Large-scale transposition mutagenesis of *Streptomyces coelicolor* identifies hundreds of genes influencing antibiotic biosynthesis

Zhong Xu\textsuperscript{a}, Yemin Wang\textsuperscript{a}, Keith F. Chater\textsuperscript{b}, Hong-Yu Ou\textsuperscript{a}, H. Howard Xu\textsuperscript{c}, Zixin Deng\textsuperscript{a}, Meifeng Tao\textsuperscript{a}\textsuperscript{*}

State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic & Developmental Sciences, and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200030, China\textsuperscript{a}; Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK\textsuperscript{b}; Department of Biological Sciences, California State University, Los Angeles, CA, USA\textsuperscript{c}

Running Head: Genome-wide Tn5 transposition in *Streptomyces coelicolor*

* Address correspondence to Meifeng Tao, tao_meifeng@sjtu.edu.cn

Science Building, Shanghai Jiao Tong University, 1954 Huashan Road, Shanghai

200030, P. R. China
Abstract

Gram-positive *Streptomyces* bacteria produce thousands of bioactive secondary metabolites including antibiotics. To systematically investigate genes affecting secondary metabolism, we developed a hyperactive transposase-based Tn5 transposition system and employed it to mutagenize the model species *Streptomyces coelicolor*, leading to the identification of 51,443 transposition insertions. These insertions were distributed randomly along the chromosome except for some preferred regions associated with relatively low GC content in the chromosomal core. Base composition of the insertion site and its flanking sequences compiled from the 51,443 insertions inferred a 19-bp expanded target site surrounding the insertion site, with a slight nucleic acid base preference in some positions, suggesting a relative randomness of Tn5 transposition targeting in the high GC *Streptomyces* genome. From the mutagenesis library, 724 mutants involving 365 genes had altered levels of production of the tripyrrole antibiotic undecylprodigiosin (RED), including 17 genes in the RED biosynthetic gene cluster. Genetic complementation revealed that most of the insertions (more than two-thirds) were responsible for the changed antibiotic production. Genes associated with branched chain amino acid biosynthesis, DNA metabolism, and protein modification affected RED production, and genes involved in signaling, stress, and transcriptional regulation were overrepresented. Some insertions caused dramatic changes in RED production, identifying future targets for strain improvement.

Importance

High-GC gram positive streptomycetes and related actinomycetes have provided more than 100 clinical drugs used as antibiotics, immunosuppressants, and antitumors. Their
genomes harbour biosynthetic genes for many more unknown compounds with potential as future drugs. Here we developed a useful genome-wide mutagenesis tool for the study of secondary metabolism and its regulation, based on transposon Tn5. Using *Streptomyces coelicolor* as a model strain, chromosomal insertion was relatively random, except for some hotspots, though there was evidence of a slightly preferred 19-bp target site. We then used prodiginine production as a model to systematically survey genes affecting antibiotic biosynthesis, providing a global view of antibiotic regulation. The analysis revealed 348 genes that modulate antibiotic production, among which more than half act to reduce production. These might be valuable targets in future investigations of regulatory mechanisms, in strain improvement, and in the activation of silent biosynthetic gene clusters.

Key words: genome-wide, transposition mutagenesis, *Streptomyces coelicolor*, antibiotic biosynthesis, prodiginine

**INTRODUCTION**

*Streptomyces* species generate many clinical drugs with antimicrobial, immunosuppressive, anticancer, and antihelminthic activities (1). The production of these secondary metabolites is determined by specific biosynthetic gene clusters, and it is tightly regulated by pathway-specific regulators, global regulators, cellular metabolic and/or physiological cues, and environmental signals. Therefore, wild type *Streptomyces* strains produce only limited amounts of secondary metabolites under laboratory conditions, despite possessing large numbers of biosynthetic gene clusters as revealed by genome sequencing projects (2, 3). In order to rationally engineer high yielding producers for important reagents and to activate the silent biosynthetic potential for new
pharmaceuticals, there is a great need for a genome-wide view of genes modulating secondary metabolism.

*Streptomyces coelicolor* A3(2), the best studied *Streptomyces* strain, readily produces two types of colored antibiotics, a benzoisochromanequinone polyketide antibiotic actinorhodin (ACT) and tripyrrole compounds undecylprodigiosin and streptorubin B (RED), which can act as very useful models for studying the biosynthesis and regulation of antibiotic production. In addition, RED belongs to the group of prodiginines that exhibit a wide range of biological activities, including uses as antibacterial, antifungal, anticancer, antimalarial, and immunosuppressive agents (4, 5). The biosynthetic pathway of RED has been extensively studied (6-13). RED is generated via a hybrid non-ribosomal peptide synthase (NRPS) - type I polyketide synthase (PKS) pathway from L-serine, L-proline, glycine, and acetyl-CoA. Global transcriptional regulatory genes, such as *afsRS, wbiA, dasR, abaA* and *rrdA* regulate the production of RED (14-18).

Many other genes also modulate RED production, such as the RNase J gene *SCO5745* (19), and the membrane protein genes *sarA* and *SCO4174* (20, 21). However, until now there has been no systematic and comprehensive genetic study of genes modulating RED production, even 15 years after the sequencing of the *S. coelicolor* genome, which encodes more than 7,800 genes (2).

Genome-wide transposition is a widely used method for systematic genetic studies in other bacteria (22-27). Unfortunately, early reported *Streptomyces* transposons derived from Tn5, *mariner*, or IS493, had limited use, mainly due to low transposition frequency and/or non-random mutations (28, 29). Instead, Tn5- and *mariner*-based transposons were used to mutagenize genomic libraries of *S. coelicolor* in *E. coli*, and the mutated
alleles were reintroduced into *Streptomyces* via homologous recombination (30-32).

Recently, a codon optimized hyperactive Tn5-based transposition system (33, 34) was
developed and shown to efficiently and randomly mutagenize *Streptomyces* genomes in vivo. However, the temperature-sensitive replicative vector used to deliver the transposon had to be cured by temperature shift-up. Moreover, the temporary replication of the transposon-delivery vector in the target cell can lead to multiple insertion events in one cell and therefore complicate the interpretation of the phenotypic alteration in the resulting mutants (28). Other transposon delivery vectors recently developed for in vivo transposition in *Streptomyces* included a *mariner*-based transposon Himar1 system (35), and a *Nocardia asteroides* IS204-based system (36).

In order to understand the modulation of the RED biosynthesis in *S. coelicolor*, herein we further developed a codon-optimized hyperactive Tn5-based transposition system, characterized genome-wide Tn5-transposition in *S. coelicolor*, and screened for genes affecting biosynthesis of RED. Remarkably, more than 300 chromosomal genes besides the RED biosynthetic genes were found to have such effects.

