

SCIENTIFIC REPORTS



OPEN

Comparative Genomic and Regulatory Analyses of Natamycin Production of *Streptomyces lydicus* A02

Huiling Wu¹, Weicheng Liu¹, Lingling Shi¹, Kaiwei Si³, Ting Liu¹, Dan Dong¹, Taotao Zhang¹, Juan Zhao¹, Dewen Liu¹, Zhaofeng Tian¹, Yuesen Yue², Hong Zhang¹, Bai Xuelian¹ & Yong Liang³

Streptomyces lydicus A02 is used by industry because it has a higher natamycin-producing capacity than the reference strain *S. natalensis* ATCC 27448. We sequenced the complete genome of A02 using next-generation sequencing platforms, and to achieve better sequence coverage and genome assembly, we utilized single-molecule real-time (SMRT) sequencing. The assembled genome comprises a 9,307,519-bp linear chromosome with a GC content of 70.67%, and contained 8,888 predicted genes. Comparative genomics and natamycin biosynthetic gene cluster (BGC) analysis showed that BGC are highly conserved among evolutionarily diverse strains, and they also shared closer genome evolution compared with other *Streptomyces* species. Forty gene clusters were predicted to involve in the secondary metabolism of A02, and it was richly displayed in two-component signal transduction systems (TCS) in the genome, indicating a complex regulatory systems and high diversity of metabolites. Disruption of the *phoP* gene of the *phoR-phoP* TCS and *nsdA* gene confirmed phosphate sensitivity and global negative regulation of natamycin production. The genome sequence and analyses presented in this study provide an important molecular basis for research on natamycin production in *Streptomyces*, which could facilitate rational genome modification to improve the industrial use of A02.

Streptomyces lydicus A02 is a Gram-positive filamentous actinomycete that produces natamycin (also known as pimaricin), which is the only member of the polyene antibiotic family that does not exert its antifungal action by forming pores and permeabilizing the plasma membrane^{1,2}. It inhibits the growth of fungi via the immediate inhibition of amino acid and glucose transport across the plasma membrane. Due to its broad spectrum of antifungal activity and naturally occurring fungal resistance to natamycin being exceptionally rare, natamycin is widely used in the treatment of fungal keratitis, as a food preservative and as an antifungal agent in agriculture^{1,3,4}. In addition to *S. lydicus*, three other species of *Streptomyces* are known to produce natamycin, i.e., *S. natalensis*, *S. gilvosporeus*, and *S. chattanoogensis*^{3,5,6}. Despite extensive studies on their natamycin biosynthetic pathways, the underlying mechanisms of natamycin production and the regulatory role at the genomic level in these *Streptomyces* species remain unclear.

The natamycin biosynthetic gene cluster (BGC) was first characterized in *S. natalensis* ATCC 27448³. It contains 5 giant multi-enzyme proteins (PimS0-PimS4) and 11 enzymes involved in polyene tailoring or export (PimA-PimK). Additionally, a series of regulators (PimR, PimM, and PimT) are associated with the biosynthesis of natamycin⁷⁻¹⁰. In 2011, the natamycin BGC of *S. chattanoogensis* L10 was characterized¹¹. A comparative cluster analysis of the two strains revealed the presence of different architectures and indicated the occurrence of the evolution of antibiotic BGCs.

Several studies have reported the role of various regulators in natamycin biosynthesis, such as the pathway-specific regulators PimM (ScnRII in *S. chattanoogensis* and SlnM in *S. lydicus* A02)^{1,9,12}, PimR⁸,

¹Institute of Plant and Environment Protection, Beijing Academy of Agriculture and Forestry Sciences, Beijing, 100097, China. ²Beijing Research and Development Center for Grass and Environment, Beijing Academy of Agriculture and Forestry Sciences, Beijing, 100097, China. ³BGI-Shenzhen, Shenzhen, Guangdong, 518083, China. Correspondence and requests for materials should be addressed to W.L. (email: liuwich@163.com) or Y.Y. (email: yusen2008@sina.com)

Genome Size (bp)	9,307,519
GC Content (%)	70.67
Gene Number (#)	8,888
Total Length (bp)	8,035,014
Gene Average Length (bp)	904
Gene Length/Genome (%)	86.33
GC Content in Gene Region (%)	68.15
Intergenic Region Length (bp)	1,272,505
GC Content in Intergenic Region (%)	68.15
Intergenic Region Length/Genome (%)	13.67
Tandem Repeat Number	2,217
Total Length (bp)	99,927
Range of Repeat Size (bp)	4-828
Tandem Repeat Length/Genome (%)	1.0736
Minisatellite DNA Number	1,451
Microsatellite DNA Number	247
rRNA Number	18
tRNA Number	68
sRNA Number	14

Table 1. Characteristics of the *S. lydicus* A02 genome assembly using single-molecule real-time sequencing.

and PimT¹⁰; the phosphopantetheinyl transferases¹³; the pleiotropic regulator SngA¹⁴; the *phoR-phoP* two-component system (TCS)¹⁵; the pleiotropic regulator AdpAch¹², and the γ -butyrolactone autoregulator receptor SngR¹⁶. Cloning of new regulatory genes and exploration of novel regulatory mechanisms are still ongoing; however, the complete genome sequence of a natamycin-producing strain has not been reported to date.

The wild-type strain A02 displayed obvious differences in morphology, pigments and natamycin production compared to *S. natalensis* ATCC 27448 and *S. chattanoogaensis* NRRL B-2255. A higher natamycin-producing capacity than that of *S. natalensis* ATCC 27448 and *S. chattanoogaensis* NRRL B-2255 was detected in A02 (Supplementary Fig. S1). We have constructed a series of engineered strains, including overexpression of the *slnM* gene with different promoters (ermEp* promoter, native promoter and dual promoters), heterologous expression of *Vitreoscilla* haemoglobin and coexpression of *Vitreoscilla* haemoglobin and *Bacillus megaterium* glucanase, to improve the natamycin yield or antifungal metabolites in strain A02^{1,17}.

In this study, we completed the whole-genome sequencing of *S. lydicus* A02 and extended these studies to the genomes of eight other *Streptomyces* species – *S. venezuelae* ATCC 10712, *S. bingchenggensis* BCW-1, *S. auratus* AGR0001, *S. avermitilis* MA-4680, and *S. coelicolor* A3(2) – each of which has more than 400 proteins matched to the A02 genome according to the gene annotation. *Streptomyces lydicus* 103 has not been reported to be able to produce natamycin, but it is the only corresponding species genome available from public databases¹⁸. The natamycin-producing strains are *S. natalensis* ATCC 27448 and *S. chattanoogaensis* NRRL ISP5002, both genome sequences of which were recently produced by whole-genome shotgun sequencing.

The phylogenetic trees were constructed based on 16 S rRNA, the core-pan genome and gene families. We identified the gene synteny across the whole genome between A02 and the other eight *Streptomyces* species and provide a comparative analysis of natamycin BGC in three natamycin-producing strains. The role of *phoR-phoP* TCS and the global negative regulatory gene *nsdA* in natamycin production was also evaluated. The genome sequence and analyses presented in this study provide important evidence of the molecular basis for further research on natamycin and other valuable secondary metabolites produced by strain A02.

Results and Discussion

Genome sequencing and assembly of *S. lydicus* A02. Initially, a total of 2,507 Mb of raw data were produced from the HiSeq. 2000 sequencing platform, from 43,361 reads. The assembly consisted of 1 contig of 9,300,345 bp (Supplementary Table S1). After single-molecule real-time sequencing (SMRT) and data clean-up, 593.1 Mb of sequence data were obtained, with an average sub-read length of 7.3 kb (Supplementary Table S2). During preassembly, 4 kb was chosen as the cutoff for recruiting seed reads, and 383 Mb of seed reads were generated. After read correction, a 263 Mb preassembly with an average read length of 6.6 kb was obtained (Supplementary Fig. S2). The final assembly was composed of a chromosomal scaffold of 9,307,519 bp (70.67% G + C) (Table 1, Supplementary Fig. S3) without any plasmid.

