeno-Eckwall, E.C. & Schottel, J.L. (1998) Quantitative relationships among thaxtomin A production, potato scab severity, and fatty acid composition in *Streptomyces*. *Can J Microbiol* **44**(8), 768-76.

Kirk, J.L., Beaudette, L.A., Hart, M., Moutoglis, P., Klironomos, J.N., Lee, H. & Trevors, J.T. (2004) Methods of studying soil microbial diversity. *J Microbiol Methods* **58**(2), 169-88.

Kitani, S., Yamada, Y. & Nihira, T. (2001) Gene replacement analysis of the butyrolactone autoregulator receptor (FarA) reveals that FarA acts as a novel regulator in secondary metabolism of *Streptomyces lavendulae* FRI-5. *J Bacteriol* **183**(14), 4357-63.

Kobayashi, N., Alam, M., Nishimoto, Y., Urasawa, S., Uehara, N. & Watanabe, N. (2001) Distribution of aminoglycoside resistance genes in recent clinical isolates of *Enterococcus faecalis*, *Enterococcus faecium* and *Enterococcus avium*. *Epidemiol Infect* **126**(2), 197-204.

Kolsto, A.B., Tourasse, N.J. & Okstad, O.A. (2009) What sets *Bacillus anthracis* apart from other *Bacillus* species? *Annu Rev Microbiol* **63**, 451-76.

Kunisawa, T. (2003) Gene arrangements and branching orders of gram-positive bacteria. *J Theor Biol* **222**(4), 495-503.

Lacey, S.L., Mehmet, S. & Taylor, G.W. (1995) Inhibition of *Helicobacter pylori* growth by 4-hydroxy-2-alkyl-quinolines produced by Pseudomonas aeruginosa. *J Antimicrob Chemother* **36**(5), 827-31.

Lambert, P.A. (2005) Bacterial resistance to antibiotics: modified target sites. *Adv Drug Deliv Rev* **57**(10), 1471-85.

Lambert, T., Gerbaud, G. & Courvalin, P. (1994) Characterization of the chromosomal aac(6')-Ij gene of *Acinetobacter sp*. 13 and the aac(6')-Ih plasmid gene of *Acinetobacter baumannii*. *Antimicrob Agents Chemother* **38**(9), 1883-9.

Lancaster, L.E., Savelsbergh, A., Kleanthous, C., Wintermeyer, W. & Rodnina, M.V. (2008) Colicin E3 cleavage of 16S rRNA impairs decoding and accelerates tRNA translocation on *Escherichia coli* ribosomes. *Mol Microbiol* **69**(2), 390-401.

Lanoot, B., Vancanneyt, M., Hoste, B., Vandemeulebroecke, K., Cnockaert, M.C., Dawyndt, P., Liu, Z., Huang, Y. & Swings, J. (2005) Grouping of streptomycetes using 16S-ITS RFLP fingerprinting. *Res Microbiol* **156**(5-6), 755-62.

Laskin, A.I. & Lechevalier, H.A. (Eds.) (1988) *CRC Handbook of Microbiology, Volume IX, Part A Antibiotics.* Boca Raton, Florida: CRC Press, Inc.

Lazzarini, A., Cavaletti, L., Toppo, G. & Marinelli, F. (2000) Rare genera of actinomycetes as potential producers of new antibiotics. *Antonie van Leeuwenhoek* **78**(3-4), 399-405.

Leclercq, R. & Courvalin, P. (1997) Resistance to glycopeptides in enterococci. *Clin Infect Dis* **24**(4), 545-54; quiz 555-6.

Levesque, C., Piche, L., Larose, C. & Roy, P.H. (1995) PCR mapping of integrons reveals several novel combinations of resistance genes. *Antimicrob Agents Chemother* **39**(1), 185-91.

Lewis, K. (2005) Persister cells and the riddle of biofilm survival. *Biochemistry (Mosc)* **70**(2), 267-74.

Lewis, K. (2008) Multidrug tolerance of biofilms and persister cells. *Curr Top Microbiol Immunol* **322**, 107-31.

Liebner, S., Harder, J. & Wagner, D. (2008) Bacterial diversity and community structure in polygonal tundra soils from Samoylov Island, Lena Delta, Siberia. *Int Microbiol* **11**(3), 195-202.

Lin, J.T., Connelly, M.B., Amolo, C., Otani, S. & Yaver, D.S. (2005) Global transcriptional response of *Bacillus subtilis* to treatment with subinhibitory concentrations of antibiotics that inhibit protein synthesis. *Antimicrob Agents Chemother* **49**(5), 1915-26.

Liras, P. & Martin, J.F. (2006) Gene clusters for beta-lactam antibiotics and control of their expression: why have clusters evolved, and from where did they originate? *Int Microbiol* **9**(1), 9-19.

Loria, R., Coombs, J., Yoshida, M., Kers, J. & Bukhalid, R. (2003) A paucity of bacterial root diseases: *Streptomyces* succeeds where others fail. *Physiological and Molecular Plant Pathology* **62**(2), 65-72.

Loria, R., Kers, J. & Joshi, M. (2006) Evolution of plant pathogenicity in *Streptomyces*. *Annu Rev Phytopathol* **44**, 469-87.

Losos, J.B. (2008) Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecol Lett* **11**(10), 995-1003.

Lottmann, J. & Berg, G. (2001) Phenotypic and genotypic characterization of antagonistic bacteria associated with roots of transgenic and non-transgenic potato plants. *Microbiol Res* **156**(1), 75-82.

Lovette, I.J. & Hochachka, W.M. (2006) Simultaneous effects of phylogenetic niche conservatism and competition on avian community structure. *Ecology* **87**(7 Suppl), S14-28.

Ludwig, W., Strunk, O., Klugbauer, S., Klugbauer, N., Weizenegger, M., Neumaier, J., Bachleitner, M. & Schleifer, K.H. (1998) Bacterial phylogeny based on comparative sequence analysis. *Electrophoresis* **19**(4), 554-68.

Lugtenberg, B.J., Dekkers, L. & Bloemberg, G.V. (2001) Molecular determinants of rhizosphere colonization by *Pseudomonas*. *Annu Rev Phytopathol* **39**, 461-90.

MacGowan, A.P. & Wise, R. (2001) Establishing MIC breakpoints and the interpretation of in vitro susceptibility tests. *J Antimicrob Chemother* **48 Suppl 1**, 17-28.

Madigan, M.T. & Martinko, J.M. (2005) *Brock Biology Of Microorganisms.* Prentice Hall.

Magnet, S., Arbeloa, A., Mainardi, J.L., Hugonnet, J.E., Fourgeaud, M., Dubost, L., Marie, A., Delfosse, V., Mayer, C., Rice, L.B. & Arthur, M. (2007) Specificity of L,D-transpeptidases from gram-positive bacteria producing different peptidoglycan chemotypes. *J Biol Chem* **282**(18), 13151-9.

Maplestone, R.A., Stone, M.J. & Williams, D.H. (1992) The evolutionary role of secondary metabolites--a review. *Gene* **115**(1-2), 151-7.

Marshall, C.G., Lessard, I.A., Park, I. & Wright, G.D. (1998) Glycopeptide antibiotic resistance genes in glycopeptide-producing organisms. *Antimicrob Agents Chemother* **42**(9), 2215-20.

Martin, J.F., Casqueiro, J. & Liras, P. (2005) Secretion systems for secondary metabolites: how producer cells send out messages of intercellular communication. *Curr Opin Microbiol* **8**(3), 282-93.

Martin, N.L. & Beveridge, T.J. (1986) Gentamicin interaction with *Pseudomonas aeruginosa* cell envelope. *Antimicrob Agents Chemother* **29**(6), 1079-87.

Martin, P., Jullien, E. & Courvalin, P. (1988) Nucleotide sequence of *Acinetobacter baumannii aphA-6* gene: evolutionary and functional implications of sequence homologies with nucleotide-binding proteins, kinases and other aminoglycoside-modifying enzymes. *Mol Microbiol* **2**(5), 615-25.

Martinez, J.L., Fajardo, A., Garmendia, L., Hernandez, A., Linares, J.F., Martinez-Solano, L. & Sanchez, M.B. (2009a) A global view of antibiotic resistance. *FEMS Microbiol Rev* **33**(1), 44-65.

Martinez, J.L., Sanchez, M.B., Martinez-Solano, L., Hernandez, A., Garmendia, L., Fajardo, A. & Alvarez-Ortega, C. (2009b) Functional role of bacterial multidrug efflux pumps in microbial natural ecosystems. *FEMS Microbiol Rev* **33**(2), 430-49.

Marttinen, P., Baldwin, A., Hanage, W.P., Dowson, C., Mahenthiralingam, E. & Corander, J. (2008) Bayesian modeling of recombination events in bacterial populations. *BMC Bioinformatics* **9**, 421.

Mascher, T., Margulis, N.G., Wang, T., Ye, R.W. & Helmann, J.D. (2003) Cell wall stress responses in *Bacillus subtilis*: the regulatory network of the bacitracin stimulon. *Mol Microbiol* **50**(5), 1591-604.

Massova, I. & Mobashery, S. (1998) Kinship and diversification of bacterial penicillin-binding proteins and beta-lactamases. *Antimicrob Agents Chemother* **42**(1), 1-17.

Matyar, F., Kaya, A. & Dincer, S. (2008) Antibacterial agents and heavy metal resistance in Gram-negative bacteria isolated from seawater, shrimp and sediment in Iskenderun Bay, Turkey. *Sci Total Environ* **407**(1), 279-85.

Maus, C.E., Plikaytis, B.B. & Shinnick, T.M. (2005a) Molecular analysis of cross-resistance to capreomycin, kanamycin, amikacin, and viomycin in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **49**(8), 3192-7.

Maus, C.E., Plikaytis, B.B. & Shinnick, T.M. (2005b) Mutation of *tlyA* confers capreomycin resistance in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **49**(2), 571-7.

Mayfield, C.I., Williams, S.T., Ruddick, S.M. & L., H.H. (1972) Studies on the ecology of actinomycetes in soil IV. Observations on the form and growth of streptomycetes in soil. *Soil Biol Biochem* **4**(1), 79-86.

Mazel, D. & Davies, J. (1999) Antibiotic resistance in microbes. *Cell Mol Life Sci* **56**(9-10), 742-54.

McCarthy, A.J. & Williams, S.T. (1991) Actinomycetes as agents of biodegradation in the environment - a review. In *8th International Symp on Biology of Actinomycetes (Isba-8)*, pp. 189-192. Madison, Wi: Elsevier Science Bv.

McManus, P.S., Stockwell, V.O., Sundin, G.W. & Jones, A.L. (2002) Antibiotic use in plant agriculture. *Annu Rev Phytopathol* **40**, 443-65.

McNeil, M.M. & Brown, J.M. (1994) The medically important aerobic actinomycetes: epidemiology and microbiology. *Clin Microbiol Rev* **7**(3), 357-417.

Meharg, A.A. & Killham, K. (1988) A comparison of carbon flow from pre-labelled and pulse-labelled plants. *Plant and Soil* **112**(2), 225-231.

Meier, A., Kirschner, P., Bange, F.C., Vogel, U. & Bottger, E.C. (1994) Genetic alterations in streptomycin-resistant *Mycobacterium tuberculosis*: mapping of mutations conferring resistance. *Antimicrob Agents Chemother* **38**(2), 228-33.

Melancon, P., Lemieux, C. & Brakier-Gingras, L. (1988) A mutation in the 530 loop of *Escherichia coli* 16S ribosomal RNA causes resistance to streptomycin. *Nucleic Acids Res* **16**(20), 9631-9.

Metcalfe, A.C., Krsek, M., Gooday, G.W., Prosser, J.I. & Wellington, E.M. (2002) Molecular analysis of a bacterial chitinolytic community in an upland pasture. *Appl Environ Microbiol* **68**(10), 5042-50.

Mincer, T.J., Jensen, P.R., Kauffman, C.A. & Fenical, W. (2002) Widespread and persistent populations of a major new marine actinomycete taxon in ocean sediments. *Appl Environ Microbiol* **68**(10), 5005-11.

Mindlin, S.Z., Petrova, M.A., Bass, I.A. & Gorlenko, Z.M. (2006) Origin, evolution, and migration of drug resistance genes. *Russ J Genet* **42**(11), 1257-1271.

Mindlin, S.Z., Soina, V.S., Petrova, M.A. & Gorlenko, Z.M. (2008) Isolation of antibiotic resistance bacterial strains from Eastern Siberia permafrost sediments. *Russian Journal of Genetics* **44**(1), 27-34.

Mingeot-Leclercq, M.P., Glupczynski, Y. & Tulkens, P.M. (1999) Aminoglycosides: activity and resistance. *Antimicrob Agents Chemother* **43**(4), 727-37.

Mingeot-Leclercq, M.P. & Tulkens, P.M. (1999) Aminoglycosides: nephrotoxicity. *Antimicrob Agents Chemother* **43**(5), 1003-12.

Misawa, N., Satomi, Y., Kondo, K., Yokoyama, A., Kajiwara, S., Saito, T., Ohtani, T. & Miki, W. (1995) Structure and functional analysis of a marine bacterial carotenoid biosynthesis gene cluster and astaxanthin biosynthetic pathway proposed at the gene level. *J Bacteriol* **177**(22), 6575-84.

Moazed, D. & Noller, H.F. (1987) Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature* **327**(6121), 389-94.

Modolell, J. & Vazquez (1977) The inhibition of ribosomal translocation by viomycin. *Eur J Biochem* **81**(3), 491-7.

Monciardini, P., Sosio, M., Cavaletti, L., Chiocchini, C. & Donadio, S. (2002) New PCR primers for the selective amplification of 16S rDNA from different groups of actinomycetes. *Fems Microbiol Ecol* **42**(3), 419-429.

Morningstar, A., Gaze, W.H., Tolba, S. & Wellington, E.M.H. (2006) Evolving gene clusters in soil bacteria. In *SGM Symposium 66*, pp. 201-222. Edited by N.A. Logan, H.M. Lappin-Scott & P.C.F. Oyston. Cambridge University Press.

Morse, B.K., Brown, M.S., Gagne, J.W., McArthur, H.A., McCormick, E.L., Murphy, T.K., Narrol, M.H., Perry, D.A., Smogowicz, A.A., Wax, R.G. & Wong, J.W. (1997) Production of tuberactinamine A by *Streptomyces griseoverticillatus var. tuberacticus* NRRL 3482 fed with (S)-2-aminoethyl-L-cysteine. *J Antibiot (Tokyo)* **50**(8), 698-700.

Mueller, U.G. & Gerardo, N. (2002) Fungus-farming insects: multiple origins and diverse evolutionary histories. *Proc Natl Acad Sci U S A* **99**(24), 15247-9.

Muggia, L., Schmitt, I. & Grube, M. (2008) Purifying selection is a prevailing motif in the evolution of ketoacyl synthase domains of polyketide synthases from lichenized fungi. *Mycol Res* **112**(Pt 2), 277-88.

Mummey, D., Holben, W., Six, J. & Stahl, P. (2006) Spatial stratification of soil bacterial populations in aggregates of diverse soils. *Microb Ecol* **51**(3), 404-11.

Musser, J.M. (1995) Antimicrobial agent resistance in mycobacteria: molecular genetic insights. *Clin Microbiol Rev* **8**(4), 496-514.

Nagaya, A., Takeyama, S. & Tamegai, H. (2005) Identification of aminotransferase genes for biosynthesis of aminoglycoside antibiotics from soil DNA. *Biosci Biotechnol Biochem* **69**(7), 1389-93.

Neeno-Eckwall, E.C., Kinkel, L.L. & Schottel, J.L. (2001) Competition and antibiosis in the biological control of potato scab. *Can J Microbiol* **47**(4), 332-40.

Nesvera, J., Hochmannova, J. & Patek, M. (1998) An integron of class 1 is present on the plasmid pCG4 from gram-positive bacterium *Corynebacterium glutamicum*. *FEMS Microbiol Lett* **169**(2), 391-5.

