The evolution of development in *Streptomyces* analysed by genome comparisons

Keith F. Chater & Govind Chandra

John Innes Centre, Norwich Research Park, Colney, Norwich, UK

Correspondence: Keith F. Chater, Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK. Tel.: +44 1603 450297; fax: +44 1603 450778; e-mail: keith.chater@bbsrc.ac.uk

Received 21 February 2006; revised 10 April 2006; accepted 12 April 2006.

First published online 22 June 2006.

DOI:10.1111/j.1574-6976.2006.00033.x

Editor: Simon Cutting

**Keywords**

comparative genomics; evolution of development; sporulation; actinobacteria; *Streptomyces*.

**Abstract**

There is considerable information about the genetic control of the processes by which mycelial *Streptomyces* bacteria form spore-bearing aerial hyphae. The recent acquisition of genome sequences for 16 species of actinobacteria, including two streptomycetes, makes it possible to try to reconstruct the evolution of *Streptomyces* differentiation by a comparative genomic approach, and to place the results in the context of current views on the evolution of bacteria. Most of the developmental genes evaluated are found only in actinobacteria that form sporulating aerial hyphae, with several being peculiar to streptomycetes. Only four (\(\text{whiA, whiB, whiD, crgA}\)) are generally present in nondifferentiating actinobacteria, and only two (\(\text{whiA, whiG}\)) are found in other bacteria, where they are widespread. Thus, the evolution of *Streptomyces* development has probably involved the stepwise acquisition of laterally transferred DNA, each successive acquisition giving rise either to regulatory changes that affect the conditions under which development is initiated, or to changes in cellular structure or morphology.

**Introduction: morphology, ecology, phylogeny and evolutionary history of streptomycetes**

Streptomycetes are among the most complex of bacteria (Chater & Losick, 1997). They grow, usually in soil, as branching thread-like hyphae to form a vegetative or substrate mycelium. This growth habit, coupled with the activities of an abundance of extracellular hydrolytic enzymes, helps them to gain access to the nutrients locked up in the insoluble polymers of soil. As the substrate mycelium becomes nutrient-limited, it goes on to manifest two properties for which streptomycetes are particularly famous: the production of antibiotics and other secondary metabolites; and the formation of reproductive aerial branches, which are at least partially parasitic on the mycelium from which they emerge. The aerial hyphae form aseptate tip compartments up to c. 100 \(\mu\)m long, often containing many tens of copies of the genome. When such a compartment stops growing, it undergoes synchronous multiple septation to give a string of unigenomic prespore compartments. The cell walls of prespore compartments undergo thickening and rounding up, and usually accumulate spore pigment, as they turn into chains of spores. Unlike the highly resistant endospores of bacilli, *Streptomyces* spores do not show conspicuous resistance to heat, but they do survive well in including dry conditions, and they presumably facilitate dispersal by wind and water, and via animals. Although most streptomycetes do not seem to sporulate in submerged culture, a few can (including some strains of *Streptomyces griseus* and *Streptomyces venezuelae*). Presumably, differences in the detailed developmental regulatory circuitry are involved in this departure from the norm, but at the time of writing, there is insufficient published information for us to consider this interesting question here.

Phylogenetically, streptomycetes are part of the Actinobacteria, the class of gram-positive and morphologically diverse bacteria that have DNA comparatively rich in G+C. Among free-living actinobacteria, G+C content ranges from 54% in some corynebacteria to more than 70% in streptomycetes, but in the obligate pathogen *Tropheryma whipplei* it is less than 50%. Figure 1 is a phylogenetic tree of sequenced actinobacteria based on 16S rRNA gene sequence comparisons. It appears that the last common ancestor of the actinobacteria lived about 2–1.5 Ga, when evidence suggests that the atmosphere first began to be oxygenated (Embley & Stackebrandt, 1994). Streptomycetes are believed to have originated about 440 Ma, apparently immediately following the invasion of the land by green plants, and coinciding with a rapid rise of oxygen in the atmosphere towards present...
levels (Heckman et al., 2001). Reflecting this timing, streptomyces are particularly well adapted to the utilization of refractory plant remains, are aerobic, and have a great number of genes and systems associated with oxidative stress (Lee et al., 2005). Earthworms and arthropods such as springtails that consume decaying plant remains, including the associated microorganisms, appeared soon after the first land plants, so it is quite possible that they exerted other major selective pressures on Streptomyces evolution. For example, the frequent passage of most soil through earthworms (Darwin, 1881) must have selected for resistance of Streptomyces spores to the chemically and enzymically harsh conditions in worm guts. The frequent consumption of small soil animals by vertebrates (such as reptiles, small mammals and birds) in more recent evolutionary time, while it may have contributed to the dispersal and diversification of streptomyces, could not have been a fundamental element in their early evolution.

**Comparative genomics of actinobacteria: the bigger picture**

In the last few years, genome sequences have been obtained for diverse actinobacteria (19 complete and annotated genomes, 16 species, 10 genera at the time of writing), as indicated in Table 1. The analysis in this review is largely confined to these sequenced genomes, about which more comprehensive statements can be made. Of the two sequences from streptomycetes, one is from the academic model organism Streptomyces coelicolor A3(2) (note that this strain is improperly named, being more similar to the Streptomyces violaceoruber clade; but the name is too deeply enshrined in the literature to justify making the correction), and the other is from Streptomyces avermitilis, an important antibiotic producer (Bentley et al., 2002; Ikeda et al., 2003). These two species are phylogenetically fairly well separated from each other and from Streptomyces griseus, another species that has been important in the study of development, and whose genome sequence is currently being determined (S. Horinouchi, pers. commun.). The genome sequence of another species, the plant pathogen Streptomyces scabies, is also available, though currently not annotated (ftp://ftp.sanger.ac.uk/pub/pathogens/ssc/Sts.db). This makes it possible to use comparative genomics to attempt to trace the evolution of Streptomyces development, and in doing this, perhaps to deepen our understanding of developmental mechanisms.

For such an approach, criteria must first be adopted for whether related genes can be considered to be orthologues (i.e. are the direct, functionally equivalent derivatives of a gene present in the last common ancestor of the species being compared) or are merely paralogues (related through ancestral gene duplication). A first criterion is sequence relatedness between gene products. One would expect most (though not all) orthologous genes to show a fairly consistent degree of amino-acid identity in comparisons between particular organisms. This ‘orthology value’ could be determined by initially making assumptions about obvious candidates for orthology, such as genes for RNA polymerase core enzyme subunits, or genes for the biosynthesis of amino acids etc. (as long as such candidate genes are represented only once in the genomes). We compared seven such gene products among S. coelicolor, S. avermitilis, Thermobifida fusca, Corynebacterium glutamicum and Mycobacterium tuberculosis (Table 2). Within this set of orthologues, comparisons between the two streptomycetes indicated a range of 88–97% amino-acid identity, while identities between S. coelicolor and the other organisms ranged from 69–82% (T. fusca) to 52–71% (C. glutamicum). On the other hand, when sets of multiple paralogues are encoded within one organism, members of each set are generally well under 50% identical. We therefore used 50% as a threshold value when considering orthology in the developmental genes discussed in this article.

