

Comparative genomics and transcriptomics analyses provide insights into the high yield and regulatory mechanism of Norvancomycin biosynthesis in *Amycolatopsis orientalis* NCPC 2-48

Xingxing Li

Institute of Medicinal Biotechnology

Cong Zhang

Institute of Medicinal Biotechnology

Ying Zhao

North China Pharmaceutical Group

Xuan Lei

Institute of Medicinal Biotechnology

Zhibo Jiang

Institute of Medicinal Biotechnology

Xuexia Zhang

North China Pharmaceutical Group

Zhihui Zheng

North China Pharmaceutical Group

Shuyi Si

Institute of Medicinal Biotechnology

Lifei Wang

Institute of Medicinal Biotechnology

Bin Hong (✉ binhong69@hotmail.com)

Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences <https://orcid.org/0000-0001-6244-8298>

Research

Keywords: Norvancomycin, *Amycolatopsis orientalis*, Comparative transcriptomics, Biosynthesis regulation

DOI: <https://doi.org/10.21203/rs.3.rs-42758/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Norvancomycin has been widely used in clinic to treat against MRSA (Methicillin-resistant *Staphylococcus aureus*) and MRSE (methicillin-resistant *Staphylococcus epidermidis*) in China. *Amycolatopsis orientalis* NCPC 2–48, a high yield strain derived from *A. orientalis* CCCC 200066, has been applied in industrial large-scale production of norvancomycin by North China Pharmaceutical Group. However, the potential high-yield and regulatory mechanism involved in norvancomycin biosynthetic pathway has not yet been addressed.

Results: Here we sequenced and compared the genomes and transcriptomes of *A. orientalis* CCCC 200066 and NCPC 2–48. These two genomes are extremely similar with an identity of more than 99.9%, and no duplication and structural variation was found in the norvancomycin biosynthetic gene cluster. Comparative transcriptomic analysis indicated that biosynthetic gene cluster of norvancomycin, as well as some primary metabolite pathways for the biosynthetic precursors of norvancomycin were generally upregulated. *AoStrR1* and *AoLuxR1*, two cluster-situated regulatory genes in norvancomycin cluster, were 23.3-fold and 5.8-fold upregulated in the high yield strain at 48 h, respectively. Over-expression of *AoStrR1* and *AoLuxR1* in CCCC 200066 resulted in an increase of norvancomycin production, indicating their positive role in norvancomycin biosynthesis. Furthermore, *AoStrR1* can regulate the production of norvancomycin by directly interacting with at least 8 promoters of norvancomycin biosynthetic genes or operons.

Conclusion: Our results suggested that the mechanism of high yield of NCPC 2–48 can be ascribed to increased expression level of norvancomycin biosynthetic genes in its cluster as well as the genes responsible for the supply of its precursors. And the norvancomycin biosynthetic genes are positively regulated by *AoStrR1* and *AoLuxR1*, of them *AoStrR1* was the ultimate pathway-specific regulator for the norvancomycin production. These results are helpful for further clarification of the holistic and pathway-specific regulatory mechanism of norvancomycin biosynthesis in the industrial production strain.

Background

Glycopeptide antibiotics, exhibiting outstanding activity against Gram-positive pathogens, are a class of widely known natural compounds produced by Actinomycetes with typical representatives of vancomycin, balhimycin, teicoplanin, A40926 *etc.* [1]. Some of glycopeptide antibiotics have been approved for clinical use to treat persistent infections by Gram-positive multi-resistant pathogens since vancomycin was first approved in 1958 [2]. Norvancomycin is an important glycopeptide antibiotics of vancomycin group, and has been used as a first-line empiric antibiotic therapy to prevent and treat the intracranial infections of MRSA (Methicillin-resistant *Staphylococcus aureus*) and MRSE (methicillin-resistant *Staphylococcus epidermidis*) in China since 1969, although its structure was finally determined in the 1980s. Norvancomycin has a closely similar chemical structure with vancomycin, only lacking an N-methyl group at the N-terminus of the polypeptide, and shows a comparable antibacterial spectrum and activity to those of vancomycin. Norvancomycin was first discovered as vancomycin from the

Amycolatopsis orientalis CPCC 200066 isolated from soil samples from Guizhou Province, China in 1959, and then was commercially developed by North China Pharmaceutical Group. The fermentation potency of norvancomycin in the original strain CPCC 200066 was about 200 µg/ml. After series of physical and chemical mutagenesis, a high yield norvancomycin strain (*A. orientalis* NCPC 2–48) was obtained from *A. orientalis* CPCC 200066 by North China Pharmaceutical Group. Using the patented fermentation medium, the fermentation potency of high-yield strain could be up to more than 6,000 µg/ml. Although *A. orientalis* NCPC 2–48 has been successfully applied to industrial large-scale production, the potential high-yield mechanism and its biosynthetic regulatory mechanism remain obscure.

The biosynthetic gene clusters (BGCs) for some known glycopeptide antibiotics were reported, including vancomycin (*vcm*) [3], balhimycin (*bal*) [4], A40926 (*dbv*) [5], teicoplanin (*tcp*) [6, 7], pekiskomycin (*pek*) [8] and kistamicin (*kis*) [9] *etc.* In recent years, more new glycopeptide antibiotic biosynthetic gene clusters are discovered by genome mining strategies [10, 11]. In our previous studies, we reported the genome sequence of *A. orientalis* CPCC 200066 and localized the biosynthetic gene cluster of norvancomycin (*nvcm*) [12] which display a high resemblance with reported gene clusters of *vcm* in *A. orientalis* sub sp. *orientalis* KCTC 9412^T [3]. More than forty genes were identified in the *nvcm* cluster, including four resistance proteins, three large NRPSs, several post-assembly tailoring enzymes, and a series of biosynthetic proteins for the supply of non-proteinaceous amino-acid precursors. Combining bioinformatics and biochemical studies, most biosynthetic steps of glycopeptide antibiotics are gradually deciphered [13, 14]. Since chemical structure of norvancomycin is closely similar to vancomycin, with only an *N*-methyl group absence at the N-terminus of the polypeptide, we speculate that the biosynthesis of norvancomycin also includes three steps by analogy with vancomycin [15, 16]. In the first step, the unnatural amino acids and vancosamine need to be synthesized for substrate preparation. β-hydroxytyrosine (Bht) is catalyzed by VcmD, OxyD and Vhp; 3,5-dihydroxyphenylglycine (Dpg) is catalyzed by DpgA/B/C/D and HpgT; 4-hydroxyphenylglycine (Hpg) is catalyzed by Pdh, HmaS, Hmo and HpgT; vancosamine is catalyzed by VasA/B/C/D/E. In the next step, the norvancomycin heptapeptide structure is assembled by three NRPSs (VcmA/B/C). In the final step, the post-modification stage, halogenation, glycosylation and cyclization of heptapeptide skeletons are performed under the action of halogenase (Vhal), glycosyltransferases (GtfD, GtfE) and oxidases (OxyA/B/C). Although *N*-methyltransferase (Vmt) gene is present in the *nvcm* cluster, the *N*-methylation is not present in norvancomycin.

The regulatory mechanisms of glycopeptide antibiotics biosynthetic clusters are still not completely characterized. The VanR-VanS pair, a two-component regulatory system, controls the transcription of glycopeptide antibiotics resistance genes in some glycopeptide antibiotics producers [17]. The VanRS homologues have been found in the A40926 cluster (Dbv6 and Dbv22) [5], balhimycin cluster (VnIRS) [18] and teicoplanin cluster (Tei2 and Tei3, also named as Tcp7 and Tcp6) [19]. Nevertheless, the vancomycin and norvancomycin cluster lacks VanRS although components of self-resistance system, *e.g.*, *vanHAXY*, are present in the *vcm* and *nvcm* cluster.

StrR-family regulators are most commonly present in glycopeptide antibiotics biosynthetic clusters as a specific pathway regulator. Bbr in balhimycin cluster binds to five promoter regions within *bal* cluster *in vitro*, and these genes are responsible for dehydrovancosamine biosynthesis (*dvaA*), oxidative cross-link (*oxyA*), transportation (*tba*), sodium-proton antiporter (*orf7*) and self-regulation (*bbr*) [20]. Likewise, Dbv4 and Tcp28 (also named as Tei15*) were proved to be positive regulators in the A40926 cluster and teicoplanin cluster, respectively [21–23]. In A40926 producing strain, Dbv4 positively regulates the transcription of the gene operons involved in Dpg biosynthesis and heptapeptide modification step (oxidative cross-link, halogenation, glycosylation and acylation) by directly binding to the upstream regions of *dbv14* and *dbv30* [21, 22]. Tcp28 (Tei15*) positively regulates teicoplanin production and six binding sites in promoter regions within *tei* cluster were identified *in vitro* [23]. Three of them proved to be transcriptionally dependent on Tcp28, whose function involves peptide assembly (*teiA*), glycosylation (*tei 2**) and regulation (*tei 16**) [24]. In addition, the proposed palindromic sequence (GTCCAa_{N17}TtGGAC) for Bbr-homologous regulator-binding site is also conserved within *bal*, *dbv* and *tcp* BGCs [20, 21, 23].

