

# The Sequence of a 1.8-Mb Bacterial Linear Plasmid Reveals a Rich Evolutionary Reservoir of Secondary Metabolic Pathways

Marnix H. Medema<sup>1,2</sup>, Axel Trefzer<sup>3</sup>, Andriy Kovalchuk<sup>1</sup>, Marco van den Berg<sup>3</sup>, Ulrike Müller<sup>3</sup>, Wilbert Heijne<sup>3</sup>, Liang Wu<sup>3</sup>, Mohammad T. Alam<sup>2</sup>, Catherine M. Ronning<sup>4</sup>, William C. Nierman<sup>4</sup>, Roel A. L. Bovenberg<sup>3,5</sup>, Rainer Breitling<sup>2,6</sup>, and Eriko Takano<sup>\*,1</sup>

<sup>1</sup>Department of Microbial Physiology, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Haren, The Netherlands

<sup>2</sup>Groningen Bioinformatics Centre, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Haren, The Netherlands

<sup>3</sup>DSM Biotechnology Center, DSM Food Specialties B.V., Delft, The Netherlands

<sup>4</sup>J. Craig Venter Institute, Rockville, Maryland

<sup>5</sup>Centre for Synthetic Biology, University of Groningen, Groningen, The Netherlands

<sup>6</sup>Faculty of Biomedical and Life Sciences, University of Glasgow, Glasgow, United Kingdom

\*Corresponding author: E-mail: e.takano@rug.nl.

**Accepted:** 23 March 2010

The Whole Genome Shotgun project of *S. clavuligerus* strain ATCC 27064 has been deposited at DDBJ/EMBL/GenBank under the accession ADGD00000000. The version described in this paper is the first version, ADGD01000000.

## Abstract

Plasmids are mobile genetic elements that play a key role in the evolution of bacteria by mediating genome plasticity and lateral transfer of useful genetic information. Although originally considered to be exclusively circular, linear plasmids have also been identified in certain bacterial phyla, notably the actinomycetes. In some cases, linear plasmids engage with chromosomes in an intricate evolutionary interplay, facilitating the emergence of new genome configurations by transfer and recombination or plasmid integration. Genome sequencing of *Streptomyces clavuligerus* ATCC 27064, a Gram-positive soil bacterium known for its production of a diverse array of biotechnologically important secondary metabolites, revealed a giant linear plasmid of 1.8 Mb in length. This megaplasmid (pSCL4) is one of the largest plasmids ever identified and the largest linear plasmid to be sequenced. It contains more than 20% of the putative protein-coding genes of the species, but none of these is predicted to be essential for primary metabolism. Instead, the plasmid is densely packed with an exceptionally large number of gene clusters for the potential production of secondary metabolites, including a large number of putative antibiotics, such as staurosporine, moenomycin,  $\beta$ -lactams, and enediynes. Interestingly, cross-regulation occurs between chromosomal and plasmid-encoded genes. Several factors suggest that the megaplasmid came into existence through recombination of a smaller plasmid with the arms of the main chromosome. Phylogenetic analysis indicates that heavy traffic of genetic information between *Streptomyces* plasmids and chromosomes may facilitate the rapid evolution of secondary metabolite repertoires in these bacteria.

**Key words:** genome sequence, chromosomal evolution, natural products, *Streptomyces clavuligerus*.

## Introduction

Prokaryotic chromosomes were originally thought to lack many of the features that characterize eukaryotic chromosomes, such as linearity and the possession of telomeres (Bendich and Drlica 2000). However, during recent decades, a few bacterial taxa have been shown to have linear chro-

mosomes that contain telomere-like structures (Hinnebusch and Tilly 1993). Among these are the high-GC, Gram-positive, soil-dwelling *Streptomyces* bacteria (Chen et al. 1994; Hopwood 2006), renowned for their capacity to produce a vast array of natural products. Their linear chromosomes are relatively long (8–10 Mb) and consist of

© The Author(s) 2010. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.5>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

a conserved “core” of 5–6 Mb and variable “arm” regions (Hopwood 2006). Nearly all genes that are likely to be unconditionally essential are located within the core. Interestingly, besides possessing linear chromosomes, streptomycetes also often contain linear plasmids, extrachromosomal DNA molecules that replicate independently from the main chromosome and contain their own telomere-like structures (Chater and Kinashi 2007; Chen 2007). Evidence is accumulating that the evolution of streptomycete chromosomes and plasmids can be very dynamic because of their relative instability (Volf and Altenbuchner 1998; Chen et al. 2002; Widenbrant et al. 2007).

Substantial effort in the study of streptomycetes is focused on the secondary metabolites they produce. Many bacterial secondary metabolites or their derivatives are used as antimicrobial agents, whereas some are used as anti-tumor drugs, immunosuppressive agents or cholesterol-lowering drugs (Bode and Muller 2005; Gullo et al. 2006; Newman and Cragg 2007). More than half of all known antibiotics originate from the streptomycetes (Berdy 1995; Challis and Hopwood 2003), and many more clearly remain to be discovered. A statistical analysis of antibiotic discovery has predicted that streptomycetes may have the capacity to produce as many as  $10^5$  secondary metabolites (Watve et al. 2001). The finding that a single genus can carry such a massive number of secondary metabolite biosynthetic gene clusters is correlated with the location of many of them in the arm regions of the chromosome, which are extremely variable between species. In a few cases, such clusters have been shown to be plasmid-borne. Thus, the 365-kb plasmid SCP1 of *Streptomyces coelicolor* carries the biosynthetic gene cluster for methylenomycin A (Kirby and Hopwood 1977; Bentley et al. 2004), and the 210-kb plasmid pSLA-2 of *Streptomyces rochei* carries four secondary metabolite biosynthetic gene clusters: three polyketide synthase (PKS) clusters (for lankacidin, lankamycin, and a mithramycin-like compound) and a carotenoid biosynthetic cluster (Mochizuki et al. 2003).

One of the most important industrial streptomycete species is *Streptomyces clavuligerus*. Multiple natural products from this organism have been structurally and functionally characterized, and two of these are used in the clinic, the  $\beta$ -lactam antibiotic cephamycin C (Alexander and Jensen 1998), and the  $\beta$ -lactamase inhibitor clavulanic acid (Saudagar et al. 2008). Clavulanic acid is widely used in combination with the semisynthetic  $\beta$ -lactam amoxicillin to treat various bacterial infections (Brogden et al. 1981). The biosynthetic genes for both compounds are encoded by a single intensively studied gene supercluster (Liras et al. 2008). Several clavam antibiotics closely related to these  $\beta$ -lactam ring-containing compounds are also produced by *S. clavuligerus* and are encoded by separate gene clusters (Evans et al. 1983; Tahlan, Anders, and Jensen 2004; Tahlan, Park, and Jensen 2004; Tahlan, Park, Wong, et al. 2004; Tahlan et al. 2007;

Zelyas et al. 2008). At least three additional antibiotics have been reported to be produced by *S. clavuligerus*, although no biosynthetic genes for these have been described: the pyrrothine class antibiotic holomycin (Kenig and Reading 1979; Oliva et al. 2001), a tacrolimus-like macrolide (Kim and Park 2008), and a compound related to the nucleoside antibiotic tunicamycin (Kenig and Reading 1979).

Here, we describe the genome sequence of *S. clavuligerus* ATCC 27064 and show that this species has a unique 1.8-Mb linear megaplasmid, which is densely packed with 25 putative secondary metabolite gene clusters, in addition to its 6.8-Mb chromosome, which also contains 23 such clusters. Some of these clusters strongly resemble known antibiotic gene clusters, others appear to be completely novel, and a number show extravagant features that have never been observed before. The megaplasmid found in *S. clavuligerus* is by far the largest linear megaplasmid ever sequenced, and its gene complement sheds light on the rapid and dynamic evolution of secondary metabolite repertoires in bacteria, in addition to being a rich and compact potential source of novel bioactive metabolites.

## Materials and Methods

### Genome Sequencing and Assembly

The genome of *S. clavuligerus* ATCC 27064 was sequenced and assembled by random shotgun sequencing. Sanger sequencing of shotgun libraries with insert sizes of 3, 10, and 50 kb was accomplished using ABI 3700 sequencers as described by Venter et al. (2001). Sequences were assembled using Celera Assembler (Levy et al. 2007). Genome assembly was facilitated and validated with an optical restriction map (Latreille et al. 2007) acquired from OpGen (Madison Wisconsin). Finally, as many gaps as possible were filled using sequences from a cosmid library created in-house at the DSM Biotechnology Center. We estimate the final genome assembly to be >99.7% complete.

### Gene Prediction, Functional Annotation, and Comparative Analysis

Straightforward gene prediction by Glimmer3 (Delcher et al. 2007) appeared to give a large number of false positives and negatives. The high GC content of *Streptomyces* species leads to the presence of many long open reading frames (ORFs) which are not actual genes but only exist by coincidence because of the low number of AT-rich stop codons. These long reading frames often are not actual genes. Therefore, we complemented Markov model-guided prediction by Glimmer3 with mapping of BlastP searches of all putative ORFs to the nonredundant (nr) protein database (using an in-house Python script) and by analysis of the GC frame plots in Artemis (Rutherford et al. 2000) to manually identify all putative genes accurately.

