The Genome Sequence of *Streptomyces lividans* 66 Reveals a Novel tRNA-Dependent Peptide Biosynthetic System within a Metal-Related Genomic Island

Pablo Cruz-Morales[1], Erik Vijgenboom[2], Fernanda Iruegas-Bocardo[1], Geneviève Girard[2], Luis Alfonso Yáñez-Guerra[1], Hilda E. Ramos-Aboites[1], Jean-Luc Pernodet[3], Jozef Anné[4], Gilles P. van Wezel[2], and Francisco Barona-Gómez[1,*]

[1]Evolution of Metabolic Diversity Laboratory, Laboratorio Nacional de Genómica para la Biodiversidad (Langebio), Cinvestav-IPN, Irapuato, México

[2]Leiden Institute of Chemistry, Molecular Biotechnology, Leiden, The Netherlands

[3]Institut de Génétique et Microbiologie, Université Paris-Sud, Orsay, France

[4]Laboratory of Bacteriology, Rega Institute for Medical Research, Katholieke Universiteit Leuven, Leuven, Belgium

*Corresponding author: E-mail: fbarona@langebio.cinvestav.mx.

Accepted: May 18, 2013

Data deposition: The project has been deposited at DDBJ/EMBL/GenBank BioProject id PRJNA168962 under the accession APVM00000000.

Abstract

The complete genome sequence of the original isolate of the model actinomycete *Streptomyces lividans* 66, also referred to as 1326, was deciphered after a combination of next-generation sequencing platforms and a hybrid assembly pipeline. Comparative analysis of the genomes of *S. lividans* 66 and closely related strains, including *S. coelicolor* M145 and *S. lividans* TK24, was used to identify strain-specific genes. The genetic diversity identified included a large genomic island with a mosaic structure, present in *S. lividans* 66 but not in the strain TK24. Sequence analyses showed that this genomic island has an anomalous (G + C) content, suggesting recent acquisition and that it is rich in metal-related genes. Sequences previously linked to a mobile conjugative element, termed plasmid SLP3 and defined here as a 94 kb region, could also be identified within this locus. Transcriptional analysis of the response of *S. lividans* 66 to copper was used to corroborate a role of this large genomic island, including two SLP3-borne "cryptic" peptide biosynthetic gene clusters, in metal homeostasis. Notably, one of these predicted biosynthetic systems includes an unprecedented nonribosomal peptide synthetase—tRNA-dependent transferase biosynthetic hybrid organization. This observation implies the recruitment of members of the leucyl/phenylalanyl-tRNA-protein transferase family to catalyze peptide bond formation within the biosynthesis of natural products. Thus, the genome sequence of *S. lividans* 66 not only explains long-standing genetic and phenotypic differences but also opens the door for further in-depth comparative genomic analyses of model *Streptomyces* strains, as well as for the discovery of novel natural products following genome-mining approaches.

Key words: bacterial next-generation genome sequencing, *Streptomyces* comparative genomics, copper homeostasis, L/F tRNA transferase, peptide biosynthesis.

Introduction

Streptomyces are Gram-positive soil-dwelling bacteria. Most members of the genus have saprophytic, free-living lifestyles, and contend for resources with several other organisms in oligotrophic environments (Hodgson 2000). Streptomycetes are mycelial organisms that grow as hyphae, which frequently branch to form an intricate vegetative mycelium. At the time of nutrient depletion, the vegetative mycelium differentiates to form reproductive structures called aerial hyphae, which are eventually converted into chains of spores (Schwedock et al. 1997). Typically, the production of secondary metabolites or natural products (NPs) correlates temporally to this phase of the life cycle (van Wezel and MacDowell 2011). Many of these NPs have relevant biomolecular activities, including most of the antibiotics used in medicine. In recent years, however, the discovery of antibiotics with novel classes of chemical structures from bacterial sources has been scarce. Fortunately, investigation of the chemical proficiency of streptomycetes has been revitalized by the advent of bacterial genome

sequencing and the development of novel sequence-based NP-discovery approaches (Challis 2008).

Streptomyces lividans 66 and *S. coelicolor* A3(2) are closely related species belonging to the *S. violaceouruber* sub-clade. Both species have been adopted as model organisms of the genus for almost five decades, and as a consequence, several derived strains from both organisms have been obtained. *S. coelicolor* was originally selected mainly for its copious production of blue (actinorhodin) and red (prodiginines) pigments, which were used as phenotypic markers for early genetic studies in the biosynthesis of NPs (Hopwood 1999). *Streptomyces lividans* was selected mainly for accepting methylated DNA, making this organism an ideal cloning host. This feature, together with its low endogenous protease activity, has granted *S. lividans* a prominent role as a host for expression of heterologous proteins—including complete NPs biosynthetic pathways—within both industrial and scientific settings (Anné et al. 2012).

Besides the distinctive traits that made *S. coelicolor* A3(2) and *S. lividans* 66 model organisms, other differences amongst these strains have been reported. *Streptomyces lividans* produces the same pigments as *S. coelicolor*, but only under certain conditions, and in contrast with *S. coelicolor*, it has been shown to lack agarase activity (Kieser et al. 2000) and to thiolate its DNA (Zhou et al. 2005). Moreover, *S. lividans* has been shown to be tolerant to high concentrations of mercury (Nakahara et al. 1985), whereas copper is required for its development (Keijser et al. 2000; Worrall and Vijgenboom 2010). Interestingly, these features are unique to the parental strain 66, equivalent to strain 1326, but are absent from or less prominent in the plasmid-less strain TK24, which was isolated after UV mutagenesis and protoplast regeneration (Hopwood et al. 1983).

During the early genetic characterization of *S. lividans*, the presence of two plasmids, termed SLP2 and SLP3, was inferred (Hopwood et al. 1983). However, only SLP2, a 50 kb linear molecule, has been physically isolated and characterized using pulsed field gel electrophoresis (Chen et al. 1993) and DNA sequencing (Huang et al. 2003). In contrast, SLP3 has remained elusive to date, despite clear early genetic evidence of its existence. Amongst this evidence, the *mer* genes responsible for resistance to mercury in *S. lividans* have been unambiguously linked to SLP3, confirming the mobile and conjugative nature of this element (Sedlmeier and Altenbuchner 1992). Moreover, an amplifiable sequence termed AUD2 has been linked to the *mer* genes and thus to SLP3 plasmid (Eichenseer and Altenbuchner 1994).

In addition to the report of the genome sequence of *S. coelicolor* M145, a strain obtained from *S. coelicolor* A3(2) that lacks its natural plasmids (Bentley et al. 2002), a draft genome sequence of *S. lividans* TK24 has been released and used for metabolic flux analysis (D'Huys et al. 2012). Genomic hybridization experiments using an *S. coelicolor* M145 microarray and DNA from different *S. lividans* strains, showing the absence of several *S. coelicolor* genes from *S. lividans*, have also been reported (Jayapal et al. 2007; Lewis et al. 2010). Most of these genes are clustered within genomic islands, although it was also concluded that genetic variation among these strains might actually span throughout the entire chromosome (Lewis et al. 2010). A good quality *S. lividans* genome sequence from a parental strain should provide further insights into the evolutionary processes leading to these intriguing genotypic and phenotypic differences.