The second criterion for orthology takes advantage of similarity of gene arrangement along the chromosome (‘synteny’), which is usually extensive between members of the same genus. Across different genera of actinobacteria there are many patches of local synteny, and there is a
<table>
<thead>
<tr>
<th>Organism</th>
<th>Morphology, special features</th>
<th>Accession number</th>
<th>Database</th>
<th>Genome attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Streptomyces coelicolor</em></td>
<td>Mycelial, spore chains on aerial hyphae</td>
<td>AL645882</td>
<td><a href="http://streptomyces.org.uk">http://streptomyces.org.uk</a></td>
<td>8.66 Mb, 72.1% GC, 7825 genes, 146 (2%) with TTA</td>
</tr>
<tr>
<td><em>S. avermitilis</em></td>
<td>Mycelial, spore chains on aerial hyphae</td>
<td>BA000030</td>
<td><a href="http://www.bio.nite.go.jp/dogan/MicroTop?GENOME_ID=sav_G1">http://www.bio.nite.go.jp/dogan/MicroTop?GENOME_ID=sav_G1</a></td>
<td>9.02 Mb, 70.7% GC, 7575 genes, 260 (3%) with TTA</td>
</tr>
<tr>
<td><em>Thermobifida fusca</em></td>
<td>Mycelial, spores clumped on aerial hyphae</td>
<td>CP000088</td>
<td><a href="http://genome.jgi-psf.org/draft_microbes/thefu/thefu.home.html">http://genome.jgi-psf.org/draft_microbes/thefu/thefu.home.html</a></td>
<td>3.64 Mb, 67.95% GC, 3419 genes, 449 (13%) with TTA</td>
</tr>
<tr>
<td><em>Nocardia farcinica</em></td>
<td>Mycelial, substrate and aerial hyphae form fragments</td>
<td>AP006618</td>
<td><a href="http://nocardia.nih.go.jp/">http://nocardia.nih.go.jp/</a></td>
<td>6.02 Mb, 70.8% GC, 5683 genes, 242 (4%) with TTA</td>
</tr>
<tr>
<td><em>Mycobacterium avium</em></td>
<td>Irregular rods</td>
<td>AE016958</td>
<td><a href="http://www.cbc.umn.edu/ResearchProjects/AGAC/Mptb/Mptbhome.html">http://www.cbc.umn.edu/ResearchProjects/AGAC/Mptb/Mptbhome.html</a></td>
<td>4.83 Mb, 69.2% GC, 4350 genes, 589 (14%) with TTA</td>
</tr>
<tr>
<td><em>M. bovis</em></td>
<td>Irregular rods</td>
<td>BX248333</td>
<td><a href="http://www.sanger.ac.uk/Projects/M_bovis/">http://www.sanger.ac.uk/Projects/M_bovis/</a></td>
<td>4.35 Mb, 65.6% GC, 3953 genes, 1442 (36%) with TTA</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>Irregular rods</td>
<td>AL123456</td>
<td><a href="http://www.sanger.ac.uk/Projects/M_tuberculosis/http://www.tigr.org/tigr-scripts/CMR2/GenomePage3.spl?database=gmt">http://www.sanger.ac.uk/Projects/M_tuberculosis/http://www.tigr.org/tigr-scripts/CMR2/GenomePage3.spl?database=gmt</a></td>
<td>4.41 Mb, 65.6% GC, 3999/4189 genes, 1462/1489 (36%) with TTA</td>
</tr>
<tr>
<td><em>Leifsonia xyli</em></td>
<td>Rods</td>
<td>AE016822</td>
<td><a href="http://leifsonia.lbi.ic.unicamp.br">http://leifsonia.lbi.ic.unicamp.br</a></td>
<td>2.58 Mb, 67.6% GC, 2030 genes, 109 (5%) with TTA</td>
</tr>
<tr>
<td><em>Bifidobacterium longum</em></td>
<td>Irregular rods, sometimes branched</td>
<td>AE014295</td>
<td></td>
<td>2.26 Mb, 60.1% GC, 1727 genes, 450 (26%) with TTA</td>
</tr>
<tr>
<td><em>Corynebacterium efficiens</em></td>
<td>Rods</td>
<td>BA000035</td>
<td><a href="http://genome.jgi-psf.org/draft_microbes/thefu/thefu.home.html">http://genome.jgi-psf.org/draft_microbes/thefu/thefu.home.html</a></td>
<td>3.15 Mb, 63.1% GC, 2942 genes, 717 (24%) with TTA</td>
</tr>
<tr>
<td><em>C. glutamicum</em></td>
<td>Rods</td>
<td>BA000036</td>
<td><a href="http://genome.ls.kitasato-u.ac.jp/esequenced.html">http://genome.ls.kitasato-u.ac.jp/esequenced.html</a></td>
<td>3.31/3.28 Mb, 53.8% GC, 3099/3058 genes</td>
</tr>
<tr>
<td><em>C. diphtheriae</em></td>
<td>Rods</td>
<td>BX248353</td>
<td><a href="http://www.uni-bielefeld.de/">http://www.uni-bielefeld.de/</a></td>
<td>2.49 Mb, 53.4% GC, 2320 genes</td>
</tr>
<tr>
<td><em>Propionibacterium acnes</em></td>
<td>Pleomorphic rods</td>
<td>AE017283</td>
<td><a href="http://www.g2l.bio.uni-goettingen.de/projects/f_projects.html">http://www.g2l.bio.uni-goettingen.de/projects/f_projects.html</a></td>
<td>2.56 Mb, 60.0% GC, 2297 genes</td>
</tr>
<tr>
<td><em>Symbiobacterium thermophilum</em></td>
<td>Obligate symbiont; possibly not actinobacterial</td>
<td>AP006840</td>
<td><a href="http://genome.ls.kitasato-u.ac.jp/esequenced.html">http://genome.ls.kitasato-u.ac.jp/esequenced.html</a></td>
<td>1249 (53%) with TTA</td>
</tr>
<tr>
<td><em>Tropheryma whippelii</em></td>
<td>Obligate pathogen</td>
<td>BX072543</td>
<td><a href="http://www.sanger.ac.uk/Projects/T_whippelie/">http://www.sanger.ac.uk/Projects/T_whippelie/</a></td>
<td>3.57 Mb, 68.6% GC, 3337 genes</td>
</tr>
<tr>
<td><em>Mycobacterium leprae</em></td>
<td>Obligate pathogen</td>
<td>AL450380</td>
<td><a href="http://www.sanger.ac.uk/Projects/M_leprae/">http://www.sanger.ac.uk/Projects/M_leprae/</a></td>
<td>0.93 Mb, 46.3% GC, 784/808 genes, 687/718 (88%) with TTA</td>
</tr>
</tbody>
</table>

Table 1. Sequested actinobacterial genomes consulted in this article
recognizable tendency for a substantial number of genes to retain a similar relative distance from the replication origin, even though the overall organization of the chromosomes has usually undergone many rearrangements (Fig. 2). The rearrangements mostly involve deletions, insertion of newly acquired DNA, and more or less symmetrical exchanges from opposite sides in relation to the replication origin (Bentley et al., 2002).

**Table 2.** Percentage amino-acid identity of representative gene products from four actinobacteria to their orthologues in *Streptomyces coelicolor*

<table>
<thead>
<tr>
<th>Gene or enzyme</th>
<th><em>Streptomyces avermitilis</em> (70.7% GC)</th>
<th><em>Mycobacterium tuberculosis</em> (65.6% GC)</th>
<th><em>Corynebacterium glutamicum</em> (53.8% GC)</th>
<th><em>Thermobifida fusca</em> (70% GC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino-acid biosynthesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hisA</td>
<td>96.2</td>
<td>68.2</td>
<td>59.9</td>
<td>74.3</td>
</tr>
<tr>
<td>hisB</td>
<td>99.0</td>
<td>57.4</td>
<td>52.5</td>
<td>75.8</td>
</tr>
<tr>
<td>leuB</td>
<td>93.1</td>
<td>67.8</td>
<td>67.8</td>
<td>69</td>
</tr>
<tr>
<td>Intermediary metabolism</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aconitase</td>
<td>92.7</td>
<td>70.4</td>
<td>68.3</td>
<td>78.5</td>
</tr>
<tr>
<td>Malate dehydrogenase</td>
<td>94.8</td>
<td>70.8</td>
<td>64.6</td>
<td>82</td>
</tr>
<tr>
<td>Transcription</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rpoB</td>
<td>95.1</td>
<td>75.3</td>
<td>71.2</td>
<td>80.6</td>
</tr>
<tr>
<td>rpoC</td>
<td>96.8</td>
<td>71.6</td>
<td>63.6</td>
<td>78.2</td>
</tr>
<tr>
<td>Mean (‘orthology value’)</td>
<td>95.4</td>
<td>68.8</td>
<td>64.0</td>
<td>77.0</td>
</tr>
</tbody>
</table>

**Fig. 2.** Plots of synteny between some actinobacterial chromosomes. The points represent positions of genes whose products reciprocally show closest relatedness in Blast comparisons between the chromosomes. A simple diagonal line would indicate fully conserved gene arrangement in plots comparing chromosomes aligned from a comparable point (i.e. the end of linear chromosomes, which have an roughly centrally located replication origin, in the case of *Streptomyces avermitilis* against *Streptomyces coelicolor*; and the origin of replication of circular chromosomes, in the case of *Mycobacterium tuberculosis* against *Thermobifida fusca*). The extent to which the diagonal becomes a cross reflects the occurrence of symmetrical exchanges on either side of the replication origin. When circular chromosomes are compared with linear ones, synteny is expected to give inflected plots, because the coordinates of the chromosomes start from different positions. Symmetrical exchanges on either side of the replication origin are then expected to give a diamond-shaped plot.
Most of the genes that are conserved between streptomycetes and other genera are located in a central core region of the linear Streptomyces chromosomes. Thus, of the order of 1–2 Mb at each end consists mostly of laterally transferred, comparatively recently acquired, genes. The core regions of the S. coelicolor and S. avermitilis chromosomes show remarkable synteny, involving some 4000 genes. Only a very small number of exchanges between chromosome arms have taken place, all more or less symmetrically in relation to the origin of replication (Ikeda et al., 2003). The synteny is punctuated by about 2000 nonhomologous genes. This indicates that the two genomes have been repeatedly exposed to foreign DNA over evolutionary time, with the cores acquiring new genes or losing old ones on average several times in every million years in the course of the estimated 220 Myr since the last ancestor common to the two species (A. Ward, pers. commun.). Nevertheless, despite the numerous insertions and deletions that separate the species, the conserved gene set has almost entirely retained its organization. Some powerful selection must exist to maintain this, so it is interesting that there are marked differences in genome organization between different genera. Possibly, new genera arise initially in response to dramatic or cataclysmic novel selective circumstances (such as were briefly discussed in the previous section), when some major rearrangement of the genome confers an adaptive value that would not be favourable in gradual adaptation to slowly changing environmental circumstances. The extensive preservation of genomic arrangement within streptomycetes suggests that their subsequent evolution has largely involved microadaptation to environments that have changed little until the present.