Transfer RNA (tRNA) and tmRNA genes were identified using ARAGORN, and ribosomal RNAs (rRNAs) were

identified using the RNAmmer tool. For functional annotation, all proteins were first processed with the AutoFACT automatic annotation pipeline (Koski et al. 2005) through hierarchical information transfer from the best hits in the Uniref90, NCBI nr, COG, KEGG, CDD, PFAM, and SMART databases. A round of manual re-annotation followed to ascertain the accuracy of the annotation.

Genome alignment of the DSM and Broad Institute draft genomes was performed using the Mauve genome alignment program (Darling et al. 2004). Clusters of orthologous groups were constructed using the OrthoMCL software (Li et al. 2003), using cutoffs of  $e$  value  $<1 \times 10^{-05}$  and identity  $>60\%$ . Dotplots for genome comparison were created with Gepard (Krumisiek et al. 2007). GC skew plots were created using GenSkew (<http://mips.gsf.de/services/analysis/genskew/>).

Phylogenetic analysis of plasmid-encoded transposase proteins was performed as follows. The *S. clavuligerus* transposase sequences were blasted against the nr protein database with BlastP. For each query sequence, the 25 best blast hits were aligned using Muscle 3.6 (Edgar 2004), the alignment was quality trimmed with Gblocks 0.91b (Talavera and Castresana 2007), and an approximate maximum likelihood phylogenetic tree was generated using FastTree 2.1.1 (<http://www.microbesonline.org/fasttree/>). Finally, all the hits incorporated in the trees were blasted against a self-assembled database of *Streptomyces* chromosomes to identify transposases from *Streptomyces* chromosomal termini (distance  $<1$  Mb to either end of the chromosome) using BlastX. Query sequences present in a monophyletic taxon comprised of at the most three proteins, in which at least one of the two other proteins was a transposase from a *Streptomyces* chromosomal terminus, were considered as positive hits.

### Metabolic Model Construction

Published genome-scale metabolic models of *S. coelicolor*, *Escherichia coli*, and *Saccharomyces cerevisiae* (Duarte et al. 2004; Borodina et al. 2005; Feist et al. 2007) were used together with a genome-scale model of *Penicillium chrysogenum* (DSM, unpublished data) as a reference to construct the metabolic model of *S. clavuligerus*. Gene-reaction associations were putatively assigned to *S. clavuligerus* genes based on sequence homology to genes included in the reference genome-scale metabolic models. “Dead-end” metabolites (i.e., metabolites that are neither consumed nor produced under any conditions) were identified, and gaps were closed, if possible, by Blast searches on the *S. clavuligerus* genome with enzymes from the KEGG database that could connect dead-end metabolites. The gene-reaction associations were further manually verified and curated using the Enzyme Commission (EC) numbers from the functional annotation to correct for erroneous gene-reaction associations. The metabolic model was checked for biomass formation *in silico* under minimal growth medium

conditions (glycerol, ammonia, phosphate, and sulfate) using the COBRA toolbox (Becker et al. 2007). An SBML version of the final model is available from the authors upon request.

## Results and Discussion

### Architecture of the *S. clavuligerus* Genome

Pulse-field gel electrophoresis of restriction-digested fragments in the nineties lead to an estimated size of 6.8 Mb for the chromosome of *S. clavuligerus* (Chen et al. 1994). Genome sequencing and assembly confirmed this estimate but, remarkably, a second large contig of 1.8 Mb was identified as well.

The optical restriction map that we obtained indicated that this second contig represents a separate replicon. This was confirmed by almost perfectly symmetrical cumulative GC skew plots of both contigs (supplementary fig. S1, Supplementary Material online), characteristic for bacterial chromosomes and plasmids (Grigoriev 1998). Origins of replication (Oris) were identified near the peak of the plots for both replicons on the basis of homology.

The identification of the two large contigs was later independently confirmed by a prepublication draft sequence from the Broad Institute (GenBank accession number ABJH00000000), although our assembly of the 6.8-Mb replicon differed from theirs by a large inversion at virtually identical rRNA operons (supplementary fig. S2, Supplementary Material online). The optical restriction map that we obtained and our GC skew analysis indicated that our assembly is most probably correct.

The discovery of a second large replicon in *S. clavuligerus* is not altogether surprising, as Kirby (1978) suspected the presence of a plasmid when he found that a genetic element without linkage to the chromosome influenced holomycin production. Intriguingly, pulse-field gel electrophoresis in the closely related *S. clavuligerus* strain NRRL 3585—the type strain which should be identical to strain ATCC 27064—has shown the presence of three plasmids (pSCL1–3) ranging in size between 11 and 430 kb (Netolitzky et al. 1995). However, only a 7-kb replicon matching to the smallest of these plasmids—the 11-kb plasmid pSCL1 (Wu and Roy 1993)—was identified in our assembly. No replicons of the sizes reported for plasmids pSCL2 or pSCL3 were identified in our assembly, and the nucleotide sequence of pSCL2 (Wu et al. 2006) did not match to any part of our contigs in a BlastN analysis. This implies that the evolution of streptomycete genomes by plasmid acquisition, loss, or recombination is extremely dynamic.

We identified 7,281 putative protein-encoding genes on the two large replicons, as well as six rRNA operons, 72 tRNA genes, and 14 pseudogenes (table 1). Strikingly, the sum of all genomic features of the two replicons matches very well to the typical make-up seen in other *Streptomyces* genomes.

**Table 1**Genome Architecture of *Streptomyces clavuligerus* ATCC 27064

	<i>S. clavuligerus</i>			<i>Streptomyces coelicolor</i>	<i>S. avermitilis</i>	<i>S. griseus</i>
	Chromosome	pSCL4	Total	Chromosome	Chromosome	Chromosome
Sequence length	6.760 Mb	1.796 Mb	8.556 Mb	8.668 Mb	8.546 Mb	9.026 Mb
GC%	72%	72%	72%	72%	72%	71%
rRNA operons	6	0	6	6	6	6
tRNAs	66	7	73	63	68	66
tmRNAs	1	0	1	1	1	1
Coding sequences	5,700	1,581	7,281	7,825	7,574	7,138
Secondary metabolite clusters	23	25	48	20	30	34

### The 1.8-Mb Replicon Has the Hallmarks of a Giant Linear Plasmid

We set out to characterize and compare the two replicons. First of all, the central region of the 6.8-Mb replicon had very large regions of conserved synteny to the chromosomes of the four completed *Streptomyces* genomes, whereas the smaller replicon had no regions with significant homology to these chromosomes (supplementary fig. S3, Supplementary Material online). Moreover, the large replicon had an origin of replication typical of *Streptomyces* chromosomes, including *dnaA* and *dnaN* (SCLAV\_2911–2912). The 6.8-Mb replicon thus appears to represent a typical *Streptomyces* chromosome. Interestingly, it is significantly smaller than other published *Streptomyces* chromosomes, which range in size between 8.5 and 9.0 Mb (Bentley et al. 2002; Ikeda et al. 2003; Ohnishi et al. 2008) (table 1). Comparative analysis shows that the conserved core region of *Streptomyces* chromosomes is present in *S. clavuligerus*, yet the typical large chromosomal arms are not (supplementary fig. S3, Supplementary Material online).

As stated in the previous section, the smaller replicon appeared to have its own Ori, highly homologous to those of the *S. coelicolor* plasmid SCP1 and the *S. rochei* plasmid pSLA2-L, two *Streptomyces* plasmids that are known to harbor secondary metabolite gene clusters. The Ori lies at around 945 kb, adjacent to close homologues of the plasmid-type DNA primase/replication proteins Orf1 and Orf2 (SCLAV\_p0889 and SCLAV\_p0888) from the *S. coelicolor* SCP1 plasmid (Redenbach et al. 1999) and an additional DNA primase/polymerase (SCLAV\_p0890).

Both replicons have genes that encode ParAB chromosome partitioning proteins near their predicted Ori (<10 kb). The *parAB* genes of the 6.8-Mb replicon (SCLAV\_2901–2902) strongly resemble the *S. griseus* and *S. avermitilis* chromosomal *parAB* genes, whereas the *parAB* genes of the 1.8-Mb replicon (SCLAV\_p0884–0885) are plasmid-type and strongly resemble the *parAB* genes on the SCP1 plasmid of *S. coelicolor* A3(2) (supplementary fig. S4, Supplementary Material online).

The smaller replicon also has its own *tap* and *tpg* genes, which encode proteins necessary for telomere replication of

*Streptomyces* chromosomes and plasmids (Bao and Cohen 2001, 2003). No *tap* or *tpg* genes were identified on the main chromosome, as for the *Rhodococcus* sp. RHA1 genome (McLeod et al. 2006). The chromosome therefore seems to depend on the *tap/tpg* genes of the smaller replicon. However, there could also be a different, unknown, system for telomere replication; the recent identification of the *tac/tpc* system (not present on the *S. clavuligerus* chromosome) of the *S. coelicolor* SCP1 plasmid shows that the *tap/tpg* system is not universal for streptomycete chromosomes and plasmids (Huang et al. 2007).