Genome annotation and general features of *S. lydicus* A02. There are 8,888 protein CDSs accounting for 86.33% of the complete genome. A total of 3,921 putative proteins were matched to the genome sequence of *S. auratus* AGR0001, which produces neophoslactomycin A and possesses potent biocontrol activity against certain phytopathogenic fungi¹⁹. Additionally, 864 hypothetical proteins have no match to any known protein databases. More than 400 putative proteins were matched to five different *Streptomyces* species, *S. venezuelae* ATCC 10712, *S. bingchenggensis* BCW-1, *S. auratus* AGR0001, *S. avermitilis* MA-4680, and *S. coelicolor* A3(2). A

Species	Genome (Mp)	GC content	Gene number	Average CDS (bp)	Plasmids	tRNA	rRNA operon
<i>S. lydicus</i> A02	9.3	70.67%	8,888	904	0	68	6
<i>S. auratus</i> AGR0001	7.85	71.5%	7,102	NA	0	66	8
<i>S. bingchenggensis</i> BCW-1	11.94	70.8%	10,107	1,031	0	66	6
<i>S. natalensis</i> ATCC 27448	8.65	70.6%	7,564	NA	0	70	3
<i>S. avermitilis</i> MA-4680	9.03	70.7%	7,582	1,027	1	68	6
<i>S. hygroscopicus</i> 5008	10.15	71.9%	8,849	952	2	68	6
<i>S. coelicolor</i> A3	8.67	72.1%	7,825	991	2	63	6
<i>S. cattleya</i> 8057	8.09	72.99%	7,580	NA	1	64	6
<i>S. venezuelae</i> 10712	8.23	72.4%	7,536	NA	0	67	7
<i>S. pristinaespiralis</i> 25486	8.13	69.7%	7,025	NA	0	65	7
<i>S. chattanoogensis</i> NRRL ISP-5002	9.13	70.9%	8031	NA	0	67	10
<i>S. lydicus</i> 103	8.2	NA	7128	NA	0	68	7

Table 2. Comparison of key genomic features among the twelve fully assembled genomes of *Streptomyces*.

total of 68 tRNA-encoding genes and six rRNA operons (six 5 S rRNAs, six 23 S rRNAs and six 16 S rRNAs) were found in the linear chromosome.

A comparative analysis was performed on the basic genome features of A02 and 11 previously sequenced *Streptomyces* strains (Table 2). The genome of *S. lydicus* A02 is smaller than the previously sequenced linear genome of *S. bingchenggensis* BCW-1 (11.9 Mbp) and *S. hygroscopicus* 5008 (10.15 Mbp) but larger than those of other *Streptomyces* strains^{19–23}. However, it has a gene number that is lower than that of only *S. bingchenggensis* BCW-1. The genome sequence of *S. lydicus* A02 has a G + C content that is comparable to that of *S. avermitilis*, which is higher than *S. pristinaespiralis* 25486 and *S. natalensis* ATCC 27448 but lower than the other strains. Among all these strains, *S. lydicus* A02 has the smallest average CDS size (904 bp).

The distribution of genes in *S. lydicus* A02 into COG functional categories is presented in Supplementary Table S3. CRISPRs compose a type of immune system that is widespread in bacteria and archaea to protect cells against foreign genetic elements (e.g., plasmids and phages)²⁴. Three credible and several questionable CRISPR loci were detected in the chromosome of strain *S. lydicus* A02. The components of the CRISPR elements were identified, and we conducted comparative analysis of the repeats and spacers with those identified in the GenBank database, prediction of the leader region in the CRISPR locus and Cas genes and related analyses (Supplementary Fig. S4).

Phylogenetic analyses of streptomycetes. To infer the phylogenetic relations of A02 and other *Streptomyces* strains, we constructed phylogenetic trees based on 16 S rRNA, gene family and core-pan genome. In the phylogenetic tree based on the 16 S rRNA gene sequences (Fig. 1A), 10 strains were primarily grouped into two subgroups. Strain A02 was positioned into a single clade that is equally closely related to *S. lydicus* 103 and *S. chattanoogensis* NRRL ISP5002, which were tightly grouped. Two other natamycin producing strains, *S. natalensis* ATCC 27448 and *S. gilvosporeus* ATCC 13326, were classified into the same group. However, the internal branches of the tree that were based on the 16 S rRNA gene sequences had low bootstrap support²⁵. Single-gene phylogenies might not always reflect the evolutionary history of a species due to the high degree of lateral gene transfer²⁶. We also reconstructed the phylogenetic trees using the core-pan genome and gene families (Fig. 1B). Both the trees exhibited similar identities, with larger bootstrap scores and higher robustness than that of 16 S rRNA. *S. gilvosporeus* was not included in this tree because its genome sequence is not available in any public database. The tree of the core-pan genome and gene families suggested that all the *Streptomyces* strains were clustered within one clade except *S. avermitilis* MA-4680. Strains A02 and *S. chattanoogensis* NRRL ISP5002 are tightly grouped, and they are equally closely related to *S. natalensis* ATCC 27448. *S. auratus* AGR0001 and *S. lydicus* 103 were classified into the same group. The MUMmer-based whole-genome alignment revealed that the sequence of strain A02 aligned at a relatively high level with *S. chattanoogensis* NRRL ISP5002, showing minimal genomic rearrangements when compared with other *Streptomyces* species (Fig. 1C). The average nucleotide identity (ANI) for 9 *Streptomyces* genome sequences was calculated using the BLAST algorithm and MUMmer alignment using the software tool JSpecies²⁷. The ANI value between A02 and *S. chattanoogensis* NRRL ISP5002 was 97.4% (Supplementary Table S4), indicating that the two strains share the closest phylogenetic relations. The ANI value between A02 and other *Streptomyces* suggested similar phylogenetic relations between A02 and other *Streptomyces* with the gene families and core-pan genome. The phylogenetic analyses of streptomycetes based on the core-pan genome, gene families and ANI value showed that *Streptomyces* species that can produce natamycin are more frequently associated with genetic recombination, which may reduce the divergence of these bacteria.

Core and pan-genome analyses. The pan-genome defines the entire genomic repertoire of a given phylogenetic clade and encodes for all possible lifestyles of its organisms, including the core genome, dispensable genome and strain-specific genes²⁸. To understand the genetic composition of the 9 streptomycetes strains in the core-pan genome more thoroughly, we clustered all 69,446 protein CDSs in 9 streptomycetes strains. All the protein CDSs were clustered into 15,404 orthologues; 5,047 (32.8%) orthologues were identified in 9 strains as the streptomycetes core genome (Fig. 2A). The 5,279 orthologues were identified as dispensable genomes, and 5,087 genes were strain-specific. *S. chattanoogensis* NRRL ISP5002 had the smallest number of strain-specific