**MATERIAL AND METHODS**

**Plasmids, primers, strains, and culture conditions.** Plasmids and strains used in this study are listed in Table 1. Primers used are listed in Table 2. *E. coli* DH5α was used as a cloning host. *E. coli* ET12567/pUZ8002 was used as a helper strain to mobilize oriTRK2-plasmid into *Streptomyces* via intergeneric conjugation (29). *S. coelicolor* X737 was derived from *S. coelicolor* M145 by removal of the entire act gene cluster via homologous recombination using pHL737 (see next section). *S. coelicolor* strain X737...
was used to construct the Tn5 transposition library. *E. coli* was grown in Luria-Bertani (LB) broth at 37°C. *S. coelicolor* strains were grown at 30°C, using Soy Flour Mannitol (SFM) agar (29) for sporulation, SFM agar supplemented with MgSO₄ (20 mM) for conjugation, YBP (18) agar for production of RED, and YEME to cultivate mycelium for genomic DNA preparation (29). Antibiotics were added when necessary at the following concentrations: apramycin (Apr), 50 μg mL⁻¹; ampicillin (Amp), 50 μg mL⁻¹; kanamycin (Km), 50 μg mL⁻¹; trimethoprim (TMP), 50 μg mL⁻¹; thiostrepton (Thio) 20 μg mL⁻¹; chloramphenicol (Cm), 25 μg mL⁻¹.

**Details of plasmid and strain construction.** pMM1, containing a 39 kb ClaI-EcoRI genomic DNA fragment of *S. coelicolor* M145, including the entire *act* gene cluster (22 kb) (37), was used to construct a gene replacement vector pHL737 for the deletion of the ACT biosynthetic gene cluster. Removal of the 22 kb *act* gene cluster from pMM1 by restriction digestion with Stul gave rise to pHL664. A 15 kb XbaI-EcoRI fragment of pHL664 was ligated with pOJ260 (38) to yield the gene replacement construct pHL737. pHL737 was introduced into *S. coelicolor* M145 by conjugation from *E. coli* ET12567/pUZ8002, according to standard method (29). Exconjugants were selected by flooding with TMP (to remove *E. coli* donor strains) and Apr at 12-14 h. Colonies were transferred to Apr-containing SFM agar medium, and spores were collected after 4-5 days incubating. These spores were subjected to three rounds of non-selective growth on antibiotic-free SFM agar medium to permit loss of pHL737, and Apr-sensitive clones without blue color (no ACT) were chosen. The *act* deletion scar sequence was amplified by PCR using primers act-up and act-down, and confirmed by DNA sequencing. The resulting *act* gene cluster deletion strain was named *S. coelicolor* X737. pHL734
(GenBank accession no. KU672723) constructed by this study was used for transposition mutagenesis in high GC Streptomyces. It contains a synthetic transposase gene \(\text{tnp}(5)\), an origin of transfer (\(\text{oriTRK2}\)), an apramycin resistance gene \(\text{aac}(3)\text{IV}\), an \(\text{E. coli}\) DNA replication origin (\(\text{ori-pUC}\)), and two mosaic end (ME) sequences \(5'\text{-CTGTCTCTTATACACATCTT-3'}\). In pHL734, the 2045 bps ME-\(\text{aac}(3)\text{IV-ori-pUC-ME}\) fragment (mini-Tn5) was amplified from pSET152 (38) using primers MEap-F and MEap-R; the 1259 bp \(\text{bla-oriTRK2}\) fragment was amplified from pIJ773 (39) using primers oriT-F and oriT-R; and the 1613 bp \(\text{PermE^*-tnp}(5)\) fragment was synthesized de novo.

pMT3 was derived from pSET152 (38) by replacing a 563 bp PspOMI-SgrAI fragment with a thiostrepton resistance gene (\(\text{tsr}\)) amplified by PCR using primers tsr-F and tsr-R. pMT3 contains an origin of transfer (\(\text{oriTRK2}\)), an integrase gene (\(\text{int}\)) and attachment site (\(\text{attP}\)) from the actinophage \(\Phi\text{C31}\), and two antibiotic resistance genes \(\text{aac}(3)\text{IV}\) and \(\text{tsr}\).

pMT3 was used to deliver wild type genes into transposition mutants for genetic complementation.

**Mutagenesis of \(S.\ coelicolor\) strains using pHL734.** pHL734 was introduced into \(S.\ coelicolor\) M145 or X737 by conjugation from \(E.\ coli\) ET12567/pUZ8002, according to the standard procedure at a donor:recipient ratio of 1:1 (29). In theory, Tn5 transposition should take place once the mini-Tn5 delivery vector pHL734 is conjugated into the recipient cell and the Tnp(5) is expressed, with the mini-Tn5 fragment being cut and pasted to the new target site, and the synthetic transposase gene \(\text{tnp}(5)\) being lost after transposition, as pHL734 cannot replicate in \(\text{Streptomyces}\). In practice, transposants (mutants) were obtained after plating the mixed \(E.\ coli\)-\(\text{Streptomyces}\) cells onto supplemented SFM, flooding at 12-14 h with TMP (to eliminate \(E.\ coli\) donor cells) and...
Apr, and incubating for 4-5 days. Transposants were transferred individually, as a patch of about 0.5-cm², to solid YBP medium to grow mycelium (for DNA extraction) and observe the production of RED. To avoid spontaneous mutants with altered yield of antibiotics, which are frequent in streptomycetes, we discarded colonies with obvious sectors, irregular shape, or tiny colony size, which had been reported to be associated with spontaneous mutations such as large deletions (40).

**Southern blot analysis.** Genomic DNA isolated from individual transposants was digested with Apal. DNA fragments were separated on an agarose gel and transferred to a positively charged nylon membrane. Mini-Tn5 insertion fragments were detected by Southern blot hybridization with mini-Tn5 DNA probe according to manufacturer's instructions (DIG DNA labeling and detection kit, Roche).

**Rescue of mini-Tn5 and identification of insertion sites in transposants.** Genomic DNA of each transposant was extracted from mycelium and digested with Apal which cuts S. coelicolor DNA on average once every 1.3 kb. No Apal site is present within mini-Tn5. The digested DNA samples were self-ligated and introduced into E. coli DH5α by chemical transformation. Apramycin resistant colonies containing mini-Tn5 with flanking chromosomal sequences were isolated. The recombinant plasmid in the resistant clone was extracted and sequenced. Primer DownS and UpS annealing to mini-Tn5 were used to determine the chromosomal sequence flanking either side of mini-Tn5.

**Tracking all transposants within the S. coelicolor X737 transposition library.** Each of the ca. 53,900 transposants constituting the S. coelicolor X737 transposition library was inoculated as a ca. 0.5-cm² patch on solid YBP medium, grown at 30°C for 4 d, and
the mycelium collected. Genomic DNA was extracted from the pooled mycelium. The transposon-chromosome junctions were tracked using the deep sequencing method “high-throughput insertion tracking by deep sequencing” (HITS) as described (41).