Genes for conjugative transfer were also detected on both replicons. Again, although the chromosome carries typical chromosomal-type *traAB* genes (SCLAV\_4235–4236), the smaller replicon has plasmid-type *traAB* genes (SCLAV\_p0254–0255) similar to those of the *S. coelicolor* plasmid SCP2 (Brolle et al. 1993).

Based on these observations, we conclude that the 1.8-Mb replicon has all the hallmarks of being a giant version of a typical *Streptomyces* linear plasmid.

### The Smaller Replicon Is a Megaplasmid Predicted to be Dispensable for the Core Metabolism of *S. clavuligerus*

Extrachromosomal genetic elements of bacteria can have quite different functions compared to the chromosomes. In some cases, they encode only accessory functions, whereas in others they have evolved to encode essential cellular functions as well. Although megaplasmids and chromosomes are part of a continuous spectrum, Bentley and Parkhill (2004) have proposed to distinguish chromosomes from megaplasmids not by their replication genes but by whether they are essential for growth of the organism and whether they carry rRNA operons. We therefore set out to predict whether the 1.8-Mb replicon is essential for growth of *S. clavuligerus*.

Strikingly, all stable RNAs necessary for primary metabolism (rRNA and tRNAs) are encoded on the main chromosome (table 1). A few putative tRNAs are encoded on the 1.8-Mb replicon, but these appear functionally redundant compared with those on the chromosome according to their

**Table 2**Genome-Scale Metabolic Model of *Streptomyces clavuligerus*

Number of reactions	1,492
Number of metabolites (internal/external)	971/202
Dimensions of S (metabolites by reactions)	1,492 × 1,173
Number of internal fluxes	1,290
Number of exchange fluxes	202
Number of chromosomal genes included	785
Number of plasmid genes included	79
Total number of genes included	864
Number of essential genes	100

amino acid specificity and anticodon sequence. To investigate whether the protein-coding genes of the replicon are likely to be involved in essential cellular functions, a *Streptomyces* core genome (Ussery et al. 2008) of 976 genes was defined, with each gene representing a cluster of orthologous genes present in all four completed *Streptomyces* genomes. This core genome had only 14 significant BlastP hits ( $e$  value  $< 1 \times 10^{-05}$  and identity  $>60\%$ ) to the 1.8-Mb replicon, distributed throughout its length, with none appearing to be required for crucial cellular processes (see supplementary table S1, Supplementary Material online). Furthermore, all 21 out of the 24 *Streptomyces*-specific signature genes (Ohnishi et al. 2008) that are present in the *S. clavuligerus* genome are located on the *S. clavuligerus* chromosome. Therefore, the 1.8-Mb replicon does not seem to encode any of the essential genetic complement of *S. clavuligerus* ATCC 27064. This is corroborated by the fact that almost all the genes with the highest codon adaptation indexes and third codon GC percentages (GC3s)—which are likely to have housekeeping functions (Wu et al. 2005)—lie on the chromosome (supplementary fig. S5, Supplementary Material online).

In order to predict whether *S. clavuligerus* could grow normally using standard carbon and nitrogen sources without the help of enzymes encoded on the 1.8-Mb replicon, we constructed a metabolic model based on our functional annotation, comparisons with known metabolic models from other organisms (Duarte et al. 2004; Borodina et al. 2005; Feist et al. 2007), and subsequent gap filling (table 2). This model successfully simulated growth of *S. clavuligerus* on various carbon and nitrogen sources. *In silico* knock-outs of all enzyme-coding genes specific to the 1.8-Mb replicon resulted in no growth defects when growth was simulated using glycerol as a carbon source. This suggests that no essential genes of primary metabolism lie on the megaplasmid.

Nonetheless, a few enzymes for primary metabolism were found to be encoded on the 1.8-Mb replicon. Among them are malate synthase and isocitrate lyase, enzymes of the glyoxylate pathway required for bacteria to grow on acetate. However, there is a malate synthase isoenzyme gene on the chromosome, which has been experimentally verified before (Chan and Sim 1998), and copies of alternative ac-

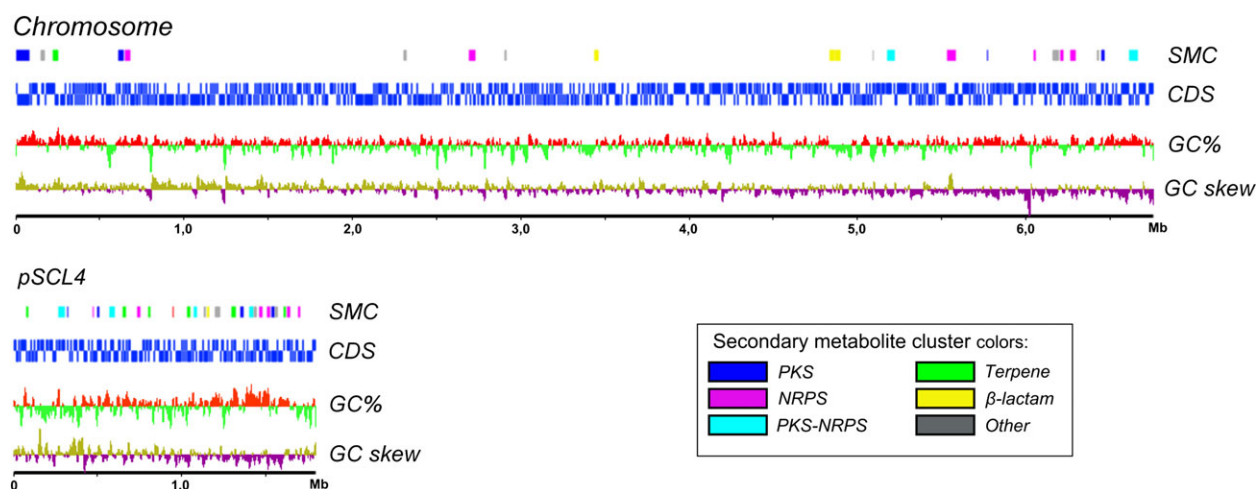
tinomycete glyoxylate pathway genes (*ccr/meaA* and *mcl1*; Akopiants et al. 2006) were also detected on the main chromosome, possibly making the plasmid-encoded copies redundant. Therefore, we predict that, even when growing on alternative carbon sources, the 1.8-Mb replicon is dispensable for primary metabolism. Experimental analysis by either curing of the whole replicon or constructing multiple targeted knock-outs should confirm this.

As the 1.8-Mb replicon does not seem to encode any functions essential to primary metabolism, the definition of Bentley and Parkhill supports the labeling of the 1.8-Mb replicon as a megaplasmid. As it is the fourth *S. clavuligerus* plasmid to be described, we name it pSCL4. It is the largest linear plasmid ever sequenced (Molbak et al. 2003). If the above definition is strictly applied, pSCL4 is actually the largest plasmid ever sequenced, as the only larger plasmid (the circular plasmid pGMI1000MP from *Ralstonia solanacearum* Salanoubat et al. 2002) should then be classified as a chromosome: it contains an important part of the metabolic core (an rRNA locus with two tRNA genes, a gene coding for the alpha subunit of DNA polymerase III, a gene for the protein elongation factor G, as well as several important enzymes of primary metabolism, including amino acid and cofactor biosynthesis), and thus is probably essential under many growth conditions (Salanoubat et al. 2002).

### The Megaplasmid Is Packed with Secondary Metabolite Gene Clusters

As the pSCL4 megaplasmid does not appear to harbor any housekeeping genes, we were interested in predicting the biological role of such a giant replicon. To our surprise, pSCL4 is packed with gene clusters putatively encoding the biosynthesis of secondary metabolites. No fewer than 25 such gene clusters, a number on the same order as observed in the chromosomes of other *Streptomyces* genomes, are dispersed throughout the plasmid (fig. 1). Together with the clusters identified on the chromosome, the total number of putative secondary metabolite gene clusters (SMC) identified in *S. clavuligerus* is 48, a number unprecedented in any bacterium (fig. 2A and B). These include 10 putative nonribosomal peptide synthetase (NRPS) gene clusters, eight putative PKS gene clusters, and six gene clusters putatively encoding NRPSs and PKSs or NRPS-PKS hybrids, as well as 12 clusters putatively encoding one or more terpene synthases or cyclases. An overview of all clusters encoding putative NRPSs and PKSs is shown in figure 2, and further details are given in supplementary table S2 (Supplementary Material online).