In addition, LuxR-family regulators have been characterized in A40926 and teicoplanin cluster. Dbv3 positively regulates A40926 production and hierarchically control the transcription of *dbv4* (StrR-family) [22]. Tcp29 (also named as Tei16*) positively regulates teicoplanin production, but in contrast to *dbv3*, Tcp29 is controlled by Tcp28 (StrR-family) [23]. However, unlike StrR-family regulator, the direct target of LuxR-family regulator in the glycopeptide antibiotic clusters remain to be elucidated, since there is no direct binding evidence currently. The bioinformatic analysis indicates that there are four putative regulators (LuxR-family, StrR-family, TetR-family and AraC-family) present in or nearby the *vcm* or *nvcm* cluster. The regulatory mechanism of the biosynthesis of vancomycin or norvancomycin has not been reported yet.

The purpose of this study is to uncover the high-yield and biosynthetic regulatory mechanism of industrial strain NCPC 2–48 through exploring the genomic and transcriptomic features compared with the original strain CPCC 200066. Here, we present comprehensive comparative genomics and transcriptomics analyses of these two strains and show that the whole biosynthetic gene cluster of norvancomycin, as well as some primary metabolite pathways for the amino acid and other precursors of norvancomycin were comprehensively upregulated. The cluster situated regulators, AoStrR1 and AoLuxR1, play a positive role in norvancomycin production, and almost all of the norvancomycin biosynthetic genes are directly controlled by AoStrR1. These results have provided insights into high yield mechanism in the industrial strain, as well as the role of cluster-specific regulators in the norvancomycin biosynthesis.

Results And Discussion

Comparative genomics analysis of norvancomycin producing strains

The complete genome of industrial producing strain *A. orientalis* NCPC 2–48 and that of original strain *A. orientalis* CPCC 200066 [12] are both a circular chromosome of 9.5 Mb with a G + C content of 68.84% (Fig. 1), only 1,303 bp shorter in the genome length of NCPC 2–48 than that of CPCC 200066. The genome analysis of industrial strain showed that it contained 8,705 genes, and the total length of genes was 8,579,274 bp, which makes up to 90.41% of genome. There are 404 tandem repeat sequence (35,253 bp), which makes up to 0.3715% of genome, 325 minisatellite DNAs, 11 microsatellite DNAs, 50 tRNAs, and 12 rRNAs.

We compared the genomes of NCPC 2–48 to CPCC 200066 and found that two genomic sequences are extremely similar with a high identity of 99.97%. There was no large fragment duplication or deletion in the entire genome of NCPC 2–48 compared with CPCC 200066. Moreover, the internal structure of the chromosome and gene order were largely conserved without rearrangements appeared in the genome of NCPC 2–48. Both strains have extremely similar codon usage of 50 tRNA genes. For the norvancomycin biosynthesis, further comparative analysis of secondary metabolism gene clusters showed that no duplication or InDels (insertions and deletions) of small fragments (≤ 50 bp) were occurred in *nvcM* biosynthetic gene cluster in the industrial strain. It suggested that the high-yield of the industrial strain is not caused by the increased copy number of the *nvcM* biosynthetic gene cluster or gene mutations within the cluster.

The genomic difference between these two strains is mainly accounted for three SVs (structural variation), including two deletions of 965 bp (SV1) and 12,212 bp (SV2) fragments and one insertion of 12,076 bp (SV3) fragment (Fig. 1, Table S3). All these SVs appear at CDSs region but far away from the *nvcM* biosynthetic gene cluster. There are 34 protein coding genes involved in these fragments of deletion and insertion, including 3 regulators, 1 transporter, 3 transposases, 9 other enzymes and 18 unknown proteins (Table S3). The coding gene information of the three SVs is shown in Table S4.

The first structural variation (SV1) occurred in gene B37_4355, which encodes α -1,2-mannosidase, a key enzyme in N-glycan processing of primary metabolic process. The 965 bp of B37_4355 sequence is deleted in genome of NCPC 2–48, corresponding to the position of 4,774,430 ~ 4,775,394 bp of original strain's chromosome.

SV2 is a deleted fragment of 12,212 bp in genome of NCPC 2–48, corresponding to the original strain's chromosome position from 7,292,954 to 7,305,165 bp, which designated gene B37_6566 through B37_6583. There are 17 possible CDSs in this region, encoding 1 regulatory protein (TetR family regulatory proteins), 2 transposases (1 excisionase and 1 integrase), 1 replication initiation protein, 1 enoyl-CoA hydratase, 1 cell division protein FtsK, 1 ABC transporter, and 10 unknown proteins. Further analysis of the sequence flanking SV2 showed that one *lacI* family transcriptional regulatory gene is located upstream. Usually, *lacI* family regulators control the expression of some key enzymes involved in carbon metabolism, and regulate the transcription of a series of downstream genes including some transcriptional factors [25]. The deletion of genes downstream *lacI* might block the regulation of *lacI* and

enhance glucose catabolism of primary metabolism, which is conducive to the precursor sugar synthesis of norvancomycin.

SV3 is a 12,076 bp of insertion, corresponding to the chromosome position of 8,365,032 bp of the original strain genome. It contains 16 possible CDSs, encoding 2 regulatory proteins (XRE family), 1 helicase, 1 transposase (excisionase), 1 flavin protein, 2 cell division protein FtsK, 1 replication initiator protein and 8 unknown proteins. Analysis of function of flanking genes revealed that there are some key enzymes of primary metabolism such as acetyl-CoA dehydrogenase and acetyl-CoA synthetase at downstream of SV3.

In addition to three fragments of deletion and insertion, there are also 216 InDels (insertions and deletions) of small fragments (≤ 50 bp) present in the genome of high-yield strain. The coding regions of some important enzymes related to primary and secondary metabolism together with transcription factors, such as AraC, LacI, LysR, TetR, MerR, YebC/PmpR, SARP family proteins and two-component regulators were present in these InDels.

In these SVs and InDels, there are no functionally known genes directly related to the biosynthetic pathway of norvancomycin and its precursors based on their functional annotation. Thus, the genomic variations of industrial strain would be difficult to give a simple explanation on its high-yield production of norvancomycin. Since some key enzymes of primary metabolism found in these SVs and InDels, or in flanking segments, we speculated that these genomic mutations may change the metabolic flow by affecting expression of the important enzymes of primary metabolism. Meanwhile, more than ten regulatory genes were detected in SVs and InDels. Along with the loss, insertion or mutation of these regulatory genes, the holistic regulation of primary or secondary metabolism and cell growth may be changed in the high-yield producer strain, which ultimately makes the overall metabolic flow more favorable for the biosynthesis of norvancomycin.

Transcriptomics profiling at norvancomycin producing strains

In order to further reveal the high yield mechanism of norvancomycin, the transcriptomic analysis of norvancomycin original strain and industrial strain were carried out at three different time points (12 h, 24 h and 48 h). RNA from NCPC 2-48 and CPCC 200066 were extracted and sequenced, and an average of 23,578,714 raw reads were generated. After removing low-quality reads, the average number of remaining clean reads was 23,523,435, and the average comparison rates of clean reads to the reference gene and reference genome were 80.54% and 96.69%, respectively. The statistics on the sequencing data for each sample is shown in Table S5. Pair-wise differentially expressed gene (DEG) analyses revealed that more than 2000 DEGs, about one-fourth of the total genes in the genome, had significantly lower or higher transcript abundance (fold change (FC) > 2 and FDR ≤ 0.001) at each time points (12 h, 24 h and 48 h) in NCPC 2-48 relative to CPCC 200066, as shown in Fig. 2A. The transcriptional levels of some genes within the *nvcm* cluster were verified by RT-qPCR (Fig. S1 in Supplementary information). The results of DEGs hierarchical clustering analysis showed that differential gene expression pattern was

similar at 24 h and 48 h (Fig. 2B), more than two-thirds of the differential genes up-regulated in the industrial strain. Interestingly, more genes were downregulated at 12 h compared to other two time points (Fig. 2B), and there are 105 genes downregulated at 12 h but upregulated at 24 h and 48 h time points. KEGG pathway search showed that 35 out of the 105 genes are located in *nvcM* cluster. The biosynthetic pathways of secondary metabolites are usually activated in a growth phase-dependent manner, so that the genes responsible for secondary metabolism coincide with the onset of stationary phase in liquid fermentation in microorganisms. In the case of high-yield strain, transcriptional levels of genes responsible for biosynthesis of norvancomycin are lower in the early stage of growth (12 h), and then upregulated abruptly from 24 h to 48 h, showing that the norvancomycin biosynthesis is more strictly controlled during the different growth stages. Due to the similarity of gene expression patterns between 24 h and 48 h, and most of DEGs included at 24 h (2,039 up-regulated and 628 down-regulated), we then analyzed the functional pathway enrichment of DEGs at 24 h based on KEGG database. Enrichment analysis of the functional categories of the transcriptome indicated that a total of 1,764 differential genes were annotated into 150 metabolic pathways, most of them related to the primary metabolism such as nitrogen metabolism and arginine biosynthesis, as well as the biosynthesis of secondary metabolites such as norvancomycin, tetracyclines and other type II polyketides, degradation of naphthalene and aromatic compounds, tyrosine and inositol phosphate metabolism processes. Top 20 of most specific KEGG enrichment results as shown in Fig. 2C.