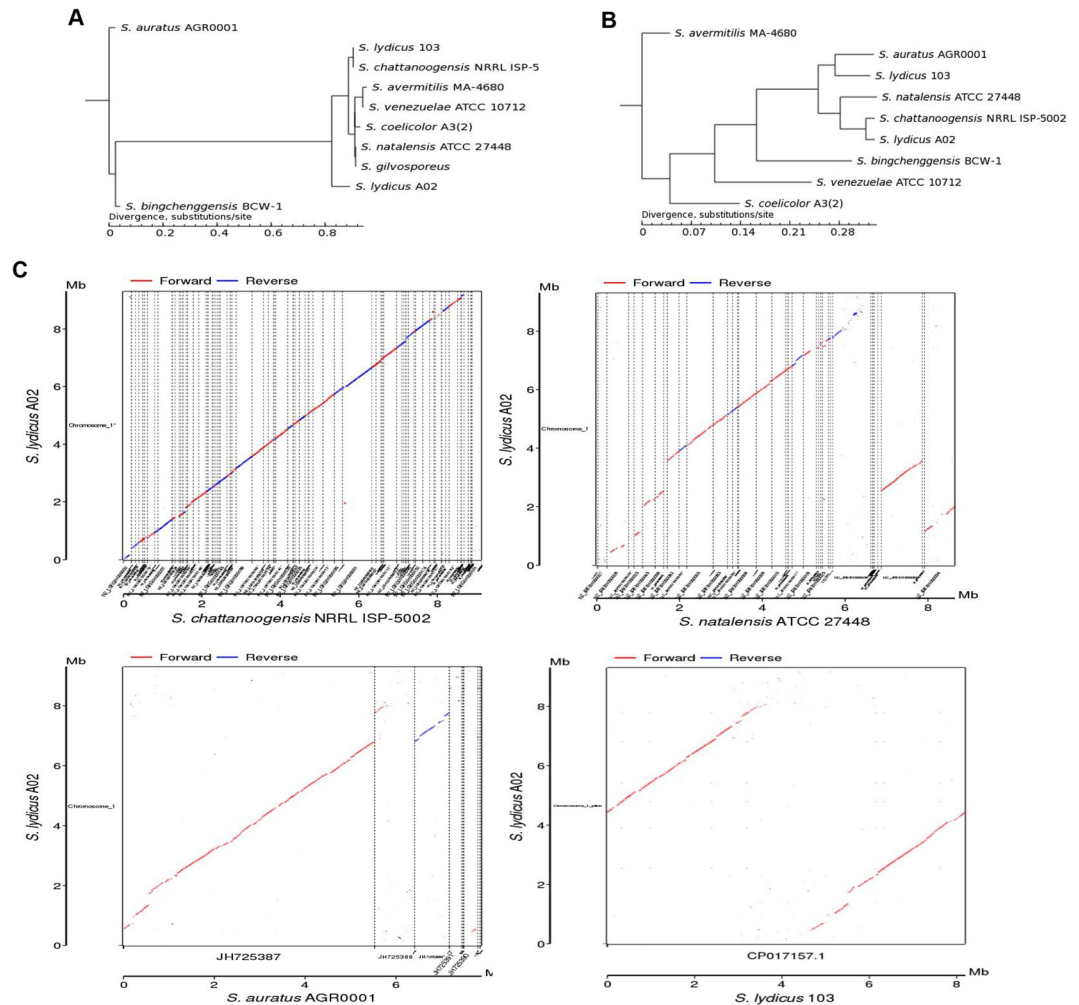


Figure 1. Phylogenetic trees for streptomycete. The phylogenetic tree is constructed by TreeBeST using PHYML based on (A) the 16S rRNA gene and (B) the phylogenetic tree analysis's result based on SNP analysis (core-fan genome and gene family).

genes because it has the closest phylogenetic relationship with A02 based on the gene family, core-pan genome and whole-genome alignment. *S. lydicus* 103 had a smaller number of strain-specific genes because of the phylogenetic relationship between *S. lydicus* 103 and A02. *S. auratus* AGR0001, as the strain with the smallest genome size in these *Streptomyces* species, matched the largest number of genes to the *S. lydicus* A02 genome but also displayed a smaller number of strain-specific genes. However, *S. bingchenggensis* BCW-1 had the largest number of strain-specific genes, perhaps because it has the largest genome size and gene numbers, although it does not have the greatest difference from the *S. lydicus* A02 genome.

To understand the relationship between the streptomycetes pan-genome size, the core genome number and the strain number, we plotted the pan-genome profile fitted curves of the nine *Streptomyces* strains. As shown in Fig. 2B, we could intuitively observe that the more genomes we added, the more new orthologue clusters were discovered, implying an open pan-genome of these *Streptomyces* strains. The number of core genomes would be expected to converge to a constant value, as judged from the slope of exponential decay (Fig. 2C). A heat map illustrated the distribution of the genes after core gene deletion (Supplementary Fig. S5).

Comparative cluster analysis of natamycin BGCs. The natamycin BGC of *S. lydicus* A02 contains 19 open reading frames (ORFs) putatively involved in natamycin biosynthesis. It contains all the ORFs reported in *S. chattanoogensis* L10, including 5 polyketide synthase genes (*slnS0*, *slnS1*, *slnS2*, *slnS3*, and *slnS4*), 11 post-tailoring genes (*slnA*, *slnB*, *slnC*, *slnD*, *slnE*, *slnF*, *slnG*, *slnH*, *slnI*, *slnJ* and *slnK*), and 2 regulator genes (*slnM* and *slnR*) (Fig. 3A).

The product of the *slnM* has 100% identity to ScnRII and 96% identity to PimM, which are predicted to encode putative RNA polymerase sigma factor in *S. chattanoogensis* L10 and *S. natalensis* ATCC 27448. The predicted gene *slnD* encodes a putative methyltransferase, which has 99% identity to ScnD in *S. chattanoogensis* L10 and 97% identity to PimD in *S. natalensis* ATCC 27448. The gene *scnRII* is the left boundary, and *scnD* is the right boundary of the natamycin BGC according to Du *et al.*¹². The deduced functions of the natamycin BGC genes and the results of similarity comparative analysis are listed in Supplementary Table S5. To facilitate a comparative

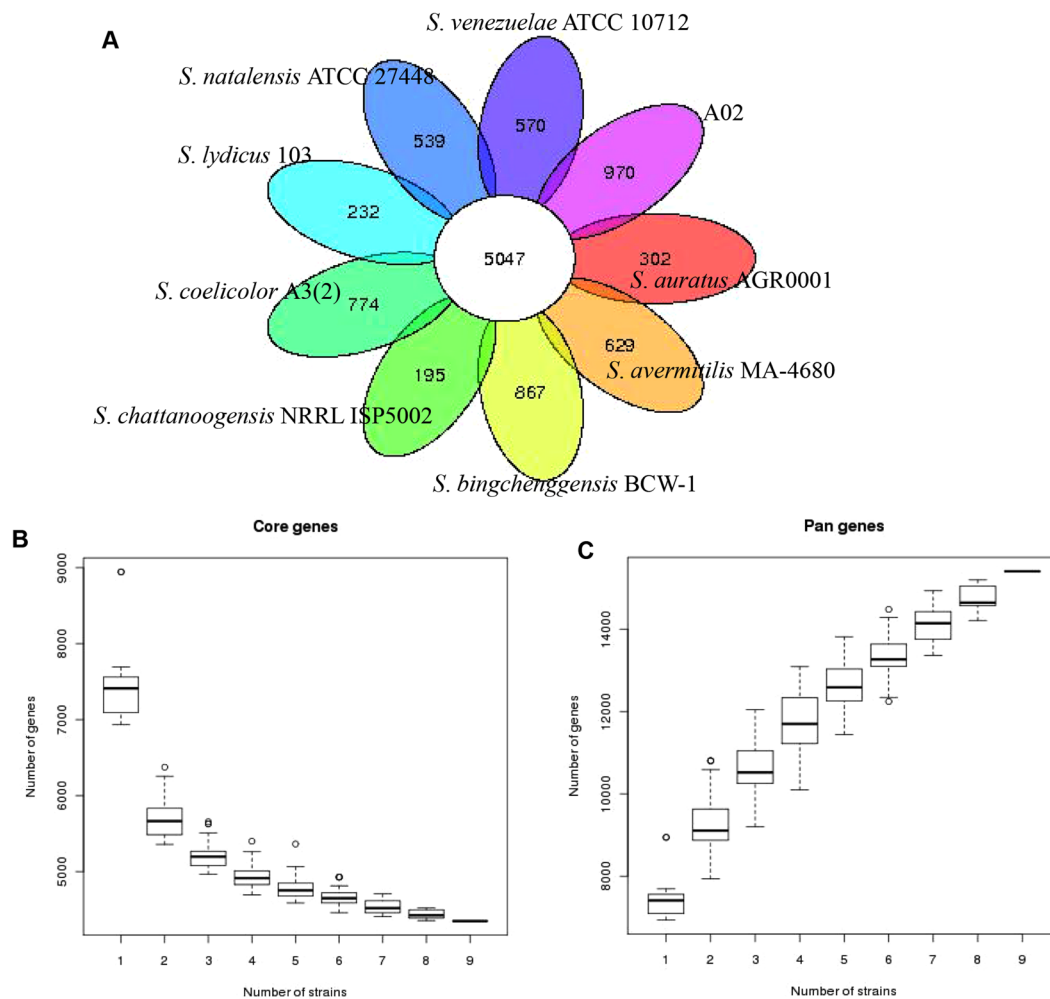


Figure 2. The pan-genome of streptomycetes. **(A)** Flower plots showing the core gene number and strain-specific gene number in 9 streptomycete strains. **(B)** Dilution curve of 9 streptomycete pan and core genomes.

analysis with the natamycin BGCs of *S. natalensis* (pim cluster) and *S. chattanoogensis* (scn cluster), we named the cluster from A02 the sln cluster.