As expected, the three known antibiotic gene clusters of *S. clavuligerus* were identified in our genome assembly. Although the supercluster encoding the clavulanic acid and cephamycin C biosynthetic pathways (SMC10-11) and one of the clavam clusters (Tahlan et al. 2007) (SMC9) lie on the main chromosome, the alanylclavam



**FIG. 1.**—Overview of the putative secondary metabolite gene clusters in the *Streptomyces clavuligerus* genome. Chromosomal maps of the *S. clavuligerus* genome, showing both the chromosome and the megaplasmid pSCL4. SMC: secondary metabolite gene clusters, shown in different colors according to their type. CDS: coding sequences. GC%: GC percentage around the midline average of 72%. GC skew: GC skew, positive (brown) or negative (purple).

cluster (SMCp13) (Zelyas et al. 2008) is located on the megaplasmid. The latter cluster has been called a “paralogous cluster” of the clavulanic acid gene cluster (Tahlan, Park, Wong, et al. 2004) and was thought to have arisen through a recent gene duplication. Interestingly, a previously unknown third cluster of genes (SMCp25) containing a more distant homolog of the clavaminic synthase gene was also identified on pSCL4. This cluster might encode the production of yet another  $\beta$ -lactam antibiotic.

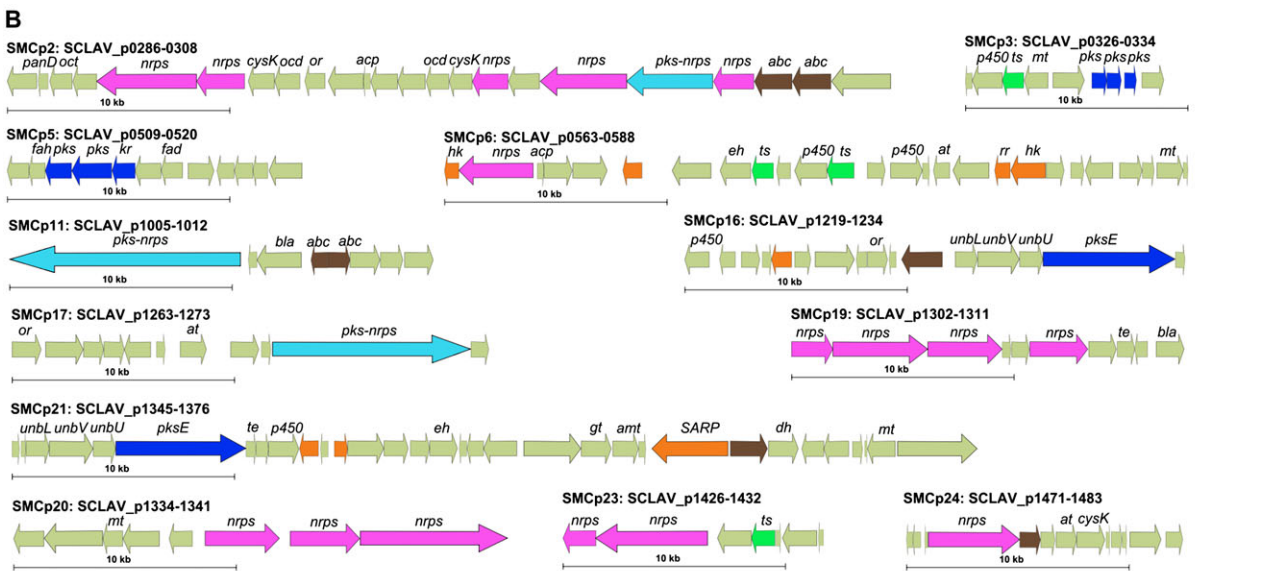
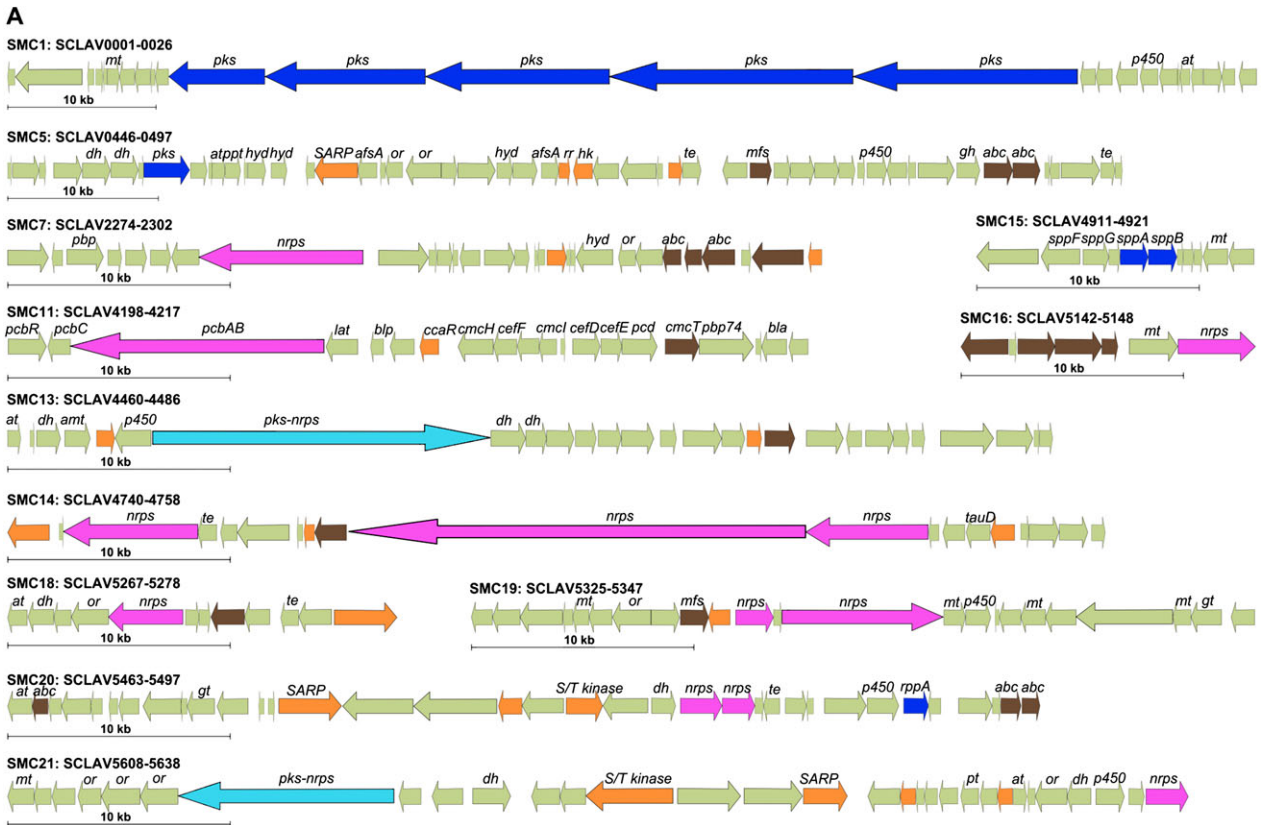
We could not predict with certainty which gene clusters are responsible for the production of holomycin and the tunicamycin-related antibiotic, given the paucity of knowledge on their biosynthesis. For the tacrolimus-like macrolide, we identified a potential gene cluster: the only macrolide PKS cluster detected, consisting of 11 modules. It is positioned near the end of one of the chromosomal arms (SMC1).

Interestingly, we also identified gene clusters potentially encoding the biosynthesis of known antibiotics on the megaplasmid. Clusters closely homologous to the staurosporine biosynthetic gene cluster of *Streptomyces* sp. TP-A0274 (Onaka et al. 2002) (SMCp14) and the moenomycin biosynthetic gene cluster from *Streptomyces ghanaensis* (Ostash et al. 2007, 2009) (SMCp18) were detected, as well as a close homolog of the IndC indigoidine blue pigment synthetase from *Streptomyces lavendulae* (Takahashi et al. 2007) (SMCp24). The fact that pSCL4 carries multiple biosynthetic gene clusters closely resembling those in relatively distantly related *Streptomyces* species, whereas the same clusters are absent in more closely related species (supplementary fig. S6, Supplementary Material online), supports the hypothesis that many secondary metabolism biosynthetic gene clusters in bacteria are acquired by horizontal gene transfer.