Although the 1–2 Mb end regions of Streptomyces chromosomes are evidently a major primary depot for receiving horizontally transferred segments of DNA, their well-documented instability is not well-suited to long-term retention of particular genes; so those genes whose selective benefits are repeatedly acted upon may well find their way into the more stable central region. There are experimental accounts of unequal exchange between chromosome ends, and between the ends of chromosomes and linear plasmids (Chen et al., 2002; Wenner et al., 2003), such as might permit the effective migration of valuable DNA towards the central region. Once a gene has migrated inwards far enough to be closer to the core than some other selectively advantageous gene, it will become less susceptible to loss involving chromosome ends. This mechanism does not preclude other routes for chromosomal gene acquisition, particularly those involving transposition or site-specific integration of specialized elements such as prophages (see final section for a discussion of the possible roles of phages in gene acquisition).

The genetic basis of developmental complexity in streptomycetes

Over several decades, some progress has been made in the genetic dissection of the complex multicellular development briefly outlined above, particularly in S. coelicolor and S. griseus (Chater, 2001; Chater & Horinouchi, 2003; Flardh & van Wezel, 2004). Several tens of genes have been identified whose products play regulatory, catalytic, organizational or structural roles in development, and there is a moderate body of information about the integrated action of these genes and gene products. Much of this information emanates from the isolation of S. coelicolor mutants that fail to complete normal development, most being defective either in aerial growth (‘bald’, or bld, mutants) (Hopwood, 1967; Merrick, 1976) or in the formation of mature grey spores on the fluffy aerial mycelium (‘white’, or wht, mutants) (Hopwood et al., 1970; Chater, 1972). Most of these mutants are affected in genes with regulatory character. The details of these genes and their interactions are explored in several recent reviews (Chater & Losick, 1997; Chater, 1998, 2001; Chater & Horinouchi, 2003; Chater, 2006), and are only briefly summarized here.

When grown near to each other on certain rich media, most of an early collection of bld mutants exhibit directional interactions that lead to aerial growth (Willey et al., 1993; Kelemen & Buttner, 1998). These interactions have been interpreted as a signalling cascade that is transmitted in the sequence bldD/bldKL/bldAH/bldGH/bldCI/bldDM/ram (Fig. 3). Although some of the more recently isolated bld mutants do not fit in this cascade, its features have helped in the formulation of developmental models (Chater, 2001). It has also been shown that some mutants that fall outside the cascade can be phenotypically corrected by pH buffering, indicating that their morphological defects are a secondary

![Fig. 3. Extracellular signalling dependent on bld genes leading to the production of surface proteins involved in aerial growth in Streptomyces coelicolor. The figure has been updated from that of Chater (1998).](image-url)
consequence of the secretion of organic acids (Viollier et al., 2001). In other species, different extracellular interactions leading to development have been described. One example, the restoration of sporulation to a bald mutant of S. alboniger by the addition of an antibiotic, pamamycin, has recently been interpreted in terms of calcium signalling (Danilenko et al., 2005); but the most extensively studied is the dependence of development in some S. griseus strains on A-factor, a member of the species-specific γ-butyrolactone signalling molecules that are often involved in switching on antibiotic biosynthesis in streptomycetes (Horinouchi, 2002). The variety of extracellular signals used during Streptomyces development may reflect some selective benefit of preventing competitors from responding to each other’s signals, and thus be a cause or reflection of speciation (Chater & Horinouchi, 2003).

Seven of the whi loci (whiA,B,D,E,G,H,I) identified in early studies of S. coelicolor sporulation (Chater, 1972) have been studied further [two ‘loci’ have not: whiF99 proved to be a whiG allele; and whiC mutants did not survive prolonged storage (C.J. Bruton and K.F. Chater, unpublished)]. Figure 4 indicates the present understanding of the regulatory interactions of some of these genes. The most clear-cut gene-to-gene interactions are the requirements of the whiH and whiI promoters for an RNA polymerase holoenzyme that contains the σ factor specified by whiG. The only known gene-to-gene interaction between the bld and whi gene cascades is the repression of whiG by BldD (Elliot et al., 2001).

In this article we mainly focus on the genes shown in Figs 3 and 4, because it has been possible to fit them into interaction networks. We also include several other developmentally and morphogenetically important genes, such as the ssgA-like gene family, mutations in some of which give a white colony phenotype. (Keijser et al., 2003; Traag et al., 2004; Noens et al., 2005). This anthology of genes was used to search the genomes of actinobacteria for likely orthologues.

**Genes implicated in the regulation of formation of aerial hyphae**

**bldA** – the determinant of the tRNA for a rare codon

In streptomycetes, the very high GC content of the DNA makes it inevitable that GC-rich codons are preferentially used. As a consequence, the third position in codons is a G or a C in about nine out of every 10 codons, providing the basis for one of the main tools used to delineate probable coding sequences in Streptomyces DNA sequences (Bibb et al., 1984; Ishikawa & Hotta, 1999). In the standard genetic code, one of six codons for leucine (UUA) has no G or C residues. Not surprisingly, this codon is rarely used in streptomycetes – in fact, it is by far the rarest codon in S. coelicolor, being used in only 2% (145/7825) of chromosomal genes (260, i.e. 3%, of the 7575 chromosomal genes of S. avermitilis contain TTA codons) (Fig. 5). These observations are of more than anecdotal interest, because the only tRNA capable of reading UUA codons efficiently is the product of a gene that can be completely deleted without diminishing growth (Lawlor et al., 1987; Leskiw et al., 1991). However, as
indicated by this gene’s designation (bldA), such mutations do have phenotypic effects – most obviously, they give rise to very easily seen bald colonies that are devoid of pigmented secondary metabolites on most media (Merrick, 1976). The complexity of this phenotype suggests that bldA may have a regulatory role in aspects of developmental and stationary phase biology. The bldA tRNA, with its 3′-AAU-5′ anticodon, should also be able to read the alternative leucine codon UUG via ‘wobble’ interaction, whereas a tRNA with the 3′-AAC-5′ anticodon cognate with the UUG codon is not expected to be able to translate UUA codons (Crick, 1966). Both tRNAs are present in streptomycetes, perhaps in part because of the need to translate UUA codons under conditions in which translation of UUA codons would be disadvantageous.

A bldA-like tRNA locus is present in all actinobacteria. In the current absence of any attempts to inactivate it in any nonstreptomycete, it is not experimentally ruled out that UUA codons are developmentally significant beyond streptomycetes. However, TTA codons are more frequent in the genomes of other actinobacteria, though those with high GC content generally have the lowest frequency (Table 1). *Nocardia farcinica*, with 70.8% GC, has TTA codons in only 4% of genes, but these include at least one important gene of primary metabolism (hisA), so we doubt whether the developmental role of bldA in streptomycetes has a close parallel in *N. farcinica*. Moreover, the distribution of TTA codons within annotated *N. farcinica* genes shows less positional bias than in *S. coelicolor* and *S. avermitilis* (Fig. 5), suggesting that they do not fulfil a similar role to that of TTA codons in streptomycetes: in *S. coelicolor*, 22% of TTA codons occur in the first 10 codons, compared to 9% in *N. farcinica* (Fuglsang, 2005). Since bldA mutants defective in aerial mycelium formation have been obtained in species as diverged as *S. coelicolor* (Merrick, 1976), *S. griseus* (Kwak et al., 1996) and *S. clavuligerus* (Trepanier et al., 2002), bldA was probably adopted as an element of the developmental biology of streptomycetes very early in their evolution. Although the focus of this article is on morphological development, we note in passing that the frequently observed effects of bldA mutations on antibiotic production are mostly exerted directly via UUA codons in the mRNAs of pathway-specific regulatory genes (Chater & Horinouchi, 2003; Chater, 2006).

In a comparison between the genomes of *S. coelicolor* and *S. avermitilis*, it turns out that only 11 orthologous genes contain TTA codons in both species (while a further 10 TTA-containing *S. coelicolor* genes have TTA-free orthologues in *S. avermitilis*). Seven of those 11 have some kind of functional annotation. One, SCO2792 (bldH), is dealt with in the next section. The others, and their predicted functions, are: SCO5495 (cyclic di-GMP phosphodiesterase); SCO1980 and 3423 (both *abaA* orfA-like, see section on *whiJ* below); SCO1434 (*cbxX*-like, *cboX*-like); SCO1242 (DNA-binding protein, WhiJ-like; see below); and SCO4114 (a sporulation-associated protein; Babcock & Kendrick, 1990). The function-unknown genes are SCO4395, 5040, 6623 and 7251. Thus, although a small number of the TTA-containing genes found in modern streptomycetes were already present early in the genus, the great majority (83% in *S. coelicolor*) have been acquired by lateral transfer in the comparatively recent evolutionary past.