The complete sequences of the three natamycin BGCs were aligned in the MAUVE program²⁹. Based on BlastP searches, the natamycin BGC of *S. lydicus* A02 has an organization that is similar to that of *S. chattanoogensis* L10 as described by Du *et al.*¹². Orthologous genes from the three natamycin BGCs are located in identical relative positions, and genes from the sln and scn clusters show a higher sequence identity than do those from the sln and pim clusters at the protein level. In contrast to the sln cluster, the pim cluster contains 2 major strain-specific regions at the two ends (Fig. 3B), i.e., pimH, putatively involved in natamycin export⁷, and pimT, involved in modulating the expression of natamycin biosynthetic genes via secretion of the natamycin inducer factor¹⁰. The scn cluster has no major strain-specific region compared with the sln cluster.

Similar to the scn cluster, the sln cluster contains a putative transposase gene (*tnp*) located between *slnL* and *slnS1*, which is likely involved in horizontal gene transfer of the natamycin BGC. Except for the putative transposase (97%), all the remaining putative proteins of the natamycin BGCs in *S. lydicus* A02 show significantly high similarity ($\geq 99\%$) to their counterparts in *S. chattanoogensis* L10, suggesting that the two strains may share the same metabolic regulatory mechanisms in the natamycin biosynthetic pathway.

Secondary metabolism. Forty secondary metabolite biosynthetic gene clusters were predicted in the complete genome of A02, including thirteen polyketide (PKS) and non-ribosomal polyketide (NRPS) clusters, six terpenoids, five lantipeptides, three siderophores, two butyrolactones, two lassopeptides, one cyanobactin, bacteriocin, ectoine, ladderane, melanin, and oligosaccharide, and three other undefined clusters (Supplementary Table S6). The total length of these gene clusters occupied 15.6% of the genome, specifically with total lengths of 1151.7 kb in the arm region and 299 kb in the core region. In contrast to the essential genes, most of the secondary metabolite biosynthetic gene clusters (30 out of 40) were in the non-core regions. The largest predicted gene cluster was for natamycin biosynthesis and was located in the 10.7 Mb right-arm region of the chromosome.

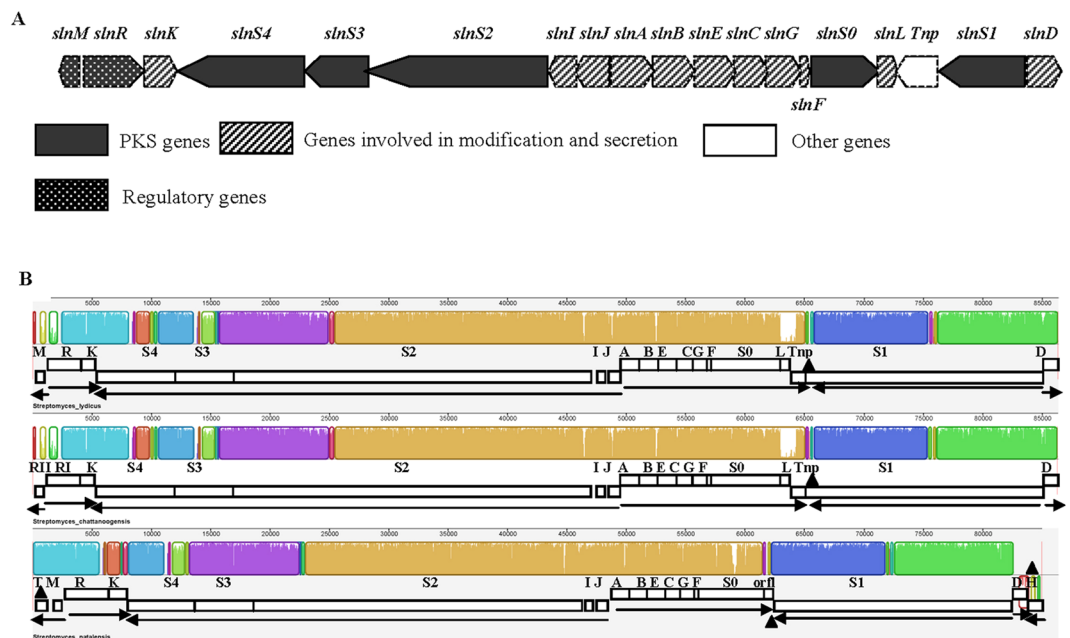


Figure 3. The natamycin biosynthetic gene cluster (BGC) of A02 (A) and comparison analysis of the natamycin BGCs of *sln*, *scn* and *pim* using the MAUVE program (B). The horizontal panels depict the *sln* (upper), *scn* (middle) and *pim* (lower) clusters. Mean sequence similarities are proportional to the heights of the red bars, and the respective scales show the sequence coordinates in base pairs. Regions with low similarities and strain-specific regions are marked with black triangles. Arrows indicate deduced transcriptional units.

Among the thirteen PKS and NRPS clusters, most of the CDSs for PKSs and NRPSs of ten clusters have orthologues in the *S. chattanoogensis* L10, with the best hitting scores by comparing against the NCBI database via BLASTP (Supplementary Table S6). The other three clusters showed no orthologues in the *S. chattanoogensis* genome, i.e., T2PKS (the first cluster) and NRPS (the sixteenth cluster) in the non-core regions and T1PKS (the twenty-ninth cluster) in the core region. Most cluster-spanning areas showed high nucleotide sequence identity to *S. lydicus* 103, except for the natamycin biosynthetic gene cluster (the thirty-fifth) and chattamycin biosynthetic gene cluster (the thirty-seventh). An analysis of the secondary metabolic gene clusters indicated a high diversity of metabolites in A02 and complex phylogenetic relationships between A02, *S. lydicus* 103 and *S. chattanoogensis* L10.

Two-component signal transduction system (TCS). Bacterial TCSs consist of a sensor histidine kinase (HK) and a response regulator/transcription factor (RR). TCSs play important roles in detecting and responding to diverse environmental stresses and cellular changes³⁰. The genome of *S. lydicus* A02 contains a large number of TCSs. Based on the BlastP analysis and conserved domains of known HKs and RRs, more than 100 *hk* and *rr* genes were discovered throughout the entire chromosome. Based on the homology-box, the topological feature of HK and the architecture of the C-terminal output domain of RR³¹, 61 TCS pairs (Supplementary Table S7) of A02 were grouped into 4 previously described subfamilies. There are 31 HK/RR pairs of the NarL subfamily, 26 pairs of the OmpR subfamily, 3 pairs of the CitB subfamily, and 1 pair of the AmtB subfamily. The complex and extensive regulatory systems indicate the high diversity of metabolites in *S. lydicus* A02.

It has been reported that phosphate control of actinorhodin and undecylprodigiosin biosynthesis in *S. lividans* is mediated by the two-component *phoR-phoP* system³². Additionally, phosphate decreases the production of pimarinic acid and negatively regulates the expression of the pimarinic biosynthesis genes in *S. natalensis*. Disruption of *phoP* resulted in increased pimarinic production and reduced sensitivity to phosphate control in *S. natalensis*¹⁵. Sequence analysis showed that the *phoR-phoP* TCS also exists in *S. lydicus* A02 (encoded by A02_3630/3632). They have a similar genetic organization and share high similarity both at the nucleotide and the amino acid levels with the *phoR-phoP* cluster found in other *Streptomyces* strains. The genes *phoR* (1,272 bp) and *phoP* (672 bp), which encode proteins of 424-aa and 224-aa, respectively, share 94/96% identity in nucleotide sequence and 97/99% similarity at the amino acid level with those in *S. natalensis* (Supplementary Table S8).

To verify the function of *phoP* in natamycin biosynthesis in *S. lydicus* A02, the *phoR-phoP* system was cloned according to the annotated sequence from *S. lydicus* A02, and a *phoP*-disrupted mutant was obtained. A 535-bp fragment internal to the chromosomal *phoP* gene was replaced with a thiostrepton resistance (*tsr*) gene, generating an apramycin-sensitive, thiostrepton-resistant mutant (named AP02) (Fig. 4A). Exconjugants with double crossovers were confirmed by PCR using PP1 and PP2 as primers and genomic DNA from A02 and AP02 as templates. The expected 672-bp *phoP* fragment was amplified from strain A02, while the PCR product from the double-crossover recombination mutant AP02 was 1.2-kb larger than that from the wild-type A02 (Fig. 4B).