Ng, W.L., Kazmierczak, K.M., Robertson, G.T., Gilmour, R. & Winkler, M.E. (2003) Transcriptional regulation and signature patterns revealed by microarray analyses of *Streptococcus pneumoniae* R6 challenged with sublethal concentrations of translation inhibitors. *J Bacteriol* **185**(1), 359-70.

Nguyen, K.T., Tenor, J., Stettler, H., Nguyen, L.T., Nguyen, L.D. & Thompson, C.J. (2003) Colonial differentiation in *Streptomyces coelicolor* depends on translation of a specific codon within the adpA gene. *J Bacteriol* **185**(24), 7291-6.

Nishimura, K., Hosaka, T., Tokuyama, S., Okamoto, S. & Ochi, K. (2007) Mutations in rsmG, encoding a 16S rRNA methyltransferase, result in low-level streptomycin resistance and antibiotic overproduction in *Streptomyces coelicolor* A3(2). *J Bacteriol* **189**(10), 3876-83.

Noguchi, N., Sasatsu, M. & Kono, M. (1993) Genetic mapping in *Bacillus subtilis* 168 of the *aadK* gene which encodes aminoglycoside 6-adenylyltransferase. *FEMS Microbiol Lett* **114**(1), 47-52.

Nogueira, T. & Springer, M. (2000) Post-transcriptional control by global regulators of gene expression in bacteria. *Curr Opin Microbiol* **3**(2), 154-8.

Noller, H.F. (1984) Structure of ribosomal RNA. *Annu Rev Biochem* **53**, 119-62.

Ochi, K. (1995) A taxonomic study of the genus *Streptomyces* by analysis of ribosomal protein AT-L30. *Int J Syst Bacteriol* **45**(3), 507-14.

Ochman, H., Lawrence, J.G. & Groisman, E.A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**(6784), 299-304.

Ogawara, H., Kawamura, N., Kudo, T., Suzuki, K.I. & Nakase, T. (1999) Distribution of beta-lactamases in actinomycetes. *Antimicrob Agents Chemother* **43**(12), 3014-7.

Ohnishi, Y., Ishikawa, J., Hara, H., Suzuki, H., Ikenoya, M., Ikeda, H., Yamashita, A., Hattori, M. & Horinouchi, S. (2008) Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J Bacteriol* **190**(11), 4050-60.

Ohnishi, Y., Kameyama, S., Onaka, H. & Horinouchi, S. (1999) The A-factor regulatory cascade leading to streptomycin biosynthesis in *Streptomyces griseus* : identification of a target gene of the A-factor receptor. *Mol Microbiol* **34**(1), 102-11.

Ohnishi, Y., Yamazaki, H., Kato, J.Y., Tomono, A. & Horinouchi, S. (2005) AdpA, a central transcriptional regulator in the A-factor regulatory cascade that leads to morphological development and secondary metabolism in *Streptomyces griseus*. *Biosci Biotechnol Biochem* **69**(3), 431-9.

Okoro, C.K., Brown, R., Jones, A.L., Andrews, B.A., Asenjo, J.A., Goodfellow, M. & Bull, A.T. (2009) Diversity of culturable actinomycetes in hyper-arid soils of the Atacama Desert, Chile. *Antonie van Leeuwenhoek* **95**(2), 121-33.

Oliva, B., Comanducci, A. & Chopra, I. (1998) Antibacterial spectra of drugs used for chemotherapy of mycobacterial infections. *Tuber Lung Dis* **79**(2), 107-9.

Olive, J.E., De Abreu, D.M., Rastogi, T., Andersen, A.A., Mittermaier, A.K., Beattie, T.L. & Collins, R.A. (1995) Enhancement of *Neurospora* VS ribozyme cleavage by tuberactinomycin antibiotics. *Embo J* **14**(13), 3247-51.

Ounissi, H., Derlot, E., Carlier, C. & Courvalin, P. (1990) Gene homogeneity for aminoglycoside-modifying enzymes in gram-positive cocci. *Antimicrob Agents Chemother* **34**(11), 2164-8.

Pan, D., Kirillov, S.V. & Cooperman, B.S. (2007) Kinetically competent intermediates in the translocation step of protein synthesis. *Mol Cell* **25**(4), 519-29.

Pang, Y., Brown, B.A., Steingrube, V.A., Wallace, R.J., Jr. & Roberts, M.C. (1994) Tetracycline resistance determinants in *Mycobacterium* and *Streptomyces* species. *Antimicrob Agents Chemother* **38**(6), 1408-12.

Paradkar, A., Trefzer, A., Chakraburtty, R. & Stassi, D. (2003) *Streptomyces* genetics: a genomic perspective. *Crit Rev Biotechnol* **23**(1), 1-27.

Payie, K.G. & Clarke, A.J. (1997) Characterization of gentamicin 2'-N-acetyltransferase from *Providencia stuartii*: its use of peptidoglycan metabolites for acetylation of both aminoglycosides and peptidoglycan. *J Bacteriol* **179**(13), 4106-14.

Peske, F., Savelsbergh, A., Katunin, V.I., Rodnina, M.V. & Wintermeyer, W. (2004) Conformational changes of the small ribosomal subunit during elongation factor G-dependent tRNA-mRNA translocation. *J Mol Biol* **343**(5), 1183-94.

Petersen, A. & Dalsgaard, A. (2003) Species composition and antimicrobial resistance genes of Enterococcus spp, isolated from integrated and traditional fish farms in Thailand. *Environ Microbiol* **5**(5), 395-402.

Phillips, L., Wellington, E.M.H. & Rees, S.B. (1994) The Distribution of Antibiotic-Resistance Patterns within Streptomycetes and Their Use in Secondary Metabolite Screening. *J Ind Microbiol* **13**(1), 53-62.

Piepersberg, W. & Distler, J. (1997) Aminoglycosides and suger components in other secondary metabolites. In *Biotecnology: Products of Secondary Metabolism*, Vol. 7Edited by H. Kleinkauf & H. von Dohren. Weinheim: VCH.

Pimentel-Elardo, S.M., Scheuermayer, M., Kozytska, S. & Hentschel, U. (2009) *Streptomyces axinellae sp. nov*., isolated from the Mediterranean sponge *Axinella polypoides* (Porifera). *Int J Syst Evol Microbiol* **59**(Pt 6), 1433-7.

Pissowotzki, K., Mansouri, K. & Piepersberg, W. (1991) Genetics of streptomycin production in *Streptomyces griseus*: molecular structure and putative function of genes *strELMB2N*. *Mol Gen Genet* **231**(1), 113-23.

Pittenauer, E., Zehl, M., Belgacem, O., Raptakis, E., Mistrik, R. & Allmaier, G. (2006) Comparison of CID spectra of singly charged polypeptide antibiotic precursor ions obtained by positive-ion vacuum MALDI IT/RTOF and TOF/RTOF, AP-MALDI-IT and ESI-IT mass spectrometry. *J Mass Spectrom* **41**(4), 421-47.

Pond, S.L. & Frost, S.D. (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**(10), 2531-3.

Pope, C.F., O'Sullivan, D.M., McHugh, T.D. & Gillespie, S.H. (2008) A practical guide to measuring mutation rates in antibiotic resistance. *Antimicrob Agents Chemother* **52**(4), 1209-14.

Powers, T. & Noller, H.F. (1990) Dominant lethal mutations in a conserved loop in 16S rRNA. *Proc Natl Acad Sci U S A* **87**(3), 1042-6.

Powers, T. & Noller, H.F. (1991) A functional pseudoknot in 16S ribosomal RNA. *Embo J* **10**(8), 2203-14.

Prior, J.L., Prior, R.G., Hitchen, P.G., Diaper, H., Griffin, K.F., Morris, H.R., Dell, A. & Titball, R.W. (2003) Characterization of the O antigen gene cluster and structural analysis of the O antigen of *Francisella tularensis subsp. tularensis*. *J Med Microbiol* **52**(Pt 10), 845-51.

Raaijmakers, J.M., Vlami, M. & de Souza, J.T. (2002) Antibiotic production by bacterial biocontrol agents. *Antonie van Leeuwenhoek* **81**(1-4), 537-47.

Raffel, S.J., Stabb, E.V., Milner, J.L. & Handelsman, J. (1996) Genotypic and phenotypic analysis of zwittermicin A-producing strains of *Bacillus cereus*. *Microbiology* **142 (Pt 12)**, 3425-36.

Ramaswamy, S. & Musser, J.M. (1998) Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. *Tuber Lung Dis* **79**(1), 3-29.

Ramon-Garcia, S., Otal, I., Martin, C., Gomez-Lus, R. & Ainsa, J.A. (2006) Novel streptomycin resistance gene from *Mycobacterium fortuitum*. *Antimicrob Agents Chemother* **50**(11), 3920-2.

Rather, P.N., Munayyer, H., Mann, P.A., Hare, R.S., Miller, G.H. & Shaw, K.J. (1992) Genetic analysis of bacterial acetyltransferases: identification of amino acids determining the specificities of the aminoglycoside 6'-N-acetyltransferase Ib and IIa proteins. *J Bacteriol* **174**(10), 3196-203.

Rather, P.N. & Orosz, E. (1994) Characterization of aarA, a pleiotrophic negative regulator of the 2'-N-acetyltransferase in *Providencia stuartii*. *J Bacteriol* **176**(16), 5140-4.

Rather, P.N., Orosz, E., Shaw, K.J., Hare, R. & Miller, G. (1993) Characterization and transcriptional regulation of the 2'-N-acetyltransferase gene from *Providencia stuartii*. *J Bacteriol* **175**(20), 6492-8.

Redenbach, M., Flett, F., Piendl, W., Glocker, I., Rauland, U., Wafzig, O., Kliem, R., Leblond, P. & Cullum, J. (1993) The *Streptomyces lividans* 66 chromosome contains a 1 MB deletogenic region flanked by two amplifiable regions. *Mol Gen Genet* **241**(3-4), 255-62.

Retzlaff, L. & Distler, J. (1995) The regulator of streptomycin gene expression, StrR, of *Streptomyces griseus* is a DNA binding activator protein with multiple recognition sites. *Mol Microbiol* **18**(1), 151-62.

Ridley, C.P., Lee, H.Y. & Khosla, C. (2008) Evolution of polyketide synthases in bacteria. *Proc Natl Acad Sci U S A* **105**(12), 4595-600.

Riesenfeld, C.S., Goodman, R.M. & Handelsman, J. (2004) Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environ Microbiol* **6**(9), 981-9.

Rintala, H., Nevalainen, A. & Suutari, M. (2002) Diversity of streptomycetes in water-damaged building materials based on 16S rDNA sequences. *Lett Appl Microbiol* **34**(6), 439-43.

Rodnina, M.V. & Wintermeyer, W. (2001) Ribosome fidelity: tRNA discrimination, proofreading and induced fit. *Trends Biochem Sci* **26**(2), 124-30.

Rodriguez, M.M., Power, P., Radice, M., Vay, C., Famiglietti, A., Galleni, M., Ayala, J.A. & Gutkind, G. (2004) Chromosome-encoded CTX-M-3 from *Kluyvera ascorbata*: a possible origin of plasmid-borne CTX-M-1-derived cefotaximases. *Antimicrob Agents Chemother* **48**(12), 4895-7.

Rogers, J., Chang, A.H., von Ahsen, U., Schroeder, R. & Davies, J. (1996) Inhibition of the self-cleavage reaction of the human hepatitis delta virus ribozyme by antibiotics. *J Mol Biol* **259**(5), 916-25.

Rong, X. & Huang, Y. (2009) Taxonomic evaluation of *Streptomyces griseus* clade using multilocus sequence analysis and DNA-DNA hybridization, with proposal to reduce 29 species and 3 subspecies to 11 genomic species. *Int J Syst Evol Microbiol* first published on Aug 5, 2009 as doi: doi:10.1099/ijs.0.012419-0.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A. & Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16**(10), 944-5.

Saadoun, I. & Gharaibeh, R. (2008) Usefulness of *strb1* and 16S rDNA-targeted PCR for detection of *Streptomyces spp*. in environmental samples. *Pol J Microbiol* **57**(1), 81-4.

Sabina, J., Dover, N., Templeton, L.J., Smulski, D.R., Soll, D. & LaRossa, R.A. (2003) Interfering with different steps of protein synthesis explored by transcriptional profiling of *Escherichia coli* K-12. *J Bacteriol* **185**(20), 6158-70.

Sacchettini, J.C., Rubin, E.J. & Freundlich, J.S. (2008) Drugs versus bugs: in pursuit of the persistent predator *Mycobacterium tuberculosis*. *Nat Rev Microbiol* **6**(1), 41-52.

Sandegren, L. & Andersson, D.I. (2009) Bacterial gene amplification: implications for the evolution of antibiotic resistance. *Nat Rev Microbiol* **7**(8), 578-88.

Sarmah, A.K., Meyer, M.T. & Boxall, A.B.A. (2006) A global perspective on the use, sales, exposure pathways, occurrence, fate and effects of veterinary antibiotics (VAs) in the environment. *Chemosphere* **65**(5), 725-759.

Savasan, S., Kaya, O., Kirkan, S. & Ciftci, A. (2008) Antimicrobial resistance in *Euterococcus faecalis* from fishes. *Ankara Univ Vet Fak Derg* **55**(2), 107-110.

Schierup, M.H. & Hein, J. (2000) Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156**(2), 879-91.

Schmitt, I. & Lumbsch, H.T. (2009) Ancient horizontal gene transfer from bacteria enhances biosynthetic capabilities of fungi. *PLoS One* **4**(2), e4437.

Schnappinger, D. & Hillen, W. (1996) Tetracyclines: antibiotic action, uptake, and resistance mechanisms. *Arch Microbiol* **165**(6), 359-69.

Schoener, T.W. (1983) Field experiments on interspecific competition. *American Naturalist* **122**(2), 240-285.

Schrey, S.D. & Tarkka, M.T. (2008) Friends and foes: streptomycetes as modulators of plant disease and symbiosis. *Antonie van Leeuwenhoek* **94**(1), 11-9.

Schroeder, R., Waldsich, C. & Wank, H. (2000) Modulation of RNA function by aminoglycoside antibiotics. *Embo J* **19**(1), 1-9.

Selva, L., Viana, D., Regev-Yochay, G., Trzcinski, K., Corpa, J.M., Lasa, I., Novick, R.P. & Penades, J.R. (2009) Killing niche competitors by remote-control bacteriophage induction. *Proc Natl Acad Sci U S A* **106**(4), 1234-8.

Sharma, A. & Rajput, S. (1996) *Salmonella* as an index of pollution of fresh-water environments. *Environ Monit Assess* **41**(1), 67-76.

Sharmeen, L., Kuo, M.Y., Dinter-Gottlieb, G. & Taylor, J. (1988) Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *J Virol* **62**(8), 2674-9.

Shaw, K.J., Hare, R.S., Sabatelli, F.J., Rizzo, M., Cramer, C.A., Naples, L., Kocsi, S., Munayyer, H., Mann, P., Miller, G.H. & et al. (1991) Correlation between

aminoglycoside resistance profiles and DNA hybridization of clinical isolates. *Antimicrob Agents Chemother* **35**(11), 2253-61.

Shaw, K.J., Rather, P.N., Hare, R.S. & Miller, G.H. (1993) Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminoglycoside-modifying enzymes. *Microbiol Rev* **57**(1), 138-63.

Shaw, K.J., Rather, P.N., Sabatelli, F.J., Mann, P., Munayyer, H., Mierzwa, R., Petrikkos, G.L., Hare, R.S., Miller, G.H., Bennett, P. & et al. (1992) Characterization of the chromosomal aac(6')-Ic gene from *Serratia marcescens*. *Antimicrob Agents Chemother* **36**(7), 1447-55.

Shea, K.M. (2003) Antibiotic resistance: What is the impact of agricultural uses of antibiotics on children's health? *Pediatrics* **112**(1), 253-258.

Skinner, R.H. & Cundliffe, E. (1980) Resistance to the antibiotics viomycin and capreomycin in the *Streptomyces* species which produce them. *J Gen Microbiol* **120**(1), 95-104.