Here, by means of a hybrid sequencing and assembly strategy, we report the complete genome sequence of *S. lividans* 66, which we use for comparative genome analyses against *S. lividans* TK24 and *S. coelicolor* M145. A large genome island specific to strain 66, which was confirmed after transcriptional analysis to be involved in metal metabolism, was identified. Furthermore, within this genomic island with an anomalous (G + C) content, the entire sequence of "plasmid" SLP3 could be identified. Remarkably, the SLP3 locus was found to encode two "cryptic" peptide biosynthetic gene clusters, including an unprecedented nonribosomal peptide synthetase—tRNA-dependent transferase (NRPS-tRNA) hybrid system. After sequence-based predictions of putative chemical structures, we postulate that this novel system directs the synthesis of a peptide involved in metal homeostasis.

Materials and Methods

Genome Sequencing and Assembly

Genomic DNA from *S. lividans* 66, obtained from the John Innes Centre culture collection (equivalent to strain 1326) was purified using standard protocols (Kieser et al. 2000) and sequenced using next generation sequencing platforms. Pyrosequencing (454, Life Sciences) was performed at the Langebio core sequencing facility (Irapuato, Mexico); ligation-based approaches (Illumina) were performed at Baseclear (Leiden, The Netherlands) and single molecule real-time sequencing (Pacific Biosystems RS) at Service XS (Leiden, The Netherlands). An initial set of contigs was obtained from Illumina and 454 reads using Celera Assembler 7.0 for Hybrid de novo assembly (Myers et al. 2000). The resulting contigs were sorted using *S. coelicolor* M145 and *S. lividans* TK24 genomes as references with R2CAT (Husemann and Stoye 2010). This assembly was improved with IMAGE 2 (Tsai et al. 2010) by iterative mapping and extension using paired-end Illumina reads. The second assembly and Illumina mate pair reads were used to obtain genome scaffolds with SSPACE (Boetzer et al. 2011). A second round of sorting with r2cat and the genome of *S. lividans* TK24 as reference yielded a single scaffold. PacBio RS (long) reads were corrected with the PacBiotoCA pipeline (Koren et al. 2012) using Illumina reads. These corrected reads were used to manually fill gaps in the main chromosomal scaffold. Oligonucleotides used for closing gaps by polymerase

chain reaction (PCR) are included in supplementary table S1, Supplementary Material online.

Genome Annotation, Mining, and Comparative Analysis

The *S. lividans* 66 genome sequence was annotated automatically with RAST (Aziz et al. 2008) inspected and improved manually. Comparative genomic analysis between the chromosomes of *S. coelicolor*, and *S. lividans* TK24 and 66, was performed using sequence comparisons with Blastn (Altschul et al. 1990), and visualized in Artemis Comparative Tools (Carver et al. 2005). A score cutoff of 200 was used to visually determine the presence or absence of genes trough the genomes. Jspecies (Richter and Roselló-Móra 2009) was used to obtain the blast-calculated average nucleotide identity (ANIb) between genomes. Annotation of NPs biosynthetic gene clusters was carefully done after BlastP searches and literature-based inferences. The domain prediction and analysis of SLI0883 was done in the PKS/NRPS Analysis Website (Bachmann and Ravel 2009).

Transcriptional Analysis

RNA-seq libraries from *S. lividans* 66 produced by Dwarakanath et al. (2012) were used for transcriptional analysis. We used the annotated genome of *S. lividans* 66 as reference after manually removing ribosomal genes. Differentially expressed genes were identified with CLC genomics work bench (CLC-Bio, Denmark). A cutoff of FDR ≤ 0.05 was used to filter the genes with significant changes. To obtain the coverage-plots, the RNAseq reads were mapped against the 66 genome using BWA (Li and Durbin 2009), and the alignments were processed with SAMtools (Li et al. 2009) for graphic representation in Artemis (Carver et al. 2005).

Phylogenetic Analysis of L/F tRNA Transferases

Homologous enzyme sequences were retrieved using BlastP (Altschul et al. 1990) from nonredundant public databases. The amino acid sequences were aligned with Muscle (Edgar 2004) and the phylogenetic relationships of selected proteins were reconstructed using MrBayes (Huelsenbeck and Ronquist 2001). Graphical representations of the phylogenetic trees were obtained with FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

Results and Discussion

Sequencing, Assembly and Annotation of the Genome of *S. lividans* 66

The genomic sequence of *S. lividans* 66 was deciphered after a combination of second- and third-generation sequencing-platforms, including pyrosequencing and ligation-based approaches, and single molecule real-time sequencing, respectively. Hybrid de novo assembly of the sequences obtained after pyrosequencing shotgun (50 Mbp) and ligated paired-end sequencing runs (996 Mbp), yielded 1,471 contigs that were sorted using the chromosomal sequence of the plasmid-less strains *S. coelicolor* M145 and *S. lividans* TK24 as references. The initial number of contigs was reduced to 999 contigs by iterative mapping and extension using the paired reads that were left out from the de novo assembly. To assist on the correct ordering, a mate-pair sequencing run, which led to 564 Mbp of sequence, was obtained. The sequence could then be organized in 154 scaffolds, leading to a genome sequence with 878 gaps. At this stage, construction of a fosmid library and gap closure after PCR was performed. The PCR-based efforts yielded an insignificant number of gaps being closed (supplementary table S1, Supplementary Material online), whereas ordering of the fosmid library is still work in progress.

As these approaches are laborious and time consuming, we aimed to explore state-of-the-art sequencing technologies that promise to massively generate long reads at a low cost (Koren et al. 2012). A library of single molecule real-time sequencing technology, which yielded 67 Mbp of sequence with a read average of 1.2 kb, was obtained. Although the amount of useful sequence was reduced to 26 Mbp after read correction, which significantly hampered the possibility of closing many gaps, the long reads obtained allowed us to confirm the correct order of the genome of 66. The assembly obtained after this process contains only 84 gaps with a depth of 170X, accounting for 99% of the predicted genome sequence of 66. Thus, a *S. lividans* 66 genomic sequence consisting of a chromosomal scaffold of 8,496,762 bp, predicted to encode SLP3, and a plasmid of 50,064 bp previously reported as SLP2 (Chen et al. 1993; Huang et al. 2003), was obtained.