Briefly, the pooled genomic DNA from the mutant library was ultrasonically sheared, and size-selected by agarose gel electrophoresis. DNA of 200-400 bp was recovered, end-blunted with NEBNext End Repair Module (NEB), and “A” base added to the 3’-end of the blunted DNA. The DNA mix was size-selected to purify fragments of 200-400 bp by agarose gel extraction, and ligated with partially complementary adapter oligos ADTn1 and ADTn. The transposon-chromosomal junctions were enriched by PCR reaction (10 cycles) with the adapter-specific primer AD-F and one of two mini-Tn5-specific primers INME-1bio and INME-2bio, both of which contain a biotin modification at their 5’-ends. The PCR product was size selected by agarose gel electrophoresis to recovered DNA of 100-200 bp. Streptavidin magnetic beads (NEB) were then used to capture the biotinylated PCR product. The resultant biotinylated single-stranded templates were tested by real time fluorescence quantitative PCR. To ensure the efficient amplification of high-GC *Streptomyces* DNA, GC buffer II (Takara) was used in all PCR reactions. The qualified single-stranded templates were PCR-amplified with primers FOR, INME-1, and INME-2 for another 8 cycles to enrich yield, then sequenced by Solexa sequencer. Sequencing reads containing the Tn5 mosaic end sequence were identified in the raw FASTA files by FastX toolkit and trimmed of the mosaic end sequence. Processed reads, typically 58 bp and 59 bp in length, were mapped to the *S. coelicolor* M145 genome sequence using bowtie software, allowing 1-2 bp mismatches inside of the 58-59 bp sequences. Custom PERL
scripts were written for specific analyses, such as to count the GC contents of the slide windows and to retrieve DNA sequences of the hot regions, the insertion sites and the slide windows.

**Statistical Analysis of Transposon Distribution.** Binomial tests were utilized to assess the randomness of Tn5 insertions in the *S. coelicolor* X737 genome. The X737 genome sequences were divided into bins of equal size and each insertion event was assigned to a bin based on its chromosomal location. Thus the chromosome was divided into 8,645 bins of 1-kb and 4,322 bins of 2-kb for the binomial tests of the distribution of the 51,443 insertions and the 25,694 (or 25,749) insertions in the top strand (or the bottom strand) respectively. The number of insertions in each bin was tabulated and a binomial test was performed for each bin, yielding the probability of \( q \) insertions in any bin in \( n \) chromosomal insertions, where the probability of one insertion in any bin in one chromosomal insertion being \( P \) \( (P = 1/\text{number of bins}) \) (42).

We also created simulated transposon insertions by using a random number generator to assess the transposition hot regions and gaps of Tn5 insertions in the *S. coelicolor* genome. Libraries constituted by 51,443 and 25,749 random numbers within a range of 1 - 8,645,000, corresponding to each site of the *S. coelicolor* M145 genome, were generated to simulate random insertion in the genome, top strand, and/or bottom strand respectively.

**Measurement of RED production.** Mini-Tn5 insertion mutants and their parent strains were cultured on solid YBP medium for 84 h and an equal weight of each culture (500 mg) was cut and removed from the plates. 500 \( \mu \)L of absolute methanol was added to the agar culture, which was subsequently broken up using a homogenizer (5,000 rpm, 15
s; twice). Samples were centrifuged to remove the suspended solids (12,000×g, 5 min), and the absorbance of supernatants at 530 nm was measured (43). The relative RED production of each mutant was calculated from the ratio (mutant530/wild type530). Two biological replicates were performed for each strain.

Genetic complementation. All genes were amplified by PCR from the genomic DNA of S. coelicolor M145 with gene-specific primers. The PCR product were ligated into pGEM-T Easy TA vector (Promega), confirmed by DNA sequencing, and sub-cloned into the pSET152-derived integrative vector pMT3. The resulting plasmids (Table 1) harboring the wild type genes were introduced into the appropriate S. coelicolor transposants, and Thio resistant clones were selected.

RESULTS

Efficient transposition by a mini-Tn5 transposon delivery vector encoding a synthetic hyperactive transposase. To mutagenize S. coelicolor, we constructed pHL734 to deliver a mini transposon of Tn5 (mini-Tn5), which includes an Apr resistance gene and an E. coli DNA replication origin bounded by two Tn5-mosaic ends (ME) (Fig. 1A). Besides the mini-Tn5, pHL734 harbors an origin of transfer (oriTRK2) to initiate conjugative transfer helped by RK2-derived helper plasmids such as pUZ8002, and a mutated Tn5 transposase gene tnp(5), codon-optimized for expression in Streptomyces. Most importantly, the tnp(5) gene was engineered to carry five point mutations (E54K, M56A, P242A, E345K, L372P) documented to cause high transposition efficiency (44-47). A strong constitutive promoter PermE* guaranteed the expression of tnp(5) in Streptomyces. The mini-Tn5 inserts in the Streptomyces chromosome can be rescued...
from the target site by restriction digestion and religation to afford a replicative plasmid in
E. coli, which can be used for the identification of the site of mini-Tn5 insertion.

We tested the ability of pHL734 to mutagenize S. coelicolor M145. Transfer of
pHL734 into M145 by conjugation typically yielded ca. 200 apramycin resistant
exconjugants per 3×10⁹ spores per Petri dish. To validate the transposition events, ten
exconjugants were tested by Southern blot. In all mutants, a single restriction fragment
hybridized with the mini-Tn5 probe (Fig. 1B), indicating a single transposition event in
each mutant. We also randomly selected 50 exconjugants, rescued the
mini-transposons in E. coli, sequenced the rescued plasmids, and analyzed the DNA
sequences joined to both ME sites. All exconjugants contained a 9 bp target site
duplication associated with Tn5 transposition, indicating that the exconjugants were all
bono fide transposition mutants. In addition, all insertions were at different and
recognizable positions in the S. coelicolor M145 genome.

Some of the S. coelicolor M145 transposition mutants showed phenotypic alteration,
such as overproducing or abolishing the blue pigmented antibiotic ACT, or deficiency in
antibiotic production and morphological differentiation (bald) (Fig. 1C). ACT production
generally obscures production of the red pigmented antibiotic RED, so an S. coelicolor
strain X737 was constructed from S. coelicolor M145 by deleting the whole gene cluster
(22 kb) for the biosynthesis of ACT, permitting RED production to be observed at the
colony level. Transfer of pHL734 into S. coelicolor X737 by conjugation also yielded ca.
200 Apr resistant exconjugants per Petri dish, and mutants overproducing RED or
abolishing RED production were readily observed (Fig. 1C).
Detailed characterization of genome-wide transposition in *S. coelicolor*. To further characterize genome-wide transposition in *S. coelicolor*, and especially to investigate genes influencing RED production, a library with approximately 53,900 mutants was generated by transposition mutagenesis of *S. coelicolor* X737. These mutants were pooled and subjected to “high-throughput insertion tracking by deep sequencing” (HITS) (41). This produced 4,583,374 reads that were matched to the *S. coelicolor* genome, identifying 51,443 insertions in the non-repetitive chromosomal regions, with 25,694 running from 5' to 3' with the top strand of *S. coelicolor* chromosome and 25,749 running in the opposite direction.