Smith, J.J., Tow, L.A., Stafford, W., Cary, C. & Cowan, D.A. (2006) Bacterial diversity in three different Antarctic Cold Desert mineral soils. *Microb Ecol* **51**(4), 413-21.

Snelling, A.M., Hawkey, P.M., Heritage, J., Downey, P., Bennett, P.M. & Holmes, B. (1993) The use of a DNA probe and PCR to examine the distribution of the aac(6')-Ic gene in *Serratia marcescens* and other gram-negative bacteria. *J Antimicrob Chemother* **31**(6), 841-54.

Sousa, S.A., Moreira, L.M., Wopperer, J., Eberl, L., Sa-Correia, I. & Leitao, J.H. (2007) The *Burkholderia cepacia bceA* gene encodes a protein with phosphomannose isomerase and GDP-D-mannose pyrophosphorylase activities. *Biochem Biophys Res Commun* **353**(1), 200-6.

Spratt, B.G., Hanage, W.P., Li, B., Aanensen, D.M. & Feil, E.J. (2004) Displaying the relatedness among isolates of bacterial species -- the eBURST approach. *FEMS Microbiol Lett* **241**(2), 129-34.

Springer, B., Kidan, Y.G., Prammananan, T., Ellrott, K., Bottger, E.C. & Sander, P. (2001) Mechanisms of streptomycin resistance: selection of mutations in the 16S rRNA gene conferring resistance. *Antimicrob Agents Chemother* **45**(10), 2877-84.

Sprusansky, O., Stirrett, K., Skinner, D., Denoya, C. & Westpheling, J. (2005) The bkdR gene of *Streptomyces coelicolor* is required for morphogenesis and antibiotic production and encodes a transcriptional regulator of a branched-chain amino acid dehydrogenase complex. *J Bacteriol* **187**(2), 664-71.

Sreevatsan, S., Pan, X., Stockbauer, K.E., Williams, D.L., Kreiswirth, B.N. & Musser, J.M. (1996) Characterization of *rpsL* and *rrs* mutations in streptomycin-resistant *Mycobacterium tuberculosis* isolates from diverse geographic localities. *Antimicrob Agents Chemother* **40**(4), 1024-6.

Srinivasan, V., Nam, H.M., Sawant, A.A., Headrick, S.I., Nguyen, L.T. & Oliver, S.P. (2008) Distribution of tetracycline and streptomycin resistance genes and class 1 integrons in Enterobacteriaceae isolated from dairy and nondairy farm soils. *Microb Ecol* **55**(2), 184-93.

Stabb, E.V., Jacobson, L.M. & Handelsman, J. (1994) Zwittermicin A-producing strains of *Bacillus cereus* from diverse soils. *Appl Environ Microbiol* **60**(12), 4404-12.

Stach, J.E., Maldonado, L.A., Ward, A.C., Goodfellow, M. & Bull, A.T. (2003) New primers for the class Actinobacteria: application to marine and terrestrial environments. *Environ Microbiol* **5**(10), 828-41.

Stackebrandt, E. & Goebel, B.M. (1994) A Place for DNA-DNA Reassociation and 16s Ribosomal-Rna Sequence-Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic Bacteriology* **44**(4), 846-849.

Stackebrandt, E., Rainey, F.A. & WardRainey, N.L. (1997) Proposal for a new hierarchic classification system, Actinobacteria classis nov. *Int J Sys Bacteriol* **47**(2), 479-491.

Stackebrandt, E., Witt, D., Kemmerling, C., Kroppenstedt, R. & Liesack, W. (1991) Designation of streptomycete 16S and 23S rRNA-based target regions for oligonucleotide probes. *Appl Environ Microbiol* **57**(5), 1468-77.

Stage-Zimmermann, T.K. & Uhlenbeck, O.C. (1998) Hammerhead ribozyme kinetics. *Rna* **4**(8), 875-89.

Stahley, M.R. & Strobel, S.A. (2006) RNA splicing: group I intron crystal structures reveal the basis of splice site selection and metal ion catalysis. *Curr Opin Struct Biol* **16**(3), 319-26.

Staley, J.T. (2006) The bacterial species dilemma and the genomic-phylogenetic species concept. *Philos Trans R Soc Lond B Biol Sci* **361**(1475), 1899-909.

Stanhope, M.J., Lefebure, T., Walsh, S.L., Becker, J.A., Lang, P., Pavinski Bitar, P.D., Miller, L.A., Italia, M.J. & Amrine-Madsen, H. (2008) Positive selection in penicillin-binding proteins 1a, 2b, and 2x from *Streptococcus pneumoniae* and its correlation with amoxicillin resistance development. *Infect Genet Evol* **8**(3), 331-9.

Stein, T., Borchert, S., Conrad, B., Feesche, J., Hofemeister, B., Hofemeister, J. & Entian, K.D. (2002) Two different lantibiotic-like peptides originate from the ericin gene cluster of *Bacillus subtilis* A1/3. *J Bacteriol* **184**(6), 1703-11.

Stern, S., Powers, T., Changchien, L.M. & Noller, H.F. (1988) Interaction of ribosomal proteins S5, S6, S11, S12, S18 and S21 with 16 S rRNA. *J Mol Biol* **201**(4), 683-95.

Steven, B., Leveille, R., Pollard, W.H. & Whyte, L.G. (2006) Microbial ecology and biodiversity in permafrost. *Extremophiles* **10**(4), 259-67.

Stone, M.J. & Williams, D.H. (1992) On the evolution of functional secondary metabolites (natural products). *Mol Microbiol* **6**(1), 29-34.

Sundin, G.W. & Bender, C.L. (1996) Dissemination of the *strA-strB* streptomycin-resistance genes among commensal and pathogenic bacteria from humans, animals, and plants. *Mol Ecol* **5**(1), 133-43.

Sylvia, D.M., Fuhrmann, J.J., P.G., H. & Zuberer, D.A. (2005) *Principles and Applications of Soil Microbiology.* Upper Saddle River, NJ: Pearson.

Szabo, I., Penyige, A., Barabas, G. & Barabas, J. (1990) Effect of aminoglycoside antibiotics on the autolytic enzyme of *Streptomyces griseus*. *Arch Microbiol* **155**(1), 99-102.

Szabo, I., Penyige, A., Barabas, G., Szabo, G. & Dinya, Z. (1989) Production of a streptomycin-Park nucleotide complex by *Streptomyces griseus*. *Antimicrob Agents Chemother* **33**(1), 58-62.

Szaflarski, W., Vesper, O., Teraoka, Y., Plitta, B., Wilson, D.N. & Nierhaus, K.H. (2008) New features of the ribosome and ribosomal inhibitors: non-enzymatic recycling, misreading and back-translocation. *J Mol Biol* **380**(1), 193-205.

Taber, H.W., Mueller, J.P., Miller, P.F. & Arrow, A.S. (1987) Bacterial uptake of aminoglycoside antibiotics. *Microbiol Rev* **51**(4), 439-57.

Takano, E. (2006) Gamma-butyrolactones: Streptomyces signalling molecules regulating antibiotic production and differentiation. *Curr Opin Microbiol* **9**(3), 287-94.

Takano, E., Chakraburtty, R., Nihira, T., Yamada, Y. & Bibb, M.J. (2001) A complex role for the gamma-butyrolactone SCB1 in regulating antibiotic production in Streptomyces coelicolor A3(2). *Mol Microbiol* **41**(5), 1015-28.

Tamegai, H., Kuki, T., Udagawa, Y., Aoki, R., Nagaya, A. & Tsukada, S. (2006) Exploration of genes that encode a carbocycle-forming enzyme involved in biosynthesis of aminoglycoside antibiotics from the environmental DNA. *Biosci Biotechnol Biochem* **70**(7), 1711-6.

Taniguchi, H., Chang, B., Abe, C., Nikaido, Y., Mizuguchi, Y. & Yoshida, S.I. (1997) Molecular analysis of kanamycin and viomycin resistance in *Mycobacterium smegmatis* by use of the conjugation system. *J Bacteriol* **179**(15), 4795-801.

Tate, R.L. (2000) *Soil Microbiology (2nd ed.).* New York: John Wiley & Sons Inc.

Thomas, M.G., Chan, Y.A. & Ozanick, S.G. (2003) Deciphering tuberactinomycin biosynthesis: isolation, sequencing, and annotation of the viomycin biosynthetic gene cluster. *Antimicrob Agents Chemother* **47**(9), 2823-30.

Tohyama, H., Okami, Y. & Umezawa, H. (1987) Nucleotide sequence of the streptomycinphosphotransferase and amidinotransferase genes from *Streptomyces griseus*. *Nucleic Acids Res* **15**(4), 1819-33.

Tolba, S. (2004) Distribution of streptomycin resistance and biosynthesis genes in streptomycetes recovered from different soil sites and the role of horizontal gene transfer in their dissemination. *PhD Thesis*. Coventry: University of Warwick.

Tolba, S., Egan, S., Kallifidas, D. & Wellington, E.M.H. (2002) Distribution of streptomycin resistance and biosynthesis genes in streptomycetes recovered from different soil sites. *Fems Microbiology Ecology* **42**(2), 269-276.

Tomono, A., Tsai, Y., Yamazaki, H., Ohnishi, Y. & Horinouchi, S. (2005) Transcriptional control by A-factor of *strR*, the pathway-specific transcriptional activator for streptomycin biosynthesis in *Streptomyces griseus*. *J Bacteriol* **187**(16), 5595-604.

Torsvik, V., Daae, F.L., Sandaa, R.A. & Ovreas, L. (1998) Novel techniques for analysing microbial diversity in natural and perturbed environments. *J Biotechnol* **64**(1), 53-62.

Trefzer, A., Pelzer, S., Schimana, J., Stockert, S., Bihlmaier, C., Fiedler, H.P., Welzel, K., Vente, A. & Bechthold, A. (2002) Biosynthetic gene cluster of simocyclinone, a natural multihybrid antibiotic. *Antimicrob Agents Chemother* **46**(5), 1174-82.

Trieu-Cuot, P. & Courvalin, P. (1986) Evolution and transfer of aminoglycoside resistance genes under natural conditions. *J Antimicrob Chemother* **18 Suppl C**, 93-102.

Trower, M.K. & Clark, K.G. (1990) PCR cloning of a streptomycin phosphotransferase (aphE) gene from *Streptomyces griseus* ATCC 12475. *Nucleic Acids Res* **18**(15), 4615.

Trujillo, M.E. & Goodfellow, M. (2003) Numerical phenetic classification of clinically significant aerobic sporoactinomycetes and related organisms. *Antonie van Leeuwenhoek* **84**(1), 39-68.

Ubeda, C., Maiques, E., Knecht, E., Lasa, I., Novick, R.P. & Penades, J.R. (2005) Antibiotic-induced SOS response promotes horizontal dissemination of pathogenicity island-encoded virulence factors in staphylococci. *Mol Microbiol* **56**(3), 836-44.

Ul-Hassan, A. (2007) Horizontal gene transfer and evolution of the streptomyces chromosome. *PhD Thesis*. Coventry: University of Warwick.

Vakulenko, S.B. & Mobashery, S. (2003) Versatility of aminoglycosides and prospects for their future. *Clin Microbiol Rev* **16**(3), 430-50.

Valenzuela-Encinas, C., Neria-Gonzalez, I., Alcantara-Hernandez, R.J., Estrada-Alvarado, I., Zavala-Diaz de la Serna, F.J., Dendooven, L. & Marsch, R. (2009) Changes in the bacterial populations of the highly alkaline saline soil of the former lake Texcoco (Mexico) following flooding. *Extremophiles* **13**(4), 609-21.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G.F., Chater, K.F. & van Sinderen, D. (2007) Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol Mol Biol Rev* **71**(3), 495-548.

Verdu, M. & Pausas, J.G. (2007) Fire drives phylogenetic clustering in Mediterranean Basin woody plant communities. *Journal of Ecology* **95**(6), 1316-1323.

Vining, L.C. (1992) Secondary metabolism, inventive evolution and biochemical diversity--a review. *Gene* **115**(1-2), 135-40.

VMD (2007) Sales of antimicrobial products authorised for use as veterinary medicines, antiprotozoals, antifungals, growth promoters and coccidiostats, in the UK in 2006. Vol. 2009.

Vogtli, M. & Hutter, R. (1987) Characterisation of the hydroxystreptomycin phosphotransferase gene (*sph*) of *Streptomyces glaucescens*: nucleotide sequence and promoter analysis. *Mol Gen Genet* **208**(1-2), 195-203.

Volff, J.N. & Altenbuchner, J. (2000) A new beginning with new ends: linearisation of circular chromosomes during bacterial evolution. *FEMS Microbiol Lett* **186**(2), 143-50.

von Ahsen, U. & Noller, H.F. (1993) Footprinting the sites of interaction of antibiotics with catalytic group I intron RNA. *Science* **260**(5113), 1500-3.

von Wintzingerode, F., Gobel, U.B. & Stackebrandt, E. (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* **21**(3), 213-29.

Vos, S., Berrisford, D.J. & Avis, J.M. (2002) Effect of magnesium ions on the tertiary structure of the hepatitis C virus IRES and its affinity for the cyclic peptide antibiotic viomycin. *Biochemistry* **41**(17), 5383-96.

Vujaklija, D., Horinouchi, S. & Beppu, T. (1993) Detection of an A-factor-responsive protein that binds to the upstream activation sequence of *strR*, a regulatory gene for streptomycin biosynthesis in *Streptomyces griseus*. *J Bacteriol* **175**(9), 2652-61.

Waksman, S.A. & Woodruff, H.B. (1940) The soil as a source of microorganisms antagonistic to disease-producing bacteria. *J Bacteriol* **40**(4), 581-600.

Wallace, S.T. & Schroeder, R. (1998) In vitro selection and characterization of streptomycin-binding RNAs: recognition discrimination between antibiotics. *RNA* **4**(1), 112-23.

Wallis, M.G., Streicher, B., Wank, H., von Ahsen, U., Clodi, E., Wallace, S.T., Famulok, M. & Schroeder, R. (1997) In vitro selection of a viomycin-binding RNA pseudoknot. *Chem Biol* **4**(5), 357-66.

Wang, R., Alexander, R.W., VanLoock, M., Vladimirov, S., Bukhtiyarov, Y., Harvey, S.C. & Cooperman, B.S. (1999a) Three-dimensional placement of the conserved 530 loop of 16 S rRNA and of its neighboring components in the 30 S subunit. *J Mol Biol* **286**(2), 521-40.

Wang, Y., Zhang, Z.S., Ruan, J.S., Wang, Y.M. & Ali, S.M. (1999b) Investigation of actinomycete diversity in the tropical rainforests of Singapore. *J Ind Microbiol Biotechnol* **23**(3), 178-187.

Wang, Z.L., Zhang, J.G. & Liu, X.M. (2000) Competition and coexistence in spatially subdivided habitats. *J Theor Biol* **205**(4), 631-9.

Wank, H., Clodi, E., Wallis, M.G. & Schroeder, R. (1999) The antibiotic viomycin as a model peptide for the origin of the co-evolution of RNA and proteins. *Orig Life Evol Biosph* **29**(4), 391-404.

Wank, H., Rogers, J., Davies, J. & Schroeder, R. (1994) Peptide antibiotics of the tuberactinomycin family as inhibitors of group I intron RNA splicing. *J Mol Biol* **236**(4), 1001-10.

Wank, H. & Schroeder, R. (1996) Antibiotic-induced oligomerisation of group I intron RNA. *J Mol Biol* **258**(1), 53-61.

Wellington, E.M., Stackebrandt, E., Sanders, D., Wolstrup, J. & Jorgensen, N.O. (1992) Taxonomic status of *Kitasatosporia*, and proposed unification with *Streptomyces* on the basis of phenotypic and 16S rRNA analysis and emendation of *Streptomyces* Waksman and Henrici 1943, 339AL. *Int J Syst Bacteriol* **42**(1), 156-60.