After automatic annotation and manual curation of the chromosomal scaffold of *S. lividans* 66, 8,152 open reading frames were predicted, of which 8,083 encode proteins (named after the prefix SLI) and 69 for RNAs (table 1). The relative differences of RNAs among *S. lividans* 66 (69), *S. lividans* TK24 (63), and *S. coelicolor* M145 (83) may be related to assembly and annotation difficulties of low (G + C) sequences. The annotated genome sequence of *S. lividans* 66 was deposited at DDBJ/EMBL/GenBank (BioProject id PRJNA168962, accession APVM00000000) and was used to perform whole-genome comparative analysis between closely related strains. These analyses lead to the discovery of 241 genes that are absent from *S. lividans* TK24, of which 136 genes are conserved between *S. lividans* 66 and *S. coelicolor* M145. Moreover, 367 genes are specific to *S. lividans* with respect to *S. coelicolor* (supplementary table S2, Supplementary Material online). Besides these differences, the syntenic regions between the three genomes have an average nucleotide identity of 99%. As discussed further, the vast amount of these genes is held within genomic islands.

Table 1
Selected Genomic and Phenotypic Features of Strains Compared

Strain	66	TK24	M145
Chromosome length	8,496,762	8,318,010	8,667,507
Contigs	85	333	1
(G + C) content	72.2	72.2	72.1
RNAs	69	63	83
Proteins	8,083	7,551	7,825
Metal sensitivity[a]	Cu^R Hg^R As^R Zn^R	Cu^S Hg^S As^S Zn^S	Cu^S Hg^S As^R Zn^S
Accession number	APVM00000000	ACEY00000000	AL645882.2

[a]Metal concentrations used to define resistance (R) or sensitivity (S) are provided in supplementary figure S1, Supplementary Material online, Hg resistance was previously reported by Nakahara et al. (1985) and Sedlmeier and Altenbuchner (1992).

Differential Genomic Islands between *S. lividans* and *S. coelicolor*

Five genomic islands (>25 kb; scoGI), and 18 genomic islets (<25 kb), previously identified in *S. coelicolor* M145 after microarray-based experiments as absent from *S. lividans* (Jayapal et al. 2007; Lewis et al. 2010) could be confirmed. Moreover, of the genes present in *S. lividans* 66 but not in *S. coelicolor* M145, 299 (81%) are contained within four novel genomic islands termed sliGI-1 to sliGI-4 (fig. 1). The obvious reason why these *S. lividans* islands could not be identified by previous efforts using microarray-based approaches is that specific oligonucleotides for the entire gene diversity of these strains could not be designed. Thus, our results complement previous results (Lewis et al. 2010), which concluded that the genetic differences among these strains span throughout their entire chromosomes. In addition, our sequence data show that the differential genomic islands of *S. coelicolor* and *S. lividans*, as summarized in table 2, account for the majority of the genetic diversity amongst the strains compared.

SliGI-1, located at the left-arm of the chromosome, is the largest of the islands identified herein, with a total of 245 kb encoding 241 predicted genes (SLI0867–SLI1107). Of these genes, 138 are conserved in M145 within a single locus. Remarkably, however, this locus is actually at the right-arm of the chromosome of M145 (fig. 1), implying a recombination event and a chromosomal rearrangement in this strain as previously suggested (Weaver et al. 2004). Whether this recombination took place during plasmid curation of *S. coelicolor* A3(2), or earlier in the evolution of this strain, is an interesting question that remains to be answered.

Moreover, sliGI-1 is the only genomic island absent from TK24. Therefore, we propose that sliGI-1 was deleted during laboratory adaptation of TK24, either during plasmid curing or afterwards, consistent with relaxation of purifying selection, as the functions encoded within this locus may not be required under laboratory conditions. Interestingly, the whole sliGI-1 region has an atypical (G + C) content of 68%, which is significantly lower than the average 71.2% for *Streptomyces* genomes as estimated herein using a total of 36 selected genomes (supplementary table S3, Supplementary Material online), and is similar to the value found in plasmid SLP2 (Huang et al. 2003). This implies recent acquisition and a dynamic nature of sliGI-1.

SliGI-1 Hosts Plasmid SLP3

Early genetic and phenotypic evidence, together with a detailed sequence analysis and annotation of sliGI-1, allowed us to identify the plasmid SLP3 therein and obtain its complete nucleotide sequence. Two lines of reasoning led us to the identification of SLP3 as a region integrated into the chromosome of *S. lividans* 66 at the 5′ end of sliGI-1.

First, mercury resistance, which is encoded by the *mer* operon (SLI0946–SLI0953), is absent from SLP3 minus strains, such as TK24. It has been shown that the mercury resistance phenotype can be rescued after conjugation between mercury sensitive strains, that is, TK64, with *mer* plus strains, such as TK19 and 66 (Sedlmeier and Altenbuchner 1992). Interestingly, a homologous *mer* operon has been found in the giant linear plasmids pRJ3L and pRJ28 from *Streptomyces* CHR3T and CHR28T, respectively (Ravel et al. 1998), and in plasmid pPSED01 isolated from *Pseudonocardia dioxanivorans* (Sales et al. 2011). The above actinobacteria were actually isolated from environments with heavy metal pollution, and when plasmids pRJ3L and pRJ28 were used to transform *S. lividans* TK24, the exconjugants became resistant to mercury. Indeed, integration of the *mer* operon into the chromosome of *S. lividans* has been previously confirmed (Ravel et al. 1998).

Second, Eichenseer and Altenbuchner (1994) described a 92 kb amplifiable element from 66, called AUD2, which contained the *mer* operon responsible for mercury resistance. AUD2 is flanked by two insertion elements, called IS1373 (864 bp), which encode functional recombinases or InsA proteins (Volff and Altenbuchner 1997). Two identical sequences predicted to encode transposases were found at the edges of a 94 kb region, which includes the *mer* operon. Thus, we propose to define the region spanning from SLI0868 to SLI0957, which encode two identical InsA paralogs, as the SLP3 amplifiable and mobile element of *S. lividans* 66 (fig. 2). Interestingly, in some *S. coelicolor* strains, the SLP3-borne *mer* operon was found not to be amplifiable, which has