In particular, the visualization of the transcriptome (Fig. 2D) showed that in the *nvcM* cluster, all key enzymes related to the biosynthesis of respective unit substrates such as Bht (VcmD, OxyD, Vhp), Hpg (Pdh, HmaS, HmO, HpgT), Dpg (DpgA/B/C/D, HpgT) and vancosamine (VasA/B/C/D/E), heptapeptide assemblage (VcmA/B/C) and post-modifications (OxyA/B/C, Vhal, GtfD/E) were transcribed significantly higher in high yield strain than that of original strain at 24 h and 48 h (Fig. 3, Table S6). It suggested that the increased transcriptional level of norvancomycin's biosynthetic genes (1.9 ~ 8.6 fold upregulated at 24 h, and 3.0 ~ 18.3 fold upregulated at 48 h) directly promote high yield of norvancomycin in NCPC 2–48. In addition, primary pathways for amino acid (Leu, Asn, Tyr) and glucose production were also upregulated (Fig. 3), including genes B37_4517 (EC 4.2.3.5, FC = 3.9) and B37_7337 (EC 5.4.99.5, FC = 2.7) for prephenate (the precursor of Hpg); B37_6997 (EC 1.3.1.12, FC = 6.8), B37_3479 (EC 2.6.1.1, FC = 2.8) and B37_6785 (EC 2.6.1.9, FC = 3.1) for Tyr (the precursor of Bht); B37_7779 (EC 6.4.1.2, FC = 2.3) for Malonyl-CoA (Dpg's precursor); B37_4701 (2.6.1.42, FC = 3.5), B37_8154 (2.2.1.6, FC = 2.2) and B37_2225 (2.3.3.13, FC = 5.9) for Leu; B37_7110 (EC 6.3.5.4, FC = 6.2) and B37_3479 (EC 2.6.1.1, FC = 2.8) for Asn and its precursor Asp synthesis; B37_7117 (RfbA, EC 2.7.7.24, FC = 5.1) for TDP-D-Glucose (vancosamine precursor). The results suggested that every step of the whole pathway of the biosynthesis of norvancomycin was significantly upregulated in the industrial production strain, from the abundant supply of amino acid and glucose precursors to the NRPS assembling and the post modification of the glycopeptide antibiotic.

AoStrR1 and AoLuxR1 positively regulate the biosynthesis of norvancomycin

The expression of antibiotic biosynthetic genes is usually regulated by cluster-specific regulators within the gene cluster. There are four putative regulatory genes (*AoLuxR1*, *AoStrR1*, *AoTetR1*, *AoAraC1*) located within or adjacent to the *nvcm* cluster in the norvancomycin producing strains. The homologues of the four genes also present in a vancomycin producing strain *A. orientalis* KCTC 9412^T (Fig. S2). But in another vancomycin producing strain *A. keratiniphila* HCCB 10007, only homologues of *AoLuxR1*, *AoStrR1* and *AoTetR1* are present (Fig. S2). Differential expression analysis revealed that the regulatory genes *AoLuxR1* and *AoStrR1* were significantly upregulated in the high-yield strain, with the same trend as the structural genes of *nvcm* clusters (Fig. 4A, Table S6). The transcriptional levels of these two regulatory genes increased since 24 h and were 23.3-fold and 5.8-fold at 48 h, respectively (Fig. 4A, Table S6). *AoTetR1* and *AoAraC1* located near the *nvcm* cluster showed 2 ~ 3 fold higher transcription level at 12 h, 24 h and 48 h in the industrial strain than that in the original strain, but in a different trend as the structural genes of *nvcm* clusters, and the transcription level (FPKM value) were much lower than that of other genes in *nvcm* cluster (Fig. 4A, Table S6).

The *AoStrR1* gene is located in the *nvcm* cluster (7,787,412 ~ 7,788,377 bp), with a total length of 966 bp, encoding the AoStrR1 protein of 321 amino acids. Aligning the amino acid sequences of 22 BGCs of known glycopeptide antibiotics from GenBank (Table S7) showed that the homologues of AoStrR1 appeared in almost all searched clusters (Fig. 4B, Table S8), with only exception of corbomycin, the newly discovered compound [11]. These amino acid sequences were used to build phylogenetic tree by MEGA-X program [26] based on Neighbor-Joining method. Compared with the clades of Tcp28 (from teicoplanin BGC), AoStrR1 tended to group with the regulators encoded in BGCs for vancomycin, decaplanin, ristocetin, ristomycin A, avopsrcin, nogabecin, keratinimicin, balhimycin, chloroeremomycin, pekiskomycin and A40926 (Fig. 4B). Among them, Bbr and Dbv4, which was confirmed as a positive regulator in balhimycin [20] and A40926 [22] cluster respectively, are highly homologous to the AoStrR1 protein with a consistency of 84% and 80%, respectively (Table S8). Thus, we speculated that AoStrR1 might be a pathway-specific regulator of norvancomycin biosynthetic genes, and the increased transcriptional level of *AoStrR1* may be an important factor to trigger production of norvancomycin.

The *AoLuxR1* gene is close to the right border of *nvcm* cluster (7,724,319 ~ 7,724,996 bp), with a total length of 678 bp, encoding the AoLuxR1 protein of 225 amino acids. There are 14 possible LuxR-like regulators identified in or adjacent to the 22 glycopeptide BGCs (Fig. 4C, Table S9). The LuxR phylogenetic tree revealed two main clades (Fig. 4C). AoLuxR1 appeared to be related to the regulators encoded in BGCs for vancomycin, decaplanin, ristocetin, ristomycin A, keratinimicin and nogabecin, with a consistency of 83% ~ 98% (Fig. 4C, Table S9). The other clade included Dbv3 (from A40926 BGC) and Tcp29 (from teicoplanin BGC), which have larger size (more than 500 amino acids) and share low consistency with AoLuxR1 (Fig. 4C, Table S9). These results suggested that the function or regulatory target of AoLuxR1 may be different from well-characterized Dbv3 or Tcp29. Although the AoLuxR1 homologous gene was not found in the well known balhimycin BGC in *A. balhimycina* DSM 5908 (there is no adjacent ORF sequences available in GenBank), it is highly conserved in the reported two vancomycin producing strains (Fig. S2, Table S9).

In order to determine the regulatory function of AoStrR1 and AoLuxR1, we constructed *AoStrR1* and *AoLuxR1* over-expression plasmids based on pULVK2A vector [27], under its native promoter or *ermE**_p, a strong constitutive promoter, respectively, and then conjugated into *A. orientalis* CPC 200066. Norvancomycin yield in the fermentation broths of the AoStrR1 / AoLuxR1 over-expressing strains were detected by HPLC and LC-MS. Fermentation results showed both AoStrR1 and AoLuxR1 genes could increase norvancomycin production, especially under the *ermE**_p promoter, which led to a 2 ~ 5 times higher norvancomycin yield (Fig. 4DE). These results indicated that AoStrR1 and AoLuxR1 both acted as activators for norvancomycin biosynthesis.

AoStrR1 binds to most of promoter regions in the *nvc* cluster

Cluster-specific regulatory proteins generally activate the transcription by binding to the promoter regions of one or more structural genes within a biosynthetic gene cluster, thereby promoting the biosynthesis of secondary metabolites and ultimately increasing their yield. In order to determine the potential target genes of AoStrR1 and AoLuxR1, we tried to express and purify his-tagged AoStrR1 and AoLuxR1 in *E. coli* BL21(DE3), and then perform electrophoretic mobility shift analysis (EMSA). Unfortunately, although different expression conditions were conducted, no soluble his-tagged AoLuxR1 was detected in the supernatant of recombinant *E. coli* BL21(DE3). The soluble his-tagged AoStrR1 was expressed and purified by nickel affinity chromatography (Fig. 5A).

To identify AoStrR1 target genes, 20 intergenic regions in the norvancomycin cluster responsible for the transcription of almost all structural genes and possible regulator genes were amplified and labeled by biotin as probes (200–500 bp) for EMSA experiments. The results showed that the recombinant His₁₀-AoStrR1 could form a stable complex with the promoter regions upstream of *vanY*, *AoStrR1*, *oxyA*, *oxyB*, *vhp*, *hmaS*, *vasA* and *ald* (Fig. 5BD). Addition of 100-fold unlabeled specific competitive DNA attenuated the shift band, indicating that the bindings of His₁₀-AoStrR1 to the above probes are specific (Fig. 5BD). The sequences of above eight AoStrR1 binding sites were input into the GLAM2 [28] and MEME [29] software and the consensus binding sequence of AoStrR1 was identified as GTCCAN₁₈TTGGAC containing an incomplete palindromic sequence (Fig. 5C). The AoStrR1 binding motif sequence is highly similar to the reported consensus sequence of its homologues Bbr, Dbv4 and Tcp28 (Tei15*) [20, 21, 23]. This indicated that the StrR-like regulators may have a conservative regulatory mechanism in different glycopeptide BGCs.