The distribution of the insertions was assessed by binomial testing and simulation analysis. The genome was divided into 8,645 bins of 1 kb, in a base-by-base moving window and each of the 51,443 transposon insertions was allocated to bins. Binomial tests performed for the observed frequency of insertions in each bin indicated that 90.3% bins had a probability >0.01, and 66.8% a probability >0.05, of having resulted from random insertion. Of those with a probability <0.01, 455 bins had more than 12 inserts/kb and 382 bins had no transposon insertion, indicating probable hot regions and cold regions of insertion respectively. Binomial tests on the insertions on the top strand and the bottom strand gave similar results, in that 70.0% bins had a probability >0.05 of having resulted from random insertion, suggesting that, overall, insertions were distributed randomly between the two strands of the chromosome. The insertion distribution was also assessed by a random Monte Carlo simulation method (22, 48). Two hundred simulated libraries of 25,749 insertions were generated, each library was divided into 8,645 bins of 1 kb, and the inserts/kb ratio was calculated for each bin,
yielding a largest ratio of 16 in 200 simulated libraries, indicating that any 1-kb DNA region with >16 inserts in one strand (or one orientation) was extremely rare in a library of 25,749 insertions. We therefore defined a “hot region” as a DNA segment with more than 16 Tn5 inserts/kb in one orientation. Thirty-nine hot regions were observed (Fig. 2A). All hot regions, except for H1 and H1', were located in the core chromosomal region. Twenty six of the 39 hot regions appeared in pairs that are located at the same regions of both chromosomal strands, e.g. H1 vs H1'. The hottest region was H20', holding 174 inserts within a 2.4 kb DNA segment (73 inserts/kb). Details of the hot regions are listed in supplemental material Table S1. In addition, 2,381 insertions were located 5-, 10-, or 15-bp from other insertions, as indicated by an analysis of the insertion gaps (Fig. 2B). Such a so called “periodic fashion” has been reported before (49, 50). Interestingly, 4,423 paired-insertions were identified that were inserted in the same chromosomal sites with opposite directions, having a gap = 0 as shown in Fig. 2B. The average theoretical number of paired-insertions was 153±12 (standard deviation) in simulations of 51,443 insertions, implying that the observed 4,423 paired-insertions resulted from transposition site preference. Many of the above-mentioned hot regions were rich in paired-insertions (Table S1).

Potential transposition “cold regions” were estimated by a comparison of the observed insertion gap sizes with simulated gap sizes calculated from libraries of 51,443 insertions (Fig. 2C). The observed gap distribution curve skewed significantly from the simulated curve at the gap size of >1.0 kb. In fact, there were 136 simulated gaps >1.0 kb, while there were 648 actual gaps >1.0 kb. In addition, the simulation suggested that a gap larger than 1.9 kb would be rare assuming a random insertion distribution, yet there were
79 gaps >1.9 kb in the X737 transposition library, which summed up to 203 kb of chromosomal DNA. Therefore most of these large gaps, particularly those larger than 1.9 kb, reflected either the cold regions of transposition insertion, or the presence of genes essential for growth under the conditions used (Table S2).

Many of the hot regions were associated with relatively low-GC content, in that 25 out of 39 hot regions had a GC content lower than 70% (Table S1). This might reflect an insertion preference for lower GC DNA segments that was indicated in statistical boxplots (Fig. 2D).

**Base composition analysis of the site of transposition insertion and flanking sequences.** The transposition insertion sites and flanking sequences of the 51,443 insertions were compiled by calculating the base composition at each position. The results are shown in Fig. 3 and Table S3, where the site of insertion corresponds to 0, and the 9-bp target site that is duplicated during Tn5 transposition corresponds to -4 to 4. Striking palindromic symmetry of base composition is observed across the 9-bp target site and it is expanded for 5-bp at either sides, recalling the previously reported 19-bp symmetric structure obtained by in vitro Tn5 mutagenesis of a mammalian cDNA library, which has a GC content of 55% (42). Based on the base composition at each position, and with the notice of significance increase of A or T at certain positions, a 19-bp consensus/bias sequence was proposed for Tn5 insertion in high GC S. coelicolor as 5'-gc(g/a)c(g/a)^g(c/t)(c/t)c(a/t)g(g/a)(g/a)c^3', where the caret (^) indicates the boundaries of the 9-bp duplicated target site (Table S3). This sequence preference was most obvious in the composition table compiled from the 4,423 paired-insertions (Table S4). Nevertheless, it should be pointed out that the biased bases are not in an
absolutely high frequency at any position and other bases are always present, therefore
lower case letters are used here to denote the consensus/bias sequence and to indicate
the relative randomness of target selection.

**Genome-wide screening for genes affecting undecylprodigiosin production.** In a
library of 51,443 random inserts in an 8.7 Mb chromosome, genes longer than 400 bp
have more than 90% likelihood of being insertionally inactivated, according to the
algorithm of Liberati *et al* (22). More than 86% of annotated *S. coelicolor* genes are
larger than 400 bp (2), legitimising use of our pHL734-generated library in a
genome-wide search for genes affecting secondary metabolism.

*S. coelicolor* X737 holds the entire gene cluster for the biosynthesis of the red
pigmented antibiotics (RED) from acetic acid and three amino acid precursors (Fig. 4A),
but lacks the genes for production of the blue/red (depending on pH) antibiotic ACT. We
were therefore able to use the X737-Tn5 library to study genes that affect the production
of RED, in the absence of any masking by ACT. We found 383 mutants producing more
RED, and 341 mutants producing less RED, than the parent strain X737. The variations
of RED production of all these mutants were confirmed by measuring UV-absorbance at
530 nm of extracts of cultures grown on YBP agar for 84 h. The insertion sites of all
mutants were located by mini-Tn5 rescuing and DNA sequencing. These 724 mutants
are listed in Supplemental dataset 1 in the supplemental material, which includes relative
production of RED, insertion sites, inactivated genes or operons, and annotation of the
gene products.

Among the 341 RED abolished/decreased mutants, 107 inserts were in the *red* gene
cluster (Fig. 4B), including hits in all biosynthetic genes for one intermediate
4-methoxy-2,2’-bipyrrrole-5-carbaldehyde (MBC), 4 of 6 biosynthetic genes (redP, R, Q, and L) for another intermediate 2-undecylpyrrole (2-UP), the condensation gene redH, and two pathway specific activator genes redD and redZ. Insertions into the function-unknown genes redS and redT abolished RED production as well, although this might reflect polar effects on the downstream redUVZ genes. Among the 107 red mutants, insertions were not found in six red genes, SCO5880 (redY), SCO5893 (redK), SCO5894 (redJ), SCO5897 (redG), SCO5898 (redF), and SCO5899 (redE). All these findings are consistent with literature reports. Thus, although it has been reported that a redK-disrupted mutant did not produce undecylprodigiosin or streptorubin B, the mutant accumulated a hydroxylated undecylprodiginine derivative instead (12), which might be also red-colored; disruption of redJ did not completely abolish RED production (51); a redG mutant produced one of the RED compounds (undecylprodigiosin) (11); and redF, redY, and redE have no assigned function in the RED biosynthetic pathway.

A further 617 inserts located outside the red cluster were found to affect RED production. These included 214 insertions in 161 genes that reduced/abolished RED production (a further 20 RED-decreasing inserts were in intergenic regions), and 362 insertions in 192 genes that increased RED production (and 21 RED-increasing inserts in intergenic regions). These RED-decreasing or RED-increasing insertions were scattered along the chromosome, and were slightly more frequent in the chromosomal "core" region (Fig. 4C and 4D).

**Analysis of chromosomal genes modulating RED production.** An important question associated with transposition mutagenesis is whether the phenotypic alteration in a mutant is caused by the transposon insertion or by a spontaneous mutation elsewhere.
To evaluate the association of the changed RED production with the insertion, 13 RED-changing mutants representing insertions in 13 genes were selected to perform in-trans complementation. The genes chosen were putative regulatory genes, previously not known to influence antibiotic production, apart from *wblA*. As shown in Fig. 5, RED production of ten (more than two-thirds) of the RED-changing mutants was restored to the same level as the parent strain X737 upon reintroduction of the corresponding wild-type alleles delivered by a chromosomal integrative vector. Thus, by extension, the variation of RED production in more than two-thirds of all RED-changing mutants was likely to have been caused by insertional gene inactivation.