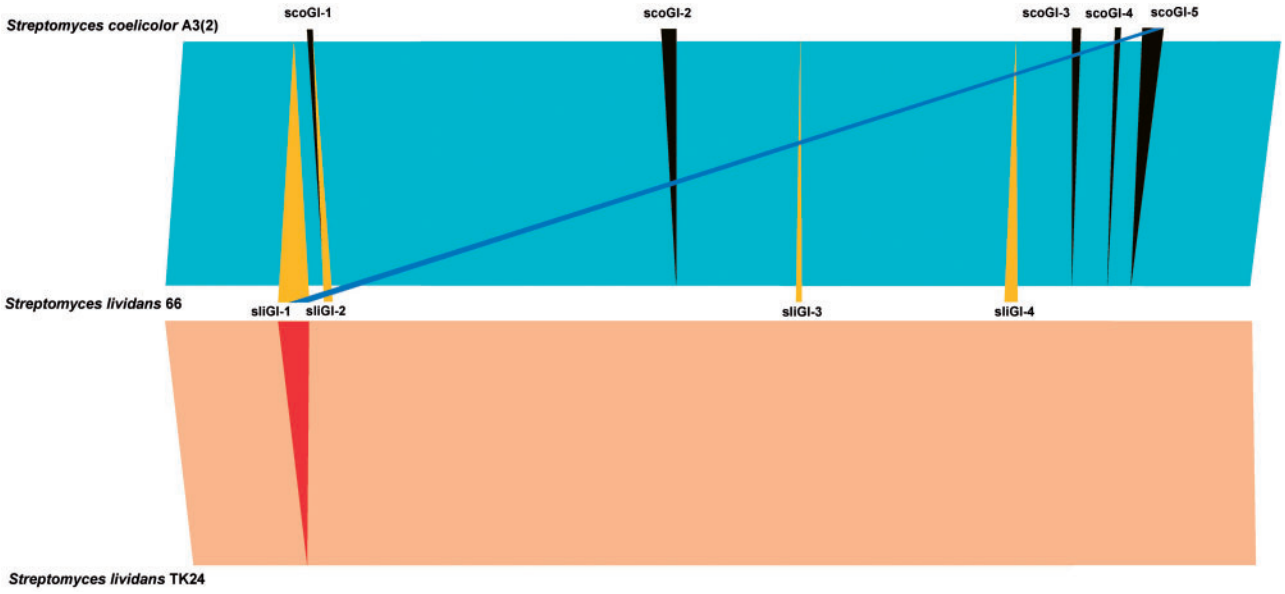

Fig. 1.—Whole-genome comparison of *S. lividans* and *S. coelicolor*. Strain 66 specific genome islands (sliGI) are shown in yellow triangles, and strain M145 (ScoGI) specific genome islands are shown in black triangles. A rearrangement between sliGI-1 and scoGI-5 is shown with a blue line, and the red triangle shows loss of sliGI-1 in TK24.

Table 2
Genomic Islands of *S. lividans* and *S. coelicolor*

GI[a]	66 Genes	TK24 Genes	A3(2) Genes	Total Length (bp)	G + C (%)
sliGI-1	*SLI0867–SLI1107 (SLP3SLI0867–SLP3SLI0956)*	—	*SCO6808–SCO6837*[b] *SCO6841–SCO6948*[b,c]	243,698	68.7
scoGI-1		—	*SCO0979–SCO1000*	26,482	70.5
sliGI-2	*SLI1222–SLI1282*	*SSPG06610–SSPG06554*	—	59,860	68.8
scoGI-2	*SLI6339–SLI6369*[d]	*SSPG01650–SSPG01677*[d]	*SCO3437–SCO3539*	107,313	68.84
sliGI-3	*SLI4708–SLI4745*	*SSPG03232–SSPG03202*	—	27,136	67.35
sliGI-4	*SLI6282–SLI6386*	*SSPG01732–SSPG01632*	*SCO3509–SCO3533*[d]	92,503	67.74
scoGI-3	—	—	*SCO6353–SCO5405*	57,444	68.59
scoGI-4	—	—	*SCO6625–SCO6641*	30,332	68.61
scoGI-5	—	—	*SCO6806–SCO6953*	153,811	69.02

[a]*S. lividans* (sliGI) and *S. coelicolor* (scoGI).
[b]Within scoGI-5.
[c]Several genes were differentially lost within this syntenic region.
[d]Genes present in sliGI-4 and ScoGI-2.

been previously explained by the fact that these strains only have one IS1373 element (Nakahara et al. 1985; Kieser et al. 2000).

SliGI-1 Is a Locus Rich in Metal-Related Genes

Annotation of SLP3 predicted 90 genes, including two "cryptic" peptide biosynthetic gene clusters (SLI0883–SLI0896 and SLI0915–SLI0920). The first is an unprecedented NRPS-tRNA biosynthetic gene cluster, which is discussed later. The second cluster codes for a lantibiotic biosynthetic system, also found conserved in *S. coelicolor* M145 and in *S. zinciresistens*, a taxonomically distant strain isolated from zinc and copper mine tailings (Lin 2011; Lin et al. 2011). Remarkably, both biosynthetic systems are physically linked to a duplicate copper response system, consisting of a CopZ2 chaperon (SLI0895), a CsoR2 response regulator (SLI0893) and a CopA2 ATPase (SLI0896). Within these *cop* genes, a CsoR-like operator sequence, homologous to the sequence recognized by the main copper response regulator CsoR recently identified in *S. lividans* 66 (Dwarakanath et al. 2012), could also be identified (figs. 2 and 3).

Analysis of the remaining sliGI-1 genes also led to the identification of metabolic functions with a role on metal homeostasis. A duplicate NADH-ubiquinone reductase system