*bldH, SCO2792* (= SAV5261), is the main target through which *bldA* affects differentiation

One of the 11 TTA-containing genes common to *S. coelicolor* and *S. avermitilis* is *adpA*, or *bldH* (Nguyen et al., 2003;
This gene was first studied in *S. griseus* (Ohnishi et al., 1999), in which it also contains a TTA codon, in the same position within the gene as in the other two orthologues. In *S. coelicolor*, this TTA codon is the major route through which *bldA* influences development, and AdpA is apparently not involved in regulating antibiotic production; while in *S. griseus* the influence of *bldA* on antibiotic production is partially mediated via *adpA*. This suggests that the way in which *adpA* is integrated into the whole physiology of its particular host has some species-specificity, a notion reinforced by the regulation of *adpA* itself. Thus, in *S. griseus*, *adpA* expression is directly repressed by a protein called ArpA (‘A-factor receptor protein’), repression being lifted in response to an extra-cellular accumulation of the γ-butyrolactone A-factor, a small membrane-diffusible signalling molecule, which diffuses back into cells and interacts with ArpA. In *S. coelicolor*, a comparable γ-butyrolactone signalling system seems to have no marked effects on morphological differentiation or *adpA* expression (Horinouchi, 2002; Kato et al., 2002; Takano et al., 2003; Takano et al., 2005). The differences in regulation of *adpA* are possibly reflected in divergences in upstream DNA sequences; an overall similarity between these regions of *S. coelicolor* and *S. griseus* is interrupted at the region known to bind ArpA and impart A-factor-dependence in *S. griseus* (Takano et al., 2003; Kato et al., 2005).

From these observations, it seems that the cascade from *bldA* via *adpA* to differntiation is a more ancient aspect of *Streptomyces* evolution than the integration of γ-butyrolactones into their biology. γ-Butyrolactones are themselves mainly confined to streptomycetes, among which they are widespread. They exhibit considerable diversity, and may well have coevolved with antibiotic biosynthetic gene sets that have been laterally transferred among streptomycetes (Takano et al., 2005).

In *S. griseus*, *adpA* is autoregulatory, contributing to its own transcriptional repression (Kato et al., 2005). Four out of six of the AdpA-binding sites of the *adpA* promoter of *S. griseus* are conserved in the *adpA* promoter regions of *S. coelicolor* and *S. avermitilis*, so autoregulation may be a general and ancient function of AdpA (Kato et al., 2005).

Although AdpA orthologues are known only in streptomycetes, AdpA belongs to a subfamily within the family of activator/repressor proteins whose best-known member is AraC of *E. coli* (Yamazaki et al., 2004; Kato et al., 2005). AraC-like proteins are characterized by the features of a dual helix-turn-helix DNA-binding domain typically located far from the N-terminus. In the case of AdpA, this domain is from aminoacids 240–332. Residues 57–195 conform to a Pfam domain DJ-1 PfpI that is involved in dimerization (Yamazaki et al., 2004). The rather large size of AdpA (around 400 aa) suggests that it may respond to some other molecule, but there are no experimental data addressing this possibility.

Extensive studies in the laboratory of S. Horinouchi have revealed a substantial regulon of target genes for AdpA in *S. griseus*, and a consensus AdpA binding site has been defined. Among the target genes are several for proteases and protease inhibitors (Ohnishi et al., 2005; Kim et al., 2005). It seems likely that one of the main ways in which AdpA influences development is via the extracellular interactions of proteases and protease inhibitors (Chater, 2006). Many streptomycetes produce extracellular proteases related to trypsin, the activity of which is held in check by binding to small proteinaceous inhibitors, such as leupeptin (Taguchi et al., 1993; Kim & Lee, 1995). In some cases it has been found that specific proteolysis of the inhibitor by a second protease is used to activate the trypsin-like protease, which is then involved in activating development (Kim & Lee, 1995). Since AdpA-dependent extracellular protease: inhibitor interactions are found in phylogenetically diverse species, it is likely that they are an ancient feature of streptomycetes (but among other sequenced actinobacterial genomes, only that of *T. fusca* apparently encodes such a protease inhibitor). The interactions may be important in the re-use of the biomass of the vegetative mycelium that provides nutrients for growth of the aerial mycelium (Manterca et al., 2006). It will be interesting to find out whether these interactions have any connection with the extracellular signalling cascade discussed earlier for *S. coelicolor* (Fig. 3), and whether species-specificity is involved, as is often the case for extracellular signalling.

**Several other bld genes appear to be confined to actinomycetes with sporulating aerial mycelium**

In the first two steps in the cascade shown in Fig. 3, *bldJ*, which is still uncharacterized, is required for the production of signal 1, an extracellular oligopeptide that has been partially characterized (Nodwell & Losick, 1998). Signal 1 is believed to be imported by a multicomponent ATP-dependent transport system specified by the genes of the *bldK* locus (SCOS112–5116; Nodwell et al., 1996). The *bldK* locus is absent from other actinobacterial genomes. However, the next three *bld* genes ‘downstream’ of the *bldA/adpA* stage in the extracellular signalling cascade are found not only in streptomycetes, but also in *T. fusca*, the only other actinomycete among those with known genome sequences that has a well-developed sporulating aerial mycelium (Fig. 6). This is surprising, in view of the rather wide apparent phylogenetic distance between these two genera based on 16S rRNA gene (Fig 1), though the comparatively high orthology value for *S. coelicolor* and *T. fusca* (Table 2) suggests that the two genera may be more closely related than Fig. 1
indicates. There is currently no information bearing on whether the *T. fusca* *bld* gene orthologues play a role in aerial growth.

### **bldG**, SCO3549 (*=SAV4614*)

This gene encodes a member of a protein family found exclusively in gram-positive bacteria, and with multiple representatives in many of them (there are 13 in *S. coelicolor*) (Bignell et al., 2000). In the characterized cases, these proteins are known as antagonists of anti-σ factors. The interaction between the two proteins depends on the phosphorylation state of the anti-anti-σ factor. When phosphorylated, it does not bind to the anti-σ factor efficiently, so the latter protein is free to bind to, and hold inactive, its target σ factor, whereas the unphosphorylated anti-anti-σ seqeusters the anti-σ, thus releasing the σ factor to interact with RNA polymerase and activate a specific set of promoters. σ factors of a particular subclass are regulated in this way, the best characterized being σB of *B. subtilis* (Chen et al., 2003).

The phosphorylation state of the anti-anti-σ is determined both by the kinase activity of the anti-σ factor and by dephosphorylation by a sensor phosphatase; and a phosphorylated form of BldG has been demonstrated (Hesketh et al., 2002; Bignell et al., 2003). In sporulating bacilli and clostridia, some of these systems are involved in forespore development (Errington, 2003) and others, notably σ3, in stress responses (Chen et al., 2003). In those cases, the *bldG*-like gene is clustered with the target anti-σ factor gene and the cognate σ factor determinant; there are six examples of such clusters in the *S. coelicolor* genome. However, although *bldG* is adjacent to a gene (SCO3548, *=SAV4615*) for a probable anti-σ factor, the two genes are not located close to a recognizable σ factor determinant. It remains completely unknown whether BldG is involved in regulating activity of any of the nine σ B-like proteins encoded in the *S. coelicolor* genome. In view of the evidence of interplay between stress responses and differentiation in *S. coelicolor* (Chater, 2001; Kelemen et al., 2001; Viollier et al., 2003; Lee et al., 2005), it may well be that particular σ factors are subject to regulation by more than one anti-σ system.

The *T. fusca* orthologues of *bldG* (Tfus0984, 77% identity) and its partner gene are at one end of a region containing 12 syntenously arranged genes. Several of those adjacent genes (but not the *bldG* pair) show widespread synteny among actinobacteria.

### **bldC**, SCO4091 (*=SAV4130*)

This gene encodes a small protein, identical in *S. coelicolor* and *S. avermitilis*, consisting of little more than a MerR family DNA-binding domain (Hunt et al., 2005). The *bldC* noncoding upstream region is conserved for more than 400 bp between *S. coelicolor* and *S. avermitilis* (> 90% identity, interrupted by one nonrelated segment about 310 bp upstream of the coding region in each case). The 400 bp includes (at least in *S. coelicolor*) 145 bp of transcribed leader sequence. The length of this conserved region is quite surprising in view of the apparently constitutive expression of *bldC* (Hunt et al., 2005). Although there are other genes encoding single-domain MerR-like proteins in these and other genomes, none of them is closely related to *bldC* except the *bldC* orthologue (Tfus1491, 75% aminoacid identity) in *T. fusca* (Hunt et al., 2005). Tfus1491 is located at one end of a stretch of six genes that show synteny with a region of the *S. coelicolor* genome (with two additional genes in the streptomycete). Three of these adjacent genes are present and syntenous in most actinobacteria. Figure 7 shows the extent of synteny in this genomic region, as a typical example of the relative genome structures of *S. coelicolor*, *S. avermitilis*, *T. fusca* and other actinobacteria.

### **bldD**, SCO1489 (*=SAV6861*)

BldD is a regulatory protein that possesses a putative DNA-binding domain similar to those of Xre-like regulators, but
The outlying member of the actinobacteria (Accession no. NZ_AAEF01000025) The gene is regulated differently. The T. fusca bldD protein in vitro sequence near the bldD gene is present in T. fusca (see above), may have roles in the stress responses of mycelial organisms. Such a role was previously suggested for Streptomyces bld genes, in a model in which the extracellular signal cascade was considered as a succession of indications that all possible alternative responses to hard times had been tried before commitment to reproductive aerial growth, which is the ‘last resort’ (Chater, 2001). However, like the

bld genes at the downstream end of the extracellular signalling cascade are peculiar to streptomycetes

The bldG, bldC and bldD genes, which are also found in T. fusca (see above), may have roles in the stress responses of mycelial organisms. Such a role was previously suggested for Streptomyces bld genes, in a model in which the extracellular signal cascade was considered as a succession of indications that all possible alternative responses to hard times had been tried before commitment to reproductive aerial growth, which is the ‘last resort’ (Chater, 2001). However, like the genes involved in the start of the extracellular cascade (bldK, bldA, bldH/adpA), those at its finish (bldM, bldN, ramC-SABR) appear to be confined to streptomycetes (Fig. 6). Conceivably, then, they fulfil roles that are specifically appropriate to the particular features of Streptomyces aerial growth.