In addition, we inferred that genes with more than one insertional mutant changing RED production in the same way, *i.e.* either all increasing or all reducing, could reliably be attributed a role in modulating RED production. Many peaks in the RED-decreasing mutation distribution plot were generated by multiple cases of insertions in single genes (Fig. 4C). For instance, five genes had more than 5 RED-decreasing insertions: *SCO3390* for a two component sensor kinase, *SCO3404* for a cell division protease FtsH2, *SCO3673* for an iron-sulphur binding reductase, *SCO5264* for a hypothetical protein, and *SCO4069* for the previously reported putative membrane protein SarA (20). Therefore products of these five genes are particularly reliable candidates for up-modulators for RED production. In addition, 15 genes had more than one RED-decreasing insertion, strongly suggesting that these genes positively affected RED production as well. In total, 20 reliable RED up-modulators were identified here, including three previously reported RED up-modulators *sarA* (20), *bldG* (52), and *whiJ* (53) (Table
S5). A further 141 candidate RED up-modulators were identified by single insertions (Supplemental dataset 1).

The highest peaks in the RED-increasing mutation distribution plot (Fig. 4D), each with more than 10 inserts, were: two clusters of genes, cmdB-F and two-component regulatory genes SCO3008/SCO3012/SCO3013; and two single genes, NAD$^+$-dependent glutamate dehydrogenase gene SCO2999 and a ATP/GTP binding protein gene SCO5677. Other reliable down-modulators included 42 genes with more than one RED-increasing insert, 13 of which have been reported previously [wblA (54), bldM (55), bldN (56), rmdA (57), relA (58), ohkA (59), cutRS (60), nsdB (61), SCO4174 (21), SCO5745 (19), and adpA (62)] (see Table S5). A further 140 potential RED down-modulators were identified by single insertions (Supplemental dataset 1).

The reliable RED modulators, including 20 up-modulators and 52 down-modulators, were classified, according to their annotated functions or previously described functions, into 8 functional groups: morphological development (5 genes), cell envelope biosynthesis (2 genes), DNA maintenance (10 genes), RNA processing (3 genes), protein modification (5 genes), amino acid metabolism (7 genes), signaling integration (25 genes), and others (15 genes) (Table S5). The most highly represented class comprised 25 genes related to stress response, signaling, and transcription regulation, of which 15 were observed for the first time to affect RED production. Four of the seven amino acid metabolism genes were branched-chain amino acid synthetic genes. More unexpectedly, ten genes for DNA replication, repair, and transfer were implicated.

Some of the insertions in reliable modulatory genes changed RED production dramatically compared to the parent strain X737. Production was increased by more...
than 10-fold in some mutants of five DNA transfer genes (SCO4127-4129, SCO5339, and SCO5677), four signaling and regulatory genes (SCO3008, SCO3012-3013, SCO3579), and one amino acid metabolism gene (SCO2999). Production was decreased dramatically in some mutants of four signaling and regulatory genes (SCO2168, SCO3664, SCO4069, SCO5264), and one respiratory gene (SCO1082).

DISCUSSION

In this study we have developed a Tn5 system for efficient transposition in the Gram positive *Streptomyces* and conducted a genome-wide mutagenesis in *S. coelicolor*. Analysis of the 51,443 transposition insertions of one library indicated that their distribution was roughly random along the chromosome, except for some transposition hot regions. Analysis of the base composition of the insertion site-flanking sequences led to the identification of a 19-bp expanded insertion target site with a palindromic consensus/biased sequence of 5'−gcrccygyycwcgrc′ygygc-3', reflecting preferences of certain bases at certain positions for the target-binding in a positively charged groove of the Tn5 transposase (63, 64). The middle base of the target site was previously found to be N for high-GC *S. coelicolor* genomic DNA library (50) and W for other DNA mutagenesis (42, 49, 50). In this study, the AT composition reached 53.7% at the mid-position of the target site, which is about twice of the average AT-content of the genomic background. We attributed this to a preference for A/T by Tn5 transposition. However, the base preference across the expanded target site is moderate since other bases were present at some level in all positions, giving a degree of randomness in target selection sufficient for biotechnological application in *Streptomyces*. 
In *Streptomyces*, antibiotics are produced as secondary metabolites using precursors derived from primary metabolism devoted to cellular maintenance, growth, and proliferation. Thus secondary metabolism is tightly controlled, giving rise to the very low level, even non-production, of most secondary metabolites of wild-type strains under diverse laboratory conditions. Taking RED biosynthesis as a model system, in the last two decades many genes have been implicated in the regulation of secondary metabolism in *S. coelicolor* (65, 66). In our systematic transposon mutagenesis we identified 724 mutants affecting RED production, and a total of 348 genes were implicated as candidate determinants of RED production level, in addition to the 17 red biosynthetic genes.

Seventy-two RED-changing genes located outside the RED biosynthetic gene cluster were represented by more than one mutant (Table S5), making it highly likely that all of these genes are important in modulating antibiotic biosynthesis. Furthermore, in tests of the association of the phenotype change with the mini-Tn5 insertion, the RED production of 10 of 13 mutants was restored to the parental level by the introduction of the appropriate wild type gene. This indicated that the mini-Tn5 insertion was responsible for the changed antibiotic production in the majority of the RED-changing mutants, while less than one-third of RED-changing mutants were false positives in which an unknown spontaneous mutation caused the phenotype (but we note that negative results in these complementation tests could arise from other causes, including the production of a dominant negative phenotype by a truncated fragment of the gene product, or gene position effects). On the basis of these complementation results, the 283 RED-changing
genes identified by single insertions have a probability of more than two thirds of being
genuine antibiotic modulating genes.

It should be noticed that mutants of some genes with similar functional annotation had
different effects on RED production. For example, both SCO5673 and SCO6012 were
annotated as chitinase, insert of SCO5673 showed RED-decreasing, while insert of
SCO6012 showed RED-increasing; both SCO3550 and SCO5166 were annotated as
helicase, insert of SCO3550 showed RED-decreasing, while insert of SCO5166 showed
RED-increasing; Inserts in methyltransferase genes SCO2317, SCO4504 and SCO5895
showed RED-decreasing, but insert in methyltransferase gene SCO2170 showed
RED-increasing. These cases may indicate the limitations of the annotations, for
example proteins with similar functional annotation may not be involved in the same
reaction. In addition, some insertions in different sites of one gene had contradictory
effects on RED production, for example, SCO3577 and SCO3061 (supplemental dataset
1), the function of these genes should be considered with caution.

It is unlikely that all antibiotic modulating genes were represented in the X737
transposition library. For example, genes that are important only in other cultivation
conditions would be overlooked. In addition, some genes, especially smaller ones, might
be missed for statistical reasons. Nevertheless, among the 348 genes that appeared to
affect the RED biosynthesis under the conditions employed, 288 had not previously been
identified as antibiotic modulators. Seventy encoded hypothetical and unknown proteins,
so it is hard to infer their modes of action at this stage. Nevertheless, our data provide a
more integrative/global view on regulation/modulation of antibiotics biosynthesis than
has previously been possible, presenting many new starting points for future
investigation of the complex regulatory mechanism controlling RED production. To maximize confidence in this new perspective, the following discussions were based on the genes represented more than once in the library screening (see Tables S5 in the supplemental material for details).