Werner, G., Hildebrandt, B. & Witte, W. (2001) Aminoglycoside-streptothricin resistance gene cluster *aadE-sat4-aphA-3* disseminated among multiresistant isolates of *Enterococcus faecium*. *Antimicrob Agents Chemother* **45**(11), 3267-9.

Wiener, P., Egan, S., Huddleston, A.S. & Wellington, E.M. (1998) Evidence for transfer of antibiotic-resistance genes in soil populations of streptomycetes. *Mol Ecol* **7**(9), 1205-16.

Williams, D.H., Stone, M.J., Hauck, P.R. & Rahman, S.K. (1989) Why are secondary metabolites (natural products) biosynthesized? *J Nat Prod* **52**(6), 1189-208.

Williams, S.T., Goodfellow, M., Alderson, G., Wellington, E.M., Sneath, P.H. & Sackin, M.J. (1983) Numerical classification of *Streptomyces* and related genera. *J Gen Microbiol* **129**(6), 1743-813.

Wink, J.M. (2009) Compendium of Actinobacteria. Vol. 2009.

Woese, C.R. & Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74**(11), 5088-90.

Wright, G.D. (2005) Bacterial resistance to antibiotics: enzymatic degradation and modification. *Adv Drug Deliv Rev* **57**(10), 1451-70.

Wright, G.D. (2007) The antibiotic resistome: the nexus of chemical and genetic diversity. *Nat Rev Microbiol* **5**(3), 175-86.

Wright, G.D. & Thompson, P.R. (1999) Aminoglycoside phosphotransferases: proteins, structure, and mechanism. *Front Biosci* **4**, D9-21.

Xu, L., Li, Q. & Jiang, C. (1996) Diversity of soil actinomycetes in Yunnan, China. *Appl Environ Microbiol* **62**(1), 244-248.

Yamada, T., Nagata, A., Ono, Y., Suzuki, Y. & Yamanouchi, T. (1985) Alteration of ribosomes and RNA polymerase in drug-resistant clinical isolates of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **27**(6), 921-4.

Yamamoto, S., He, Y., Arakawa, K. & Kinashi, H. (2008) Gamma-butyrolactone-dependent expression of the *Streptomyces* antibiotic regulatory protein gene *srrY* plays a central role in the regulatory cascade leading to lankacidin and lankamycin production in *Streptomyces rochei*. *J Bacteriol* **190**(4), 1308-16.

Yamazaki, H., Tomono, A., Ohnishi, Y. & Horinouchi, S. (2004) DNA-binding specificity of AdpA, a transcriptional activator in the A-factor regulatory cascade in *Streptomyces griseus*. *Mol Microbiol* **53**(2), 555-72.

Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**(8), 1586-91.

Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A.M. (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**(1), 431-49.

Yates, J.L. (1979) Role of ribosomal protein S12 in discrimination of aminoacyl-tRNA. *J Biol Chem* **254**(22), 11550-4.

Yeh, K.C., To, K.Y., Sun, S.W., Wu, M.C., Lin, T.Y. & Chen, C.C. (1994) Point mutations in the chloroplast 16s rRNA gene confer streptomycin resistance in *Nicotiana plumbaginifolia*. *Curr Genet* **26**(2), 132-5.

Yergeau, E., Newsham, K.K., Pearce, D.A. & Kowalchuk, G.A. (2007) Patterns of bacterial diversity across a range of Antarctic terrestrial habitats. *Environ Microbiol* **9**(11), 2670-82.

Yim, G., Wang, H.H. & Davies, J. (2006) The truth about antibiotics. *Int J Med Microbiol* **296**(2-3), 163-70.

Yim, G., Wang, H.H. & Davies, J. (2007) Antibiotics as signalling molecules. *Philos Trans R Soc Lond B Biol Sci* **362**(1483), 1195-200.

Yin, X., O'Hare, T., Gould, S.J. & Zabriskie, T.M. (2003) Identification and cloning of genes encoding viomycin biosynthesis from *Streptomyces vinaceus* and evidence for involvement of a rare oxygenase. *Gene* **312**, 215-24.

Youssef, N.H. & Elshahed, M.S. (2009) Diversity rankings among bacterial lineages in soil. *ISME J* **3**(3), 305-13.

Zhang, Z., Wang, Y. & Ruan, J. (1997) A proposal to revive the genus *Kitasatospora* (Omura, Takahashi, Iwai, and Tanaka 1982). *Int J Syst Bacteriol* **47**(4), 1048-54.

Zhi, X.Y., Li, W.J. & Stackebrandt, E. (2009) An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class Actinobacteria, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. *Int J Syst Evol Microbiol* **59**(Pt 3), 589-608.

10 Appendix – Supplementary Figures

10.1 Percentage identity of resistance genes

Percent Identity

Divergence

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1		99.8	72.8	73.0	73.2	73.0	72.6	72.8	72.6	75.4	73.4	76.7	76.0	74.9	75.4
2	0.2		72.8	73.0	73.2	73.0	72.6	72.8	72.6	75.4	73.4	76.7	76.0	74.9	75.4
3	34.0	34.0		99.6	99.4	99.8	99.4	99.6	99.4	83.2	82.9	82.9	83.2	83.4	83.2
4	33.7	33.7	0.4		99.8	99.8	99.4	99.6	99.4	83.4	83.1	83.1	83.4	83.6	83.4
5	33.4	33.4	0.6	0.2		99.6	99.3	99.4	99.3	83.6	83.2	83.2	83.6	83.8	83.6
6	33.7	33.7	0.2	0.2	0.4		99.6	99.8	99.6	83.4	83.1	83.1	83.4	83.6	83.4
7	34.3	34.3	0.6	0.6	0.8	0.4		99.4	99.3	83.1	82.7	82.7	83.1	83.2	83.1
8	34.0	34.0	0.4	0.4	0.6	0.2	0.6		99.8	83.2	82.9	82.9	83.2	83.4	83.2
9	34.3	34.3	0.6	0.6	0.8	0.4	0.8	0.2		83.1	82.7	82.7	83.1	83.2	83.1
10	30.0	30.0	19.2	18.9	18.7	19.0	19.5	19.2	19.5		92.7	95.5	98.7	95.2	99.6
11	33.1	33.1	19.8	19.5	19.3	19.5	20.0	19.8	20.0	7.7		92.7	93.3	92.7	92.7
12	28.1	28.1	19.7	19.4	19.2	19.5	20.0	19.7	20.0	4.6	7.7		96.1	96.3	95.5
13	29.2	29.2	19.2	18.9	18.7	18.9	19.4	19.2	19.4	1.3	7.1	4.0		95.7	98.7
14	30.8	30.8	19.0	18.7	18.5	18.7	19.2	19.0	19.2	5.0	7.7	3.9	4.4		95.2
15	30.0	30.0	19.2	18.9	18.7	19.0	19.5	19.2	19.5	0.4	7.7	4.6	1.3	5.0	
16	28.6	28.6	19.5	19.2	18.9	19.2	19.7	19.5	19.7	8.7	8.2	7.3	8.1	7.5	8.7
17	29.2	29.2	18.7	18.5	18.2	18.5	19.0	18.7	19.0	6.0	7.9	4.8	5.8	5.4	6.0
18	28.7	28.7	19.5	19.2	19.0	19.2	19.8	19.5	19.8	7.9	7.7	7.1	7.7	7.1	7.9
19	29.7	29.7	19.5	19.2	18.9	19.2	19.7	19.5	19.7	6.7	8.3	5.2	6.2	5.8	6.7
20	29.8	29.8	20.0	19.7	19.5	19.8	20.3	20.0	20.3	8.9	7.9	7.9	7.7	7.9	8.9
21	30.6	30.6	19.6	19.3	19.0	19.3	19.8	19.6	19.8	8.7	5.2	7.9	8.1	7.9	8.7
22	26.0	26.0	21.6	21.4	21.1	21.4	21.9	21.6	21.9	20.1	21.1	19.3	19.8	20.6	20.1
23	30.3	30.0	32.0	32.3	32.0	32.3	32.8	32.5	32.8	27.8	29.5	27.9	28.6	30.3	28.1
24	19.1	18.9	29.0	28.7	28.4	28.7	29.3	29.0	29.3	26.2	27.0	25.6	26.2	26.2	26.2
25	30.0	30.0	23.9	23.6	23.4	23.6	24.2	23.9	24.2	20.6	20.8	20.8	20.6	20.8	20.6
26	28.7	28.4	27.5	27.3	27.0	27.3	27.8	27.5	27.8	23.4	23.1	23.4	24.1	24.4	23.4
27	29.5	29.3	26.5	26.2	25.9	26.2	26.8	26.5	26.8	22.8	22.8	23.3	23.6	24.3	22.8
28	22.1	21.8	29.4	29.2	29.5	29.2	29.7	29.4	29.7	28.1	29.2	27.3	28.1	29.2	28.1
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**

Figure 10.1. Percentage identity and divergence of partial *strA* nucleotide sequences. The following strains were omitted from the table due to them having an identical sequence to another strain: DSM 40627, DSM 40653, DSM 40654, DSM 40657, DSM 40658, DSM 40659, DSM 40660, DSM 40670, DSM 40759, DSM 40855, DSM 40878, CR13, CB162, CB163, NBRC 13350 to DSM 40236^{T}; DSM 40776 to DSM 40757; CB145 to CB142; CB153, CB154 to CB147. **Figure continues on the next page.**

Percent Identity

16	17	18	19	20	21	22	23	24	25	26	27	28		
76.4	76.0	76.4	75.6	75.0	75.0	78.2	75.2	83.2	75.4	76.4	75.8	81.0	**1**	DSM40236
76.4	76.0	76.4	75.6	75.0	75.0	78.2	75.4	83.4	75.4	76.5	76.0	81.2	**2**	Z34
83.1	83.6	83.1	83.1	82.1	83.1	81.4	74.1	76.2	79.7	77.1	77.8	75.8	**3**	DSM40265
83.2	83.8	83.2	83.2	82.3	83.2	81.6	73.9	76.4	79.9	77.3	78.0	76.0	**4**	DSM40757
83.4	84.0	83.4	83.4	82.5	83.4	81.8	74.1	76.5	80.1	77.5	78.2	75.8	**5**	CW12
83.2	83.8	83.2	83.2	82.3	83.2	81.6	73.9	76.4	79.9	77.3	78.0	76.0	**6**	CW15
82.9	83.4	82.9	82.9	81.9	82.9	81.2	73.6	76.0	79.5	76.9	77.7	75.6	**7**	DW15
83.1	83.6	83.1	83.1	82.1	83.1	81.4	73.7	76.2	79.7	77.1	77.8	75.8	**8**	E989
82.9	83.4	82.9	82.9	81.9	82.9	81.2	73.6	76.0	79.5	76.9	77.7	75.6	**9**	E1013
91.8	94.2	92.6	93.7	91.1	91.8	82.5	76.9	78.0	82.1	80.1	80.4	76.7	**10**	DSM40136
92.4	92.6	92.7	92.2	92.0	95.0	81.8	75.8	77.5	81.9	80.3	80.4	76.0	**11**	DSM40707
93.1	95.3	93.3	95.0	92.0	92.6	83.1	76.9	78.4	81.9	80.1	80.1	77.3	**12**	DSM40939
92.4	94.4	92.7	94.0	92.2	92.4	82.7	76.4	78.0	82.1	79.5	79.9	76.7	**13**	666
92.9	94.8	93.3	94.4	92.0	92.6	82.1	75.2	78.0	81.9	79.3	79.3	76.0	**14**	AR23
91.8	94.2	92.6	93.7	91.1	91.8	82.5	76.7	78.0	82.1	80.1	80.4	76.7	**15**	DW21
	91.2	95.0	90.9	90.7	91.4	83.4	76.0	77.7	81.4	79.9	79.9	76.2	**16**	CB142
9.4		90.9	99.3	92.2	92.6	82.5	76.0	78.8	83.4	80.3	80.6	77.7	**17**	CB147
5.3	9.8		90.3	91.1	92.2	84.0	77.5	78.8	82.7	80.1	79.9	76.4	**18**	CB158
9.8	0.8	10.4		91.8	92.2	81.9	75.6	78.2	82.9	79.7	80.1	77.1	**19**	CB172
9.4	7.7	9.0	8.1		90.9	83.1	76.7	78.0	82.9	81.8	81.8	77.1	**20**	E996
9.2	7.9	8.3	8.3	9.2		81.6	74.5	76.4	80.6	79.7	80.1	75.8	**21**	RB063
18.9	20.1	18.1	20.8	18.6	21.4		78.0	80.4	83.4	82.5	81.9	79.9	**22**	DSM40155
29.2	29.2	27.0	29.8	27.3	31.5	26.2		79.3	76.0	81.2	79.1	79.0	**23**	DSM40263
26.7	25.1	25.1	25.9	25.4	28.7	22.8	24.3		78.8	79.5	78.4	84.9	**24**	CR50
21.6	18.8	19.8	19.6	18.9	22.6	18.9	29.2	25.2		80.3	79.7	76.4	**25**	DSM40093
23.6	23.1	23.4	23.9	20.4	23.9	20.1	21.9	24.1	23.1		95.2	80.6	**26**	DSM40005
23.6	22.6	23.6	23.3	20.3	23.3	20.8	24.7	25.7	23.8	5.0		79.9	**27**	DSM40221
28.9	26.7	28.6	27.5	26.7	29.4	23.6	24.9	17.1	28.6	22.6	23.6		**28**	DSM40480
16	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**		

Percent Identity

Divergence

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1		70.4	70.9	70.9	69.8	70.4	70.4	74.9	72.6	75.4	76.0	73.2	74.9	74.3
2	37.9		99.4	98.9	98.3	98.9	98.3	81.6	81.0	82.7	81.6	82.1	81.6	80.4
3	37.0	0.6		99.4	98.9	99.4	98.9	82.1	81.6	83.2	82.1	82.7	82.1	81.0
4	37.0	1.1	0.6		98.3	98.9	98.9	82.7	82.1	83.8	82.7	83.2	82.7	81.6
5	38.8	1.7	1.1	1.7		98.3	97.8	81.6	80.4	82.1	81.0	81.6	81.0	79.9
6	37.9	1.1	0.6	1.1	1.7		99.4	81.6	81.0	82.7	81.6	82.1	81.6	80.4
7	37.9	1.7	1.1	1.1	2.3	0.6		81.6	81.0	82.7	81.6	82.1	81.6	80.4
8	30.9	21.4	20.6	19.9	21.4	21.4	21.4		92.7	97.2	97.8	94.4	98.9	91.6
9	34.3	22.1	21.4	20.6	22.9	22.1	22.1	7.7		93.3	92.7	91.1	92.7	91.6
10	30.0	19.9	19.1	18.4	20.6	19.9	19.9	2.9	7.1		97.2	96.1	97.2	91.6
11	29.2	21.4	20.6	19.9	22.1	21.4	21.4	2.3	7.7	2.9		94.4	97.8	91.1
12	33.4	20.6	19.9	19.1	21.4	20.6	20.6	5.8	9.6	4.0	5.8		94.4	89.9
13	30.9	21.4	20.6	19.9	22.1	21.4	21.4	1.1	7.7	2.9	2.3	5.8		91.6
14	31.7	22.9	22.1	21.4	23.6	22.9	22.9	9.0	9.0	9.0	9.6	10.9	9.0	
15	30.0	19.1	18.4	17.7	19.9	18.4	18.4	5.2	7.7	4.0	5.2	8.3	5.2	10.9
16	30.9	22.9	22.1	21.4	23.6	22.9	22.9	7.7	7.7	8.3	7.7	9.0	7.7	5.8
17	31.7	20.6	19.9	19.1	21.4	19.9	19.9	6.5	9.0	5.2	6.5	9.6	6.5	12.2
18	30.9	26.0	25.2	24.4	26.8	26.0	26.0	10.2	10.2	8.3	8.3	10.2	10.2	11.5
19	34.3	20.6	19.9	19.1	21.4	20.6	20.6	7.7	4.0	7.7	8.3	10.2	7.7	8.3
20	21.4	25.2	24.4	24.4	26.0	25.2	25.2	24.4	25.2	21.4	22.9	24.4	24.4	22.1
21	30.0	36.1	37.0	37.0	37.9	37.9	37.9	33.4	35.2	34.3	34.3	36.1	35.2	36.1
22	18.4	33.4	32.6	32.6	34.3	33.4	33.4	30.9	30.0	30.0	29.2	31.7	30.9	33.4
23	27.6	26.0	25.2	25.2	26.8	26.0	26.0	25.2	25.2	22.9	24.4	25.2	25.2	26.0
24	30.9	28.4	27.6	27.6	29.2	28.4	28.4	23.6	22.1	20.6	23.6	24.4	23.6	26.0
25	33.4	29.2	28.4	28.4	30.0	29.2	29.2	23.6	22.1	20.6	23.6	24.4	23.6	26.8
26	21.4	34.3	33.4	33.4	35.2	34.3	34.3	28.4	27.6	26.0	26.8	30.9	28.4	29.2
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**