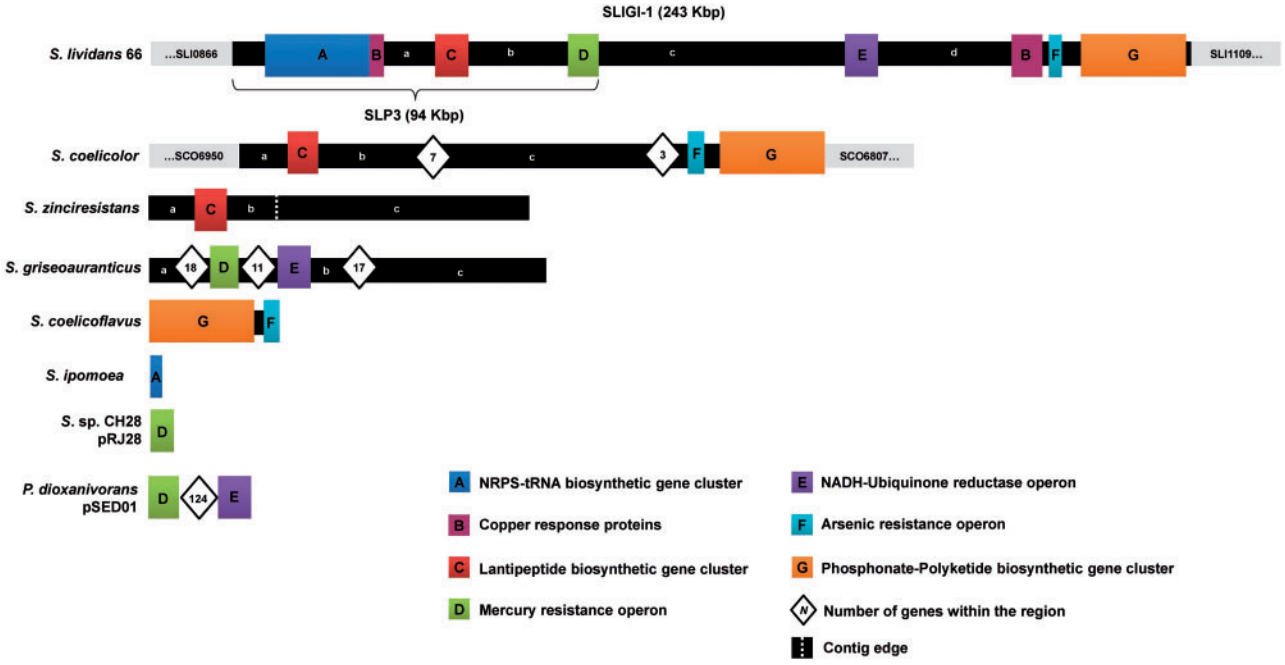

Fig. 2.—Mosaic structure of Genomic Island 1. SliGI-1 is represented as a thick black continuous line. Synteny blocks encoding metabolic functions implicated in metal response are shown with different colors and in uppercase letters. Regions that show certain degree of conservation between synteny blocks are marked with lowercase letters. The SLP3 mobile element is highlighted with a key. Conserved similar regions found in the genomes or plasmids of other actinomycetes are shown in the following lines.

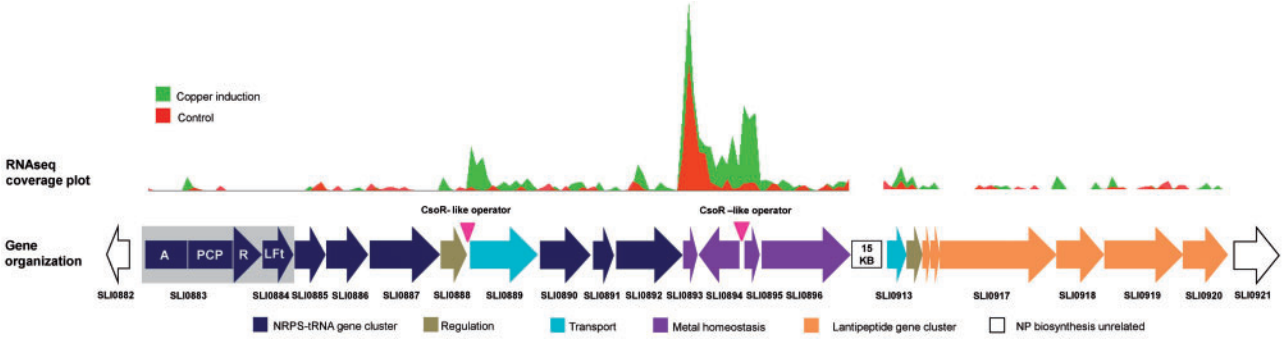

Fig. 3.—Transcriptional response to copper of the SLP3-borne peptide biosynthetic system. The coverage plot of the mapped RNAseq reads shows induction of transport and copper homeostasis proteins. The localization of two CsoR-like operators is also shown as pink triangles. The NRPS-tRNA peptide biosynthetic system also found in *S. ipomoea*, including its domain organization, is highlighted in a gray box.

(SLI1027–SLI1034) known to reduce Cu(II) into Cu(I) (Rodríguez-Montelongo et al. 2006) could be identified. The gene context of this extra copper-reducing system could be found conserved in pPSED01 plasmid of *P. dioxanivorans*, as well as in the genome of *S. griseoaurianticus* (fig. 2). Other sliGI-1 metal-related genes include the previously reported *Streptomyces* arsenic resistance *ars* operon (SLI1077–SLI1080; Wang et al. 2006). The sliGI-1 *ars* operon is actually paralogous to a functional system conserved in the core region of the chromosome of many streptomycetes, including *S. lividans* and *S. coelicolor* (Hänel et al. 1989). Interestingly, the duplicate *ars* operon is physically linked to a polyketide biosynthetic gene cluster (SLI1088–SLI1103) that is conserved in M145 and in *S. coelicoflavus*. Finally, other copper-related genes could be found within SLI1047–SLI1077, including two multi-copper oxidases (SLI1053 and SLI1071); extra copies of *copZ* and *copA* (SLI1063–SLI1064); and a paralog (SLI1067) of the previously characterized SLI4214 copper metallochaperone, or Sco1, involved in developmental switch and cytochrome c oxidase activity (Blundell et al. 2013) (fig. 2).

The high density of metal-related genes contained within sliGI-1, which seems to have a mosaic structure with elements from both closely and distantly related organisms, strongly hints toward a generalized functional role of this locus in metal homeostasis. Furthermore, the mosaic structure of sliGI-1, together with its anomalous (G + C) content, strongly argues in favor of recent acquisition. This suggests that *S. lividans* 66 originally inhabited a niche where metals played a role, and this may be related to loss of the entire locus in TK24. Indeed, previous reports (Hänel et al. 1989; Ravel et al. 1998), and our own data, demonstrate that TK24 is sensitive to metals, including zinc, arsenic, mercury, and copper. TK24 metal sensitivity contrasts with the high metal tolerance exhibited by 66 for all four metals tested, and by M145 for zinc and arsenic (table 1 and supplementary fig. S1, Supplementary Material online).

Expression of SliGI-1 Metal-Related Genes in Response to Copper

To corroborate the possible involvement of sliGI-1 in metal homeostasis the transcriptional response of *S. lividans* 66 under copper induction was assessed. For this analysis, we used RNAseq data that we have recently reported (Dwarakanath et al. 2012). The data sets analyzed come from cultures of 66 that were induced after addition of Cu(II) at a concentration of 400 μM for 2 h during early log phase, and control cultures without added copper. After mapping, the available RNA reads against our *S. lividans* 66 genome sequence, something that could not be done previously due to the lack of a proper reference sequence, differential expression of 137 genes throughout the entire chromosome was confirmed (fold change ≥ 2, FDR corrected P value < 0.05). Of these genes, 29 were induced and 108 were repressed (supplementary fig. S2, Supplementary Material online).

Of the observed copper-induced genes, 11 of them are actually encoded within sliGI-1 (table 3), accounting for 38% of the total number of genes that are induced within only 2.9% of the chromosome. Strikingly, in contrast, none of the repressed genes are encoded within sliGI-1, supporting the proposed role of this locus in metal homeostasis. In agreement with our annotation, the induced genes include the earlier-mentioned duplicate copper response system, consisting of a CopZ2 chaperon (SLI0895), a CsoR2 response regulator (SLI0893), a CopA2 ATPase (SLI0896), and their predicted CsoR-like operator sequences (fig. 3). In the light of these results, the physical association of these copper-responsive genes with the two SLP3-borne "cryptic" peptide biosynthetic gene clusters (including SLI0883–SLI0887 upstream the CsoR-like operator) suggests a functional link between these predicted metabolites and metal homeostasis.