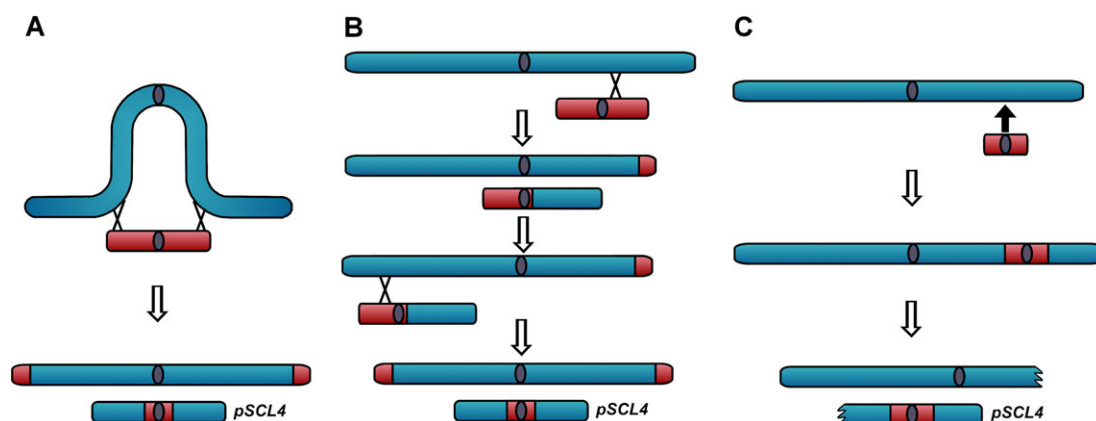
Putative gene clusters that might encode unknown products were also identified on both the megaplasmid and the chromosome. These include two enediyne-type PKS clusters on the megaplasmid (SMCp16 and SMCp21), which have the typical *unbLVU* genes encoding the biosynthetic machinery for the core enediyne structure. The clusters are quite similar to the biosynthetic gene clusters of C-1027 (Liu et al. 2002) and calicheamycin (Ahlert et al. 2002). Both *S. clavuligerus* clusters fall into the phylogenetic subgroup of nine-membered enediyne polyketides that was recently identified (Udwary et al. 2007) (supplementary fig. S7, Supplementary Material online), although we should note that recombination might have occurred between the two clusters (we observed some extremely similar regions in the multiple sequence alignments of these PKSs) which might distort the phylogenetic analysis.

The second enediyne-type PKS cluster lies very close to an NRPS gene cluster (SMCp20) with a very unusual feature that has not been observed before: the module containing the thioesterase is fused C-terminally to a major facilitator-type transporter. This may be used to export the end product from the cell immediately after assembly. Such a mechanism could have the advantage that the transport efficiency of the substance is much higher, and the cell can synthesize highly toxic compounds without these poisoning the producing cell.

Yet another remarkable cluster (SMC14) on the chromosome has one of its putative NRPS modules fused to a  $\beta$ -lactamase domain (PFAM number PF00144). No significant resemblance of this domain to any experimentally studied protein was detected, so we can only speculate about its function: it might, for example, act as a transpeptidase or bind to, but not hydrolyze, a  $\beta$ -lactam compound that is then attached to the peptide synthesized by



<i>abc</i> : ABC transporter	<i>hk</i> : histidine kinase	<i>pks</i> : polyketide synthase		NRPS		Terpene Synthase
<i>acp</i> : acyl-carrier protein	<i>hyd</i> : hydrolase	<i>pksE</i> : enediylne polyketide synthase		PKS		Regulatory Protein
<i>afsA</i> : butyrolactone biosynthesis gene	<i>kr</i> : ketoreductase	<i>ppt</i> : phosphopantetheinyl transferase		PKS-NRPS		Transport Protein
<i>amt</i> : aminotransferase	<i>mfs</i> : major facilitator class transporter	<i>pt</i> : phosphotransferase				
<i>at</i> : acyltransferase	<i>mt</i> : methyltransferase	<i>rppA</i> : chalcone synthase				
<i>bla</i> : beta-lactamase	<i>nrps</i> : non-ribosomal peptide synthetase	<i>rr</i> : response regulator				
<i>cefDE</i> : cephamycin biosynthesis genes	<i>ocd</i> : ornithine cyclodeaminase	<i>S/T kinase</i> : serine / threonine kinase				
<i>cmcIT</i> : cephamycin transporters	<i>oct</i> : ornithine carbamoyltransferase	<i>SARP</i> : Streptomyces antibiotic regulatory protein				
<i>cysK</i> : cysteine synthase	<i>or</i> : oxidoreductase	<i>sppABFG</i> : spore pigment gene homologs				
<i>dh</i> : dehydrogenase	<i>p450</i> : cytochrome p450	<i>ts</i> : thioesterase				
<i>eh</i> : epoxide hydrolase	<i>pbp</i> : penicillin-binding protein	<i>te</i> : terpene synthase				
<i>gh</i> : glycosyl hydrolase	<i>pcbABC</i> : penicillin biosynthesis homologs	<i>unbLVU</i> : enediylne biosynthesis genes				
<i>gt</i> : glucosyltransferase						



**FIG. 3.**—Three scenarios for the evolution of megaplasmid pSCL4. (A) A double crossover between the chromosomal core and the core of a smaller plasmid. (B) Two consecutive recombination events, the first yielding an asymmetrical plasmid and the second a symmetrical one. (C) Integration of a relatively small plasmid into a chromosomal arm and subsequent breaking off of the arm to yield an independent replicon.

the NRPS. Generally, these domains are of high importance in  $\beta$ -lactam producing organisms, as they can catalyze the opening and hydrolysis of the  $\beta$ -lactam ring of  $\beta$ -lactam antibiotics and can thus provide resistance to the strain's own antibiotics. Twenty-two proteins carrying a predicted  $\beta$ -lactamase domain were detected on the chromosome or plasmid (supplementary table S3, Supplementary Material online).

### Cross-regulation Takes Place between the Megaplasmid and the Chromosome

Expression of secondary metabolite genes can be regulated in a variety of ways. One important regulatory mechanism involves small signaling molecules called  $\gamma$ -butyrolactones (Takano 2006). In *S. clavuligerus*, a gene encoding a  $\gamma$ -butyrolactone receptor protein (Scar/Brp) was recently identified and shown to regulate clavulanic acid and cephamycin C production (Kim et al. 2004; Santamarta et al. 2005). We identified this gene as the megaplasmid-encoded SCLAV\_p0894, which appeared to be the only  $\gamma$ -butyrolactone receptor protein-encoding gene in the entire genome. However, four ScbA/AfsA-like putative butyrolactone biosynthetic proteins were detected: three on the chromosome (SCLAV\_0463, SCLAV\_0471, SCLAV\_2310) and one on the megaplasmid (SCLAV\_p0812). Based on phylogenetic analysis (supplementary fig. S8, Supplementary Material online), SCLAV\_2310 is the most likely candidate for being involved in butyrolactone biosynthesis. The fact that the only  $\gamma$ -butyrolactone receptor protein is encoded on the megaplasmid is remarkable, as all other characterized

$\gamma$ -butyrolactone receptors are chromosomally encoded. Moreover, it means that Brp transregulates several factors on the chromosome (at least the cephamycin C and clavulanic acid gene clusters). Because in a phylogenetic analysis, Brp was shown to cluster together with all known chromosomally encoded  $\gamma$ -butyrolactone receptors (Nishida et al. 2007), one is tempted to speculate that this gene was originally derived from the main chromosome and became plasmid-borne through recombination or transposition. Interestingly, two of the AfsA domain-containing butyrolactone biosynthetic proteins (SCLAV\_0463 and SCLAV\_0471) lie in a PKS gene cluster (SMC5), which consequently might be regulated by butyrolactone signaling.

Other regulatory genes were also identified. Fifty sigma factor genes were found, of which 43 lie on the chromosome and seven on pSCL4. Fifty-one paired two-component regulatory systems were identified, of which 44 are encoded on the chromosome and seven on the megaplasmid. Also, 21 orphan response regulator genes and eight orphan histidine kinase genes were identified. The two-component systems include at least five (chromosomally encoded) very close homologues of systems that have been observed to be involved in the regulation of antibiotic biosynthesis: AfsQ12 (SCLAV\_3812–3813) (Ishizuka et al. 1992), CutRS (SCLAV\_4778–4779) (Chang et al. 1996), RapA12 (SCLAV\_4312–4313) (Lu et al. 2007), AbrB12 (SCLAV\_1381–1382) and AbiA123 (SCLAV\_3595–3597). Additionally, 26 serine/threonine kinase genes were detected, of which only two are on the megaplasmid. Finally, 20 *Streptomyces* antibiotic regulatory protein (SARP) genes were detected, of which eight

**FIG. 2.**—Secondary metabolite biosynthetic gene clusters: a detailed picture. Putative secondary metabolite gene clusters (SMCs) encoding NRPSs and PKSs on (A) the chromosome and (B) the megaplasmid pSCL4 of *Streptomyces clavuligerus* ATCC 27064. Different core and accessory genes are depicted in different colors as indicated. Cluster borders are approximate and were estimated on the basis of functional annotations and putative operon structures. Details on the putative secondary metabolite gene clusters not shown in this figure (encoding  $\beta$ -lactams, terpene synthases, lantibiotics, etc.) can be found in supplementary table 2 (Supplementary Material online).



are on the megaplasmid. Seven SARPs are encoded in or very near secondary metabolism gene clusters.