Figure 10.2. Percentage identity and divergence of partial StrA amino acid sequences. The following strains were omitted from the table due to them having an identical sequence to another strain: DSM 40627, DSM 40653, DSM 40654, DSM 40657, DSM 40658, DSM 40659, DSM 40660, DSM 40670, DSM 40759, DSM 40855, DSM 40878, CR13, CB162, CB163, Z34, NBRC 13350 to DSM 40236^{T}; DSM 40757, DSM 40776 to CW15; CB145 to CB142; CB154, CB153 to CB147. **Figure continues on the next page.**

Percent Identity

15	16	17	18	19	20	21	22	23	24	25	26		
75.4	74.9	74.3	74.3	72.6	81.6	75.4	83.8	77.1	74.9	73.2	81.6	**1**	DSM40236
83.2	80.4	82.1	77.7	82.1	78.8	71.5	73.2	78.2	76.5	76.0	72.6	**2**	DSM40265
83.8	81.0	82.7	78.2	82.7	79.3	70.9	73.7	78.8	77.1	76.5	73.2	**3**	DSM40757
84.4	81.6	83.2	78.8	83.2	79.3	70.9	73.7	78.8	77.1	76.5	73.2	**4**	CW12
82.7	79.9	81.6	77.1	81.6	78.2	70.4	72.6	77.7	76.0	75.4	72.1	**5**	DW15
83.8	80.4	82.7	77.7	82.1	78.8	70.4	73.2	78.2	76.5	76.0	72.6	**6**	E989
83.8	80.4	82.7	77.7	82.1	78.8	70.4	73.2	78.2	76.5	76.0	72.6	**7**	E1013
95.0	92.7	93.9	89.9	92.7	79.3	73.2	74.9	78.8	79.9	79.9	76.5	**8**	DSM40136
92.7	92.7	91.6	89.9	96.1	78.8	72.1	75.4	78.8	81.0	81.0	77.1	**9**	DSM40707
96.1	92.2	95.0	91.6	92.7	81.6	72.6	75.4	80.4	82.1	82.1	78.2	**10**	DSM40939
95.0	92.7	93.9	91.6	92.2	80.4	72.6	76.0	79.3	79.9	79.9	77.7	**11**	666
92.2	91.6	91.1	89.9	90.5	79.3	71.5	74.3	78.8	79.3	79.3	74.9	**12**	AR23
95.0	92.7	93.9	89.9	92.7	79.3	72.1	74.9	78.8	79.9	79.9	76.5	**13**	DW21
89.9	94.4	88.8	88.8	92.2	81.0	71.5	73.2	78.2	78.2	77.7	76.0	**14**	CB142
	89.4	98.9	89.4	92.2	79.9	72.1	76.0	80.4	80.4	80.4	78.2	**15**	CB147
11.5		88.3	88.8	91.1	80.4	73.2	76.0	79.3	78.8	79.3	76.0	**16**	CB158
1.1	12.9		88.3	91.1	78.8	70.9	74.9	79.3	79.3	79.3	77.1	**17**	CB172
10.9	11.5	12.2		88.8	80.4	71.5	74.3	77.7	79.3	79.3	76.5	**18**	E996
8.3	9.6	9.6	11.5		79.3	72.1	74.9	78.8	79.9	79.9	76.5	**19**	RB063
23.6	22.9	25.2	22.1	24.4		75.4	77.7	83.8	79.3	78.2	79.9	**20**	DSM40155
35.2	33.4	37.0	35.2	35.2	30.0		76.5	73.7	76.5	75.4	78.2	**21**	DSM40263
29.2	29.2	30.9	30.9	30.9	26.8	28.4		78.8	76.0	75.4	82.1	**22**	CR50
22.9	24.4	24.4	26.0	25.2	18.4	32.6	25.2		78.8	77.7	77.1	**23**	DSM40093
22.9	25.2	24.4	23.6	23.6	24.4	28.4	29.2	25.2		94.4	78.8	**24**	DSM40005
22.9	24.4	24.4	23.6	23.6	26.0	30.0	30.0	26.8	5.8		76.5	**25**	DSM40221
26.0	29.2	27.6	27.6	28.4	23.6	26.0	20.6	27.6	25.2	28.4		**26**	DSM40480
15	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**		

Percent Identity

	1	2	3	4	5	6	7	8	9	10	11	12		
1		99.0	99.0	99.0	98.4	79.6	79.8	79.8	76.2	76.4	76.6	76.6	1	DSM40257
2	1.0		99.6	99.6	99.0	80.0	80.2	80.6	76.6	76.8	77.0	77.0	2	DSM40058
3	1.0	0.4		99.6	99.0	80.2	80.4	80.8	76.8	77.0	77.2	77.2	3	DSM40083
4	1.0	0.4	0.4		99.4	80.0	80.2	80.6	76.6	76.8	77.0	77.0	4	DSM40938
5	1.6	1.0	1.0	0.6		79.4	79.8	80.0	77.0	76.8	77.0	77.0	5	CB157
6	23.9	23.4	23.1	23.4	24.2		98.6	97.8	81.0	81.2	81.0	81.0	6	DSM40066
7	23.7	23.1	22.8	23.1	23.7	1.4		98.8	81.0	81.2	81.0	81.0	7	DSM40598
8	23.7	22.6	22.3	22.6	23.4	2.3	1.2		80.2	80.4	80.2	80.2	8	DSM40864
9	28.8	28.2	27.9	28.2	27.6	22.1	22.1	23.2		99.2	98.8	99.0	9	DSM40236
10	28.5	27.9	27.6	27.9	27.9	21.8	21.8	22.9	0.8		99.2	99.0	10	DSM40659
11	28.2	27.6	27.3	27.6	27.6	22.1	22.1	23.2	1.2	0.8		99.8	11	DSM40654
12	28.2	27.6	27.3	27.6	27.6	22.1	22.1	23.2	1.0	1.0	0.2		12	Z34
	1	2	3	4	5	6	7	8	9	10	11	12		

Divergence

Figure 10.3. Percentage identity and divergence of partial *vph* nucleotide sequences. The following strains were omitted from the table due to them having an identical sequence to another strain: DSM 40463 to DSM 40066^{T}; DSM 40627, DSM 40660, DSM 40657, DSM 40855, DSM 40653, DSM 40658, DSM 40670, DSM 40759, NBRC13350, CB162, CB163, CR13 to DSM 40236^{T}; DSM 40878 to DSM 40654

Percent Identity

	1	2	3	4	5	6	7	8	9		
1		99.4	99.4	78.9	78.9	80.1	72.9	72.9	72.3	1	DSM40257
2	0.6		98.8	79.5	79.5	80.7	73.5	73.5	72.9	2	DSM40083
3	0.6	1.2		78.3	78.9	79.5	72.9	72.9	72.3	3	CB157
4	24.8	24.0	25.7		97.0	97.0	78.3	78.9	78.3	4	DSM40066
5	24.8	24.0	24.8	3.1		98.8	78.9	79.5	78.9	5	DSM40598
6	23.2	22.3	24.0	3.1	1.2		78.3	78.9	78.3	6	DSM40864
7	33.7	32.7	33.7	25.7	24.8	25.7		99.4	98.2	7	DSM40236
8	33.7	32.7	33.7	24.8	24.0	24.8	0.6		98.8	8	DSM40659
9	34.6	33.7	34.6	25.7	24.8	25.7	1.8	1.2		9	DSM40654
	1	2	3	4	5	6	7	8	9		

Divergence

Figure 10.4. Percentage identity and divergence of partial *vph* nucleotide sequences. The following strains were omitted from the table due to them having an identical sequence to another strain: DSM 40058^{T}, DSM 40938^{T} to DSM 40257; DSM 40066^{T} to DSM 40463; DSM 40627, DSM 40660, DSM 40657, DSM 40855, DSM 40653, DSM 40658, DSM 40670, DSM 40759, NBRC13350, CB162, CB163, CR13 to DSM 40236^{T}; DSM 40878, Z34 to DSM 40654

10.2 Phylogenetic trees of individual housekeeping genes

All the trees in this section were constructed using the neighbour-joining method; the numbers besides the branches indicate the percentage bootstrap value of 10000 replicates. The scale bar indicates 10% nucleotide or amino acid dissimilarity. ^{T} denotes a species' type strain.

Figure 10.5. Phylogenetic tree of the nucleotide sequence of the argininosuccinate lyase gene a*rgH*

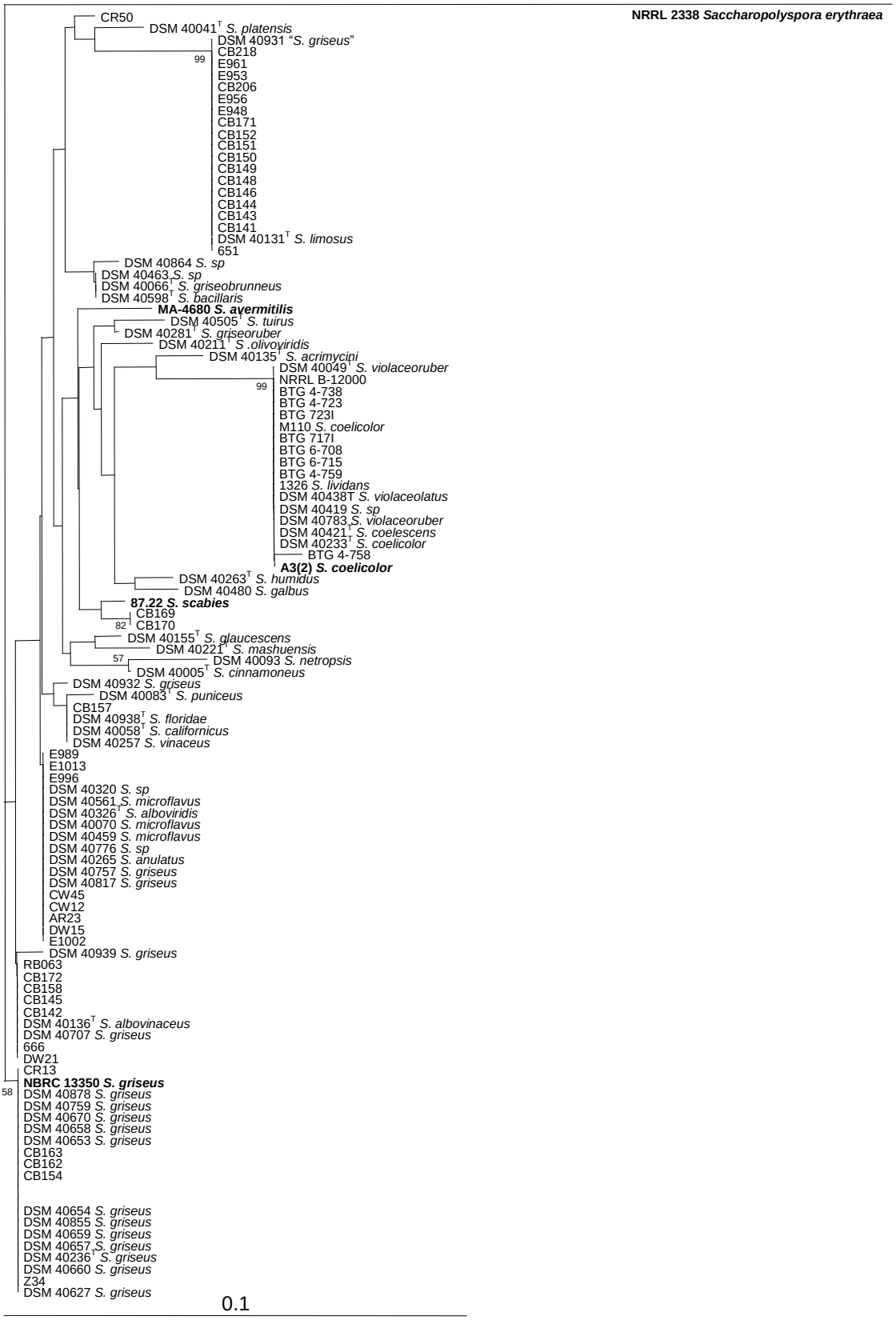


Figure 10.6. Phylogenetic tree of the amino acid sequence of the argininosuccinate lyase gene a*rgH*

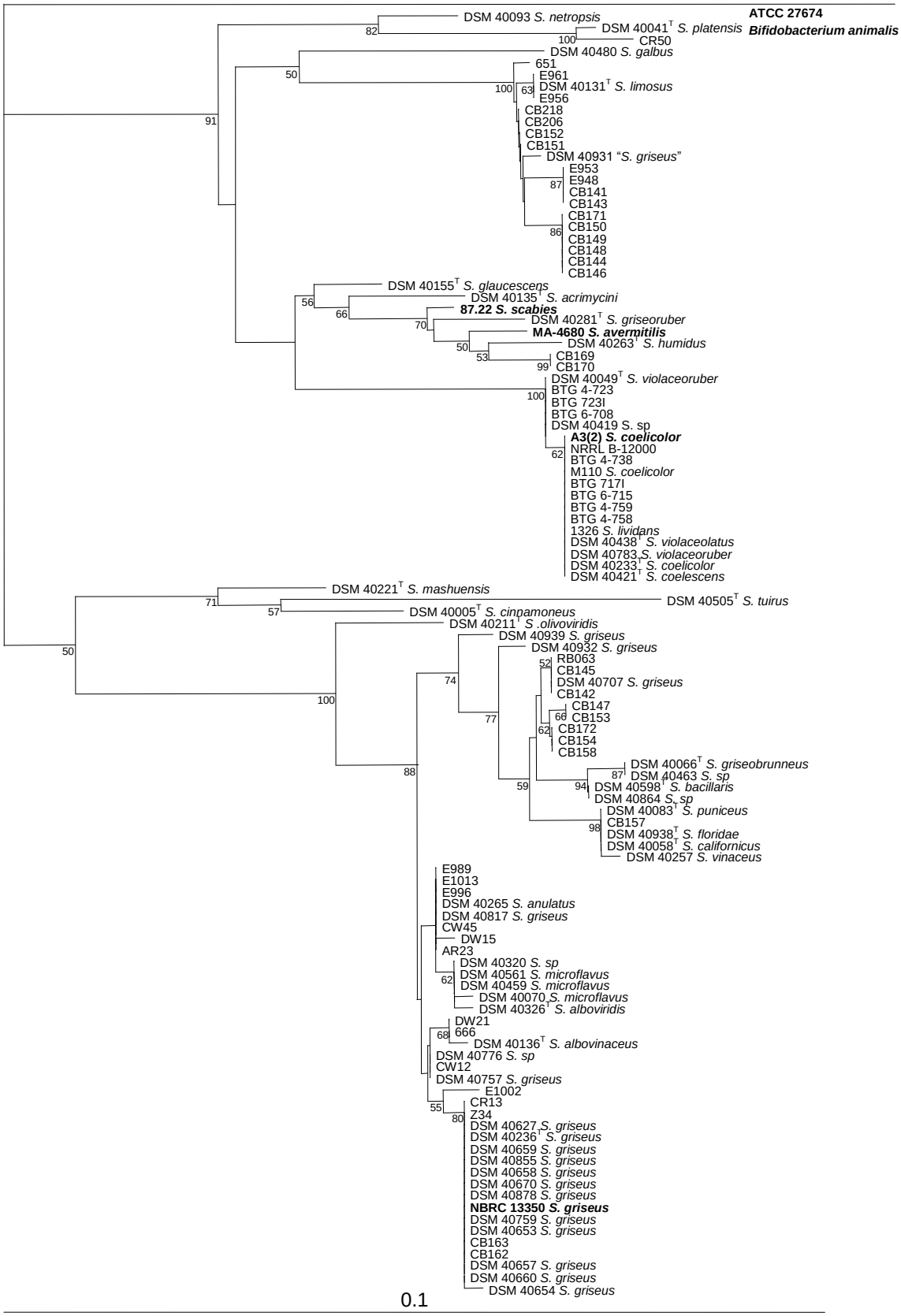


Figure 10.7. Phylogenetic tree of the nucleotide sequence of the ATP synthase beta chain gene *atpD*