BldM is an aberrant orphan response regulator (Molle & Buttner, 2000), and BldN an ECF σ factor (Bibb et al., 2000). It appears that σbldN is directly required for the transcription of one of two promoters of bldM (Bibb et al., 2000). An important consequence of the expression of bldM is the activation of the chp genes encoding the morphogenetic chaplin proteins (Elliot et al., 2003; see below). Interestingly, some bldM and bldN alleles give rise to a white aerial mycelium phenotype (Ryding et al., 1999), suggesting that they are both needed at more than one stage of development. Both of the S. avermitilis orthologues show considerable conservation in their upstream noncoding regions with their S. coelicolor equivalents: in the case of bldM this extends for about 350 bp, showing 88% identity with the exception of one short mismatched segment (20 bp in S. avermitilis, 10 bp in S. coelicolor); while the homology upstream of bldN extends for about 160 bp of 94% identity. The same segment is about 84% identical in the S. griseus orthologue of bldN, called adsA, which is a direct target of AdpA (= BldH) (Yamazaki et al., 2000); and bldN transcription is also bldH-dependent in vivo in S. coelicolor (Bibb et al., 2000). Curiously, much of the divergence is in the segment shown in vitro to be protected by AdpA in S. griseus, which, unusually, is within the transcribed region upstream of adsA (Yamazaki et al., 2000). In S. coelicolor, bldN loses its temporal control in a bldD mutant, being strongly expressed even early in vegetative growth; and the equivalent segment of the bldN upstream region of S. coelicolor binds to BldD protein in vitro (Elliot et al., 2001). However, it is not clear if AdpA and BldD both interact with the bldN upstream region of either organism. A second BldD-binding site for bldN corresponds to the segment of the promoter region expected to interact with RNA polymerase (Elliot et al., 2001), which is extensively conserved among the three species. We suppose that the temporal expression of bldN reflects both repression by BldD and activation by AdpA in all three species, and that the regulation of bldN by these two proteins is likely to be an ancient trait of streptomycetes. However, divergence in the shared BldD-/AdpA-binding DNA segment may cause the fine details of any competition or interaction between the two proteins to differ in S. griseus from those in S. coelicolor and S. avermitilis.
The ramCSABR genes, SCO6681–6685 ( = SAV7499–7503)

A second important morphogenetic endpoint of the cascade, that seems to bypass bldN and bldM and the production of chaplins, is the five-gene ramCSABR cluster ( = amfTSBAR in S. griseus) (Chater & Horinouchi, 2003; Ueda et al., 2005; Willey et al., 2006). [There is a paradox in the literature here, since it was reported that a bldM mutant behaved the same as a bldD mutant in the extracellular complementation cascade (Molle & Buttner, 2000); but these experiments may be complicated by the fact that some of the mutants used to define the cascade have different genetic backgrounds.] The first step in ram cluster expression is the transcription of ramR, which encodes a response regulator that contains most of the residues conserved in phosphorylation pockets that are the targets of histidine regulator that contains most of the residues conserved in these experiments may be complicated by the fact that some of the mutants used to define the cascade have different genetic backgrounds.] The first step in ram cluster expression is the transcription of ramR, which encodes a response regulator that contains most of the residues conserved in phosphorylation pockets that are the targets of histidine regulator that contains most of the residues conserved in the literature here, since it was reported that a bldM mutant behaved the same as a bldD mutant in the extracellular complementation cascade (Molle & Buttner, 2000); but these experiments may be complicated by the fact that some of the mutants used to define the cascade have different genetic backgrounds.] The first step in ram cluster expression is the transcription of ramR, which encodes a response regulator that contains most of the residues conserved in phosphorylation pockets that are the targets of histidine regulator that contains most of the residues conserved in

The differentiation of aerial hyphae into spores

A section of the sporulation regulatory network is found only in streptomyces

Four of the regulatory genes shown in Fig. 4 that are needed for aerial hyphae to form mature grey spores are absent from other actinobacteria (though one of them, whiG, is found in many nonactinobacterial bacteria, including E. coli and B. subtilis). Three of the four genes (whiG, H, I) form a particular regulatory subsection of the early whi gene network, and the fourth (sigF) encodes a σ factor for late gene expression.

whiG, SCO5621 ( = SAV2630)

When aerial hyphae begin to grow, it appears that they are developmentally indeterminate. They acquire a specific developmental fate when the sporulation-specific WhiG σ factor becomes active (Mendez & Chater, 1987; Chater et al., 1989) (perhaps partially as a result of the release of whiG from repression by BldD; see above). WhiG is an orthologue of the σ factors that, in many bacteria, direct transcription of some of the genes associated with motility and chemotaxis (Tan et al., 1998). Although the S. coelicolor chromosome contains a prophage-like insertion upstream of whiG that is absent from S. avermitilis, the genes downstream are conserved and syntenous between the species. The whiG genes of several other streptomycetes have been shown to play a similar role in committing aerial hyphae to sporulation, and, where sequenced, all are highly conserved in their coding regions and in their promoters at least from –1 to –45, the region containing one of the two characterized BldD-binding sites in the whiG upstream region of S. coelicolor (Soliveri et al., 1993; Kormanec et al., 1994; Elliot et al., 2001; Catakli et al., 2005).

whiH, SCO5819 ( = SAV2445) and whil, SCO6029 ( = SAV2230)

Two sporulation regulatory genes, whiH and whil, are well-established targets for σWhiH RNA polymerase holoenzyme (Ryding et al., 1998; Ainsa et al., 1999; Kormanec et al., 1999). There are significant matches between S. coelicolor and S. avermitilis in the upstream regions for each of these genes. Each gene encodes a member of a well known and very widespread family of regulatory DNA-binding proteins. WhiH is in the family containing GntR, many of whose members respond to small organic acids, prompting speculation that WhiH responds to a change in concentration of such a metabolite during aerial mycelium growth (Ryding et al., 1998). WhiH appears to induce the strong

FEMS Microbiol Rev 30 (2006) 651–672 © 2006 Federation of European Microbiological Societies Published by Blackwell Publishing Ltd. All rights reserved
developmentally controlled fisZp2 promoter to bring about sporulation septation (Flardh et al., 2000).

WhiI resembles the response regulators associated with bacterial two-component systems (but has important differences in the normally conserved phosphorylation pocket, which are likewise changed in the S. avermitilis orthologue) (Ainsa et al., 1999). There is circumstantial evidence that different genes are induced or repressed by different forms of WhiI, and it is postulated that WhiI (and perhaps other orphan response regulators, including BldN and RamR; see above) responds to a phosphorylated intermediary metabolite rather than being phosphorylated by a histidine protein kinase (Y. Tian, K. Fowler, K.C. Findlay and K.F. Chater, in preparation). WhiI may induce the chromosome condensation that accompanies sporulation (Ainsa et al., 1999).

**sigF, SCO4035 (= SAV4185)**

The σ factor encoded by sigF is required for the late stages of sporulation (Potuckova et al., 1995; Kormanec et al., 1996; Kelemen et al., 1996). Thus, a sigF mutant undergoes sporulation septation, but the resulting spore chains are irregular, thin-walled and more or less unpigmented, and contain uncondensed DNA. There is no detectable transcription of sigF in whiA, B, G, H, I or J mutants, but the reason for this is unknown (Kelemen et al., 1996). The S. aureofaciens orthologue is also dependent on at least whiG (Kormanec et al., 1996), indicating that sigF fulfils the same role in diverse species, and therefore probably acquired this role early in Streptomyces evolution. Interestingly, σ^F_ is a member of the same subgroup of gram-positive-specific σ factors as the two forespore σs of B. subtilis and the B. subtilis stress-responsive σ^F, two of which are controlled by anti-σ/anti-anti-σ cascades (see bldG above). However, no genes for such a cascade are located in the vicinity of sigF or its orthologue in S. avermitilis. No direct targets for σ^F_ have yet been identified, but one of the whiE promoters (for orfVIII) is inactive in a sigF mutant (Kelemen et al., 1998; Kelemen & Buttner, 1998).