Based on the EMSA results, AoStrR1 might be the ultimate overall positive regulator responsible for norvancomycin production by binding directly to almost all promoters in *nvc* cluster, including those of the genes and operons responsible for Bht biosynthesis (Vhp), Hpg biosynthesis (HmaS), vancosamine biosynthesis (VasA), linear heptapeptide cyclization (OxyAB) and self-resistance of norvancomycin (VanY). In addition, AoStrR1 did not bind to the promoters of the other 3 regulators (*AoLuxR1*, *AoTetR1*, *AoAraC1*) described here but its own promoter. When comparing the regulatory targets of those reported

StrR-like regulators in different glycopeptide BGCs, it is interesting that except for Tcp28 (Tei15*), other three regulators (AoStrR1, Bbr and Dbv4) control one common biosynthetic step, heptapeptide cyclization (*oxyA*), and both AoStrR1 and Bbr could bind to promoter region of genes responsible for specialized amino sugar biosynthesis (*vasA / dvaA*) in their respective gene clusters. Different from Dbv4 and Tcp28 (Tei15*), Bbr and AoStrR1 are able to bind to its own upstream region. It suggested that the positive feedback mechanism of AoStrR1 might be responsible for the significant upregulation of norvancomycin production in the industrial strain. Moreover, among the four StrR-family regulators discussed above, only AoStrR1 could bind to the upstream of *vanY*, part of the putative self-resistance genes. Given that self-resistance regulatory system VanRS is lack in *nvcm* cluster, it is tempting to speculate that AoStrR1 is somehow involved in the regulation of self-resistance.

Conclusion

In this study, we compared the genomes of original norvancomycin producing strain CPCC 200066 and industrial strain *A. orientalis* NCPC 2–48. Three SVs and some InDels of small fragments (≤ 50 bp) were found in the genome of high-yield strain compared with the original strain, and it is difficult to give a clear clue on its high-yield mechanism of norvancomycin. We further compared the transcriptomes of CPCC 200066 and NCPC 2–48 in three time points to identify differential expression genes involved in the production of norvancomycin. The results showed that the transcriptional upregulation of most *nvcm* biosynthetic genes and genes for precursor supply led to the high yield of norvancomycin in the industrial strain. Furthermore, two positive pathway-specific regulators of norvancomycin production were confirmed by overexpression of AoStrR1 and AoLuxR1. AoStrR1 positively regulates most *nvcm* genes by directly binding to multiple promoter regions in *nvcm* cluster. Thus, this study provides insights into the high yield mechanism and regulatory mechanism of norvancomycin in *A. orientalis* NCPC 2–48, which sets a foundation for future strain improvement to help overcome production capacity limitations.

Materials And Methods

Strains, plasmids and growth conditions

The norvancomycin-producing strains *A. orientalis* CPCC 200066 and NCPC 2–48 were grown at 28 °C on solid Bennet medium (peptone 0.2%, glycerol 1%, beef extract 0.1%, glucose 1%, malt extract 0.3%, yeast extract 0.1%, agar 2%, pH 7.2.) for sporulation and liquid Bennet medium for fermentation. The Mannitol soya flour (MS) medium agar [30] was used for conjugation between *A. orientalis* and *E. coli*. Tryptic soy broth liquid medium [12] was used to strain grown for isolation of genomic DNA. *E. coli* ET12567/pUZ8002 were used for conjugal transfer according to the established protocol [30]. All the strains and plasmids used in this study are listed in Table S1 and Table S2.

Genomic DNA extraction and sequencing

Genomic DNA were extracted from *A. orientalis* CPCC 200066 and NCPC 2–48 strains using the DNA extraction kit (TANBead, China) according to the manufacturer's instructions. Genomic DNA library

preparation, Illumina sequencing, chromosome assemble and annotation were carried out at Beijing Genomics Institute (Shenzhen, China), and performed as described [12]. The second-generation sequencing platform Illumina Hiseq 2000 and a third generation sequencing platform Pacbio RSII were used for high-throughput sequencing.

RNA extraction and sequencing

RNA extraction. Strains were collected at 12 h, 24 h and 48 h at the beginning of the fermentation, and total RNA were extracted using the TRIzol reagent (Invitrogen, USA) and chloroform followed by a PureLink™ RNA Mini Kit (Invitrogen, USA) according to the kit's instructions. The RNA samples from the three different time points of the original strain were named B12, B24, and B48, respectively. The RNA samples of high-yield strains were named N12, N24 and N48, respectively.

cDNA library construction and RNA-sequencing. The above 6 RNA samples were handed over by Beijing Genomics Institute (Shenzhen, China) to build cDNA library. The second-generation sequencing platform BGISEQ-500 was used for high-throughput sequencing. After removing low-quality reads, the filtered data was compared with the reference sequence using HISAT [31] and Bowtie2 [32] tools.

Differentially expressed genes (DEGs). The RSEM tool [33] was used to quantify gene expression. To eliminate the influence of gene length and sequencing quantity, the results of gene expression quantification were mapped to FPKM (fragments per kilobase of exon per million fragments mapped). We calculated the differential expression of the gene between different samples based on the FPKM value. In our analysis, differentially expressed genes (DEGs) were defined as genes with a fold change more than two times and $FDR \leq 0.001$. Hierarchical clustering analysis on DEGs were performed by cluster [34, 35] and show on javaTreeView [36]. The functional enrichment of DEGs was analyzed by KEGG [37]. The software Integrated Genome Browser [38] was used to create and view the whole genome expression on transcriptome level for each sample.

AoStrR1 and AoLuxR1 over-expression in *A. orientalis* CPCC 200066

The vector pULVK2A[27] and pULVK2A-*ermE**p which containing an *ermE**p promoter, were used for gene overexpression. The 1310 bp fragment containing the coding region of *AoStrR1* and its native promoter or 1005 bp fragment containing the coding region of *AoLuxR1* and its native promoter were amplified from *A. orientalis* CPCC 200066 genomic DNA by PCR with specific primers (Table S2), and were cloned into *HindIII*-*Bam*HI sites of pULVK2A to obtain the p2A-AoStrR1 and p2A-AoLuxR1. The 981 bp coding region of *AoStrR1* or 678 bp coding region of *AoLuxR1* were amplified from *A. orientalis* CPCC 200066 genomic DNA by PCR with specific primers (Table S2) and were cloned into *NdeI*-*Bam*HI sites of pULVK2A-*ermE**p to obtain the p2A-*ermE**p-AoStrR1 and p2A-*ermE**p-AoLuxR1. The constructed vector were introduced into *A. orientalis* CPCC 200066 by intergeneric conjugation from *E. coli* ET12567/pUZ8002 according to the established protocol [39].

Strain fermentation and norvancomycin yield detection

A. orientalis CPCC 200066 and its derivatives were cultured in 100 ml Bennet liquid medium at 28°C for 2 days, and then 10% seed culture was transferred into 100 ml fresh Bennet for continuous fermentation at 28 °C for 4 days. Each fermented supernatant was collected by centrifugation at 12,000 rpm for 10 min, and then filtered with microporous membrane of 0.22 µm and analyzed by HPLC-MS (Agilent 1290-Agilent 1956 single quadrupole MS coupled system). The HPLC conditions were as follows: Agilent Eclipse plus C18 column (250 mm × 4.6 mm, 5 µm), mobile phase Solvent A was 100% MeOH, Solvent B was water with 0.1% formic acid. The HPLC program included column elution with a linear gradient of 5 to 30% solvent A (100% MeOH) over 30 min at 25°C. The flow rate was set at 0.8 ml/min. MS spectra data were collected in the positive-ion mode in which a mass range of m/z 150 to 2,000 covered. The NVCM peak was extracted ion chromatogram (EIC) of m/z 717.9 [M + 2H]²⁺.

AoStrR1 and AoLuxR1 protein expression and purification

The AoStrR1 and AoLuxR1 coding sequence were amplified by PCR from *A. orientalis* CPCC 200066 genomic DNA using specific primers (Table S2). The amplified DNA fragment were cloned into *NdeI*-*Bam*HI sites of pET-16b (Novagen, Madison, USA) to give pET16b-AoStrR1 or pET16b-AoLuxR1, and then expressed as fusion proteins with the His₁₀-tag in *E. coli* BL21(DE3). The transformed strains were grown in LB medium until an OD₆₀₀ of 0.8 and induced by adding 0.1 mM IPTG for a further 7 h at 28°C. The bacteria were harvested by centrifugation (5,000 rpm, 10 min, 4°C), and resuspended in 25 ml binding buffer (20 mM NaH₂PO₄, 500 mM NaCl, and 20 mM imidazole, pH 7.4), subsequently, lysed by high pressure continuous flow cell cracker (Constant systems, TS 0.75 kW). The lysate was centrifuged (12,000 rpm, 20 min, 4°C) for removing the residue, and the His₁₀-tagged AoStrR1 protein present in the supernatant. A HisTrap™ FF crude kit was used to purify recombinant protein, as described by the manufacturer (GE Healthcare). Fractions eluted from the column with 500 mM imidazole were dialyzed against 1 × TGEK buffer (250 mM Tris, 50% Glycerol, 5 mM EDTA, 500 mM KCl, pH 8.0) at 4°C by the PD10 Desalting Columns (GE Healthcare), and then stored at -80°C. The concentration of His₁₀-tagged AoStrR1 was determined using BCA™ Protein Assay Kit (Pierce Biotechnology, Rockford, USA) and its purity assessed using SDS-PAGE analysis.