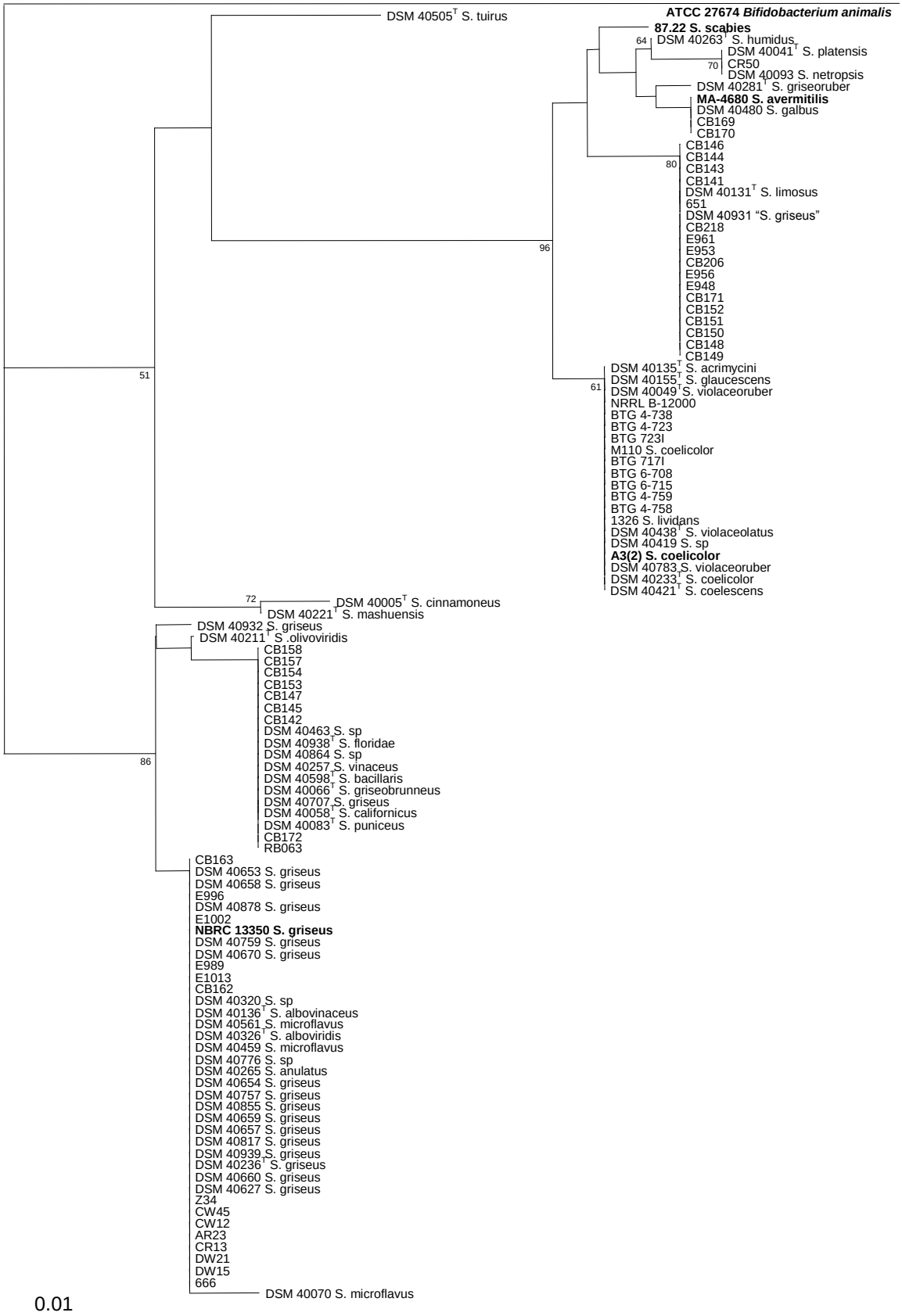


Figure 10.8. Phylogenetic tree of the amino acid sequence of the ATP synthase beta chain gene *atpD*

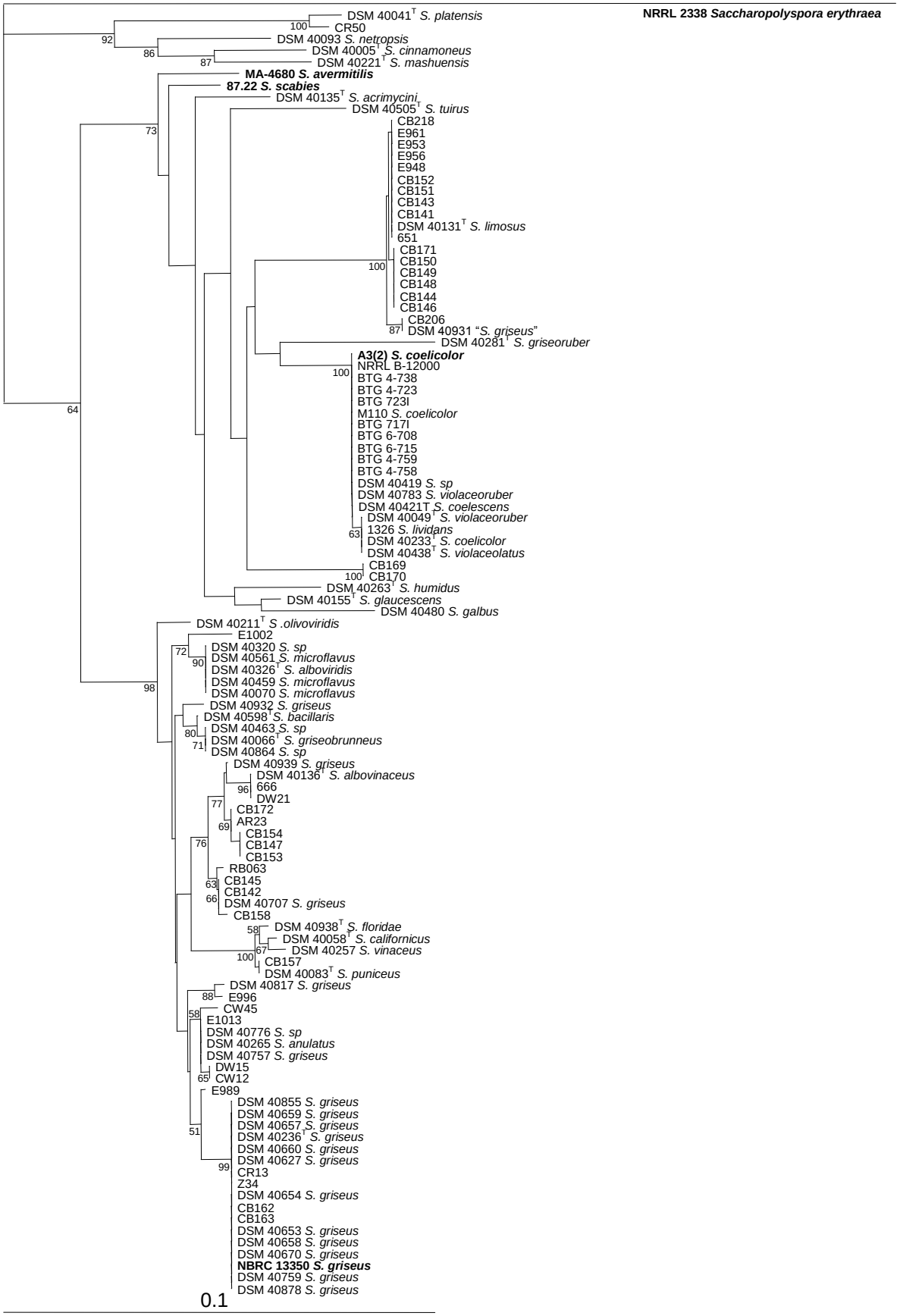


Figure 10.9. Phylogenetic tree of the nucleotide sequence of the elongation factor P gene *efp*

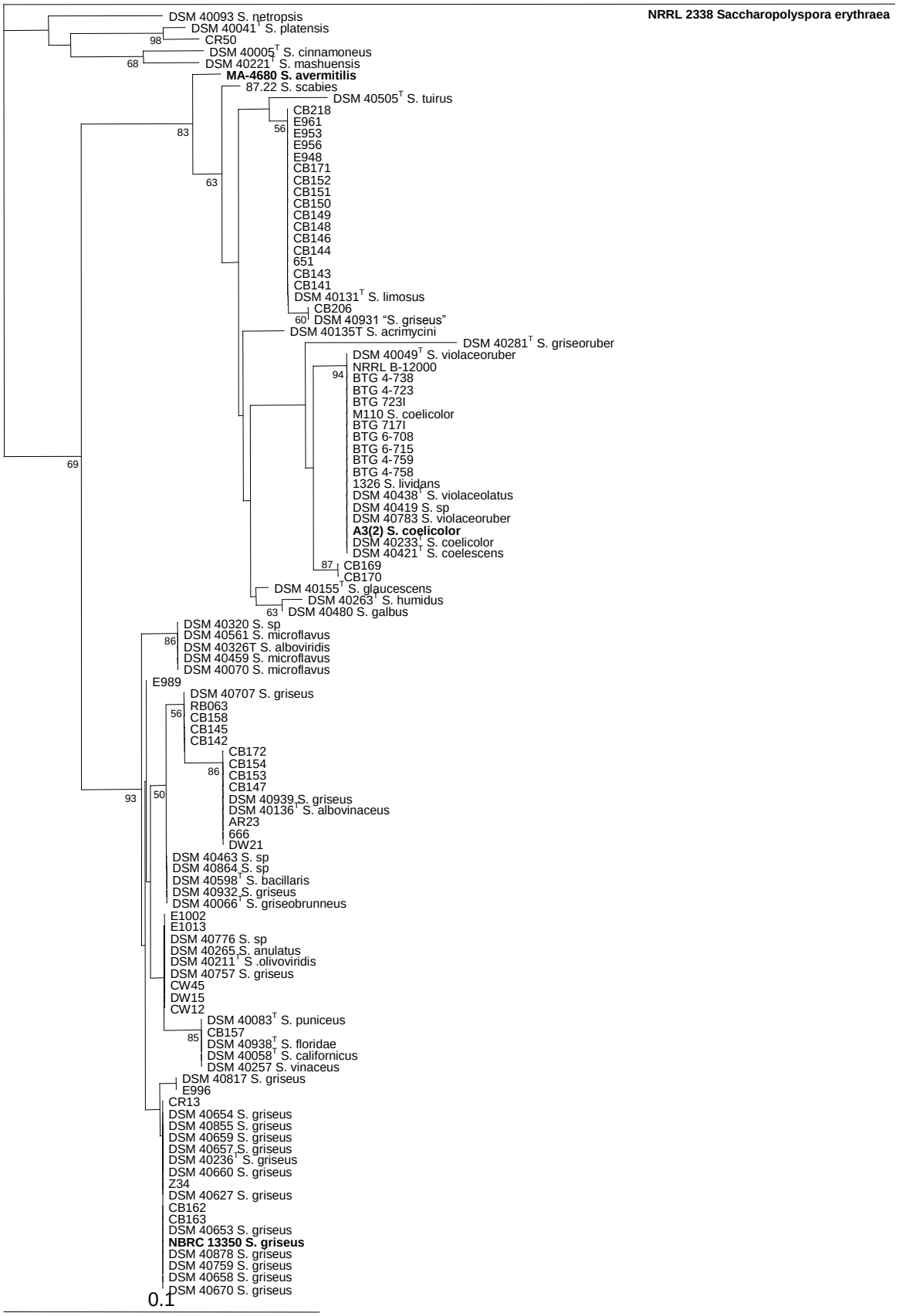


Figure 10.10. Phylogenetic tree of the amino acid sequence of the elongation factor P gene *efp*

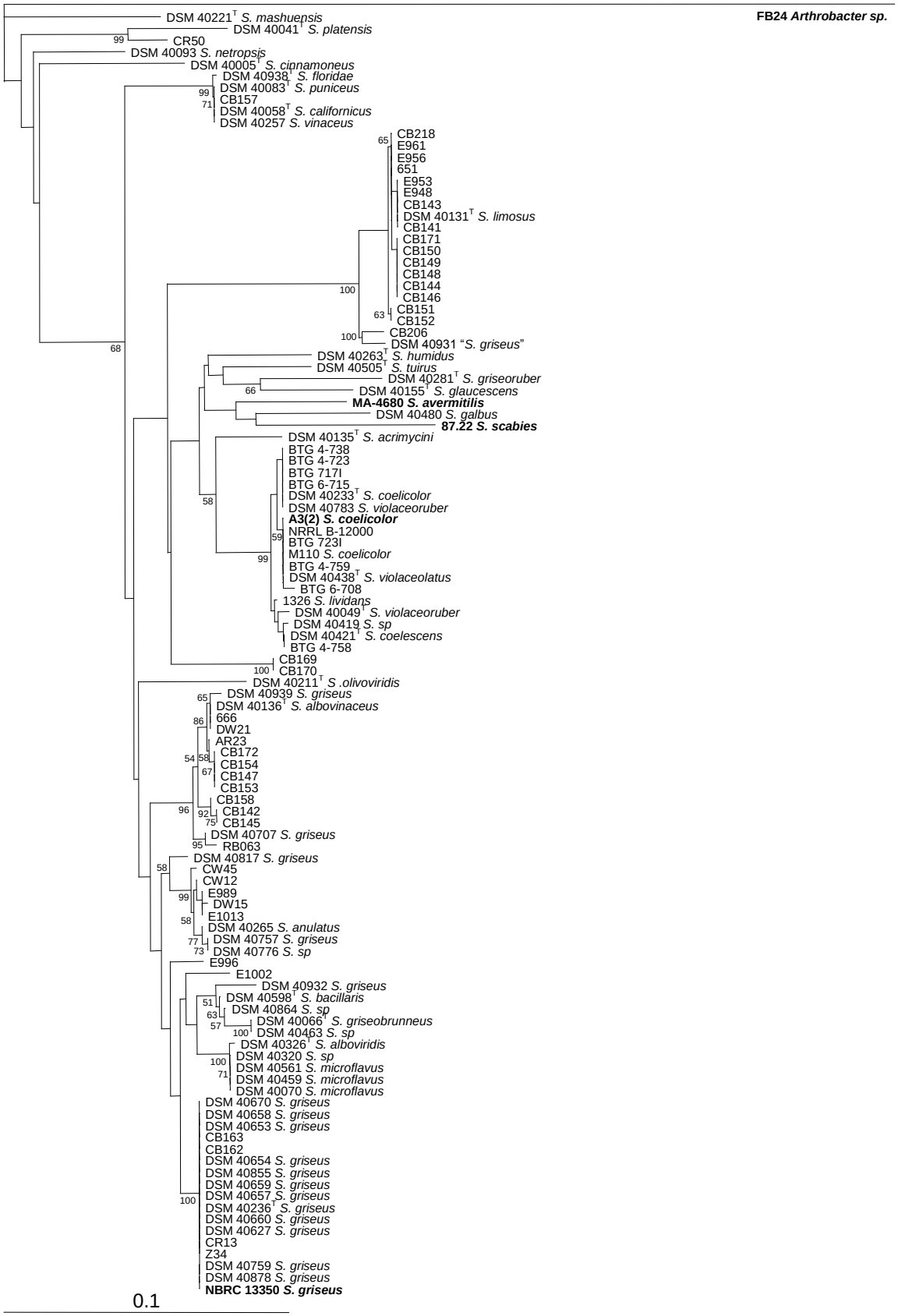


Figure 10.11. Phylogenetic tree of the nucleotide sequence of the phosphoenolpyruvate carboxylase gene *ppc*