**A section of the sporulation regulatory network involves genes that are found in morphologically simple actinobacteria**

**whiA, SCO1950 (= SAV6294), whiB, SCO3034 (= SAV5042) and whiD, SCO4767 (= SAV3216)**

whiA and whiB mutants both have unusually long and curly aerial hyphae that completely lack sporulation septa (Flardh et al., 1999). A whiB mutant of S. aureofaciens had a similar phenotype (Kormanec et al., 1998), and the whiB orthologue of Streptomyces griseocarneus complemented a whiB mutant of S. coelicolor (Soliveri et al., 1993). Thus, WhiA and WhiB are thought to be required to stop the continued extension of aerial hyphae. Probably, sporulation septation can take place only in nongrowing hyphae. It is likely that the two gene products are closely intertwined in their actions, either by an intimate regulatory interplay (they certainly influence each other’s expression) or by virtue of some protein–protein interaction (Jakimowicz et al., 2006). The developmentally controlled expression of whiA and whiB in S. coelicolor (Soliveri et al., 1992; Ainsa et al., 2000) and at least of whiB in S. aureofaciens (Soliveri et al., 1992; Kormanec et al., 1998) appears to be largely independent of whiG.

whiA is located in a gene cluster that shows significant conservation across all gram-positive bacteria. For example, homologues of whiA and the two upstream genes are similarly clustered in B. subtilis (Ainsa et al., 2000). A whiA-free version of the cluster is found in many gram-negative bacteria. There appear to be no reports of the physiological role or molecular structure and function of WhiA-like proteins in any organisms except streptomycetes. Perhaps in other bacteria too it is required to bring growth to an orderly halt, but the elimination of this function might not give rise to an obvious phenotype in unicellular organisms. whiA is expressed at a low level throughout growth by readthrough transcription extending from the upstream transcription unit, but is strongly upregulated during aerial hyphal development, from an additional promoter of its own within the immediately upstream coding sequence (Ainsa et al., 2000). Presumably whiA acquired its sporulation-specific function by mutations affecting its regulation that occurred very early in the Streptomyces lineage. The distribution of whiB-like (’wbl’) genes is quite different from that of whiA homologues. WhiB orthologues are omnipresent in free-living actinobacteria, but absent from all other organisms. They are members of a substantial family of generally small (80–120 aa) proteins whose major conserved features are a set of four cysteine residues and a short glycine- and tryptophan-rich segment (Soliveri et al., 2000). Several of these proteins, like WhiB itself, have orthologues widespread among actinobacteria, but others show species-specificity, implying that members of the family have often been acquired laterally. This is consistent with the frequent occurrence of wbl genes on plasmids and plasmids of actinomycetes – for example, the large linear plasmid SCP1 present in S. coelicolor A3(2) carries three wbl genes (Bentley et al., 2004). Several wbl genes and their products have been studied experimentally in streptomycetes and mycobacteria. One of these, whiD, is a late sporulation gene in S. coelicolor (Chater, 1972; Molle et al., 2000), while its orthologue (whmB, or whiB3) in mycobacteria is involved in a physiologically ill-defined but important role in the pathogenicity of M. tuberculosis and M. bovis (Steyn et al., 2002). Studies of the whiD/whmB/whiB3
orthologues have provided insights into the function and structure of WhiB-like proteins. Evidence that WhiB3 interacts directly with the principle σ factor of M. smegmatis (Steyn et al., 2002) may have broad relevance to the action of Wbl proteins, particularly since another wbl gene (the wblP gene of plasmid SCPI) encodes a fusion of a Wbl domain with an ECF σ factor domain (Bentley et al., 2004); and four other wbl genes of S. coelicolor and its plasmid SCPI are located very close to σ factor determinants. The physiological stimulus for the activity of Wbl proteins has been postulated to be some kind of internally generated redox change associated with the switch from growth to non-growth, based on the four conserved cysteine residues (Soliveri et al., 2000); and the idea of redox sensing has received support from evidence that WhiD contains an oxygen-sensitive 4Fe,4S cluster (Jakimowicz et al., 2005).

Other developmental genes peculiar to complex actinobacteria

The ssg gene family

The ssgA gene (SCO3926, =SAV4267) was identified first in S. griseus as a high copy-number suppressor of a hypersporulating mutant (Kawamoto & Ensign, 1995; Kawamoto et al., 1997), and was shown to be required for correct sporulation septation of S. griseus (Jiang & Kendrick, 2000), and of S. coelicolor on most media (the exception being minimal medium with mannitol as carbon source) (Van Wezel et al., 2000). It has been suggested that the different abilities of S. griseus and S. coelicolor to sporulate in submerged culture may reflect differences in ssgA regulation in the two organisms, since it is a target of AdpA (and hence is A-factor-dependent) in S. griseus, but is AdpA-independent in S. coelicolor, in which it is regulated solely by the adjacent ssgR gene (SCO3925, =SAV4268, encoding an IclR-like protein) (Traag et al., 2004). There also seem to be functional differences between SsgR and its S. griseus orthologue, SsR, since sssR did not complement an ssgR mutant.

The whi genes described above have little obvious influence on ssgA expression, but, on the other hand, whiH appears to be overexpressed in ssgA or ssgR mutants. This is an interesting observation in view of the implied involvement of ssgA in septation, since whiH has been tentatively implicated in the sporulation-specific regulation of the key cell-division gene ftsZ (see above).

Streptomyces coelicolor has six other ssgA paralogues, and S. avermitilis five (hence the term ‘SALP’ for SsgA-like proteins). In addition, at least two ssgA-like genes are present in T. fusca, and one in K. radiodurans, the gene family being otherwise absent from the database (Noens et al., 2005). One of the paralogues in S. coelicolor, ssgB (SCO1541), is also involved in sporulation: a knockout mutant was white and completely defective in sporulation septation, and, interestingly, had a striking large-colony phenotype (Keijser et al., 2003). It appears that ssgB is not needed for ssgA expression. In a comprehensive gene knockout analysis of the ssgA-like genes of S. coelicolor, it has emerged that all of the SALPs have roles in the precise determination of the pattern of peptidoglycan biosynthesis and hydrolysis that is needed for proper sporulation septation and spore morphogenesis (Noens et al., 2005).

samR, SCO2935 (=SAV5141)

The Beijing laboratory of H. Tan has studied several white colony mutants of Streptomyces ansochromogenes. Most of these have proven to be defective in orthologues of known whi genes of S. coelicolor, providing confirmation (together with similar work on S. aureofaciens from the laboratory of J. Kormanec in Bratislava) that these genes (whiG, B, A) control development in a similar manner in diverse species. In the course of this work, Tan et al. (1998) also discovered a previously unknown developmental gene, samR, mutations in which had a bald colony phenotype; but when they went on to disrupt the orthologous gene of S. coelicolor, the mutant colonies produced fluffy, nonsporulating aerial mycelium similar to that of a whiG mutant. The product of samR is a member of the IclR family of transcriptional regulators, and has no orthologues outside of streptomyces.

whiJ, SCO4543 (SAV3425), and its relatives and their satellite gene families, including bldB

The literature on whiJ is complicated in several ways. The term whiJ was first applied when Ryding (1995) succeeded in complementing the whi-77 mutant analysed by Chater and Merrick (1976) and several other previously little-studied whi mutants from the collection first isolated by Hopwood et al. (1970), from which the whiA, B, D, E, G, H and I loci had been discovered (Chater, 1972). Complementation was attributable to one gene, SCO4543, which was therefore defined as whiJ in the EMBL database entry (AF106004) (J. Ainsa, N.J. Ryding and K.F. Chater, unpublished). The product of this gene has a lambda repressor-like DNA-binding domain at its N-terminus. Subsequently, a transposon-induced whiJ mutant was found to harbour a transposon insertion in SCO4543, and it was pointed out that this gene and its two neighbours (SCO4542 and 4544) are all members of apparently Streptomyces-specific extensive gene families (Gehring et al., 2000). These are found in various combinations, often in combination with members of further gene families. One of these families includes the bldB gene cloned by Piret & Chater (1985). All of these three gene
families are peculiar to streptomycetes and, it now emerges, T. fusca. One of the clusters containing members of these families, termed abaA, was discovered in a search for genes that, at high copy number, increased antibiotic production (Fernandez-Moreno et al., 1992). The extent of the confusion surrounding these genes is illustrated by the fact that whiJ has inadvertently been used to refer to SCO4542 instead of the correct SCO4543, and abaA to refer to just one of the five genes (ORFs1–5) in the abaA locus (SCO0700–0704), in a comparison of BldB with some of its paralogues (Eccleston et al., 2002).

The phenotypes of mutations in the genes flanking whiJ are not yet known. Interestingly, a TTA-containing paralogue of one of the genes flanking whiJ (SCO4544), and of abaA orfA (SCO0702), is present in both long terminal inverted repeats of the linear plasmid SCP1 of S. coelicolor (SCP1.58c, SCP1.295), where it is associated with the genes for three spore-associated proteins (SapC, D and E) (Bentley et al., 2004). It will be rewarding to subject these gene families to more extensive analysis.

**Genes directly involved in morphogenesis and spore structure**

We have already discussed the SapB morphogenetic oligopeptide in some detail, because of its special place as an endpoint of the bld gene cascade; but several other surface components also play special – perhaps unique – roles in morphogenesis (Elliot & Talbot, 2004).