Electrophoretic mobility shift analysis (EMSA)

DNA fragments containing the promoter regions of the *nvcm* genes were obtained by PCR using primers labeled at their 5'-ends with Biotin (Table S2) and used as probes in EMSAs. The EMSAs were performed using LightShift® Chemiluminescent EMSA Kit (Thermo Scientific). Each 20 µl binding reaction consisted of 2 µl 10 × binding buffer (100 mM Tris-HCl, 500 mM KCl, 10 mM DTT, pH 7.5), 2 µg salmon sperm DNA, 20 fmol labeled probe and 1 µM purified His₁₀-tagged protein. The specific competitive reactions were carried out by adding 2 pmol competitors (unlabeled probes) apart from ingredients above. Reaction mixtures were incubated at room temperature for 20 min and then analyzed using a native 5% TBE-acrylamide gel and run at 4 °C, 100 V for 80 min. After electrophoretic transfer to nylon membrane

(Hybond-H⁺, GE Amersham), the probes were visualized by Lightshift Chemiluminescent EMSA Kit (Pierce Biotechnology), according to the manufacturer's instructions.

Bioinformatics Analyses

Amino acid sequences of proteins homologous to AoStrR1 and AoLuxR1 were retrieved using Blastp tools and were further manually curated to ensure its location within or near the glycopeptide biosynthetic gene cluster. The phylogenetic trees were calculated using the MEGA-X program based on Neighbor-Joining method [26]. The consensus motif of AoStrR1 binding sequence was represented as logos, obtained at the WebLogo website, using GLAM2 [28] and MEME [29] algorithm.

Data availability

The RNA-sequencing data are publicly available at NCBI's GenBank's repository under NCBI BioProject ID: PRJNA624813 (<https://dataview.ncbi.nlm.nih.gov/object/PRJNA624813?reviewer=qm5eugf6h4jq922e3sagtea24c>), BioSample ID: SAMN03273949 & SAMN14585755, SRA ID: SRR11529125 ~ 127 & SRR11529348 ~ 350.

Declarations

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its additional information files.

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors agreed to publish this article.

Conflict of interest

The authors declare no financial or commercial conflict of interest.

Funding

This work was supported by the National Natural Science Foundation of China (81703398, 81872780 and 81630089), the Drug Innovation Major Project of China (2018ZX09711001-007-001 and 2018ZX09711001-006-011), CAMS Innovation Fund for Medical Sciences (2016-I2M-3-012 and 2018-I2M-3-005).

Authors' contributions

XXL and CZ performed the experiments, analyzed the primary data and wrote the draft manuscript. YZ, XL and XXZ assisted with the extraction of genomic DNA. YZ and ZHZ assisted with the extraction of RNA. XL assisted with the construction of overexpression strains and the fermentation. ZBJ and SYS assisted the chemical work in this study and revised the manuscript. BH and LFW supervised the whole research work and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank CPCC (China Pharmaceutical Culture Collection) for providing the original strain CPCC 200066. We also thank Beijing Genomics Institute (Shenzhen, China) for Illumina Hiseq sequencing, assembly and annotation of the genome and transcriptome.

References

1. Kahne D, Leimkuhler C, Lu W, Walsh C. Glycopeptide and lipoglycopeptide antibiotics. *Chem Rev.* 2005; 105:425–48.
2. Van Bambeke F. Lipoglycopeptide Antibacterial Agents in Gram-Positive Infections. A Comparative Review *Drugs.* 2015;75::2073–95.
3. Jeong H, Sim YM, Kim HJ, Lee DW, Lim SK, Lee SJ. Genome Sequence of the Vancomycin-Producing *Amycolatopsis orientalis* subsp. *orientalis* Strain KCTC 9412T. *Genome Announc.* 2013; 1.
4. Pelzer S, Sussmuth R, Heckmann D, Recktenwald J, Huber P, Jung G, Wohlleben W. Identification and analysis of the balhimycin biosynthetic gene cluster and its use for manipulating glycopeptide biosynthesis in *Amycolatopsis mediterranei* DSM5908. *Antimicrob Agents Chemother.* 1999;43:1565–73.
5. Sosio M, Stinchi S, Beltrametti F, Lazzarini A, Donadio S. The gene cluster for the biosynthesis of the glycopeptide antibiotic A40926 by *nonomuraea* species *Chem Biol.* 2003;10:541–9.
6. Li TL, Huang F, Haydock SF, Mironenko T, Leadlay PF, Spencer JB. Biosynthetic gene cluster of the glycopeptide antibiotic teicoplanin: characterization of two glycosyltransferases and the key acyltransferase. *Chem Biol.* 2004;11:107–19.
7. Sosio M, Kloosterman H, Bianchi A, de Vreugd P, Dijkhuizen L, Donadio S. Organization of the teicoplanin gene cluster in *Actinoplanes teichomyceticus*. *Microbiology.* 2004;150:95–102.

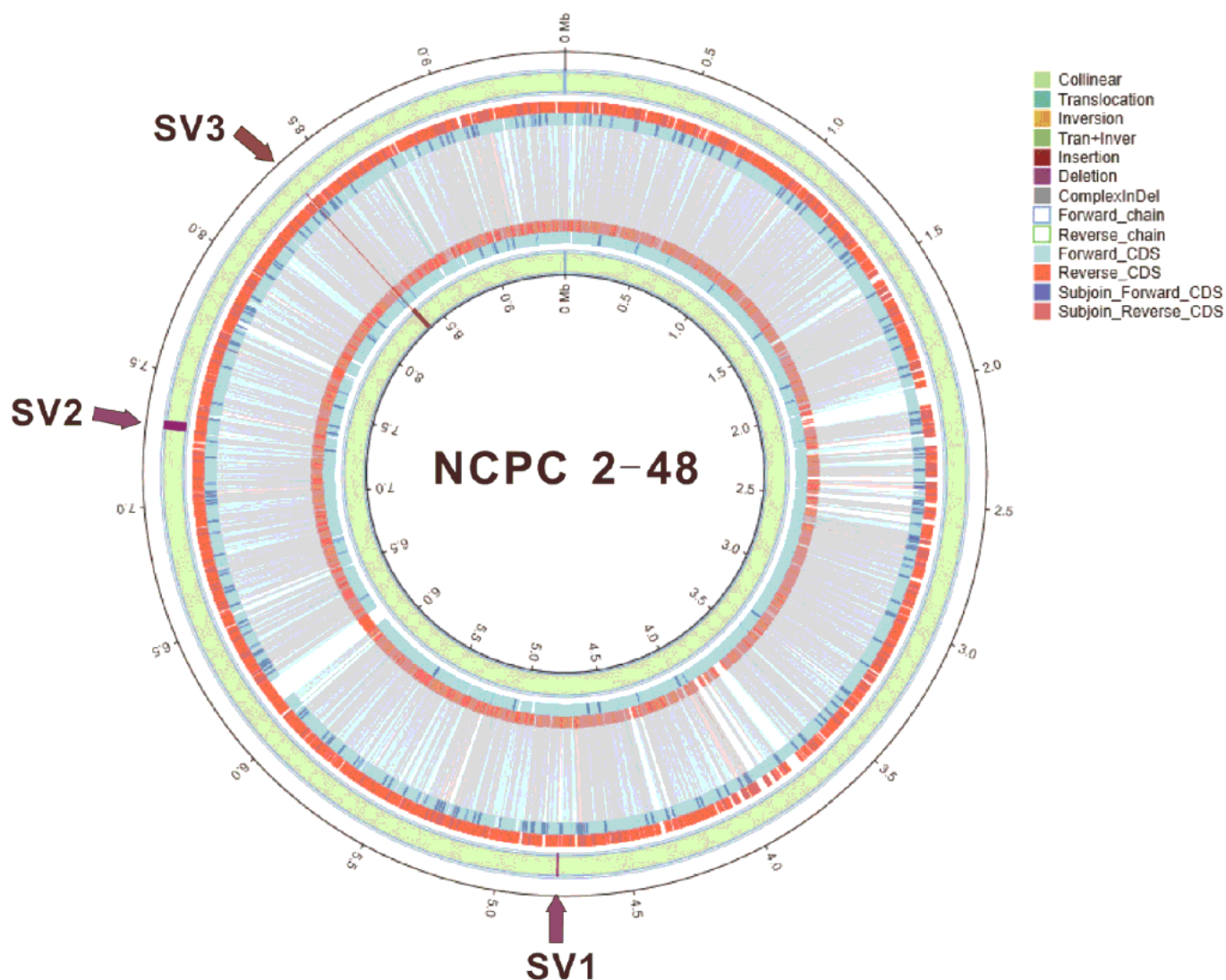
8. Thaker MN, Wang W, Spanogiannopoulos P, Waglechner N, King AM, Medina R, Wright GD. Identifying producers of antibacterial compounds by screening for antibiotic resistance. *Nat Biotechnol.* 2013;31:922–7.
9. Nazari B, Forneris CC, Gibson MI, Moon K, Schramma KR, Seyedsayamdost MR. *Nonomuraea* sp. ATCC 55076 harbours the largest actinomycete chromosome to date and the kistamicin biosynthetic gene cluster. *Medchemcomm.* 2017;8:780–8.
10. Waglechner N, McArthur AG, Wright GD. Phylogenetic reconciliation reveals the natural history of glycopeptide antibiotic biosynthesis and resistance. *Nat Microbiol.* 2019;4:1862–71.
11. Culp EJ, Waglechner N, Wang W, Fiebig-Comyn AA, Hsu YP, Koteva K, Sychantha D, Coombes BK, Van Nieuwenhze MS, Brun YV, Wright GD. Evolution-guided discovery of antibiotics that inhibit peptidoglycan remodelling. *Nature.* 2020;578:582–7.
12. Lei X, Zhang C, Jiang Z, Li X, Shi Y, Liu M, Xie Y, Wang L, Hong B. Complete genome sequence of *Amycolatopsis orientalis* CPC200066, the producer of norvancomycin. *J Biotechnol.* 2017;247:6–10.
13. Yim G, Thaker MN, Koteva K, Wright G. Glycopeptide antibiotic biosynthesis. *J Antibiot (Tokyo).* 2014;67:31–41.
14. Kaniusaite M, Tailhades J, Kittila T, Fage CD, Goode RJA, Schittenhelm RB, Cryle MJ. Understanding the early stages of peptide formation during the biosynthesis of teicoplanin and related glycopeptide antibiotics. *FEBS J.* 2020.
15. Xu L, Huang H, Wei W, Zhong Y, Tang B, Yuan H, Zhu L, Huang W, Ge M, Yang S, et al: Complete genome sequence and comparative genomic analyses of the vancomycin-producing *Amycolatopsis orientalis*. *BMC Genomics.* 2014; 15:363.
16. Hubbard BK, Walsh CT. Vancomycin assembly: nature's way. *Angew Chem Int Ed Engl.* 2003; 42:730 – 65.
17. Alduina R, Tocchetti A, Costa S, Ferraro C, Cancemi P, Sosio M, Donadio S. A Two-Component regulatory system with opposite effects on glycopeptide antibiotic biosynthesis and resistance. *Sci Rep.* 2020;10:6200.
18. Kilian R, Frasch HJ, Kulik A, Wohlleben W, Stegmann E. The VanRS Homologous Two-Component System VnRSAb of the Glycopeptide Producer *Amycolatopsis balhimycina* Activates Transcription of the vanHAXSc Genes in *Streptomyces coelicolor*, but not in *A. balhimycina*. *Microb Drug Resist.* 2016;22:499–509.
19. Beltrametti F, Consolandi A, Carrano L, Bagatin F, Rossi R, Leoni L, Zennaro E, Selva E, Marinelli F. Resistance to glycopeptide antibiotics in the teicoplanin producer is mediated by van gene homologue expression directing the synthesis of a modified cell wall peptidoglycan. *Antimicrob Agents Chemother.* 2007;51:1135–41.
20. Shawky RM, Puk O, Wietzorrek A, Pelzer S, Takano E, Wohlleben W, Stegmann E. The border sequence of the balhimycin biosynthesis gene cluster from *Amycolatopsis balhimycina* contains bbr, encoding a StrR-like pathway-specific regulator. *J Mol Microbiol Biotechnol.* 2007;13:76–88.