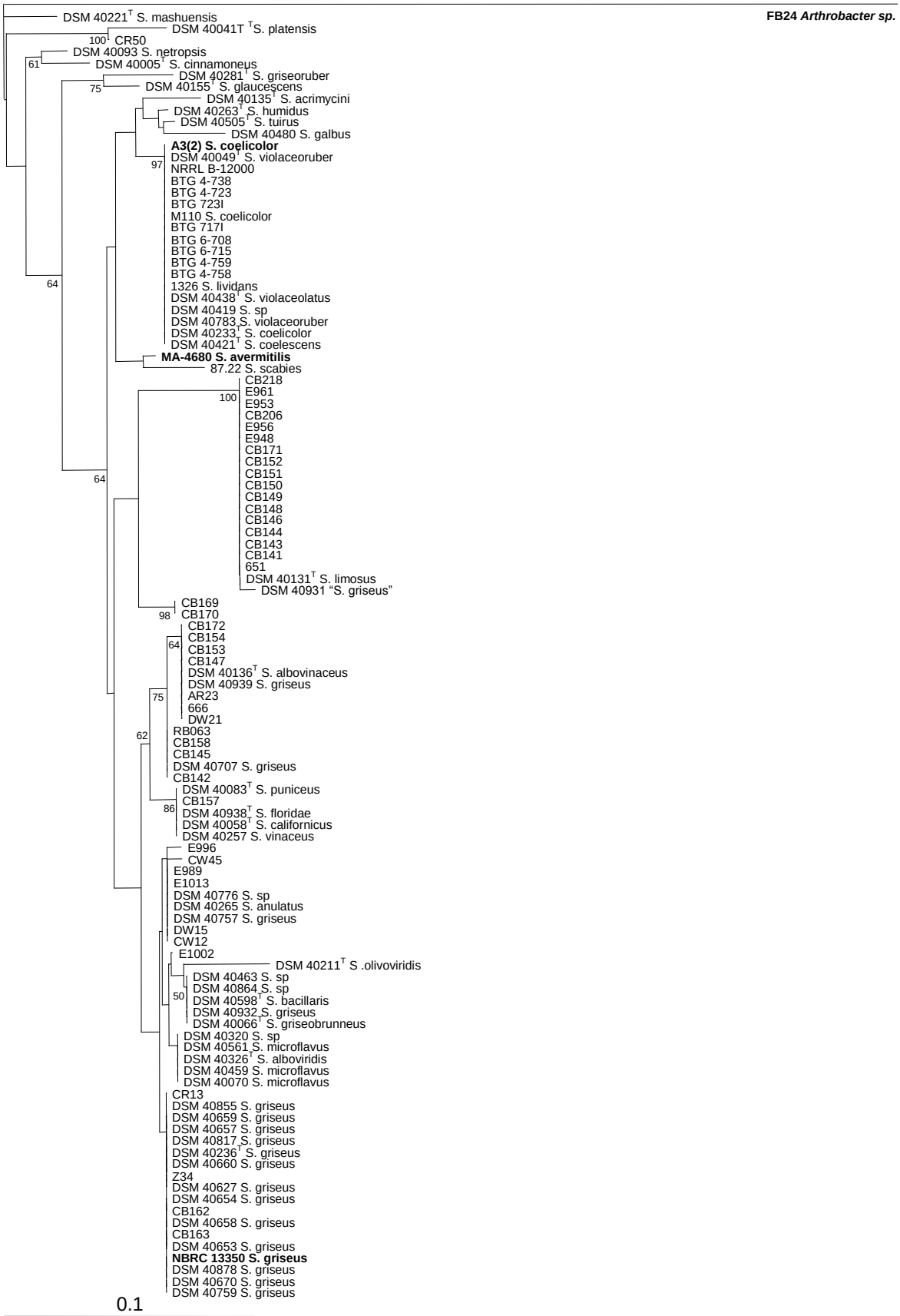


Figure 10.12. Phylogenetic tree of the amino acid sequence of the phosphoenolpyruvate carboxylase gene *ppc*

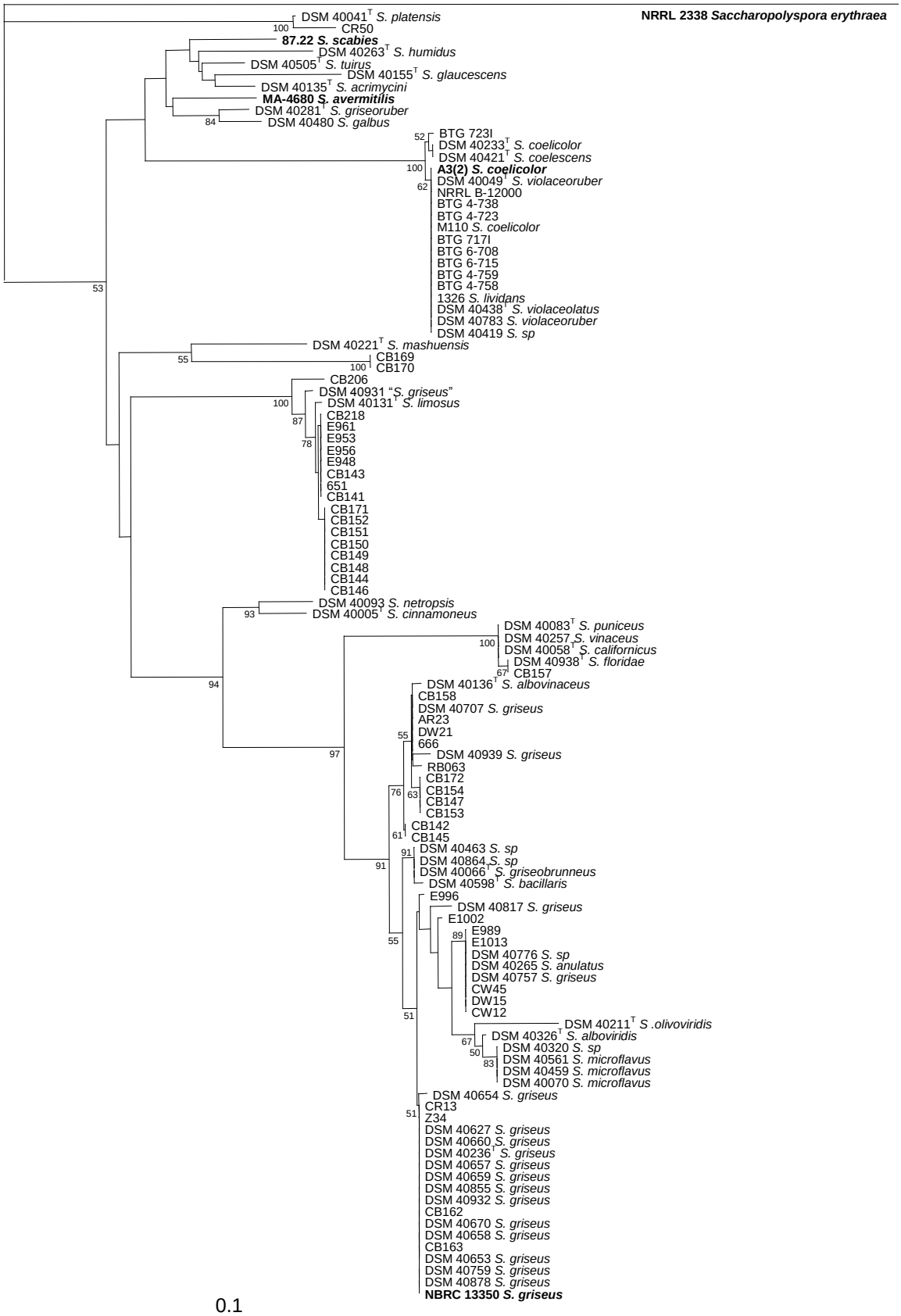


Figure 10.13. Phylogenetic tree of the nucleotide sequence of the ribosomal protein L3 gene *rplC*

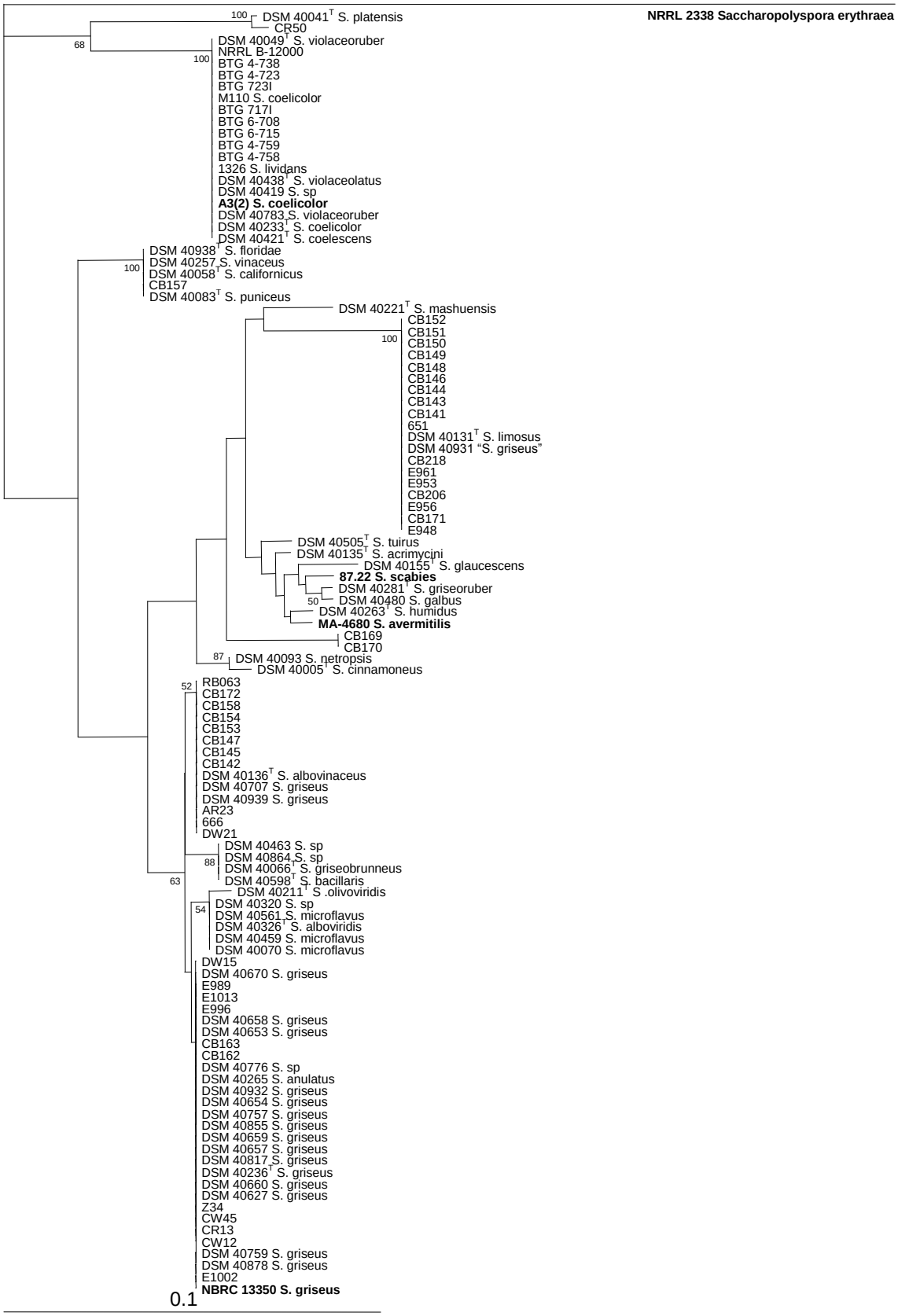


Figure 10.14. Phylogenetic tree of the nucleotide sequence of the ribosomal protein L3 gene *rplC*

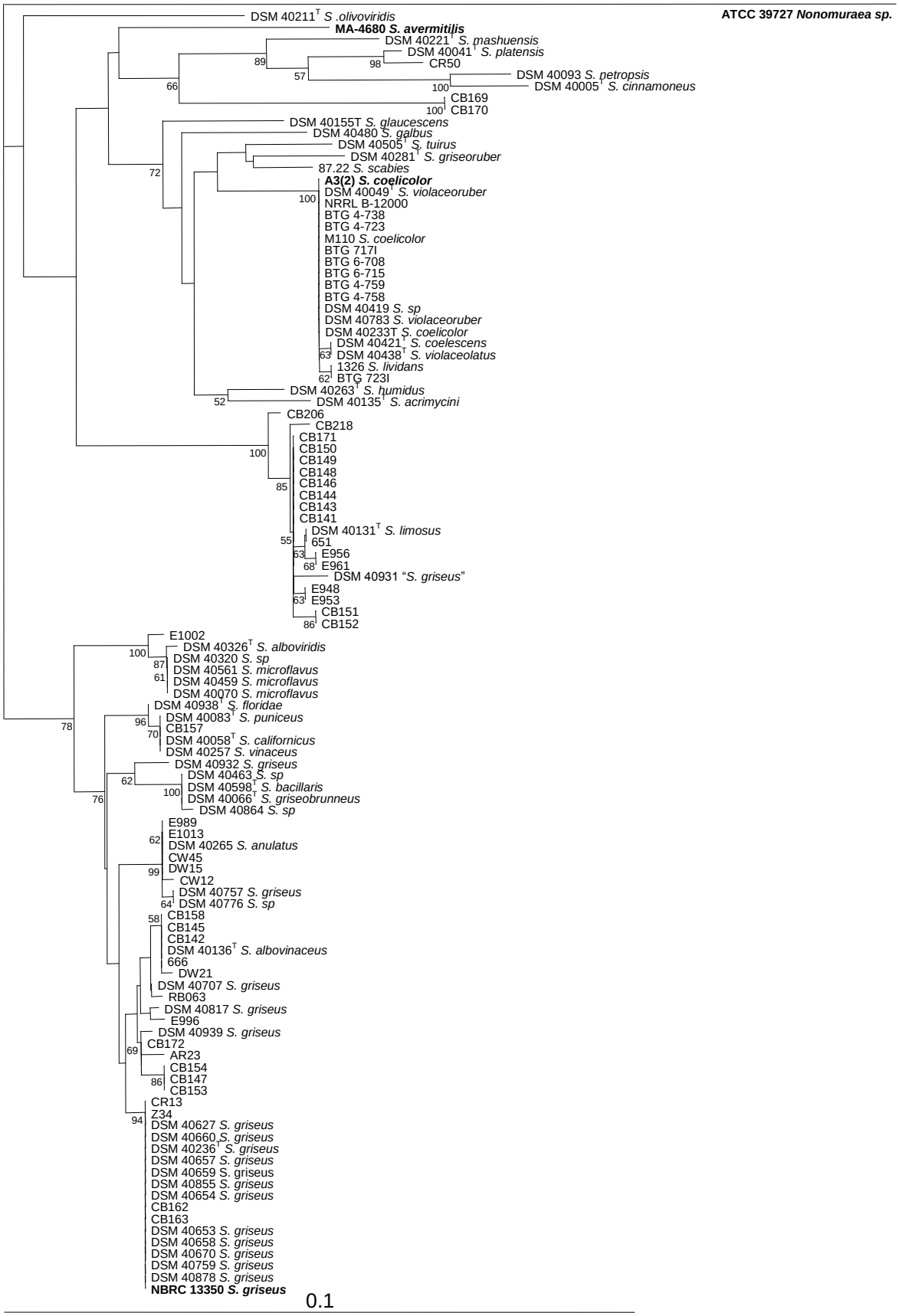


Figure 10.15. Phylogenetic tree of the nucleotide sequence of the DNA-directed RNA polymerase beta subunit gene *rpoB*