**chpA-H** [in order, SAV2716, 7257, 1674, 2717, 1800, 2705, 2699, 1675; there are *Streptomyces avermitilis* orthologues only of chpC (SAV6636), chpE (SAV6478), chpH (SAV6635)]

The growth of aerial hyphae into the air has turned out to require a partially functionally redundant family of highly surface-active proteins, called ‘chaplins’, encoded by the chp genes (Claessen et al., 2003; Elliot et al., 2003). Some of the chaplins are thought to be covalently linked by sortase enzymes to the aerial hyphal peptidoglycan, and others probably attach noncovalently to the anchored chaplins. The only known occurrence of chaplins outside streptomycetes is in the aerial mycelium-forming T. fusca, which has a single chp gene. It is considered that chaplins enable aerial hyphae to break through the surface tension barrier of a water layer that separates the substrate mycelium from direct contact with the air (Claessen et al., 2003; Elliot et al., 2003; Elliot & Talbot, 2004; Gebbink et al., 2005); but we think it is also possible that the chaplin layer acts to capture a column of water between the aerial hyphal membrane and the chaplin-containing layer, to provide an aqueous environment within which cellular growth of aerial hyphae can take place. This alternative theory might also apply to the analogous hydrophobins of mycelial fungi. The production of chaplins depends on the integrity of the extracellular cascade involving bld genes (see above).

**rdlA and rdlB** (SCO2718, 2719)

The surface of aerial hyphae of *S. coelicolor* also contains another class of amphipathic proteins, the rodlin, which, possibly interact with chaplins, and are encoded by two adjacent rdl genes (Claessen et al., 2002; Claessen et al., 2004). Although they are made only by aerial hyphae, rodlin are not essential for aerial growth, and indeed they are absent from *S. avermitilis*, apparently because of a deletion of a genomic segment including the rdl genes (they are present in many other streptomycetes, but are unknown in other organisms) (Claessen et al., 2004). Interestingly, mutants lacking rodlin (Claessen et al., 2002) or chaplins (Gebbink et al., 2005) attach less well to hydrophobic surfaces, loosely supporting the idea that aerial hyphae may serve as exploratory hyphae as well as sporophores (Yeo & Chater, 2005).

**whiE** (SCO5314–5321)

This is a cluster of eight genes, most of which encode recognizable enzymes of polyketide biosynthesis (Davis & Chater, 1990). Mutations in *whiE* genes cause a loss of spore pigment or a change in spore colour, and there is considerable information about the roles of these enzymes in the synthesis of a 24-carbon chain-length polyketide (Shen et al., 1999; Moore & Piel, 2000). However, the mature WhiE polyketide has not been characterized because it seems to be covalently attached to the spore wall. The *whiE* genes form two diverging transcription units, both of which are switched on only during sporulation. Both promoters are dependent on all the early *whi* genes, but only the one reading into orfVIII is also dependent on the late sporulation-specific SigF σ factor (Kelemen et al., 1998).

The *S. avermitilis* cluster (SAV2837–2844) and the partially sequenced cluster of *S. halstedii* are colinear with that of *S. coelicolor*. Although the clusters are flanked on both sides by quite different genes in *S. avermitilis* and *S. coelicolor*, their overall positions in the genome are conserved. The conservation of genomic position suggests that the flanking genes have been subjected to extensive flux (insertion and deletion differences), while the *whiE* genes themselves have remained anchored. This interpretation may be over-simple, in view of the results of a large-scale survey of *whiE* ORFII genes in diverse streptomycetes (Metsa-Ketela et al., 2002), which showed that the phylogeny of these strains obtained by ribosomal DNA sequencing differed substantially from that obtained by analysis of
the whiE genes. This surprising finding could mean that whiE genes have intermittently been transferred laterally between genomes, raising the question – yet to be addressed by further genomic comparisons – of whether other segments of the core genome might also have been exchanged between streptomycetes over evolutionary time.

Remarkably, the genome of S. griseus contains no whiE gene cluster (Y. Ohnishi and S. Horinouchi, pers. commun.). Instead, the pigmentation of both spores and mycelium is attributable to a form of melanin that results from the action of a cytochrome P-450 (P-450mel) on 1,3,6,8-tetrahydroxynaphthalene, itself made by the condensation of five molecules of malonyl-CoA by a type III polyketide synthase (RppA) (Funa et al., 2005). P-450mel and rppA genes are also present in S. coelicolor (albeit somewhat diverged from their S. griseus counterparts), but they do not appear to contribute to pigmentation (Izumikawa et al., 2003).

crgA, SCO3854 (SAV4331)

Like samR, crgA (Del Sol et al., 2003) affects differentiation differently in different species. It encodes a small, membrane-bound protein that inhibits premature sporulation septation in S. coelicolor. A crgA mutant showed premature aerial mycelium growth and sporulation, as well as early production of the blue antibiotic actinorhodin. On the other hand, a crgA mutant of S. avermitilis had white, coiled, mostly nonsporulating aerial mycelium. Multiple copies of crgA in both species gave white colonies in which the aerial hyphae did not undergo sporulation septation. The gene is present (and shows local synteny) in all actinomycete genomes, and it would be interesting to know whether it always affects septation.

Other genes important for development

The morphological changes associated with development involve several proteins that determine the location and components of events associated with sporulation septation. These include FtsZ (cell-division) and ParAB (chromosome segregation), both of which have complex promoter regions, with separate segments being responsible for expression during vegetative growth or differentiation (Flardh et al., 2000; Jakimowicz et al., 2006). On the other hand, a different solution to the need for expression at more than one developmental stage has been found for at least one set of enzymes active at different developmental stages: glycogen deposition at the end of vegetative growth or at the end of aerial growth involves differently regulated but almost identical operons for isoforms of four enzymes that include the glycogen branching enzyme (Schneider et al., 2000; Yeo & Chater, 2005).

Conclusion: an overview of the evolution of Streptomyces development

The pattern of occurrence of the genes discussed in this article in different genomes can be used in an attempt to reconstruct the evolution of Streptomyces development (Fig. 8). It seems that the aerial growth-stimulating system involving both SapB/AmfS-like lanthionine-containing oligopeptides and chaplins was established very early in the evolutionary emergence of streptomycetes, presumably providing a means by which a mycelium could proceed to aerial growth.
growth in a coordinated manner. The production of these molecules in *S. coelicolor* is itself dependent on other extracellular signals, most of which are still uncharacterized; but at least one of them (signal 1) is an incompletely defined oligopeptide (Nodwell & Losick, 1998). These signals within the colony indicate activity of the BldH, G, C and D regulatory proteins, as well as of the uncharacterized *bldf* gene product that is needed to generate signal 1. It has been proposed that these gene products are involved in responses to different stresses that would otherwise limit continued growth (Chater & Horinouchi, 2003): thus, only when such responses have all been activated would the ‘last resort’ of terminal differentiation (aerial growth at the expense of the accumulated substrate mycelial biomass) be undertaken. At least in this sense, the signal cascade ending in SapB accumulation might be seen to be equivalent to the intra-cellular sporulation phosphorelay in *Bacillus* spp. and their relatives, which culminates in the accumulation of phosphorylated Spo0A regulatory protein, the trigger for commitment to endospore formation (Errington, 2003; Fujita & Losick, 2005).

It seems likely that AdpA (BldH) played a key role in development early in *Streptomyces* evolution (from somewhere between 440 and 220 Ma), and that even then the UUA codon of *adpA* mRNA was being used as a route through which *bldA* influenced development. However, certain other details of *adpA* regulation show species-specificity, such as its A-factor-sensitive repressibility by ArpA in *S. griseus*. The nature and range of signal inputs affecting AdpA levels and/or activity have probably developed as part of the adaptation of different species to different ecological niches, or their adoption of different strategies for responding to a shared niche. For example, colonies of one species might sporulate comprehensively and more or less synchronously at a hint of imminent hard times, while those of another could have evolved to delay sporulation as long as possible in case an environmental change restores conditions favourable for growth, making it possible to achieve greater biomass before commitment to sporulation. Since there are orthologues of *bldG, bldC* and *bldD* in *T. fusca*, but apparently not in other completely sequenced actinobacteria, it is likely that any stress responses in which these genes take part predated the emergence of streptomycetes. *T. fusca* and streptomycetes are alone among sequenced actinobacteria in being obligately mycelial and reproducing principally via a sporulating aerial mycelium, so these stresses might conceivably be associated with this life-style. The absence of the *amf* system and of multiple SALPs, chaplins and rodlinS from all organisms except streptomycetes suggests that the acquisition of these features was associated with the development of the *Streptomyces* aerial mycelium, and that *T. fusca* may have a different means of solving the problems associated with reproduction via aerial growth, albeit probably using its single chaplin and two SALPs. It has been suggested that *Streptomyces* aerial hyphae may serve not only as sporophores, but alternatively as the means by which a mycelial organism can extend across the air space of tiny cavities in the soil habitat to reach and colonize other nearby surfaces (Yeo & Chater, 2005). This suggests, in turn, that exploratory aerial growth could have preceded the acquisition by aerial hyphae of their reproductive function, and provided the evolutionary opportunity for it. The capture of *whiG, whiH* and *whil*, followed by their adaptation to become increasingly precise regulators of the fragmentation process that became sporulation, presumably happened around the time of the birth of the first ancestor peculiar to streptomycetes. In this light, it is curious that *whiH* and *whil* are located about 200 and 400 genes away from *whiG*, in both *S. coelicolor* and *S. avermitilis*, and that the intervening regions, although not showing extensive synteny with other actinobacterial genomes, include a number of genes for which there are orthologues in most actinobacteria. If *whiG* and its targets were acquired in a single event, for example as a regulon on a plasmid, there must have been several changes in chromosomal organization in the *whiG,H,I* region quite soon after that initial acquisition event.