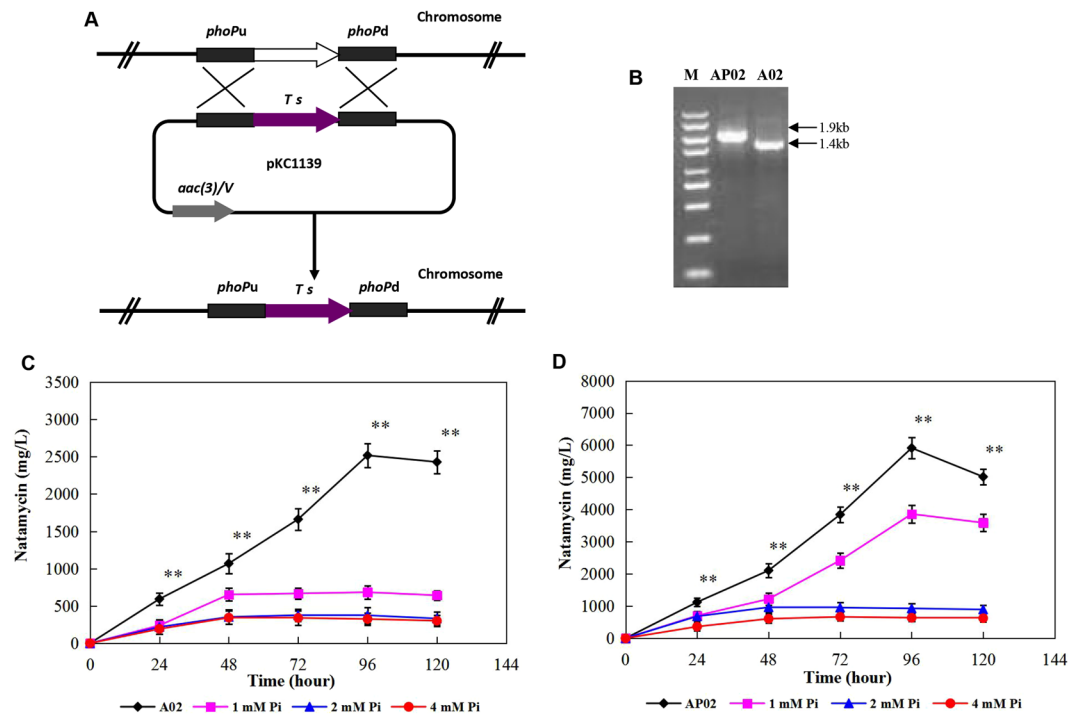


Figure 4. Construction of the *phoP* mutant and the effect of phosphate concentration on natamycin production in A02. (A) Schematic showing the deletion of *phoP* in A02. (B) PCR verification of *phoP* mutant. Lane M, DNA Marker. (C) Effect of increasing phosphate concentrations on natamycin production in A02. (D) Effect of increasing phosphate concentrations on natamycin production in AP02. The experiments were repeated three times, and the data shown are the mean \pm SE; * $P < 0.05$, ** $P < 0.01$.

These results showed that *phoP* was precisely replaced with the *tsr* gene in the homologous double-crossover recombination mutant AP02. The production of natamycin in the *phoP*-disrupted mutant was approximately 2.3-fold of the wild-type A02 in YSG medium, and it also showed reduced sensitivity to phosphate control (Fig. 4C,D). At 1 mM phosphate, the *phoP* mutant still produced approximately 65% of the natamycin level in the unsupplemented cultures in YSG medium, whereas the wild-type A02 was unable to synthesize significant amounts of natamycin at this phosphate concentration. This result indicated that, similar to *S. natalensis*, the phosphate control of natamycin biosynthesis in *S. lydicus* A02 was mediated by the two-component *phoR-phoP* system, which plays a negative regulatory role in natamycin biosynthesis.

***NsdA* in *S. lydicus* A02 and its regulatory role in natamycin biosynthesis.** The *nsdA* gene, which was first found in *S. coelicolor* A3(2) in 2006, is widely present and conserved in *Streptomyces* and negatively influences antibiotic production in multiple species of actinomycetes^{33–35}. The disruption of *nsdA* in *S. coelicolor* resulted in an overproduction of spores and three antibiotics, including actinorhodin, methylenomycin, and calcium-dependent antibiotic. An *nsdA*-disrupted mutant of *S. bingchengensis* displayed increased production of milbemycin A4 (1.5-fold) and nanchangmycin (9-fold), with more pigment and spores than its wild-type strain.

The genome sequence of *S. lydicus* A02 encodes a predicted homologue of *nsdA* (with 88% identity to *nsdA* in *S. qingfengmyceticus*) that occurs in regions having a high degree of synteny with the respective chromosomal locations in *S. qingfengmyceticus*. Based on its predicted sequences, the *nsdA* in *S. lydicus* A02 was amplified using the primers NP1 and NP2. The amplified 1.479-kb *nsdA* gene encodes a 493-amino-acid protein, which shares 80–88% identity in nucleotide sequence and 75–91% similarity at the amino acid level with *nsdA* in the other nine *Streptomyces* strains (Supplementary Table S9).

To study the role of *nsdA* in natamycin biosynthesis in *S. lydicus* A02, a 972-bp fragment internal to the chromosomal *nsdA* gene was replaced with a *tsr* gene, generating an apramycin-sensitive, thiostrepton-resistant mutant (AN02). Exconjugants with double crossovers were confirmed by PCR, using NP1 and NP2 as primers and genomic DNA from A02 and AN02 as templates. The expected 1.5-kb *nsdA* fragment was amplified from *S. lydicus* A02, while the PCR product from the double-crossover recombination mutant AN02 was 700-bp larger than that from the wild-type A02 (Fig. 5A). These results showed that *nsdA* was replaced precisely with the *tsr* gene in the homologous double-crossover recombination mutant AN02.

The *nsdA*-disrupted mutant AN02 differed from its wild-type strain *S. lydicus* A02 in both pigment production and sporulation. After 10 days of culture on PDA medium, the mutant AN02 produced less yellow pigment than the wild-type A02. Additionally, the mutant AN02 produced only a few light-grey spores, while the wild-type A02 produced abundant dark-grey spores (Fig. 5B).

An HPLC assay of the fermentation products revealed that the disruption of *nsdA* significantly increased the natamycin production of *S. lydicus* A02. The mutant AN02 produced nearly 5.0 g L⁻¹ natamycin in YSG medium