Sequence-Based Predictions of an NRPS–tRNA Biosynthetic Hybrid System

As the understanding of the NP biosynthetic logic progressed, in silico prediction of structural chemical elements has become possible, particularly for nonribosomal peptide biosynthesis (Challis 2008). The *S. lividans* 66 NRPS–tRNA biosynthetic hybrid system includes an NRPS homolog (SLI0883), which encodes a single adenylation (A) domain predicted to recognize arginine; a Phosphopantetheinyl Carrier Protein (PCP); and a Reductase (R) domain. However, no condensation (C) domain, Thioesterase (TE) domain, or additional adenylation domain, could be identified in or closely associated with this unusual biosynthetic gene cluster (SLI0883–SLI0892). The six predicted additional biosynthetic genes, all transcribed in the same direction and potentially transcriptionally coupled, encode putative tailoring enzymes (fig. 3; see supplementary table S4, Supplementary Material online, for a detailed annotation of this biosynthetic gene cluster).

Downstream of the NRPS gene, we identified a gene (SLI0883) encoding a leucyl/phenylalanyl tRNA protein transferase (SLI0884), which has 35% aa identity to the enzyme Aat (ECK0876) from *Escherichia coli*. This enzyme catalyzes the transfer of leucine or phenylalanine, from a charged aminoacyl-tRNA, to an N-terminal basic residue of a protein, usually an arginine, via the N-end rule protein degradation

Table 3
SliGI-1 Genes Induced in Response to Copper (II)

Gene	Annotation	Fold Change	FDR
SLI0889	Major facilitator superfamily MFS 1	6.46	3.57E−03
SLI0894	Heavy metal-associated domain-containing secreted protein	2.65	0.01
SLI0895	Copper chaperone	6.51	2.22E−04
SLI1026	Carbonic anhidrase	3.62	0.05
SLI1041	L,D-transpeptidase catalytic domain containing protein	3.5	0.04
SLI1043	Major facilitator superfamily transporter	3.31	8.92E−05
SLI1044	Cysteine synthase	6.34	1.04E−07
SLI1052	Multicopper oxidase	3.65	5.19E−04
SLI1063	Copper chaperone	3.4	0.03
SLI1064	Copper transporting ATPase	3.03	1.58E−03
SLI1071	Multicopper oxidase	2.22	0.03

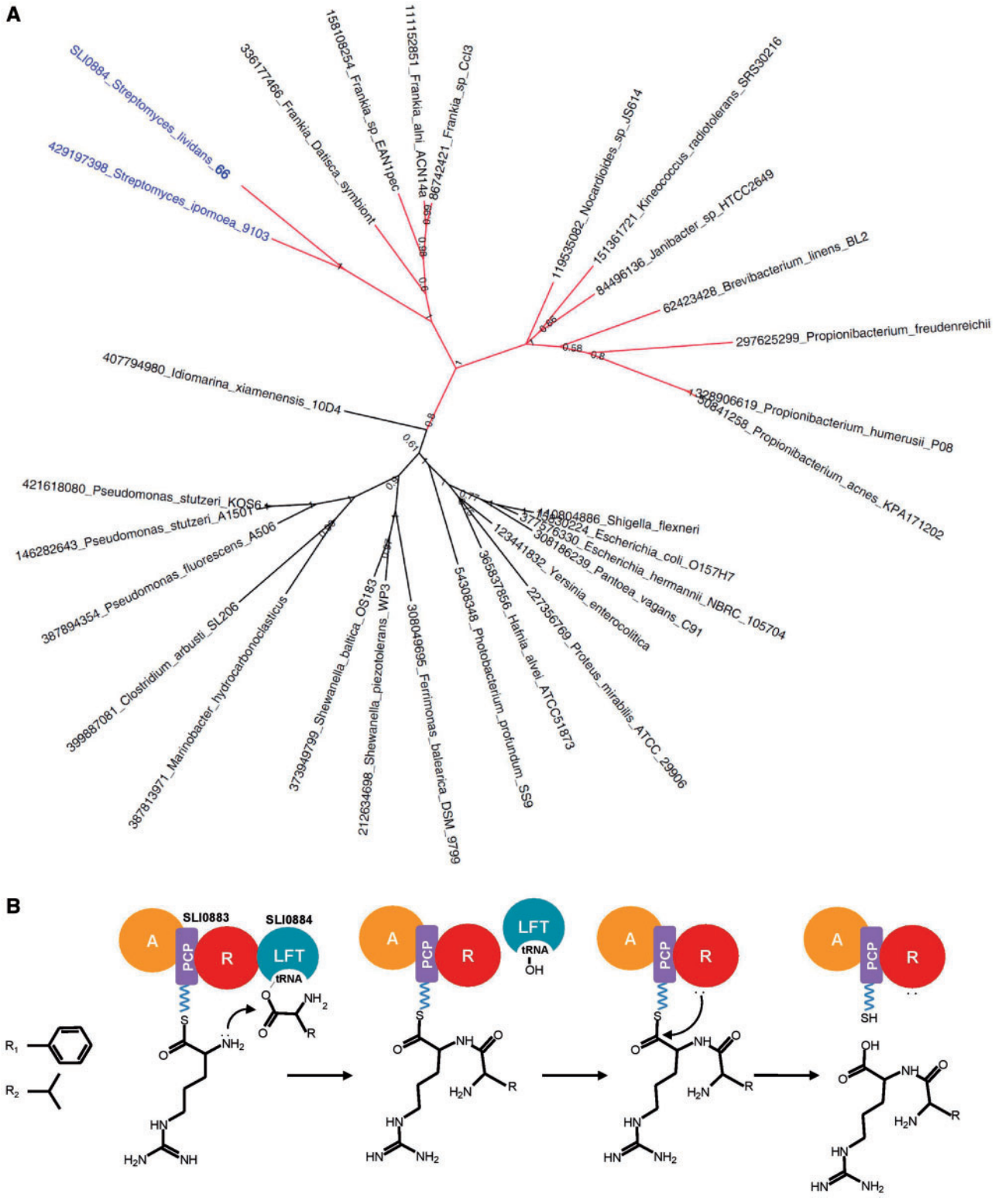

Fig. 4.—Sequence analysis of the NRPS-tRNA biosynthetic system. (*A*) Phylogenetic reconstruction of L/F tRNA transferases in *Actinobacteria* (highlighted in red) and selected homologs from other bacterial lineages. *Streptomyces* homologs are highlighted in blue. The accession numbers of each sequence are indicated before the species names. (*B*) Proposed mechanism of peptide bond formation by the NRPS-tRNA biosynthetic hybrid system.

pathway (Leibowitz and Soffer 1969; Watanabe et al. 2007). Sequence searches in the database using SLI0884 as a query sequence, followed by phylogenetic reconstructions, showed the presence of two orthologs among streptomycetes. The second ortholog (GI ZP_19189297) was found next to an orthologous NRPS with similar domain organization in the genome of *S. ipomoea* (Bioproject PRJNA183480). This suggests a recent recruitment within the context of NPs biosynthesis. The closest homologs outside *Streptomyces* could be found in *Frankia* species, although their gene context is not related to NPs biosynthesis, and thus a proteolytic role may be assumed. A similar scenario was encountered outside the *Actinobacteria* (fig. 4*A*).