If many of the genes for development have been horizontally acquired, what is their source? Most of them are absent from other known bacteria, but a huge pool of novel genes is present in bacteriophages. Presumably, many of these genes have evolved under extreme selective pressures in the face of the constant evolution of host organisms to resist death by phage infection, as well as under occasional competition between phages infecting the same host (Hendrix, 2005). This means that phages are potentially a rich source both of novel regulatory genes and of genes for surface-acting and other structural proteins. Both of these gene classes are important for the evolution of developmental and morphological complexity. Moreover, phages provide the means by which new genes can be efficiently introduced into bacterial hosts (Hendrix, 2005). At least in the case of *whiB*-like genes, there are examples in phage genomes. One can push the idea of phages as a source of developmental possibilities further, in considering how the use of the *bldA*-specified tRNA for a rare codon might have acquired specific developmental significance. Thus, a drift towards high GC content could be associated with selection against phages through codon bias, and could eventually permit a loss of the UUA-reading tRNA, and an associated complete inability to translate UUA codons that would confer resistance to phages that employ this codon. Some phages would counter this by acquiring a copy of the tRNA (tRNA genes are often present in phage genomes). By this means, the tRNA might be reintroduced into a prototype-streptomycete as part of a prophage genome. Because of continued selection for phage resistance, the
newly acquired tRNA might not be expressed abundantly during vegetative growth (the main phage-sensitive stage of bacteria), but it could be expressed without disadvantage in stationary phase, and come to be exploited by genes that are specifically useful in stationary phase, as seems to be the case for TTA-containing genes of streptomycetes (Chater, 2006).

Spores (as we know them) presumably evolved from mycelial fragments, under selective pressures that may have included the need to survive inside the guts of worms and other invertebrates that consumed the plant remains on which the early streptomycetes were growing. The ability of spores to survive in such a hostile environment would have been enhanced by the cross-linking of the aromatic spore pigment to the spore wall in a lignification-like process; it is not surprising, therefore, that \textit{whiE} genes appear to have been acquired early in the \textit{Streptomyces} lineage (notwithstanding their absence from \textit{S. griseus}: possibly the melanin pigment responsible for spore colour in that organism also ‘lignifies’ spore walls). Most antibiotics, including many other aromatic polyketides, are much more strain-specific, and have presumably been acquired more recently. Two sporulation-specific gene sets, for SapB and spore pigment, have some relatedness to genes for antibiotic production (lantibiotics and aromatic polyketides, respectively). Since their acquisition by \textit{Streptomyces} appears to have preceded the acquisition of antibiotic production genes, it is an intriguing possibility that they may have been important in the origin of such genes.

The \textit{whiAB} part of the sporulation regulatory cascade seems to have had a different evolution, involving vertically inherited (rather than laterally acquired) genes whose initially nondevelopmental regulation and regulatory roles came to be integrated into sporulation. The presence of \textit{whiA} orthologues in all gram-positive bacteria indicates a particularly ancient lineage, preceding the divergence of the low and high GC gram-positive bacteria some 2 Ga. No role for WhiA-like proteins has been found in any nonstreptomycetes, but the evidence connecting WhiA with the orderly shut-down of growth may provide a clue for studies of orthologues in other bacteria. On the other hand, \textit{whiB} orthologues and their paralogues (generally termed \textit{wbl} genes) are confined to actinobacteria. In view of evidence that the products of such genes may be redox-sensitive regulators, it is interesting that actinobacteria share another special feature of their redox regulation: they use mycothiol as a thiol-disulphide buffer, a role played by glutathione in many other bacteria (Fahey, 2001). Possibly, there is a Wbl–mycothiol association, which could have been fixed at the time when the actinobacteria emerged (about 1.5–2 Ga), coinciding with the first appearance of oxygen at significant levels in the atmosphere. Five \textit{wbl} genes (\textit{wbiB}, \textit{wbiD}, \textit{wblA}, \textit{wblC} and \textit{wblE}) are present in most actinobacteria, so \textit{wbl} genes must have diversified very early in actinobacterial evolution. One of them, \textit{wblC}, is involved in the expression of multiple antibiotic resistance in both mycobacteria and streptomycetes (Morris \textit{et al}., 2005), so, by extrapolation, it is likely that each of the five had acquired a specific physiological role at this early stage of evolution. We therefore suppose that \textit{wbiB} and \textit{wbiD} were subsequently (and probably sequentially) sequestered for sporulation in streptomycetes, rather than evolving by duplication and divergence from a common sporulation-specific progenitor during a gradual increase in the complexity of sporulation control. Thus, the evolution of the later stages of sporulation, which might be equated with an increase in the effectiveness of mycelial fragments as survival and dispersal agents, was probably associated with the acquisition of successively later-acting regulons; hence the adoption of \textit{sigF} and \textit{whiD}, each a member of a paralogous gene family, as \textit{Streptomyces} sporulation-specific controlling elements. This process may have interesting differences from the evolution of developmental complexity that has emerged from other systems. For example, in \textit{B. subtilis} endospore formation, each of the two initial compartment-specific \textit{σ} factors, \textit{σ}^{E} and \textit{σ}^{F}, is replaced in the later stages of sporulation by a closely related paralogous \textit{σ} factor (\textit{σ}^{K} and \textit{σ}^{G}, respectively) (Errington, 2003), strongly suggesting that the evolution of the late stages of sporulation, during which the spores acquire their resistance properties, involved the duplication of the sporulation-specific progenitors of the \textit{σ}^{E} and \textit{σ}^{F} genes, and their subsequent divergence. This theme of developmental evolution by the duplication and divergence of developmentally specific master control genes is also found in the homeobox genes of vertebrates and the MADS-box genes that determine floral organ identity in plants (Cohen, 1999). We note here that the multiple SALPs of streptomycetes, which play important roles in consecutive stages of cell wall reorganization during sporulation, seem to have evolved by duplication and divergence as the process of sporulation of aerial hyphae evolved (Noens \textit{et al}., 2005).

Unlike most other \textit{whi} genes, both \textit{whiA} and \textit{whiB} have two promoters, one of which is low-level and constitutive, while the other is associated with development (Soliveri \textit{et al}., 1992; Flardh \textit{et al}., 2000). The developmental promoters appear to be part of an autoactivating circuit in which both gene products affect the expression of both genes (though it is not known if this regulation is direct) (Jakimowicz \textit{et al}., 2006). Although the constitutive promoters may simply serve to provide a priming element in the autoregulatory developmental circuit, it is also possible that WhiA and WhiB serve some other physiological purpose in vegetative hyphae (perhaps similar to whatever their role is in nondifferentiating actinobacteria). Likewise, the \textit{ftsZ} and \textit{parAB} genes, both of which have somewhat different roles in vegetative growth and sporulation, also have dual promoters, one for vegetative functioning and one for sporulation.
(Flardh et al., 2000; Jakimowicz et al., 2006). Thus, it is quite common to find that genes have acquired additional RNA polymerase interaction sites as developmental complexity has evolved. In the cases of ftsZ and parAB, it seems that WhiH and WhiA/WhiB are, respectively, closely involved in this developmental activation (Flardh et al., 2000; Jakimowicz et al., 2006). No doubt many other examples of the evolutionary plasticity of promoter regions will emerge in future analysis of Streptomyces development.

Note added in press

The genome sequence of a representative of the mycelial nitrogen-fixing symbiont Frankia has very recently become available (http://genome.jgi-psf.org/draft_microbes/fra_c/fra_c.home.html). Although members of this genus are not available (http://genome.jgi-psf.org/draft_microbes/fra_c/home.html), the genome sequence of a representative of the mycelial nitrogen-fixing symbiont Frankia has very recently become available (http://genome.jgi-psf.org/draft_microbes/fra_c/fra_c.home.html). Although members of this genus are not available (http://genome.jgi-psf.org/draft_microbes/fra_c/home.html), the genome sequence of a representative of the mycelial nitrogen-fixing symbiont Frankia has very recently become available (http://genome.jgi-psf.org/draft_microbes/fra_c/fra_c.home.html).

Acknowledgements

This work was supported by grant IGF12432 from the BBSRC, a Competitive Strategic Grant to the John Innes Centre from the BBSRC, and the John Innes Foundation. We thank Alan Ward for discussions on Streptomyces phylogeny, Sueharu Horinouchi and Yasuo Ohnishi for communicating unpublished results, and David Hopwood for helpful comments on the manuscript.

References


Claessen D, Wosten HAB, vanKeulen G, Faber OG, Alves AMCR, Meijer WG & Dijkhuizen L (2002) Two novel homologous proteins of Streptomyces coelicolor and Streptomyces lividans...
are involved in the formation of the rodlet layer and mediate attachment to a hydrophobic surface. *Mol Microbiol* **44**: 1483–1492.  
Fujita M & Losick R (2005) Evidence that entry into sporulation in *Bacillus subtilis* is governed by a gradual increase in the level and activity of the master regulator Spo0A. *Genes Dev* **19**: 2236–2244.  


Soliveri JA, Gomez J, Bishai WR & Chater KF (2000) Multiple paralogous genes related to the *Streptomyces coelicolor* developmental regulatory gene *whiB* are present in