Figure 10.16. Phylogenetic tree of the amino acid sequence of the DNA-directed RNA polymerase beta subunit gene *rpoB*

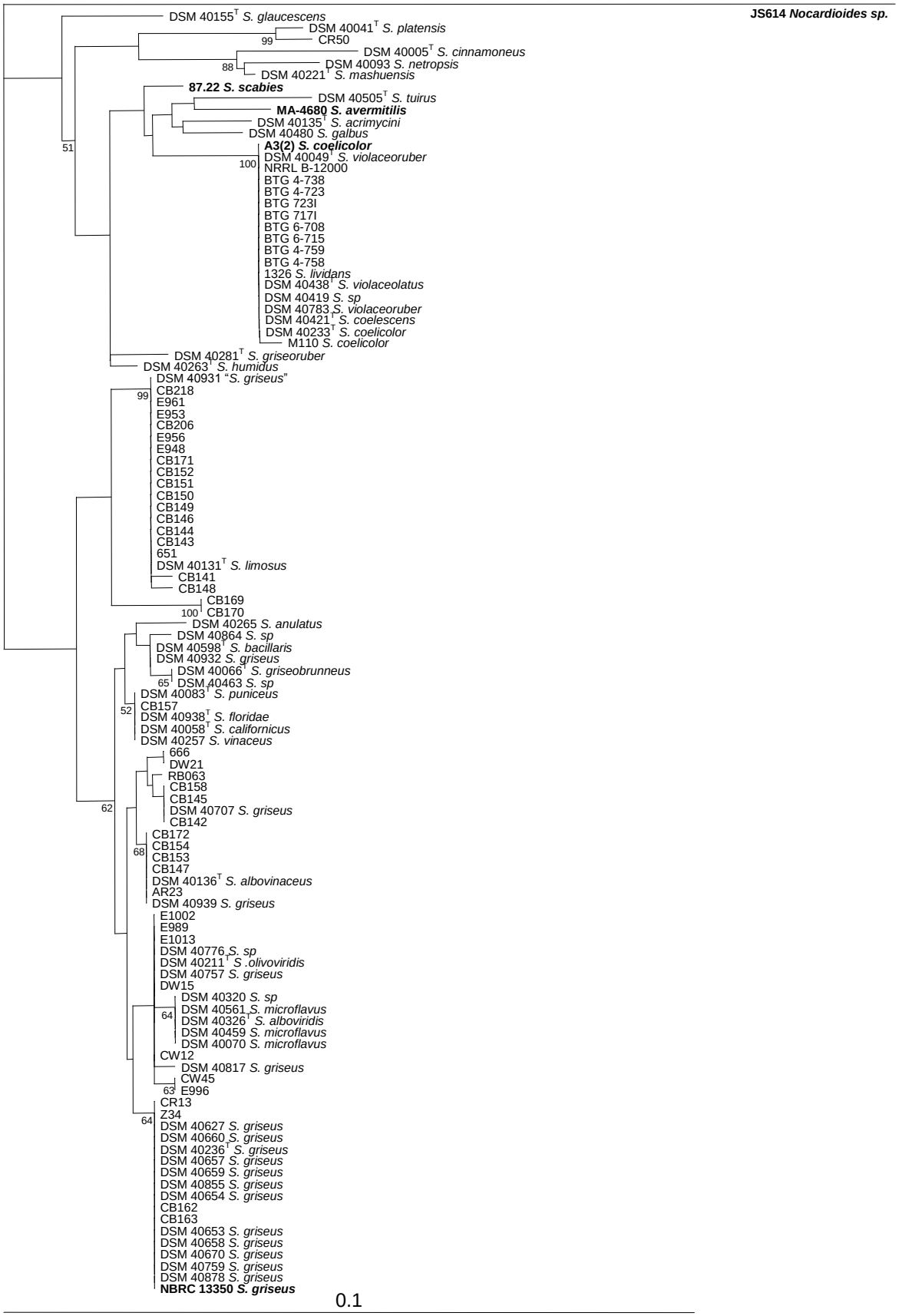


Figure 10.17. Phylogenetic tree of the nucleotide sequence of the 30S ribosomal protein S12 gene *rpsL*

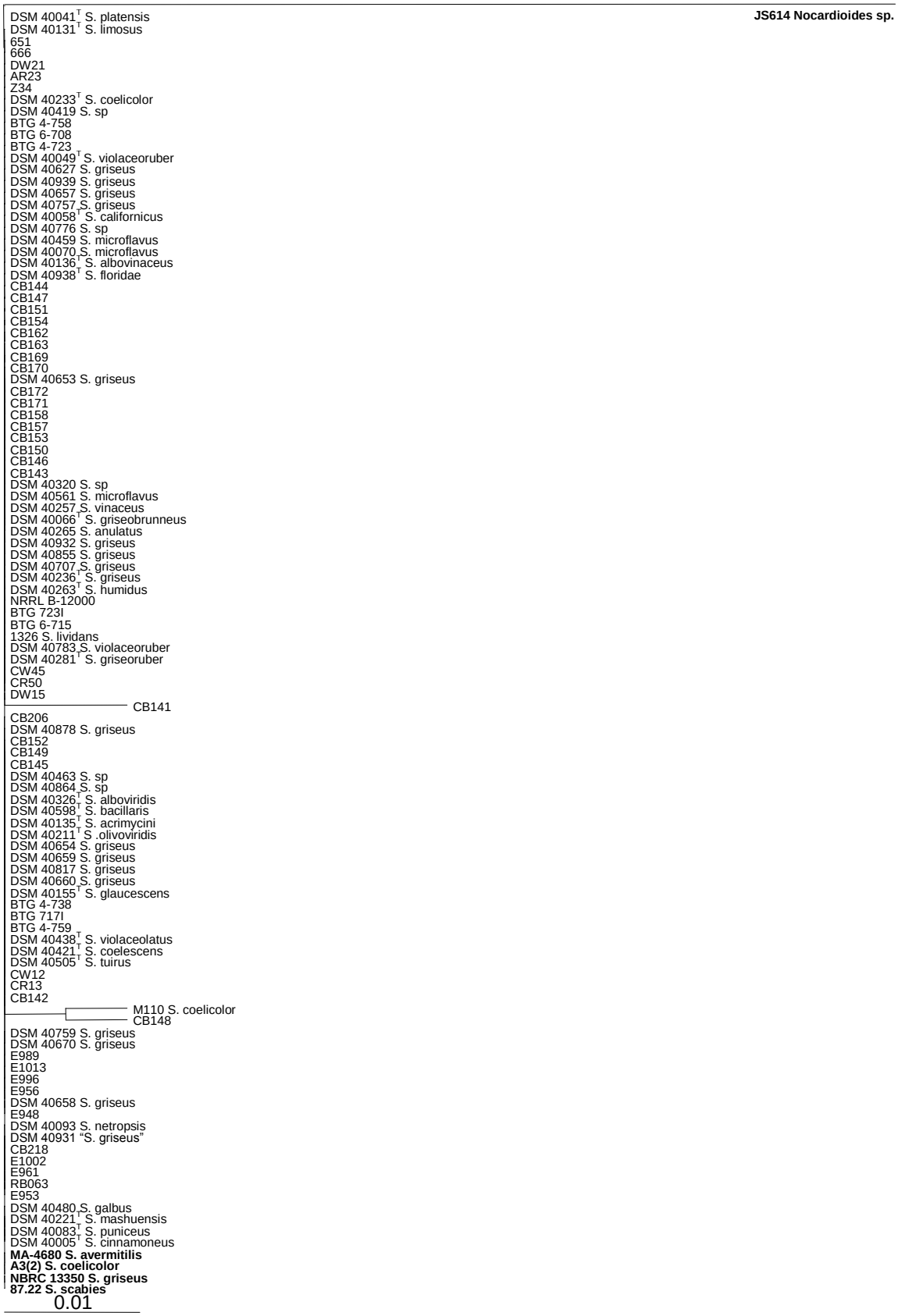


Figure 10.18. Phylogenetic tree of the amino acid sequence of the 30S ribosomal protein S12 gene *rpsL*

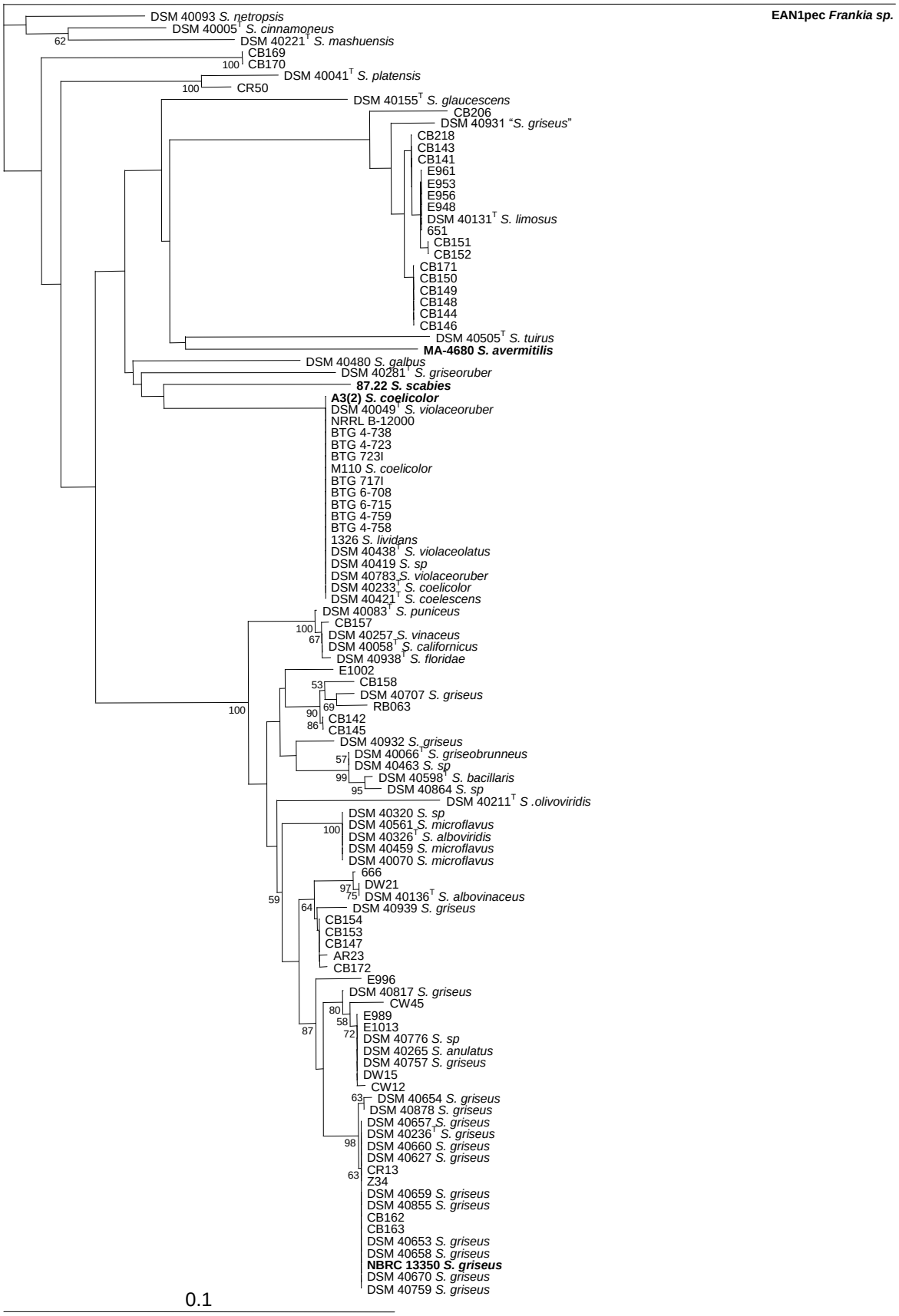


Figure 10.19. Phylogenetic tree of the nucleotide sequence of the Fe-containing superoxide dismutase gene *sodF*

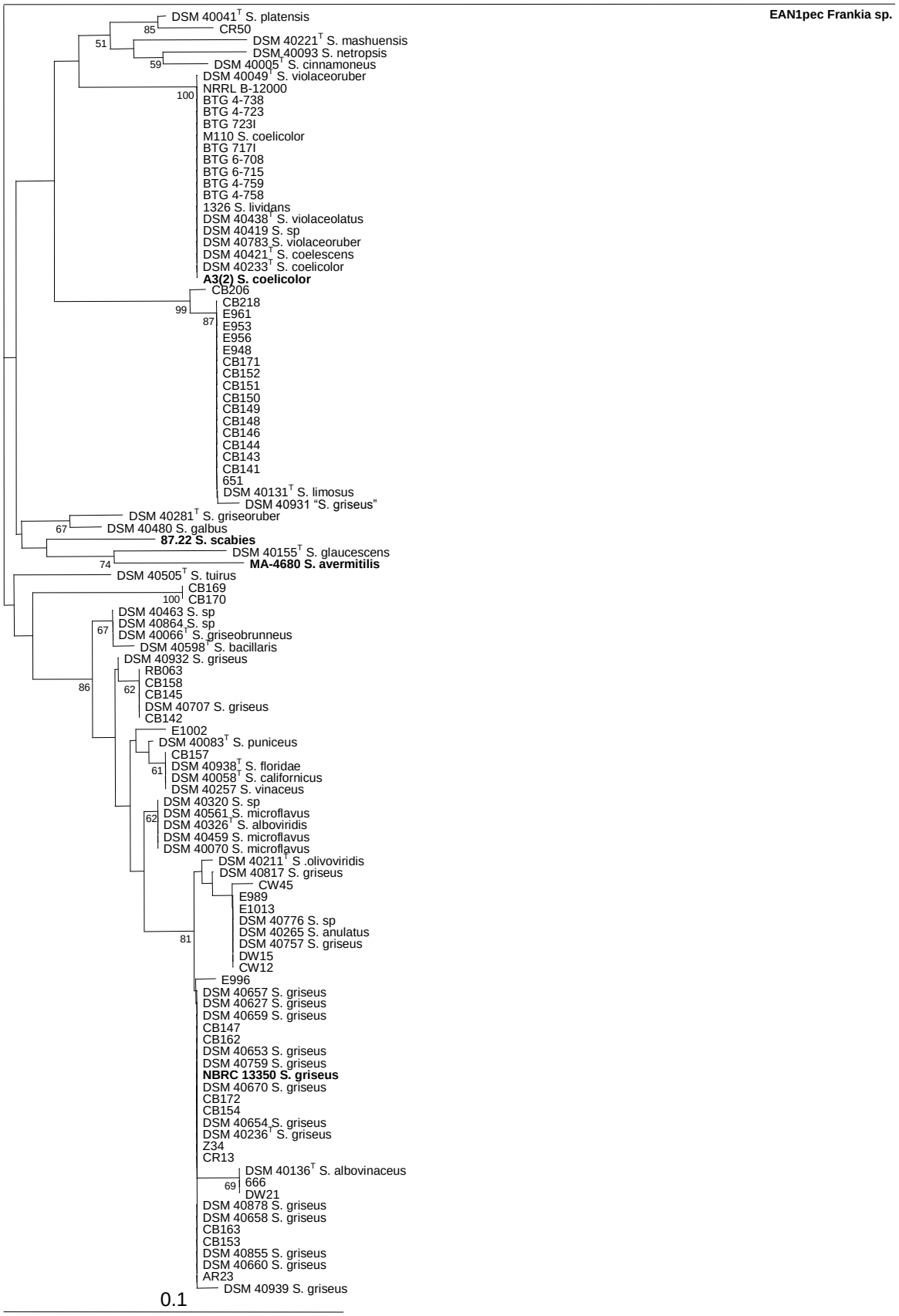


Figure 10.20. Phylogenetic tree of the amino acid sequence of the Fe-containing superoxide dismutase gene *sodF*