Mini-Tn5 mutants of eight genes involved in the synthesis of branched-chain amino acids affected RED production. Mutants of SCO3345 (ilvD), SCO2528 (leuA), and SCO5529 (leuA2) increased RED production, while mutants of SCO5553 (leuC) decreased RED production. These observations may reflect the balance between the dual roles of branched-chain amino acids as building blocks of proteins and precursors of fatty acid biosynthesis, the latter being a major pathway for acetyl-CoA consumption in Streptomyces needed both for growth and as an important precursor supply for RED biosynthesis.

Mini-Tn5 mutants in genes encoding homologues of bifunctional DNA primase/polymerase (SCO5581), recombination endonuclease VII (SCO1645), and transcriptional-repair coupling factor Mfd (SCO3109) increased RED production. Although the roles of these genes in streptomycetes have not been established, they seem to be dispensable for normal growth and development. However, it is very likely that they are involved in DNA metabolism, suggesting a link between antibiotic production and DNA metabolism. This could either reflect competition for nutrients and energy [eg. enhancing ATP supply increased RED production in Serratia marcescens (67)], or imply cross-talk between growth-related DNA metabolism involving these genes,
and the establishment of secondary metabolism in non-growing cells [DNA is replicated mainly in hyphal tips, while it has been suggested that secondary metabolism takes place in non-growing sub-apical compartments (65)].

The membrane proteins SCO4126-4131, named CmdA-F, are thought to form a type IV-like DNA translocation system on the cell membrane, based on homology of SCO4127 (CmdB) to type IV DNA transfer motor proteins, and the finding that \( \Delta cmdABCDEF \) mutants were deficient in chromosome segregation during sporulation; and it was noted that the mutants were dark blue compared to the wild type strain, indicating overproduction of ACT (68). In our study, mini-Tn5 insertions into SCO4127-4131, and into SCO5677 encoding a somewhat SCO4127-like putative DNA transfer protein, increased RED dramatically. In addition, both SCO4127-4131 and SCO5677 mutants were deficient in aerial mycelium formation and had a cracked colony surface. We propose that these DNA transfer proteins may transfer DNA into newly formed aerial hyphae initials, and that in the SCO4127/SCO5677 mutants, newly formed aerial hyphae initials fail to receive chromosomes, leading to aerial growth arrest. Cellular nutrients in the “old hyphae” may then be assigned to antibiotic production. Actually, “empty” new hyphal branch initials have been observed in Streptomyces, though in vegetative growth (69). Interestingly, mutants of tgdA (SCO4132, for secreted lytic transglycosylase homologue) showed increased RED production to the same level as the cmdB-F and SCO5677 mutants. In addition, the tgdA mutants resembled the cmdB-F and SCO5677 mutants in terms of the cracked and bald colony surface. Just as
for cmdB-F and SCO5677, we therefore speculate that SCO4132 takes part in the
initiation of new hyphal branches, particularly for the formation of aerial hyphae.

Functions concerned with the maintenance of membrane integrity are generally under
the control of an ECF sigma factor, σE. In further support for an interplay of membrane
function and antibiotic production, insertions in four genes belonging to the σE regulon
(70) had diverse but significant effects: SCO3404 (ftsH2, for ATP-dependent
metalloprotease) mutants were RED-decreasing and mutants of SCO2067 and
SCO3855 (for homologues of protein-disulfide isomerase DsbA) were RED-increasing.

Many previous reported genes involved in stress response, signal transduction, and
transcriptional regulation were identified again to modulate antibiotic production, such as
c-diGMP signaling (SCO0928) (57), ppGpp signaling (SCO1513) (58), cAMP signaling
(SCO4928) (71, 72), and transcription regulators (wblA, sarA, rrdA and cutRS) (15,18,
20, 60). Some of them regulate antibiotic production by changing the expression of the
biosynthetic gene clusters (eg. nsdA and wblA) (54, 73) or the activation of substrate (eg.
SCO1596) (59)

Besides influencing RED production, some of these genes, including SCO4127-4131
(cmdB-F), SCO5677, SCO3579 (wblA), SCO2999, and SCO4069 (sarA), have been
reported to control the production of other antibiotics such as actinorhodin (20, 32, 54, 68,
74), implying that they play global/pleiotropic roles in modulating secondary metabolism,
and may be of future use in industrial strain improvement or activation of silent
biosynthetic gene clusters that may lead to the discovery of new pharmaceuticals. In fact,
orthologs of SCO3579 (wblA) have been used in strain improvement in *Streptomyces*
ghanaensis and Streptomyces peucetius (75, 76). The additive effects of different
modulators for antibiotic production have been applied in engineering optimised hosts for
heterologous expression of antibiotic pathway gene sets (77). For improved antibiotic
production, it is likely to be favorable to combine mutations in genes that belong to
different functional classifications. It should also be noted that not all RED modulatory
genes will also affect production of other antibiotics, both because of possible specificity
in the target profiles of regulatory proteins, and because different pathways may use
different precursors.

Inserts located in non-coding regions may influence the expression of small
non-coding RNAs (ncRNAs), which can be crucial regulatory elements in gene
expression, mRNA stability, translation and protein functions. Such effects have been
suggested for antibiotic production in Streptomyces, many non-coding RNAs being
specified within antibiotic biosynthesis clusters (78). Taking into account available data
about ncRNAs in S. coelicolor (78, 79, 80), we identified two inserts in intergenic sRNA
regions (scr3437, scr3580) that influenced RED production. In addition, two inserts in the
intergenic region between SCO3842 and SCO3843 that determines a tRNA-Leu showed
increased RED production; and two inserts in the intergenic region between SCO4231
and SCO4232 showed decreased RED production, even though no insert in SCO4231
and SCO4232 was found to affect RED production, indicating a independent RED
regulatory function of this intergenic region. More evidence will be needed to prove the
functions of these non-coding regions for RED production.

ACKNOWLEDGEMENTS
This work was supported by Ministry of Science and Technology (863: 2010AA10A201; 2016YFE0101000), Ministry of Education (NCET-08-0778-533-09001), and National Natural Science Foundation (31170084) of China, and the Science and Technology Commission of Shanghai Municipality (15JC1400401). We thank Tobias Kieser, Qiannan Hu, and Xiuhua Pang for useful discussion.

REFERENCES


http://dx.doi.org/10.1038/nrmicro3033.


http://dx.doi.org/10.1128/iai.00156-10.


http://dx.doi.org/10.1101/gr.1710304.

ordered cosmid library for the model organism *Streptomyces coelicolor* A3(2).


http://dx.doi.org/10.1007/s10482-010-9518-1.


http://dx.doi.org/10.1073/pnas.170059797.


http://dx.doi.org/10.1007/s00253-009-2047-z.


http://dx.doi.org/10.1111/1574-6968.12117.


http://dx.doi.org/10.1007/s00253-012-4550-x.


http://dx.doi.org/10.1016/j.mimet.2011.10.018.