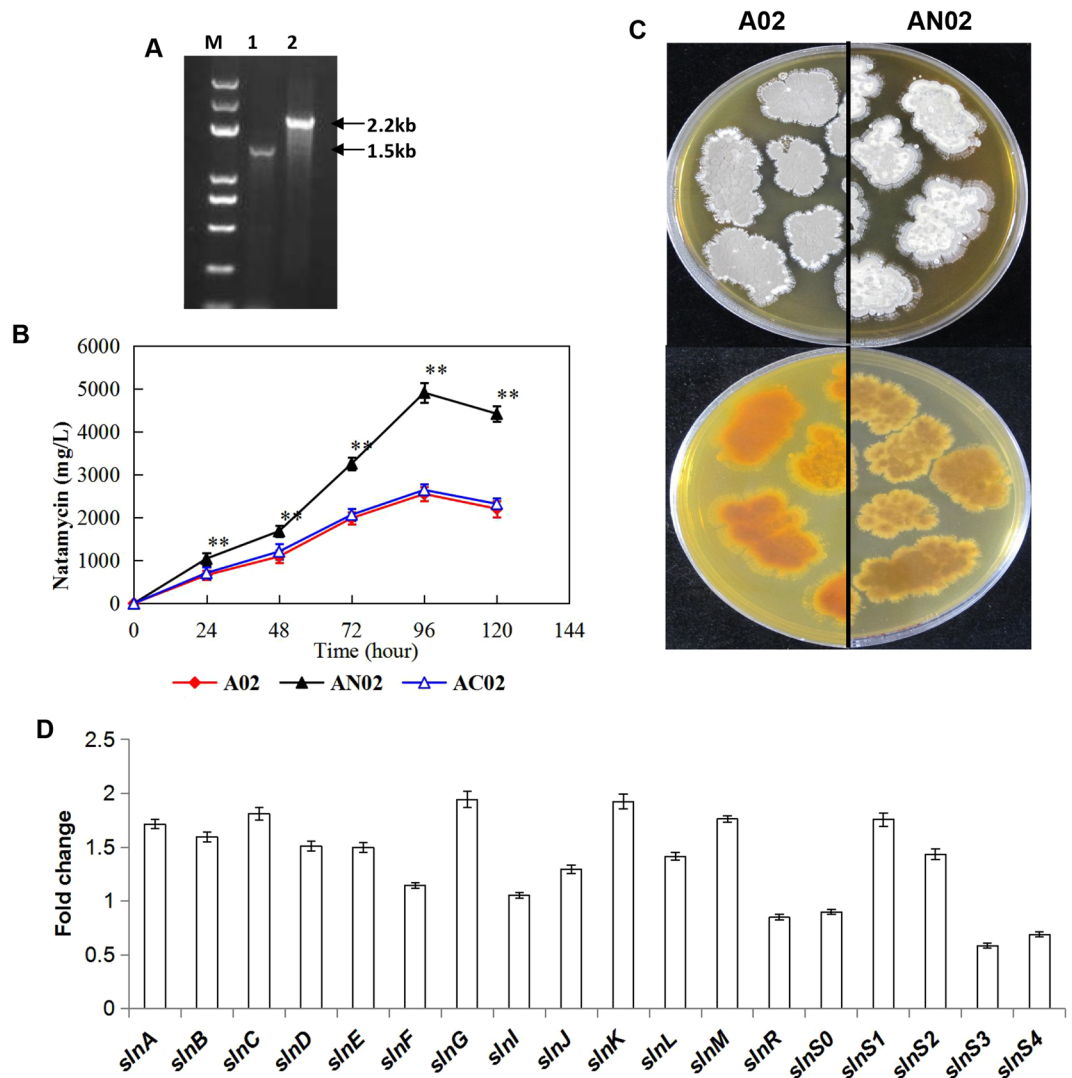


Figure 5. Effects of *nsdA* disruption on natamycin production and expression of the *sln* cluster in A02. (A) PCR verification of the *nsdA* mutant. Lane M, DNA Marker; Lane 1, Amplification with primers NP1 and NP2 in A02; Lane 2, Amplification with primers NP1 and NP2 in AN02. (B) Spore formation (upper) and pigment production (lower) of the A02 and AN02 on PDA plates for 10 days. (C) Natamycin production in the A02, AN02, and AC02. The experiments were repeated three times, and the data shown are the mean \pm SE; * $P < 0.05$, ** $P < 0.01$. (D) Expression of the genes of the *sln* cluster detected by qRT-PCR in A02 and AN02.

after 96 h, which was increased 1.9-fold compared to the wild-type A02. The changes in natamycin production and physio-morphological features were reversed to the wild-type levels by reintroducing *nsdA* into the mutant AN02 (Fig. 5C). These results showed that *nsdA* repressed the production of natamycin in *S. lydicus* A02.

In our previous study, natamycin production in *S. lydicus* was significantly enhanced (3.0-fold higher) by overexpression of the *slnM* gene in an engineered strain AM02¹. Here, we found that the transcriptional level of *nsdA* in the wild-type A02 was substantially higher than in AM02 (Supplementary Fig. S6). To test how *nsdA* regulates natamycin production in *S. lydicus*, gene expression analysis of the *sln* cluster in the wild-type A02 and *nsdA* mutant AN02 was performed by qRT-PCR. As the pathway-specific positive regulators of the *sln* cluster, the transcription of *slnM* in AN02 was much higher than in A02, while the transcript amount of *slnR* was decreased in AN02 (Fig. 5D). This transcription pattern was consistent with the previous finding for *pimM* and *pimR*, the first regulatory genes for pimarinic biosynthesis that function independently of each other¹⁰. Compared with the wild-type A02, the expression levels of most genes in the *sln* cluster were increased in AN02, with the exceptions of *slnR*, *slnS0*, *slnS3* and *slnS4*, whose transcript amounts were decreased. Thus, *nsdA* negatively regulates the expression of the *sln* cluster, directly or indirectly, to repress natamycin biosynthesis in *S. lydicus* A02, and the exact mechanism requires further study.

Conclusions

We presented the first description of the complete genome of natamycin-producing strain A02 by a combination of next-generation sequencing platforms and SMRT sequencing. Comparative genomics analysis based on the core-pan genome, gene families and ANI value showed that *Streptomyces* species that can produce natamycin are associated with genetic recombination more frequently, and they share closer genome evolution compared with other *Streptomyces* species. Comparative analysis of natamycin BGC supported findings from phylogenetic analyses that the natamycin BGCs between *S. lydicus* A02 and *S. chattanoogensis* L10 showed significantly high similarity, suggesting that the same natamycin biosynthetic regulation mechanisms are used by these two strains. The prediction of secondary metabolite biosynthetic gene clusters and TCS in the genome indicated that strain A02 possesses complex regulatory systems and a high diversity of metabolites. Regulatory analyses of the *phoR-phoP* TCS and *nsdA* gene confirmed phosphate sensitivity and the global negative regulation of natamycin production. The genome sequence and analyses presented in this study provide an important molecular basis for research into natamycin production in *Streptomyces*, which could facilitate rational genome modification to improve the industrial use of A02.

Materials and Methods

Manipulation of DNA and RNA from A02. Genomic DNA was isolated from A02 with TIANamp Bacteria DNA Kit (Tiangen, Beijing, China) and dissolved in DNase-free double-distilled water. For RNA extraction, the wild-type A02 and recombinant strains of *S. lydicus* were inoculated into YEME medium without sucrose³. Mycelia were collected after 48 h of growth, flash-frozen in liquid nitrogen, and ground into a fine powder. RNA was extracted using the RNeasy Mini Kit (Tiangen) according to the manufacturer's instructions and treated with DNase I (Tiangen) to eliminate possible chromosomal DNA contamination. The concentration and purity of the DNA and RNA were determined with a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA).

Genome sequencing and assembly. The genomic DNA of A02 was sequenced by combining next-generation sequencing platforms (Illumina paired end, 2*90-bp, and 500-bp insert size) and SMRT sequencing (Pacific Biosystems RS) by the Wuhan Institute of Biotechnology (Wuhan, China). PacBio RS (loBPng) reads were cleaned with sub-reads in the SMRT portal, and only clean reads were included in the subsequent analyses. For assembly of the SMRT sequencing reads, the longest reads were first utilized as seeds to recruit all other short reads for the construction of highly accurate preassembled reads through a consensus procedure with HGAP3³⁶. Thereafter, the preassembled reads were constructed by aligning all of the reads to each of the seed reads using BLASR³⁷. After the preassembly step, the resulting preassembled reads typically had read accuracies above 99%. Celera Assembler³⁸ was then used to assemble all of the clean reads to the preassembly, and pilon was applied to generate the best consensus sequence as the final genome sequence result. The method used for correcting Pacbio RSII assembly using the data from the Illumina MiSeq. 2000 was pilon³⁹.

Gene prediction and annotation. Putative protein-coding sequences were predicted using GLIMMER 3.0⁴⁰. CDS annotation was based on the BLASTP (2.2.26) and GO/IPR: iprscan_4.8 programs with NR⁴¹, Swiss-Prot⁴², TrEMBL⁴², COG⁴³, and KEGG⁴⁴ databases, followed by manual inspection. The tRNA and rRNA genes were predicted with tRNAscan-SE⁴⁵ and RNAmmer⁴⁶, respectively. Secondary metabolite gene clusters were predicted by antiSMASH 2.0⁴⁷. The CRISPR finder (<http://crispr.u-psud.fr/Server/>) was used for identifying CRISPR/Cas systems⁴⁸.