21. Alduina R, Lo Piccolo L, D'Alia D, Ferraro C, Gunnarsson N, Donadio S, Puglia AM. Phosphate-controlled regulator for the biosynthesis of the dalbavancin precursor A40926. *J Bacteriol.* 2007;189:8120–9.
22. Lo Grasso L, Maffioli S, Sosio M, Bibb M, Puglia AM, Alduina R. Two Master Switch Regulators Trigger A40926 Biosynthesis in *Nonomuraea* sp. Strain ATCC 39727 *J Bacteriol.* 2015;197:2536–44.
23. Horbal L, Kobylansky A, Truman AW, Zaburranyi N, Ostash B, Luzhetskyy A, Marinelli F, Fedorenko V. The pathway-specific regulatory genes, *tei15** and *tei16**, are the master switches of teicoplanin production in *Actinoplanes teichomyceticus*. *Appl Microbiol Biotechnol.* 2014; 98:9295 – 309.
24. Yushchuk O, Horbal L, Ostash B, Marinelli F, Wohlleben W, Stegmann E, Fedorenko V. Regulation of teicoplanin biosynthesis: refining the roles of *tei* cluster-situated regulatory genes. *Appl Microbiol Biotechnol.* 2019; 103:4089 – 102.
25. Swint-Kruse L, Matthews KS. Allostery in the LacI/GalR family: variations on a theme. *Curr Opin Microbiol.* 2009; 12:129 – 37.
26. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol.* 2018;35:1547–9.
27. Kumar CV, Coque JJ, Martin JF. Efficient Transformation of the Cephamycin C Producer *Nocardia lactamdurans* and Development of Shuttle and Promoter-Probe Cloning Vectors. *Appl Environ Microbiol.* 1994;60:4086–93.
28. Frith MC, Saunders NF, Kobe B, Bailey TL. Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput Biol.* 2008; 4:e1000071.
29. Bailey TL, Williams N, Misleh C, Li WW. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 2006;34::W369-73.
30. Kieser T, Bibb M, Buttner M, Chater K, Hopwood D. *Practical Streptomyces Genetics*. Norwich: The John Innes Foundation; 2000.
31. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015; 12:357 – 60.
32. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10:R25.
33. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:323.
34. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A.* 1998;95:14863–8.
35. de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software *Bioinformatics.* 2004;20:1453–4.
36. Saldanha AJ. Java Treeview—extensible visualization of microarray data. *Bioinformatics.* 2004;20:3246–8.

37. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 2008;36:D480-4.
38. Nicol JW, Helt GA, Blanchard SG Jr, Raja A, Loraine AE. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics.* 2009;25:2730–1.
39. Li X, Yu T, He Q, McDowall KJ, Jiang B, Jiang Z, Wu L, Li G, Li Q, Wang S, et al. Binding of a biosynthetic intermediate to AtrA modulates the production of lidamycin by *Streptomyces globisporus*. *Mol Microbiol.* 2015;96:1257–71.

Figures

CPCC200066



Chromosome Features	Length (bp)	G +C content (%)	Protein coding genes	rRNA genes (operons)	tRNA genes	sRNA genes
CPCC200066	9,490,992	68.83	8,643	12 (4)	50	14
NCPC 2-48	9,489,689	68.84	8,705	12 (4)	50	14

Figure 1

Comparison of genomic sequences of *A. orientalis* CPCC 200066 and NCPC 2-48. The inner cycle and outer cycle represent the genome of NCPC 2-48 and CPCC 200066, respectively. Delete region (SV1 and SV2) are indicated by purple arrows, which are deleted in genome of NCPC 2-48 with a length greater than

50 bp; Insertion region (SV3) is indicated by brown arrows, which is inserted in genome of NCPC 2-48 with a length greater than 50 bp.