70. Danese PN, Silhavy TJ. 1997. The sigma(E) and the Cpx signal transduction systems control the synthesis of periplasmic protein-folding enzymes in *Escherichia coli*. Genes Dev 11:1183-1193.


differentiation in *Streptomyces coelicolor* A3(2). J Bacteriol **188**:8368-8375. 

http://dx.doi.org/10.1128/JB.00933-06.


### Table 1 Plasmids and strains

<table>
<thead>
<tr>
<th>Plasmid and Strain</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pHL734</td>
<td>Tn5-based transposon vector; <em>Amp′, Apr′</em></td>
<td>This work</td>
</tr>
<tr>
<td>pMM1</td>
<td>Cosmid containing the <em>act</em> gene cluster; <em>Apr′</em></td>
<td>(37)</td>
</tr>
<tr>
<td>pHL664</td>
<td>pMM1 derivative with the 22-kb <em>act</em> gene cluster deleted with Stul; <em>Apr′</em></td>
<td>This work</td>
</tr>
<tr>
<td>pOJ260</td>
<td><em>E. coli-Streptomyces</em> conjugation transfer vector, <em>Apr′</em></td>
<td>(38)</td>
</tr>
<tr>
<td>pHL737</td>
<td>pOJ260 derivative, for the deletion of the entire <em>act</em> gene cluster in <em>S. coelicolor</em> M145; <em>Apr′</em></td>
<td>This work</td>
</tr>
<tr>
<td>pSET152</td>
<td><em>Streptomyces</em> integrative vector, <em>Apr′</em></td>
<td>(38)</td>
</tr>
<tr>
<td>pMT3</td>
<td>pSET152 derivative, contains thiostrepton resistance gene (<em>tsr</em>); <em>Thio′</em></td>
<td>This work</td>
</tr>
<tr>
<td>pHXZ097</td>
<td>pMT3 harboring <em>SCO1104</em></td>
<td>This work</td>
</tr>
<tr>
<td>pHXZ098</td>
<td>pMT3 harboring <em>SCO1289</em></td>
<td>This work</td>
</tr>
<tr>
<td>pHXZ100</td>
<td>pMT3 harboring <em>SCO2184</em></td>
<td>This work</td>
</tr>
<tr>
<td>pHXZ103</td>
<td>pMT3 harboring <em>SCO3008</em></td>
<td>This work</td>
</tr>
<tr>
<td>pHXZ104</td>
<td>pMT3 harboring <em>SCO3269</em></td>
<td>This work</td>
</tr>
<tr>
<td>pHXZ105</td>
<td>pMT3 harboring <em>SCO3579</em></td>
<td>This work</td>
</tr>
<tr>
<td>pHXZ106</td>
<td>pMT3 harboring <em>SCO3664</em></td>
<td>This work</td>
</tr>
<tr>
<td>pHXZ108</td>
<td>pMT3 harboring <em>SCO4025</em></td>
<td>This work</td>
</tr>
</tbody>
</table>
pHXZ109  pMT3 harboring SCO4215  This work
pHXZ112  pMT3 harboring SCO5264  This work
pHXZ113  pMT3 harboring SCO6118  This work
pHXZ114  pMT3 harboring SCO6265  This work
pHXZ115  pMT3 harboring SCO7733  This work

*Escherichia coli*

DH5α  cloning host  Invitrogen

ET12567/ pUZ8002  *dam, dcm, hsdM, hsdS, hsdR, cat, tet,*  helper strain for intergeneric conjugation;  

*Cm’, Km’* 

*Streptomyces coelicolor*

M145  Prototrophic plasmid-free derivative of wild-type strain A3(2)  (2)

X737  M145 derivative without *act* gene cluster  This work

<table>
<thead>
<tr>
<th>Name</th>
<th>Primer sequence (5’→3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEap-F</td>
<td>GAGAGGGCCCTGTCTTTATACATCTTTAGCTGCTGTTTCTCG</td>
</tr>
<tr>
<td>MEap-R</td>
<td>GAAGGGCCCGCTAGCTTATACATCTTTAGCTGCTGCTGCTG</td>
</tr>
<tr>
<td>oriT-F</td>
<td>AGGGCCCATCAACTACGTCAGTGGGCACCTTTTCG</td>
</tr>
<tr>
<td>oriT-R</td>
<td>AGGGCCCATCAACTACGTCAGTGGGCACCTTTTCG</td>
</tr>
<tr>
<td>tsr-F</td>
<td>GGGCCGATCAAGGGCGAATTACCTTACTATG</td>
</tr>
</tbody>
</table>

Table 2 Primers used in this study
<table>
<thead>
<tr>
<th>tsr-R</th>
<th>CGCCGGCGTCCGAGGAACAGAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>act-up</td>
<td>CGGCCTGCTGCGCGAGGACATCCA</td>
</tr>
<tr>
<td>act-down</td>
<td>CCGGAGGCAGTGCTGAGCAGTCCG</td>
</tr>
<tr>
<td>DownS</td>
<td>ACTGCTGAGGCTTTTCCTGCGTC</td>
</tr>
<tr>
<td>UpS</td>
<td>GAGTTAGCTCAGTCCATAGAG</td>
</tr>
<tr>
<td>ADTn1</td>
<td>pGATCGGAAGAGCGGTCAAGGCT</td>
</tr>
<tr>
<td>ADTn2</td>
<td>AGTTCTCCAGGTGCTTACGGCT</td>
</tr>
<tr>
<td>AD-F</td>
<td>CCTGCTGAAAGGGCTCTTGCCT</td>
</tr>
<tr>
<td>INME-1bio</td>
<td>bio-ACAGGAAACAGCTATGACAT</td>
</tr>
<tr>
<td>INME-2bio</td>
<td>bio-CCGCTGCCAGTCGATTTGG</td>
</tr>
<tr>
<td>INME-1</td>
<td>ACAGGAAACAGCTATGACATGAAGA</td>
</tr>
<tr>
<td>INME-2</td>
<td>CCGCTGCCAGGCAGTCGATTTGGCT</td>
</tr>
<tr>
<td>SCO1104-F</td>
<td>GCTGATGCGCGGACAGGACCGG</td>
</tr>
<tr>
<td>SCO1104-R</td>
<td>TCAACACGCAGACGCAGCGCCAG</td>
</tr>
<tr>
<td>SCO1289-F</td>
<td>CCCAGCGGATTTCTGAGGATGC</td>
</tr>
<tr>
<td>SCO1289-R</td>
<td>CGGTCGGCTGTTTCCGAGGACCGG</td>
</tr>
<tr>
<td>SCO2184-F</td>
<td>CGACTACATGCGCCAGGTCCCGC</td>
</tr>
<tr>
<td>SCO2184-R</td>
<td>CGGCCAACGGCACCAGCCAGCTCA</td>
</tr>
<tr>
<td>SCO3008-F</td>
<td>CGGACAGATGGCTACGTCCCGG</td>
</tr>
<tr>
<td>SCO3008-R</td>
<td>CGGGCTCAGGGCAGGTCCCGCCC</td>
</tr>
<tr>
<td>SCO3269-F</td>
<td>TAAGGTCGGCGCCGTCAAGCTGG</td>
</tr>
<tr>
<td>SCO3269-R</td>
<td>TACCGCCAGCGCGAGGACTCAA</td>
</tr>
<tr>
<td>SCO3579-F</td>
<td>CGGCGTTGGGTACGTGCTCCGA</td>
</tr>
<tr>
<td>Accession</td>
<td>Primer Forward</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------</td>
</tr>
<tr>
<td>SCO3579-R</td>
<td>GGGCCCCTGGTCCGGCCTCGGGG</td>
</tr>
<tr>
<td>SCO3664-R</td>
<td>AGTGCTGTGGCAGGCTCCCGCCCCTCGTC</td>
</tr>
<tr>
<td>SCO4025-F</td>
<td>CGGTTCGCGGGGCTGGCTCCCGT</td>
</tr>
<tr>
<td>SCO4025-R</td>
<td>GCGCCCGGTCCGGGTAGGGGCACC</td>
</tr>
<tr>
<td>SCO4215-F</td>
<td>CGGTTCGCGGGGCTGGCTCCCGT</td>
</tr>
<tr>
<td>SCO4215-R</td>
<td>TGGGGCGAGACCAGCTCAGTCAC</td>
</tr>
<tr>
<td>SCO5264-F</td>
<td>CAGCCGCTTCCGGAGACCCCTTCGGG</td>
</tr>
<tr>
<td>SCO5264-R</td>
<td>AGCCGGGGGGCCGGCTCCGGGTT</td>
</tr>
<tr>
<td>SCO6118-F</td>
<td>CGATGGCCGCGCGGACCCCCGGG</td>
</tr>
<tr>
<td>SCO6118-R</td>
<td>GCTCCGCGTCAGGGCTCCGGGTTG</td>
</tr>
<tr>
<td>SCO6265-F</td>
<td>GATGGCGAGGCGAGCTCGAGTCAC</td>
</tr>
<tr>
<td>SCO6265-R</td>
<td>TGGGGCGAGACCAGCTCAGTCAC</td>
</tr>
<tr>
<td>SCO7733-F</td>
<td>CCGCCGCGCAGGCAAGGCAAGGAC</td>
</tr>
<tr>
<td>SCO7733-R</td>
<td>ACAGGCTCATCGCCGGCTCTAC</td>
</tr>
</tbody>
</table>
**Figures and legends**