Given that SLI0883 and SLI0884 are potentially transcriptionally coupled, and that a single A domain linked to PCP and R domains seem unlikely to be able to produce a peptide bond, we propose that the L/F tRNA transferase homolog present in SLP3 accounts for the lack of both A and C domains. Indeed, recent mechanistic data for the *E. coli* enzyme suggest an analogous condensation mechanism as found in ribosomes and C domains (Fung et al. 2011). Thus, the enzyme product of SLI0884 would attach a Leu or Phe residue, provided by the cognate aminoacyl-tRNA, to an Arg residue, whereas the latter is bound to the PCP of SLI0883. The emerging peptide will then be released by the action of the reductase (R domain) upon the thioester group, as previously found in myxochelin biosynthesis (Li et al. 2008). A reductive cleavage of the predicted metabolite would presumably lead to an aldehyde that might then be further reduced to an alcohol functionality rather than yielding the compound containing a free carboxylic acid. The biosynthetic logic that we propose, as depicted in figure 4*B*, is compatible with the generation of a peptide with the ability to chelate metals such as copper. Moreover, this proposal complements the growing universe of aminoacyl-tRNA transferases involved in amide bond formation within peptide biosynthesis (Garg et al. 2008; Gondry et al. 2009; Zhang et al. 2011).

Final Remarks

In summary, this report provides a high-quality genome sequence of a widely used actinomycete model organism, that is, *S. lividans* 66. The sequencing and assembly strategy used herein, involving most of the commercially available sequencing platforms, provides a useful guideline for the ever-growing number of microbiological laboratories engaged in genome sequencing. Of particular relevance for genome evolutionary comparative analyses is the possibility of performing de novo assemblies, leading to a single scaffold that can be used for comparisons. As demonstrated, the proposed pipeline increases the possibility for identifying genetic diversity amongst closely related strains that otherwise would have been lost. The importance of subtle differences at both the gene-level and single-nucleotide polymorphisms can be associated to adaptive evolution, or from a more anthropocentric view, to genotypes related to desirable features within industrial settings. Indeed, the availability of a genome sequence of *S. lividans* 66 provides a framework for the biotechnological analysis of *S. lividans* TK24, which is one of the most adopted hosts for heterologous production of proteins.

Amongst the gene diversity uncovered by this study, we focused our attention to a large mobile genomic island related to metal metabolism, which harbors the elusive plasmid SLP3. This discovery opens the door for detailed analysis of the elements mediating the transfer mechanism of SLP3. Interestingly, SLP3 encodes enzymes that direct the synthesis of natural products that remain to be discovered. As these biosynthetic systems, some of them with unprecedented features, have been lost in TK24, this organism (together with gene-specific knockouts) can be used for metabolite profiling and discovery of novel NPs. The fact that even in a thoroughly investigated bacterium, such as *S. lividans*, novel NPs biosynthetic systems remain to be discovered, emphasizes the large chemical potential of *Streptomyces*.

To uncap this potential, this study provides two lessons. First, laboratory strains grown for long periods in axenic and artificial laboratory conditions may end up losing their mobile genetic elements which may contain these accessory metabolic features (Medema et al. 2010; Kinashi 2011). Second, beyond the anthropocentric use of NPs as antibiotics, these compounds do have diverse biological and physiological roles relevant in real environmental conditions (Price-Whelan et al. 2006; Yim et al. 2007). In our case, we were able to demonstrate that expression of the NP biosynthetic gene clusters of SLP3 responds to copper. These notions suggest that NPs are adaptive traits whose evolutionary implications have been largely neglected. From a biotechnological point-of-view, they also provide meaningful biological settings that can direct efforts for the discovery of novel NPs with relevant biomolecular activities in the laboratory.

Supplementary Material

Supplementary figures S1 and S2 and tables S1–S4 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The work in FBG laboratory was supported by Conacyt, Mexico (No. 82319) and Langebio institutional funds. This work was supported by CONACyT graduate scholarships to P.C.-M., F.I.-B., and L.A.Y.-G., and the Dutch Technology Foundation (STW) grant 10467 to G.G. and G.P.v.W.

Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.