Core-Pan and synteny analysis. The genes from the A02 genome were used as the gene pool. Then, the genes predicted by Query samples were BLASTed within the gene pool, and the BLAST results were filtered by their length and identity. The BLAST coverage ratios (BCR) of genes from the gene pool and Query samples were calculated separately. If the BCR values from the reference and Query sample were smaller than the setting value, then the gene from the reference was not in homology with the Query, and the gene from the Query genome was added to the gene pool. Query samples had the previous steps repeated one by one, and the final gene pool was called the pan gene pool⁴⁹.

Synteny analysis was performed according to Kurtz *et al.*⁵⁰. The query genome was mapped to the reference genome with MUMmer at the nucleic acid level. The sequence of the target bacterium was ordered according to that of the reference bacterium based on MUMmer. The x and y axes of the two-dimensional synteny graph and the upper and following axes of the linear synteny graph were then constructed after the same proportion of size reduction in the length of both sequences.

Strains and plasmids. The wild-type A02 and recombinant strains were grown at 29°C on potato dextrose agar (PDA) slants for spore formation. *S. natalensis* ATCC 27448 and *S. chattanoogensis* NRRL B-2255 were obtained from Agricultural Research Service Culture Collection. *Escherichia coli* strain DH5 α was used as a host for genetic manipulation. Non-methylating *E. coli* strain ET12567 (pUZ8002) was used for DNA conjugal transfer from *E. coli* to *S. lydicus*. Conjugation and regeneration were performed as described by Kitani *et al.*⁵¹ and Paranthaman *et al.*⁵². Plasmid pUC19 was used for routine cloning and subcloning experiments. *E. coli-Streptomyces* shuttle vector pKC1139⁵³, containing *oriT* of RK2 and an apramycin resistance gene for selection in actinomycetes and *E. coli*, was used to construct deletion mutants. The integration vector pSET152 containing Φ 31 *int* and *attP*⁵⁴ was used to introduce a single copy of *nsdA* into *S. lydicus*.

When necessary, media were supplemented with antibiotics at the following concentrations: 100 μ g mL⁻¹ apramycin to LB medium, 60 μ g mL⁻¹ to selection medium, 30 μ g mL⁻¹ to fermentation medium, and 30 μ g mL⁻¹ thiosrepton to selection medium. The synthesis of oligonucleotide primers and DNA sequencing of PCR products were performed by Invitrogen (Beijing, China), and all the primers sequences are listed in Supplementary Table S10.

Construction of a *phoP*-disrupted strain and a *phoP*-complemented strain. To obtain the *phoP*-disrupted mutant, two DNA fragments containing the upstream fragment of 1.045-kb with primers P1 and P2 and downstream fragment of 1.028-kb with primers P3 and P4 were amplified individually. The 1.695-kb *tsr* gene was PCR-amplified from pPM927 with primers TP1 and TP2. These three fragments were sequentially ligated into pKC1139, which was digested with the corresponding enzyme to generate pKC1139P and verified by PCR and restriction enzyme analysis. The recombinant vector was introduced into *S. lydicus* A02 by intergenic conjugation. The *phoP*-deleted mutant AP02 was detected by apramycin sensitivity and thiostrepton resistance tests and confirmed by PCR analysis with the primers PP1 and PP2. GenBank accession numbers for the *phoP* and *phoR* of strain A02 are T2613482 and T2613483, respectively.

Construction of an *nsdA*-disrupted strain and an *nsdA*-complemented strain. A 1.479-kb coding region of *nsdA* was PCR-amplified from A02 with primers NP1 and NP2, which were designed according to the annotated sequence of *nsdA* from *S. lydicus* A02. After being verified by DNA sequencing, the PCR product of *nsdA* was cloned into the vector pMD18-T and then sequenced.

The construction of the *nsdA*-disrupted mutant was performed according to the above method, and the related primer sequences are listed in Supplementary Table S10. To complement the *nsdA* mutant, a 2.2-kb DNA fragment carrying the promoter and coding region of *nsdA* was PCR-amplified with primers NP3 and NP4 using the genomic DNA of A02 as the template. The PCR amplicon was inserted into the *EcoRI* and *XbaI* sites of pSET152 to give pSN152. The resulting vector was introduced into the *nsdA*-disrupted mutant AN02 and further integrated into the chromosome after transformation. The transformants were selected by 60 µg mL⁻¹ apramycin and further confirmed by PCR analysis with the primers NP1 and NP2.

Quantitative analysis of natamycin production by fermentation. Spores (5×10^7 each) of *S. lydicus* from PDA agar slant were inoculated into 50 mL of a seed culture medium. The seed cultures were incubated at 29 °C for 24 h on a rotary shaker (250 rpm). Yeast-salts-glycerol (YSG) medium¹² was used for natamycin production according to Wu *et al.*¹. Following sample extraction, quantitative analysis of natamycin was carried out using JAI LC-9101 Recycling Preparative HPLC (Japan Analytical Industry, Tokyo, Japan) equipped with a JAIGEL-ODS C18 reverse phase column (Japan Analytical Industry).

Gene expression analysis by quantitative real-time PCR (qRT-PCR). qRT-PCR was used to quantify the mRNA levels of natamycin BGC. Primer pairs for qRT-PCR were used according to Lee *et al.*¹⁴. One microgram of total RNA was used for complementary DNA (cDNA) synthesis, and qRT-PCR analysis was carried out with Takara SYBR® Premix Ex Taq™ II (Tli RNaseH Plus) (Takara Beijing, China) supplemented with ROX using an ABI GeneAmp PCR System 7500. The *lysA* gene was used as the reference gene. The primers for the *slnL* gene were designed using Primer 5 software (PREMIER Biosoft International, Silicon Valley, USA) (Table S10).

Data availability. The chromosome sequence of A02 using next-generation sequencing platforms and SMRT sequencing was deposited in Genbank under the accession number CP007699.1 and CP007699.2.

References

- Wu, H. L. *et al.* *SlnM* gene overexpression with different promoters on natamycin production in *Streptomyces lydicus* A02. *J. Ind. Microbiol. Biot.* **41**, 163–172 (2014).
- te Welscher, Y. M., van Leeuwen, M. R., de Kruijff, B., Dijksterhuis, J. & Breukink, E. Polyene antibiotic that inhibits membrane transport proteins. *Proc. Natl. Acad. Sci. USA* **109**, 11156–11159 (2012).
- Aparicio, J. F., Colina, A. J., Ceballos, E. & Martín, J. F. The biosynthetic gene cluster for the 26-membered ring polyene macrolide pimaricin. *J. Biol. Chem.* **274**, 10133–10139 (1999).
- Pedersen, J. C. Natamycin as a fungicide in agar media. *Appl. Environ. Microbiol.* **58**, 1064–1066 (1992).
- Du, Y. L. *et al.* Identification of a novel *Streptomyces chattanoogensis* L10 and enhancing its natamycin production by overexpressing positive regulation ScnRII. *J. Microbiol.* **47**, 506–513 (2009).
- Li, H. D., Jin, Z. H., Zhang, H. G. & Jin, H. *Protoplast formation, regeneration and UV mutagenesis of natamycin producing Streptomyces gilvosporeus*. *Ind. Microbiol.* **38**, 43–46 (2008).
- Aparicio, J. F., Fouces, R., Mendes, M. V., Olivera, N. & Martín, J. F. A complex multienzyme system encoded by five polyketide synthase genes is involved in the biosynthesis of the 26-membered polyene macrolide pimaricin in *Streptomyces natalensis*. *Chem. Biol.* **7**, 895–905 (2000).
- Antón, N., Mendes, M. V., Martín, J. F. & Aparicio, J. F. Identification of PimR as a positive regulator of pimaricin biosynthesis in *Streptomyces natalensis*. *J. Bacteriol.* **186**, 2567–2575 (2004).
- Antón, N. *et al.* PimM, a PAS domain positive regulator of pimaricin biosynthesis in *Streptomyces natalensis*. *Microbiology* **153**, 3174–3183 (2007).
- Vicente, C. M. *et al.* PimT, an amino acid exporter controls polyene production via secretion of the quorum sensing pimaricin-inducer PI-factor in *Streptomyces natalensis*. *Microb. Cell Fact.* **8**, 33 (2009).
- Du, Y. L. *et al.* The pleiotropic regulator AdpAch is required for natamycin biosynthesis and morphological differentiation in *Streptomyces chattanoogensis*. *Microbiology* **157**, 1300–1311 (2011).
- Du, Y. L. *et al.* Identification of a novel *Streptomyces chattanoogensis* L10 and enhancing its natamycin production by overexpressing positive regulation scnRII. *J. Microbiol.* **47**, 506–513 (2009).
- Jiang, H. *et al.* Improvement of natamycin production by engineering of phosphopantetheinyl transferases in *Streptomyces chattanoogensis* L10. *Appl. Environ. Microbiol.* **79**, 3346–3354 (2013).
- Lee, K. M., Lee, C. K., Choi, S. U., Park, H. R. & Wang, Y. I. Functional analysis of a BarX homologue (SngA) as a pleiotropic regulator in *Streptomyces natalensis*. *Arch. Microbiol.* **189**, 569–577 (2008).
- Mendes, M. V. *et al.* The two-component *phoR-phoP* system of *Streptomyces natalensis*: inactivation or deletion of *phoP* reduces the negative phosphate regulation of pimaricin biosynthesis. *Metab. Eng.* **9**, 217–227 (2007).
- Lee, K. M. *et al.* Cloning and *in vivo* functional analysis by disruption of a gene encoding the γ -butyrolactone autoregulator receptor from *Streptomyces natalensis*. *Arch. Microbiol.* **184**, 249–257 (2005).
- Wu, H. L. *et al.* Heterologous coexpression of *Vitreoscilla* hemoglobin and *Bacillus megaterium* glucanase in *Streptomyces lydicus* A02 enhanced its production of antifungal metabolites. *Enzyme Microb. Techn.* **81**, 80–87 (2015).