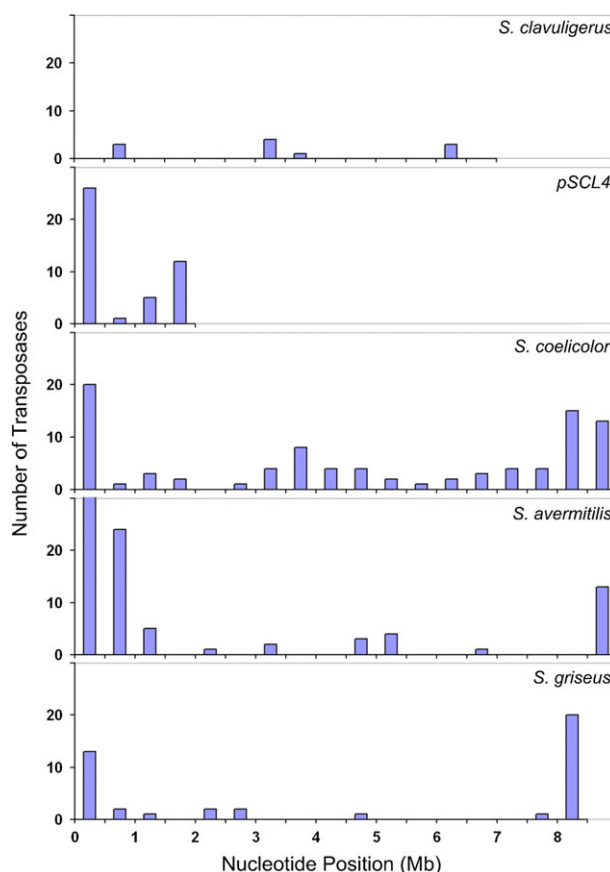
The vast array of potential regulatory mechanisms suggests that the production of secondary metabolites in *S. clavuligerus*—as in other streptomycetes—is under complex regulation, being highly tuned to the specific needs of the organism. It is striking that for all classes of regulators, the plasmid seems to encode considerably fewer compared with the chromosome, reinforcing the view that it is a highly specialized genetic element.

### Scenarios for the Evolution of pSCL4

The small size of the *S. clavuligerus* chromosome and the functional similarity between pSCL4 and *Streptomyces* chromosomal arms prompted the question how a megaplasmid-like pSCL4 could have originated. Interestingly, a megaplasmid or chromosome of very similar size (1.8 Mb) has been observed in *S. coelicolor* A3(2), after a single crossover between the 365-kb SCP1 plasmid and the chromosome (Yamasaki and Kinashi 2004). However, such a simple single crossover cannot have produced a plasmid configuration as seen in pSCL4, as it would have led to an asymmetrical replicon with the Ori located distantly from the centre (fig. 3B, upper part). The central position of the pSCL4 Ori suggests that multiple recombination events have taken place either simultaneously or consecutively. Furthermore, the small size of the main chromosome compared to those of other *Streptomyces* species suggests that the plasmid may well have been endowed with genetic regions from the chromosome arms in the recent past. Other observations that should also be taken into account are the clustering of replication genes (*parAB*, *tap/tpg*, plasmid primase/replicase genes) and other important genes such as the butyrolactone receptor gene *brp* close to the plasmid Ori, and the existence of a region close to the plasmid Ori (SCLAV\_p0926–0939) similar to chromosomal regions of many other actinomycetes but not found on the *S. clavuligerus* chromosome. These considerations suggest three possible scenarios for the origin of pSCL4 (fig. 3).

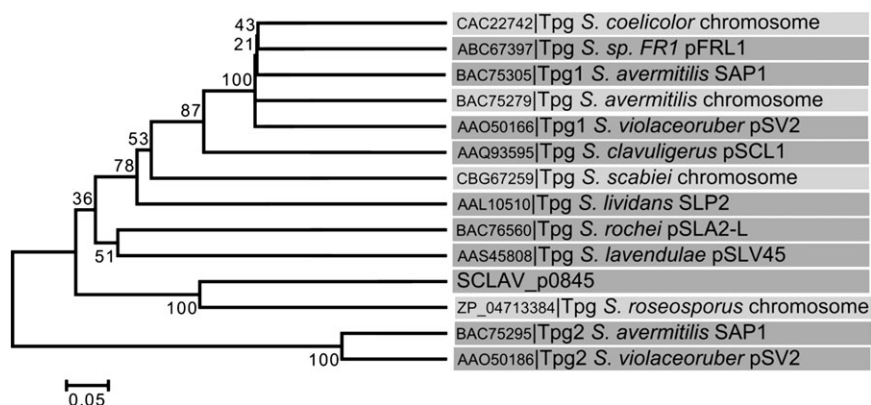
According to the first scenario, a relatively small plasmid underwent double crossing over with the chromosomal core, to yield a core chromosome with small plasmid arms and a plasmid core with large chromosomal arms. The second scenario is a variant of the first but postulates two asynchronous recombination events: first, a smaller plasmid recombined with one chromosomal arm, yielding an asymmetrical plasmid and then this situation was stabilized by another recombination with the other chromosomal arm. In the third scenario, a small integrative plasmid would have integrated into one of the chromosomal arms, after which it would have broken off (e.g., during conjugative transfer) and become a separate replicon.

The first two scenarios both seem parsimonious, as they can explain the symmetrical GC skew plot of pSCL4 without the need for extensive subsequent evolutionary fine-tuning.



**Fig. 4.**—Transposases in *Streptomyces* genomes. Transposases identified in the *Streptomyces clavuligerus* genome compared with the transposases identified in the genomes of *Streptomyces coelicolor*, *S. avermitilis*, and *S. griseus*. In contrast to the three other genomes, the chromosome of *S. clavuligerus* does not encode a large number of transposases at its ends. However, pSCL4 does encode many transposases. The fact that this feature—typical of chromosomal ends—is not present in the chromosome, yet present in the plasmid supports the hypothesis that the plasmid originated by acquiring the terminal regions of the chromosome.

Moreover, these scenarios are also favored over the third because the main chromosome is quite symmetrical despite its small size and has long arms on neither side. A final argument supporting the first and second scenarios is provided by the distribution of transposons throughout the *S. clavuligerus* genome. Normally, many transposons are present at the far termini of the main chromosomes of streptomycetes (Chen et al. 2002), yet they are absent from the termini of the *S. clavuligerus* chromosome. Instead, many transposons are present near the termini of pSCL4 (44 in total, judged by the number of transposases found; fig. 4). Intriguingly, phylogenetic analysis showed that most of these (25 of the 44) are close homologues of transposases encoded in *Streptomyces* chromosomal termini (<1-Mb distance to the ends). This suggests that the plasmid may indeed contain the former chromosomal termini of *S. clavuligerus*. Sequencing of



**Fig. 5.**—Phylogeny of Tpg proteins shows the dynamic evolution of linear plasmids and chromosomes in *Streptomyces*. Phylogenetic analysis of Tpg proteins shows that there is no correlation between the phylogenetic clustering and presence of a *tpg* gene on either the plasmid (in dark gray) or the chromosome (in light gray). This suggests that *tap* and *tpg* genes are often transferred from chromosomes to plasmids and vice versa, for example, through recombination. Phylogenies were calculated using the neighbor-joining (NJ) method in MEGA 4 (Tamura et al. 2007). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) is shown next to the branches.

the telomeres at the very ends of chromosome and megaplasmid might help to confirm this.

### Heavy Traffic between Plasmids and Chromosomal Arms in Actinomycete Bacteria

Our findings underline the dynamic evolution of actinomycete genomes. They seem to acquire or lose plasmids regularly, and the gene content of their chromosomal arms fluctuates dramatically (Choulet et al. 2006). This is clearly visible in the phylogeny of the Tpg telomere proteins (fig. 5), which shows almost complete intermixing of chromosome- and plasmid-encoded genes, signifying extensive transfer of *tpg* genes via recombination or integration. Our identification of a very large plasmid densely packed with genetic material that is typically encoded on chromosomal arms suggests that linear plasmids play a major role in this genomic flux. Because of their ability to reach large sizes and the possibility of plasmid–chromosome recombination or plasmid integration (Kinashi et al. 1992) such as described above, plasmids seem to be more important than previously thought in determining the large variability of actinomycete genomes. This hypothesis fits neatly with the observation that many actinomycete species-specific secondary metabolite clusters are located within genomic islands (Penn et al. 2009), which are often mobilized on plasmids (Dobrindt et al. 2004). The fact that the *S. clavuligerus* ATCC 27064 strain appears to have a very different genome composition in terms of plasmids than the NRRL 3585 strain (Netolitzky et al. 1995)—which originates from the same source (Higgins and Kastner 1971)—epitomizes the highly dynamic nature of actinomycete plasmids. Earlier observations of rapid recombination events between chromosomes and plasmids (Gravius et al. 1994; McLeod et al. 2006) in actinomycetes with linear chromosomes had already suggested such a conclusion.