Anné J, et al. 2012. Recombinant protein production and streptomycetes. J Biotechnol. 158(4):159–167.
Aziz RK, et al. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75.
Bachmann BO, Ravel J. 2009. Chapter 8. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. Methods Enzymol. 458:181–217.
Bentley SD, et al. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417(6885):141–147.
Blundell KL, Wilson MT, Svistunenko DA, Vijgenboom E, Worrall JA. 2013. Morphological development and cytochrome c oxidase activity in *Streptomyces lividans* are dependent on the action of a copper bound Sco protein. Open Biol. 3:120163.
Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27(4):578–579.
Carver TJ, et al. 2005. ACT: the Artemis Comparison Tool. Bioinformatics 21(16):3422–3423.
Challis GL. 2008. Mining microbial genomes for new natural products and biosynthetic pathways. Microbiology 154(Pt 6):1555–1569.
Chen CW, Yu TW, Lin YS, Kieser HM, Hopwood DA. 1993. The conjugative plasmid SLP2 of *Streptomyces lividans* is a 50 kb linear molecule. Mol Microbiol. 7(6):925–932.
D'Huys PJ, et al. 2012. Genome-scale metabolic flux analysis of *Streptomyces lividans* growing on a complex medium. J Biotechnol. 161(1):1–13.
Dwarakanath S, et al. 2012. Response to copper stress in *Streptomyces lividans* extends beyond genes under direct control of a copper-sensitive operon repressor protein (CsoR). J Biol Chem. 287(21): 17833–17847.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792–1797.
Eichenseer C, Altenbuchner J. 1994. The very large amplifiable element AUD2 from *Streptomyces lividans* 66 has insertion sequence-like repeats at its ends. J Bacteriol. 176(22):7107–7112.
Fung AW, et al. 2011. An alternative mechanism for the catalysis of peptide bond formation by L/F transferase: substrate binding and orientation. J Mol Biol. 409(4):617–629.
Garg RP, Qian XL, Alemany LB, Moran S, Parry RJ. 2008. Investigations of valanimycin biosynthesis: elucidation of the role of seryl-tRNA. Proc Natl Acad Sci U S A. 105(18):6543–6547.
Gondry M, et al. 2009. Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes. Nat Chem Biol. 5(6): 414–420.
Hänel F, Krügel H, Fiedler G. 1989. Arsenical resistance of growth and phosphate control of antibiotic biosynthesis in *Streptomyces*. J Gen Microbiol. 135(3):583–591.
Hodgson DA. 2000. Primary metabolism and its control in streptomycetes: a most unusual group of bacteria. Adv Microb Physiol. 42: 47–238.
Hopwood DA. 1999. Forty years of genetics with *Streptomyces*: from in vivo through in vitro to in silico. Microbiology 145(9):2183–2202.
Hopwood DA, Kieser T, Wright HM, Bibb MJ. 1983. Plasmids, recombination and chromosome mapping in *Streptomyces lividans* 66. J Gen Microbiol. 129(7):2257–2269.
Huang CH, et al. 2003. Linear plasmid SLP2 of *Streptomyces lividans* is a composite replicon. Mol Microbiol. 47(6):1563–1576.
Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17:754–755.
Husemann P, Stoye J. 2010. R2cat: Synteny plots and comparative assembly. Bioinformatics 26(4):570–571.
Jayapal KP, Lian W, Glod F, Sherman DH, Hu WS. 2007. Comparative genomic hybridizations reveal absence of large *Streptomyces coelicolor* genomic islands in *Streptomyces lividans*. BMC Genomics 8:229.
Keijser BJ, van Wezel GP, Canters GW, Kieser T, Vijgenboom E. 2000. The *ram*-dependence of *Streptomyces lividans* differentiation is bypassed by copper. J Mol Microbiol Biotechnol. 2:565–574.
Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA. 2000. Practical *Streptomyces* genetics. Norwich (UK): John Innes Foundation. p. 427–430.
Kinashi H. 2011. Giant linear plasmids in *Streptomyces*: a treasure trove of antibiotic biosynthetic clusters. J Antibiot (Tokyo). 64(1): 19–25.
Koren S, et al. 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nat Biotechnol. 30(7):693–700.
Leibowitz MJ, Soffer RL. 1969. A soluble enzyme from *Escherichia coli* which catalyzes the transfer of leucine and phenylalanine from tRNA to acceptor proteins. Biochem Biophys Res Commun. 36(1): 47–53.
Lewis RA, et al. 2010. Metabolic and evolutionary insights into the closely-related species *Streptomyces coelicolor* and *Streptomyces lividans* deduced from high-resolution comparative genomic hybridization. BMC Genomics 11:682.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760.
Li H, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079.
Li Y, Weissman KJ, Müller R. 2008. Myxochelin biosynthesis: direct evidence for two-and four-electron reduction of a carrier protein-bound thioester. J Am Chem Soc. 130(24):7554–7555.
Lin Y, et al. 2011. Draft genome of *Streptomyces zinciresistens* K42, a novel metal-resistant species isolated from copper-zinc mine tailings. J Bacteriol. (22):6408–6409.
Lin YB. 2011. *Streptomyces zinciresistens* sp. nov., a zinc-resistant actinomycete isolated from soil from a copper and zinc mine. Int J Syst Evol Microbiol. 61(Pt 3):616–620.
Medema MH, et al. 2010. The sequence of a 1.8-mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. Genome Biol Evol. 2:212–224.
Myers EW, et al. 2000. A whole-genome assembly of *Drosophila*. Science 287(5461):2196–2204.
Nakahara H, et al. 1985. Mercuric reductase enzymes from *Streptomyces* species and group B *Streptococcus*. J Gen Microbiol. 131(5): 1053–1059.
Price-Whelan A, Dietrich LE, Newman DK. 2006. Rethinking "secondary" metabolism: physiological roles for phenazine antibiotics. Nat Chem Biol. 2(2):71–78.
Ravel J, Schrempf H, Hill RT. 1998. Mercury resistance is encoded by transferable giant linear plasmids in two Chesapeake bay *Streptomyces* strains. Appl Environ Microbiol. 64(9):3383–3388.
Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A. 106(45): 19126–19131.
Rodríguez-Montelongo L, Volentini SI, Farías RN, Massa EM, Rapisarda VA. 2006. The Cu(II)-reductase NADH dehydrogenase-2 of *Escherichia coli* improves the bacterial growth in extreme copper concentrations and increases the resistance to the damage caused by copper and hydroperoxide. Arch Biochem Biophys. 451(1):1–7.
Sales CM, et al. 2011. Genome sequence of the 1,4-dioxane-degrading *Pseudonocardia dioxanivorans* strain CB1190. J Bacteriol. 193(17): 4549–4550.
Schwedock J, McCormick JR, Angert ER, Nodwell JR, Losick R. 1997. Assembly of the cell division protein FtsZ into ladder-like structures in the aerial hyphae of *Streptomyces coelicolor*. Mol Microbiol. 25(5):847–858.
Sedlmeier R, Altenbuchner J. 1992. Cloning and DNA sequence analysis of the mercury resistance genes of *Streptomyces lividans*. Mol Gen Genet. 236(1):76–85.

Tsai IJ, Otto TD, Berriman M. 2010. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. Genome Biol. 11(4):R41.

van Wezel GP, McDowall KJ. 2011. The regulation of the secondary metabolism of *Streptomyces*: new links and experimental advances. Nat Prod Rep. 28(7):1311–1333.

Volff JN, Altenbuchner J. 1997. High-frequency transposition of IS1373, the insertion sequence delimiting the amplifiable element AUD2 of *Streptomyces lividans*. J Bacteriol. 179(17): 5639–5642.

Wang L, et al. 2006. *ars*RBOCT arsenic resistance system encoded by linear plasmid pHZ227 in *Streptomyces* sp. Strain FR-008. Appl Environ Microbiol. 72(5):3738–3742.

Watanabe K, et al. 2007. Protein-based peptide-bond formation by aminoacyl-tRNA protein transferase. Nature 449(7164):867–871.

Weaver D, et al. 2004. Genome plasticity in *Streptomyces*: identification of 1 Mb TIRs in the *S. coelicolor* A3(2) chromosome. Mol Microbiol. 51(6): 1535–1550.

Worrall JA, Vijgenboom E. 2010. Copper mining in *Streptomyces*: enzymes, natural products and development. Nat Prod Rep. 27(5):742–756.

Yim G, Wang HH, Davies J. 2007. Antibiotics as signalling molecules. Philos Trans R Soc Lond B Biol Sci. 362(1483):1195–1200.

Zhang W, Ntai I, Kelleher NL, Walsh CT. 2011. tRNA-dependent peptide bond formation by the transferase PacB in biosynthesis of the pacidamycin group of pentapeptidyl nucleoside antibiotics. Proc Natl Acad Sci U S A. 108(30):12249–12253.

Zhou X, et al. 2005. A novel DNA modification by sulphur. Mol Microbiol. 57(5):1428–1438.

Associate editor: Kenneth Wolfe