18. Jia, N., Ding, M. Z., Luo, H., Gao, F. & Yuan, Y. J. Complete genome sequencing and antibiotics biosynthesis pathways analysis of *Streptomyces lydicus* 103. *Sci. Rep.* **7**, (2017).
19. Han, X. L. *et al.* Genome sequence of *Streptomyces auratus* strain AGR0001, a phoslactomycin-producing actinomycete. *J. Bacteriol.* **194**, 5472–5473 (2012).
20. Wang, X. J. *et al.* Genome sequence of the milbemycin-producing bacterium *Streptomyces bingchengensis*. *J. Bacteriol.* **192**, 4526–4527 (2010).
21. Wu, H. L. *et al.* Genomic and transcriptomic insights into the thermo-regulated biosynthesis of validamycin in *Streptomyces hygroscopicus* 5008. *BMC Genomics* **13**, 377 (2012).
22. Omura, S. *et al.* Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. *Proc. Natl. Acad. Sci.* **98**, 12215–12220 (2001).
23. Bentley, S. D. *et al.* Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141–147 (2002).
24. Haft, D. H., Selengut, J., Mongodin, E. F. & Nelson, K. E. A guild of 45 CRISPR associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* **1**, e60 (2005).
25. Tian, X. P. Comparative genomics analysis of *Streptomyces* species reveals their adaptation to the marine environment and their diversity at the genomic level. *Front. Microbiol.* **7**, (e1004041) (2016).
26. Marri, P. R., Hao, W. & Golding, G. B. Gene gain and gene loss in streptococcus: is it driven by habitat? *Mol. Biol. Evol.* **23**, 2379–2391 (2006).
27. Richter, M. & Rossello, M. R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. USA* **106**, 19126–19131 (2009).
28. Vernikos, G., Medini, D., Riley, D. R. & Tettelin, H. Ten years of pan-genome analyses. *Curr. Opin. Microbiol.* **23**, 148–154 (2015).
29. Darling, A. C., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394–1403 (2004).
30. Szurmant, H., Mohan, M. A., Imus, P. M. & Hoch, J. A. YycH and YycI interact to regulate the essential YycFG two-component system in *Bacillus subtilis*. *J. Bacteriol.* **189**, 3280–3289 (2007).
31. Mascher, T., Helmann, J. D. & Udden, G. Stimulus perception in bacterial signal-transducing histidine kinases. *Microbiol. Mol. Biol. Rev.* **70**, 910–938 (2006).
32. Sola-Landa, A., Moura, R. S. & Martín, J. F. The two-component *PhoR-PhoP* system controls both primary metabolism and secondary metabolite biosynthesis in *Streptomyces lividans*. *Proc. Natl. Acad. Sci. USA* **100**, 6133–6138 (2003).
33. Li, W. C. *et al.* Identification of a gene negatively affecting antibiotic production and morphological differentiation in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **188**, 8368–8375 (2006).
34. Wang, X. J., Guo, S. L., Guo, W. Q. & Xiang, W. S. Role of *nsdA* in negative regulation of antibiotic production and morphological differentiation in *Streptomyces bingchengensis*. *J. Antibiot.* **62**, 309–313 (2009).
35. Yu, Z., Wang, Q., Deng, Z. X. & Tao, M. F. Activation of silent antibiotic synthesis in *Streptomyces lividans* by disruption of a negative regulator *nsdA*, a gene conserved in *Streptomyces*. *Chin. J. Biotechnol.* **22**, 751–762 (2006).
36. Chin, C. S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569 (2013).
37. Chaisson, M. J. & Tesler, G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics* **13**, 238 (2012).
38. Myers, E. W. *et al.* A whole-genome assembly of *Drosophila*. *Science* **287**, 2196–2204 (2000).
39. Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *Plos One* **9**, e112963 (2014).
40. Delcher, A. L., Bratke, K. A., Powers, E. C. & Salzberg, S. L. Identifying bacterial genes and 14 endosymbiont DNA with Glimmer. *Bioinformatics* **23**, 673–679 (2007).
41. Wheeler, D. L. *et al.* Database resources of the National Center for Biotechnology Information: 2002 update. *Nucleic Acids Res.* **30**, 13–16 (2002).
42. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48 (2000).
43. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33–36 (2000).
44. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. & Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **38**, D355–D360 (2010).
45. Schattner, P., Brooks, A. N. & Lowe, T. M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **33**, W686–W689 (2005).
46. Lagesen, K. *et al.* RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108 (2007).
47. Weber, T. *et al.* AntiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic acids Res.* **43**, W237–W243 (2015).
48. Grissa, I., Gilles, V. & Christine, P. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic acids Res.* **35**, W52–W57 (2007).
49. Qin, J. R. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
50. Kurtz, S. A. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).
51. Kitani, S., Bibb, M. J., Nihira, T. & Yamada, Y. Conjugal transfer of plasmid DNA from *Escherichia coli* to *Streptomyces lavendulae* FRI-5. *J. Microbiol. Biotechnol.* **10**, 535–538 (2000).
52. Paranthaman, S. & Dharmalingam, K. Intergeneric conjugation in *Streptomyces peucetius* and *Streptomyces* sp. strain C5: chromosomal integration and expression of recombinant plasmids carrying the *chiC* gene. *Appl. Environ. Microbiol.* **69**, 84–91 (2003).
53. Bierman, M. *et al.* Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene* **116**, 43–49 (1992).
54. Wilkinson, C. J. *et al.* Increasing the efficiency of heterologous promoters in actinomycetes. *J. Mol. Microbiol. Biotechnol.* **4**, 417–426 (2002).

Acknowledgements

This work was supported by Beijing Natural Science Foundation (6152007), funding for training talents in Beijing City (2014000020060G180), Science and Technology Plan Project of Beijing (D151100003915003), and Scientific and Technological Innovation Capacity Construction Special Funds of Beijing Academy of Agriculture and Forestry (KJCX20170410). The authors thank Dr. George Lazarovits, Agroecology Research Services Centre, London, Canada, for technical reading.

Author Contributions

H.L.W., W.C.L. and Y.S.Y. designed the study and prepared the manuscript. L.L.S., X.L.B. and H.Z. performed the experiments. T.L., D.D., T.T.Z., J.Z., Z.F.T., K.W.S. and L.Y. collaborated in genomic sequencing and analysis. D.W.L. contributed to the collection of strains and lab works. All authors read and approved the manuscript before.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-09532-3](https://doi.org/10.1038/s41598-017-09532-3)

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017