### Conclusions

The sequencing of the genome of *S. clavuligerus* reveals the potential for producing a vast array of interesting novel secondary metabolites, some of them encoded by unusual gene configurations that have never been observed before. The fact that a major part of the secondary metabolite biosynthesis gene clusters is localized on a specialized giant linear plasmid (pSCL4), which we predict not to contain any genes essential for primary metabolism, indicates that plasmids could more often encode secondary metabolites than previously thought (Kinashi 2008). The small size of the *S. clavuligerus* chromosome and the absence of hallmarks typical for *Streptomyces* chromosomal termini, together with the remarkable presence of such terminal hallmarks on the plasmid, suggest that the megaplasmid may have originated by a double recombination of a smaller plasmid with the chromosome. There is evidence that cross-regulation of chromosomal genes by at least one plasmid-encoded regulator still occurs. Intriguingly, the deduced flow of genetic material between different replicons may not be an exception: the phylogeny of telomere replication proteins from plasmids as well as chromosomes suggests that fluxes of genetic material regularly take place between streptomycete chromosomes and plasmids. Indeed, the mobilization of secondary metabolite gene clusters onto large linear plasmids such as pSCL4, which could be vectors for horizontal gene transfer (Ravel et al. 1998, 2000), indicates that constant and extensive “open source” evolution (Frost et al. 2005) of secondary metabolite-encoding DNA regions in actinomycetes could be responsible for the large differences in secondary metabolite repertoires between different species. Approaches that “awaken” uncharacterized gene clusters will undoubtedly be pivotal to uncover the functionalities of the secondary metabolites they encode (Scherlach and Hertweck 2009). Furthermore, if the plasmid could be cured from the strain (Hsu and Chen 2010), the small

chromosome of *S. clavuligerus* may be a very interesting vehicle for synthetic biology (Keasling 2008), serving as a starting point for the construction of a “minimal streptomycete.”

## Supplementary Material

Supplementary figures S1–S8 and tables 1 and 2 are available at *Genome Biology and Evolution* online ([http://www.oxfordjournals.org/our\\_journals/gbe/](http://www.oxfordjournals.org/our_journals/gbe/)).

## Acknowledgments

This work was supported by the Dutch Technology Foundation (STW), which is the applied science division of the Netherlands Organisation for Scientific Research (NWO), and the Technology Program of the Ministry of Economic Affairs (STW 10463). R.B. is supported by an NWO-Vidi fellowship and E.T. by a Rosalind Franklin Fellowship, University of Groningen. We thank Christian Kuijlaars for the help in the initial phase of metabolic model construction. We thank David Hopwood and the anonymous reviewers for constructive comments and suggestions.

## Literature Cited

- Ahlert J, et al. 2002. The calicheamicin gene cluster and its iterative type I enediyne PKS. *Science* 297:1173–1176.
- Akopiants K, Florova G, Li C, Reynolds KA. 2006. Multiple pathways for acetate assimilation in *Streptomyces cinnamonensis*. *J Ind Microbiol Biotechnol.* 33:141–150.
- Alexander DC, Jensen SE. 1998. Investigation of the *Streptomyces clavuligerus* cephamycin C gene cluster and its regulation by the CcaR protein. *J Bacteriol.* 180:4068–4079.
- Bao K, Cohen SN. 2001. Terminal proteins essential for the replication of linear plasmids and chromosomes in *Streptomyces*. *Genes Dev.* 15:1518–1527.
- Bao K, Cohen SN. 2003. Recruitment of terminal protein to the ends of *Streptomyces* linear plasmids and chromosomes by a novel telomere-binding protein essential for linear DNA replication. *Genes Dev.* 17:774–785.
- Becker SA, et al. 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc.* 2:727–738.
- Bendich AJ, Drlica K. 2000. Prokaryotic and eukaryotic chromosomes: what's the difference? *Bioessays.* 22:481–486.
- Bentley SD, et al. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417:141–147.
- Bentley SD, et al. 2004. SCP1, a 356,023 bp linear plasmid adapted to the ecology and developmental biology of its host, *Streptomyces coelicolor* A3(2). *Mol Microbiol.* 51:1615–1628.
- Bentley SD, Parkhill J. 2004. Comparative genomic structure of prokaryotes. *Annu Rev Genet.* 38:771–792.
- Berdy J. 1995. Are actinomycetes exhausted as a source of secondary metabolites? *Biotechnologia.* 7–8:13–34.
- Bode HB, Muller R. 2005. The impact of bacterial genomics on natural product research. *Angew Chem Int Ed Engl.* 44:6828–6846.
- Borodina I, Krabben P, Nielsen J. 2005. Genome-scale analysis of *Streptomyces coelicolor* A3(2) metabolism. *Genome Res.* 15:820–829.
- Brogden RN, et al. 1981. Amoxicillin/clavulanic acid: a review of its antibacterial activity, pharmacokinetics and therapeutic use. *Drugs.* 22:337–362.
- Brolle DF, Pape H, Hopwood DA, Kieser T. 1993. Analysis of the transfer region of the *Streptomyces* plasmid SCP2. *Mol Microbiol.* 10:157–170.
- Challis GL, Hopwood DA. 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. (Suppl 100)* 2: 14555–14561.
- Chan M, Sim TS. 1998. Malate synthase from *Streptomyces clavuligerus* NRRL3585: cloning, molecular characterization and its control by acetate. *Microbiology.* 144:3229–3237.
- Chang HM, Chen M, Shieh YT, Bibb MJ, Chen CW. 1996. The *cutRS* signal transduction system of *Streptomyces lividans* represses the biosynthesis of the polyketide antibiotic actinorhodin. *Mol Microbiol.* 21:1075–1085.
- Chater KF, Kinashi H. 2007. *Streptomyces* linear plasmids: their discovery, functions, interactions with other replicons, and evolutionary significance. *Microbiol Monogr.* 7:1.
- Chen CW. 2007. *Streptomyces* linear plasmids: replication and telomeres. *Microbiol Monogr.* 7:33–61.
- Chen CW, Huang CH, Lee HH, Tsai HH, Kirby R. 2002. Once the circle has been broken: dynamics and evolution of *Streptomyces* chromosomes. *Trends Genet.* 18:522–529.
- Chen CW, et al. 1994. The linear chromosomes of *Streptomyces*: structure and dynamics. *Actinomycetologica.* 8:103–112.
- Choulet F, et al. 2006. Evolution of the terminal regions of the *Streptomyces* linear chromosome. *Mol Biol Evol.* 23:2361–2369.
- Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14:1394–1403.
- Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679.
- Dobrindt U, Hochhut B, Hentschel U, Hacker J. 2004. Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol.* 2:414–424.
- Duarte NC, Herrgard MJ, Palsson BO. 2004. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.* 14: 1298–1309.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
- Evans RH, et al. 1983. Ro 22-5417, a new clavam antibiotic from *Streptomyces clavuligerus*. II. Fermentation, isolation and structure. *J Antibiot (Tokyo).* 36:213–216.
- Feist AM, et al. 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol.* 3:121.
- Frost LS, Leplae R, Summers AO, Toussaint A. 2005. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol.* 3:722–732.
- Gravius B, et al. 1994. The 387 kb linear plasmid pPZG101 of *Streptomyces rimosus* and its interactions with the chromosome. *Microbiology.* 140:2271–2277.
- Grigoriev A. 1998. Analyzing genomes with cumulative skew diagrams. *Nucleic Acids Res.* 26:2286–2290.
- Gullo VP, McAlpine J, Lam KS, Baker D, Petersen F. 2006. Drug discovery from natural products. *J Ind Microbiol Biotechnol.* 33: 523–531.