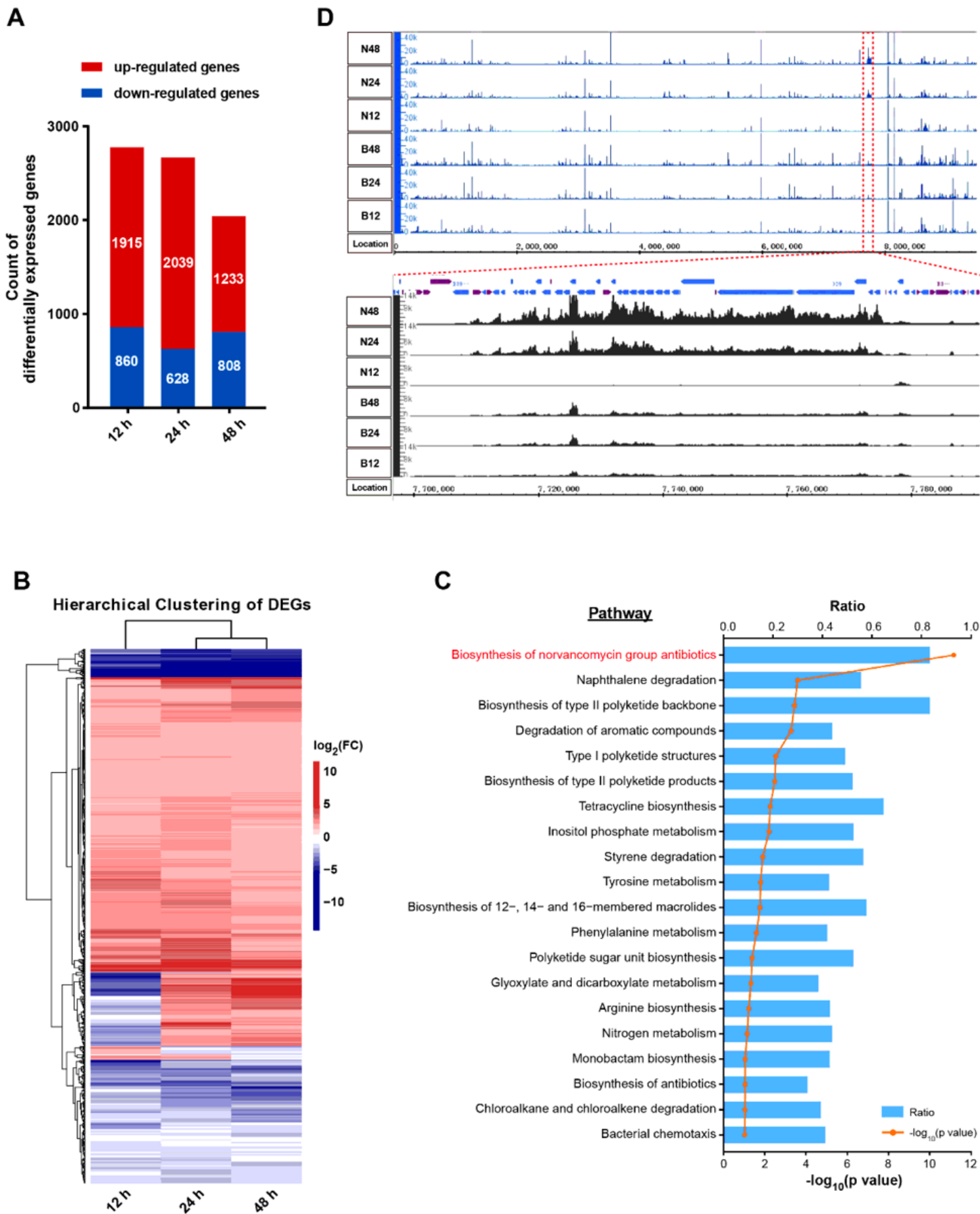
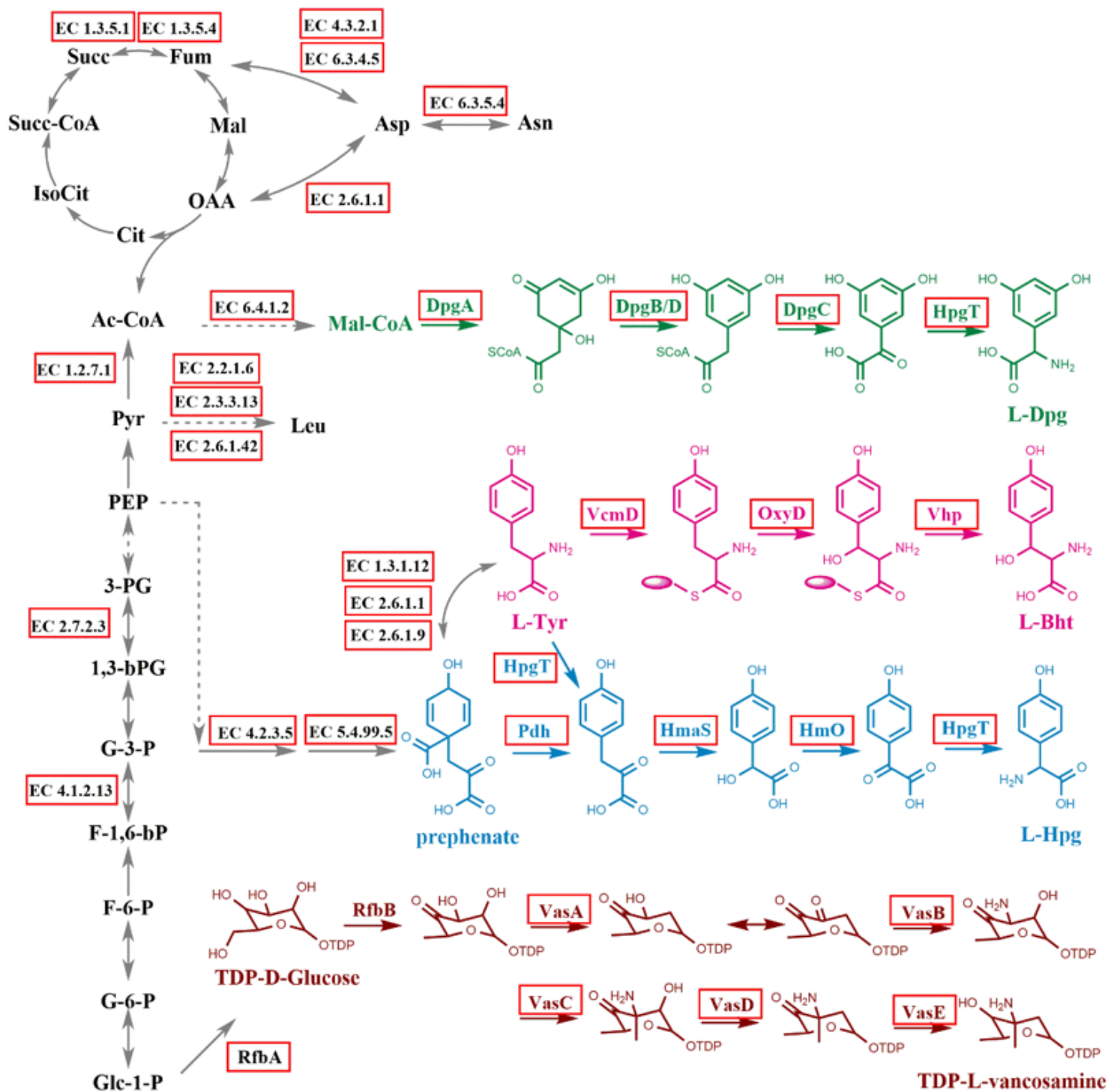


Figure 2

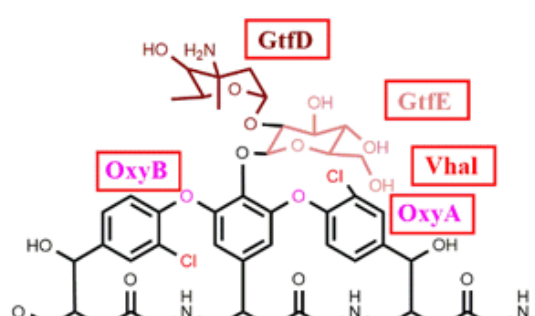
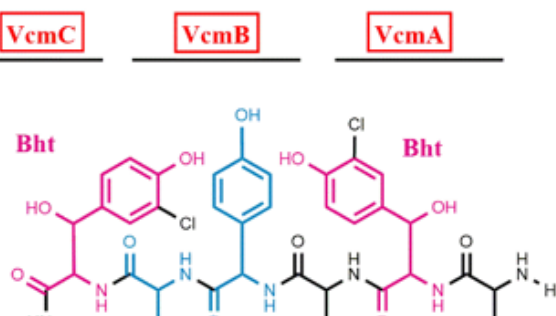
Comparative transcriptomics analysis of *A. orientalis* CPCC 200066 and NCPC 2-48. (A) NCPC 2-48's differentially expressed genes (DEGs) with greater than two-fold change (FC) and $FDR \leq 0.001$ in expression compared to the CPCC 200066 at each time points. (B) Hierarchical clustering of DEGs for each time points. Only DEGs that showed in all pairwise were used to build this heatmap. Gradient color barcode at the right top indicates $\log_2(\text{fold change})$ value. Each row represents a gene and each column represents a pairwise. Genes with similar expression value are clustered both at row and column level. (C) Statistics of pathway enrichment of DEGs in 24 h's pairwise. Ratio is the DEGs numbers annotated in this pathway term to all gene numbers annotated in this pathway term. Greater Ratio means greater intensiveness. Orange line is $-\log_{10}(\text{p-value})$. We just display the top 20 of enriched pathway terms. (D) The whole genome (top) and nvcn cluster (below) expression on transcriptome level for each sample. The samples from the three different time points (12 h, 24 h, 48 h) of the original strain CPCC 200066 were named B12, B24, and B48. The samples of high-yield strains NCPC 2-48 were named N12, N24 and N48.