**Figure 1** Transposon mutagenesis of *S. coelicolor* by the Tn5 transposon delivery vector pHL734. (A) pHL734 and the mini-Tn5 transposable unit. ME, the mosaic end sequence for Tn5 transposase recognition and cutting; ori-pUC, origin for DNA replication in *E. coli*; aac(3)IV, apramycin resistance gene; oriT, origin for intergeneric conjugation from *E. coli* to *S. coelicolor*; bla, ampicillin resistance gene; trp(5), encoding a Tn5-transposase variant with five point mutations (E54K, M56K, L372P, P242A, and E345K). (B) Verification of randomly selected transposition mutants by Southern hybridization with the mini-Tn5 DNA probe. A single and unique band is observed for each mutant. M, DNA molecular weight marker. (C) Selected transposition mutants derived from *S. coelicolor* strains M145 (WT) or X737 (*act* deletion) with alterations in the production of ACT (blue) and/or RED (red) (*OV* = overproduction). Actinorhodin (ACT) blue color is a dominant phenotype of M145 while undecylprodigiosin (RED) red color is the definitive phenotype of X737 and its derivatives under the same growth conditions.
Figure 2 Characterization of genome-wide Tn5 transposition in *S. coelicolor* X737. (A) Distribution of Tn5 insertions along the chromosome. The plots were generated by a sliding window method with the window size set at 1 kb and the step set at 0.2 kb. The core region of the chromosome is indicated by black segment while the left and right arms are in gray. oriC, origin of replication of the chromosome. Forward, the mini-Tn5 insertions run from 5' to 3' with the top strand of *S. coelicolor* chromosome; Reverse, the opposite direction of mini-Tn5 insertions in the chromosome. Insertion hot regions H1-H26 and H1'-H26' are marked. Each hot region has more than 16 insertions/kb in one direction. All hot regions, except for peaks H1 and H1', are located in the chromosomal core. (B) Statistic of small insertional gaps of the X737 transposition library. A gap between two adjacent Tn5 insertions was defined as an insertional gap. Gap size, the
distance between two adjacent insertions. When the numbers of small insertional gaps were counted individually, there were more 0-, 5-, 10-, 15-bp gaps than for other gap sizes. (C) Statistical comparison of actual insertional gap sizes of the X737 transposition library with simulated gap sizes of simulated libraries generated by random Monte Carlo method. According to their sizes, insertional gaps were grouped into bins of every 100 bp and the number of gaps in each bin was counted. Actual gaps, gaps of the 51,443 detected insertions sites. Simulated gaps, gaps of 51,443 simulated insertions. The simulation was repeated 200 times and the average numbers of gaps were plotted. (D) Boxplots of the insertion frequencies in chromosomal DNA regions of different G+C contents. G+C contents were counted from every 2 kb DNA. The transverse lines in the inter-quartile range (IQR) boxes are the medians. The IQR boxes indicate the 25th and 75th percentile respectively. The whiskers extend ± 1.5×IQR. Small black dots (•) represent outliers. The average hit/2-kb is indicated by the dashed gray line.
Figure 3 Base composition at the Tn5 insertion site and its flanking sequences compiled from 51,443 insertions. (A) Compositions of AT and GC at the Tn5 insertion site and its flanking sequences compiled from 51,443 insertions. (B) Compositions of AG (pyrimidine) and CT (purine) at the Tn5 insertion site and its flanking sequences compiled from 51,443 insertions. The insertion site corresponds to 0. The black double arrows indicate the 9-bp target site (-4 to 4). The dashed double arrows indicate the 19-bp expanded target site (-9 to 9).
Figure 4 Distribution of Tn5 insertions affecting RED production along the S. coelicolor chromosome. (A) RED biosynthetic pathway. Both products of this pathway, undecylprodigiosin and streptorubin B, are red-colored and indistinguishable in this study, therefore are regarded as RED. (B) Tn5 insertions in the red biosynthetic gene cluster. Filled arrows are red genes essential for the production of RED. Tn5 insertions are shown by vertical short lines at the insertion sites with one or two hits (mutants). (C) Distribution of 234 RED-decreasing insertions outside the red gene cluster. The 107 hits within the red gene cluster are not included. (D) Distribution of 383 RED-increasing insertions. The (C) and (D) plots were generated by a sliding window method with the window size set at 200 kb and the step set at 20 kb. The core region of the chromosome is indicated by the black segment while the left and right arms are in gray. oriC, origin of
replication of the chromosome. Peaks with more than ten insertions in single genes are emphasized.

Figure 5 Relative RED production in RED-changing regulatory gene mutants and recombinants carrying the wild type alleles. The wild type regulatory gene alleles, including a 300-bp upstream fragment, were cloned into an integrative vector and introduced into the corresponding mutants for genetic complementation. Antibiotic production levels in S. coelicolor X737 (WT) and X737/vector are used as references. Data are mean of three experiments. Error bars are standard deviations. Partial, RED reduction was partially restored. Y, yes, RED production was fully restored to the level of the wild type strain. The selected genes were newly identified by this study except for SCO3579 (wblA).