- Higgins CE, Kastner RE. 1971. *Streptomyces clavuligerus* sp. nov., a  $\beta$ -lactam antibiotic producer. *Int J Syst Bacteriol.* 21:326–331.
- Hinnebusch J, Tilly K. 1993. Linear plasmids and chromosomes in bacteria. *Mol Microbiol.* 10:917–922.
- Hopwood DA. 2006. Soil to genomics: the *Streptomyces* chromosome. *Annu Rev Genet.* 40:1–23.
- Hsu CC, Chen CW. 2010. Linear plasmid SLP2 is maintained by partitioning, intrahyphal spread, and conjugal transfer in *Streptomyces*. *J Bacteriol.* 192:307–315.
- Huang CH, et al. 2007. The telomere system of the *Streptomyces* linear plasmid SCP1 represents a novel class. *Mol Microbiol.* 63:1710–1718.
- Ikeda H, et al. 2003. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol.* 21:526–531.
- Ishizuka H, Horinouchi S, Kieser HM, Hopwood DA, Beppu T. 1992. A putative two-component regulatory system involved in secondary metabolism in *Streptomyces* spp. *J Bacteriol.* 174:7585–7594.
- Keasling JD. 2008. Synthetic biology for synthetic chemistry. *ACS Chem Biol.* 3:64–76.
- Kenig M, Reading C. 1979. Holomycin and an antibiotic (MM 19290) related to tunicamycin, metabolites of *Streptomyces clavuligerus*. *J Antibiot (Tokyo).* 32:549–554.
- Kim HS, et al. 2004. Cloning and characterization of a gene encoding the  $\gamma$ -butyrolactone autoregulator receptor from *Streptomyces clavuligerus*. *Arch Microbiol.* 182:44–50.
- Kim HS, Park YI. 2008. Isolation and identification of a novel microorganism producing the immunosuppressant tacrolimus. *J Biosci Bioeng.* 105:418–421.
- Kinashi H. 2008. Antibiotic production, linear plasmids and linear chromosomes in *Streptomyces*. *Actinomycetologica.* 22:20–29.
- Kinashi H, Shimaji-Murayama M, Hanafusa T. 1992. Integration of SCP1, a giant linear plasmid, into the *Streptomyces coelicolor* chromosome. *Gene* 115:35–41.
- Kirby R. 1978. An unstable genetic element affecting the production of the antibiotic holomycin by *Streptomyces clavuligerus*. *FEMS Microbiol Lett.* 3:283–286.
- Kirby R, Hopwood DA. 1977. Genetic determination of methylenomycin synthesis by the SCP1 plasmid of *Streptomyces coelicolor* A3(2). *J Gen Microbiol.* 98:239–252.
- Koski LB, Gray MW, Lang BF, Burger G. 2005. AutoFACT: an automatic functional annotation and classification tool. *BMC Bioinformatics.* 6:151.
- Krumsiek J, Arnold R, Rattei T. 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 23:1026–1028.
- Latreille P, et al. 2007. Optical mapping as a routine tool for bacterial genome sequence finishing. *BMC Genomics.* 8:321.
- Levy S, et al. 2007. The diploid genome sequence of an individual human. *PLoS Biol.* 5:e254.
- Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13:2178–2189.
- Liras P, Gomez-Escribano JP, Santamarta I. 2008. Regulatory mechanisms controlling antibiotic production in *Streptomyces clavuligerus*. *J Ind Microbiol Biotechnol.* 35:667–676.
- Liu W, Christenson SD, Standage S, Shen B. 2002. Biosynthesis of the enediyne antitumor antibiotic C-1027. *Science* 297:1170–1173.
- Lu Y, et al. 2007. Characterization of a novel two-component regulatory system involved in the regulation of both actinorhodin and a type I polyketide in *Streptomyces coelicolor*. *Appl Microbiol Biotechnol.* 77:625–635.
- McLeod MP, et al. 2006. The complete genome of *Rhodococcus sp.* RHA1 provides insights into a catabolic powerhouse. *Proc Natl Acad Sci U S A.* 103:15582–15587.
- Mochizuki S, et al. 2003. The large linear plasmid pSLA2-L of *Streptomyces rochei* has an unusually condensed gene organization for secondary metabolism. *Mol Microbiol.* 48:1501–1510.
- Molbak L, et al. 2003. The plasmid genome database. *Microbiology* 149:3043–3045.
- Netolitzky DJ, Wu X, Jensen SE, Roy KL. 1995. Giant linear plasmids of beta-lactam antibiotic producing *Streptomyces*. *FEMS Microbiol Lett.* 131:27–34.
- Newman DJ, Cragg GM. 2007. Natural products as sources of new drugs over the last 25 years. *J Nat Prod.* 70:461–477.
- Nishida H, Ohnishi Y, Beppu T, Horinouchi S. 2007. Evolution of  $\gamma$ -butyrolactone synthases and receptors in *Streptomyces*. *Environ Microbiol.* 9:1986.
- Ohnishi Y, et al. 2008. Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J Bacteriol.* 190:4050–4060.
- Oliva B, O'Neill A, Wilson JM, O'Hanlon PJ, Chopra I. 2001. Antimicrobial properties and mode of action of the pyrrothine holomycin. *Antimicrob Agents Chemother.* 45:532–539.
- Onaka H, Taniguchi S, Igarashi Y, Furumai T. 2002. Cloning of the staurosporine biosynthetic gene cluster from *Streptomyces sp.* TP-A0274 and its heterologous expression in *Streptomyces lividans*. *J Antibiot (Tokyo).* 55:1063–1071.
- Ostash B, et al. 2009. Complete characterization of the seventeen step moenomycin biosynthetic pathway. *Biochemistry* 48:8830–8841.
- Ostash B, Saghatelian A, Walker S. 2007. A streamlined metabolic pathway for the biosynthesis of moenomycin A. *Chem Biol.* 14:257–267.
- Penn K, et al. 2009. Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. *ISME J.* 3:1193–1203.
- Ravel J, Schrepff H, Hill RT. 1998. Mercury resistance is encoded by transferable giant linear plasmids in two Chesapeake bay *Streptomyces* strains. *Appl Environ Microbiol.* 64:3383–3388.
- Ravel J, Wellington EM, Hill RT. 2000. Interspecific transfer of *Streptomyces* giant linear plasmids in sterile amended soil microcosms. *Appl Environ Microbiol.* 66:529–534.
- Redenbach M, Bibb M, Gust B, Seitz B, Spychaj A. 1999. The linear plasmid SCP1 of *Streptomyces coelicolor* A3(2) possesses a centrally located replication origin and shows significant homology to the transposon Tn4811. *Plasmid.* 42:174–185.
- Rutherford K, et al. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16:944–945.
- Salanoubat M, et al. 2002. Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature* 415:497–502.
- Santamarta I, Perez-Redondo R, Lorenzana LM, Martin JF, Liras P. 2005. Different proteins bind to the butyrolactone receptor protein ARE sequence located upstream of the regulatory *ccaR* gene of *Streptomyces clavuligerus*. *Mol Microbiol.* 56:824–835.
- Saudagar PS, Survase SA, Singhal RS. 2008. Clavulanic acid: a review. *Biotechnol Adv.* 26:335–351.
- Scherlach K, Hertweck C. 2009. Triggering cryptic natural product biosynthesis in microorganisms. *Org Biomol Chem.* 7:1753–1760.
- Tahlan K, Anders C, Jensen SE. 2004. The paralogous pairs of genes involved in clavulanic acid and clavam metabolite biosynthesis are differently regulated in *Streptomyces clavuligerus*. *J Bacteriol.* 186:6286–6297.

- Tahlan K, et al. 2007. 5S clavam biosynthetic genes are located in both the clavam and paralog gene clusters in *Streptomyces clavuligerus*. *Chem Biol.* 14:131–142.
- Tahlan K, Park HU, Jensen SE. 2004. Three unlinked gene clusters are involved in clavam metabolite biosynthesis in *Streptomyces clavuligerus*. *Can J Microbiol.* 50:803–810.
- Tahlan K, Park HU, Wong A, Beatty PH, Jensen SE. 2004. Two sets of paralogous genes encode the enzymes involved in the early stages of clavulanic acid and clavam metabolite biosynthesis in *Streptomyces clavuligerus*. *Antimicrob Agents Chemother.* 48: 930–939.
- Takahashi H, et al. 2007. Cloning and characterization of a *Streptomyces* single module type non-ribosomal peptide synthetase catalyzing a blue pigment synthesis. *J Biol Chem.* 282: 9073–9081.
- Takano E. 2006.  $\gamma$ -Butyrolactones: *Streptomyces* signalling molecules regulating antibiotic production and differentiation. *Curr Opin Microbiol.* 9:287–294.
- Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 56:564–577.
- Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol.* 24:1596–1599.
- Udwardy DW, et al. 2007. Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc Natl Acad Sci U S A.* 104:10376–10381.
- Ussery DW, Wassenaar TM, Borini S. 2008. Microbial communities: core and pan-genomics. In: *Computing for comparative microbial genomics*. London: Springer. p. 213–228.
- Venter JC, et al. 2001. The sequence of the human genome. *Science* 291:1304–1351.
- Volff JN, Altenbuchner J. 1998. Genetic instability of the *Streptomyces* chromosome. *Mol Microbiol.* 27:239–246.
- Watve MG, Tickoo R, Jog MM, Bhole BD. 2001. How many antibiotics are produced by the genus *Streptomyces*? *Arch Microbiol.* 176: 386–390.
- Widenbrant EM, Tsai HH, Chen CW, Kao CM. 2007. *Streptomyces coelicolor* undergoes spontaneous chromosomal end replacement. *J Bacteriol.* 189:9117–9121.
- Wu G, Culley DE, Zhang W. 2005. Predicted highly expressed genes in the genomes of *Streptomyces coelicolor* and *Streptomyces avermitilis* and the implications for their metabolism. *Microbiology.* 151:2175–2187.
- Wu W, Leblanc SK, Piktel J, Jensen SE, Roy KL. 2006. Prediction and functional analysis of the replication origin of the linear plasmid pSCL2 in *Streptomyces clavuligerus*. *Can J Microbiol.* 52:293–300.
- Wu X, Roy KL. 1993. Complete nucleotide sequence of a linear plasmid from *Streptomyces clavuligerus* and characterization of its RNA transcripts. *J Bacteriol.* 175:37–52.
- Yamasaki M, Kinashi H. 2004. Two chimeric chromosomes of *Streptomyces coelicolor* A3(2) generated by single crossover of the wild-type chromosome and linear plasmid SCP1. *J Bacteriol.* 186: 6553–6559.
- Zelyas NJ, Cai H, Kwong T, Jensen SE. 2008. Alanylclavam biosynthetic genes are clustered together with one group of clavulanic acid biosynthetic genes in *Streptomyces clavuligerus*. *J Bacteriol.* 190:7957.

**Associate editor:** Eugene Koonin