I. the biosynthesis of amino acid precursors and vancosamine



II. Linear heptapeptide assembled by NRPSs

III. The post-modifications of the linear heptapeptide



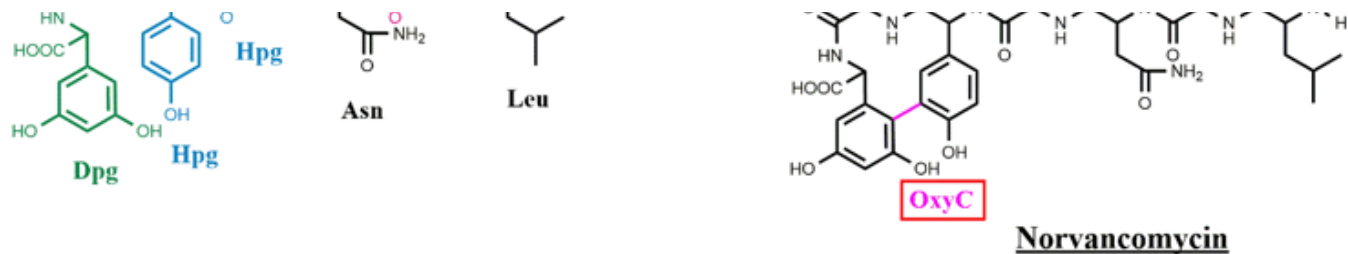
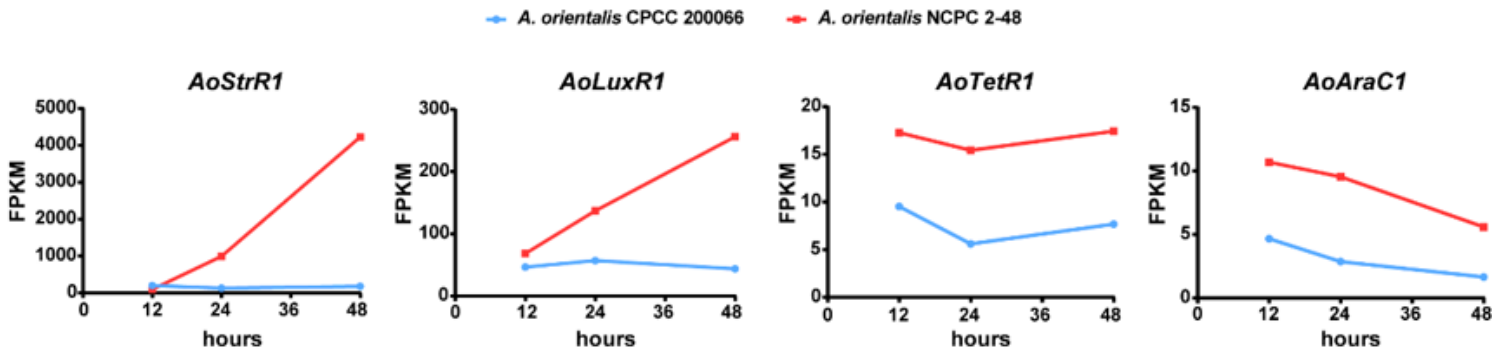
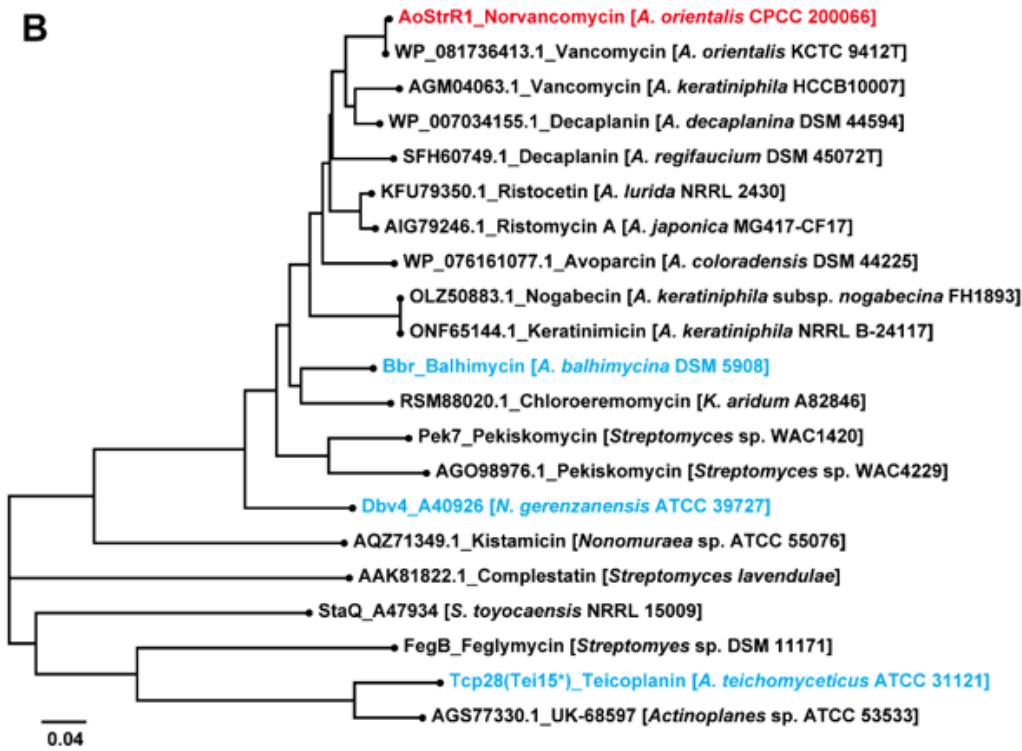
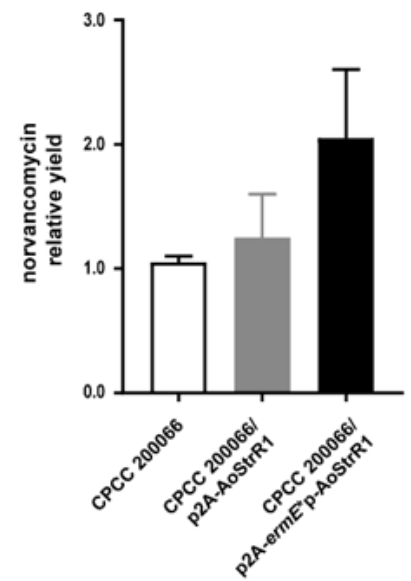
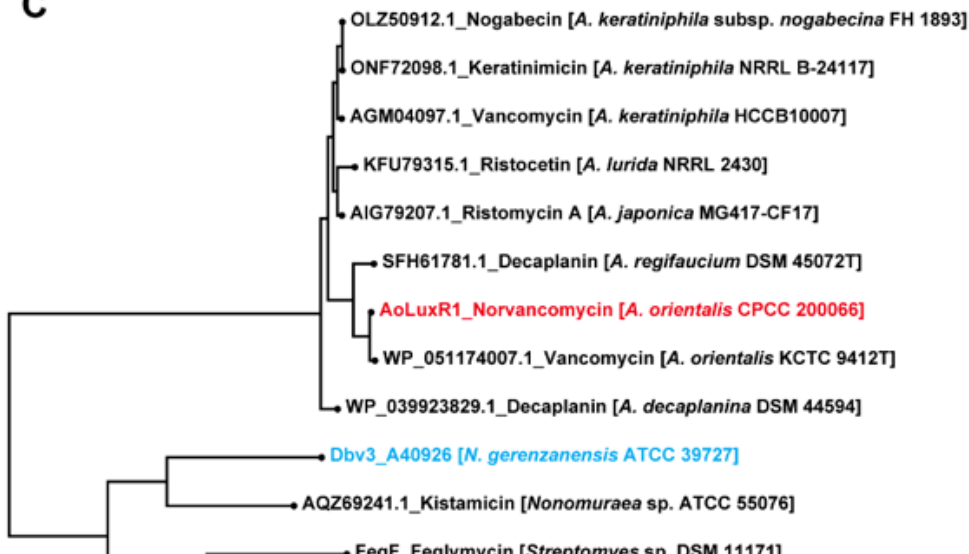
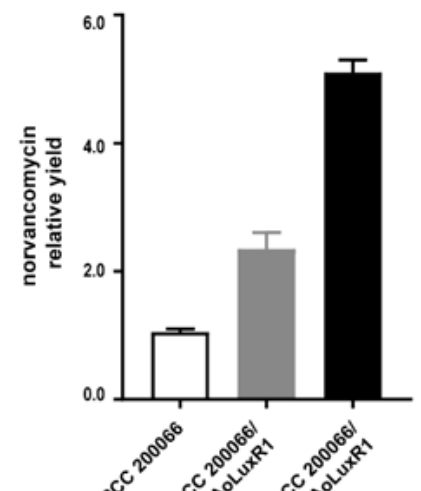


Figure 3

Metabolic pathway of norvancomycin biosynthesis up-regulated in high-yield strain. Three steps are involved in the biosynthesis of norvancomycin: I) the biosynthesis of amino acid precursors, II) The modified amino acids are assembled to form linear heptapeptide by NRPSs, III) The post-modifications of the linear heptapeptide. Red box at each gene indicates that its transcriptional level was up-regulated (two-fold change and $FDR \leq 0.001$) in NCPC 2-48 at 24 h.

A**B****D****C****E**

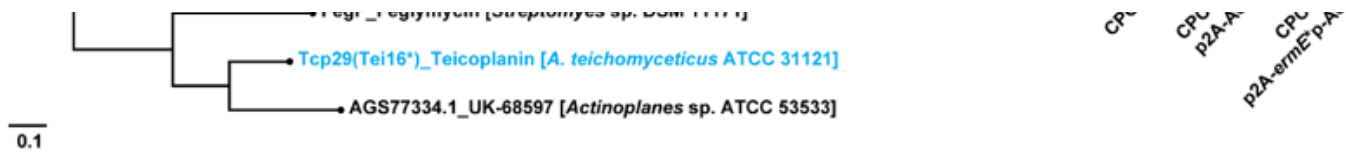


Figure 4

AoStrR1 and AoLuxR1 are cluster-specific positive regulators for norvancomycin biosynthesis. (A) The expression level of four potential regulatory genes at each time points on RNA-seq assay. The gene expression quantification is based on the FPKM (fragments per kilobase of exon per million fragments mapped) value. (B, C) The phylogenetic tree of known and putative StrR-like and LuxR-like transcriptional regulators encoded in glycopeptide BGCs. The AoStrR1 and AoLuxR1 in norvancomycin BGC are labeled with red. Previously experimentally studied regulators are labeled with blue. The final topology was derived from 1000 bootstraps. (D, E) Norvancomycin yield in the fermentation product of the original strain CPC 200066 and AoStrR1 / AoLuxR1-overexpressed strains. p2A and p2A-ermE*p represent two overexpressed strains which contain its native promoter or ermE*p strong promoter. Values are presented as means + SEM (two biological repeats for each strain).

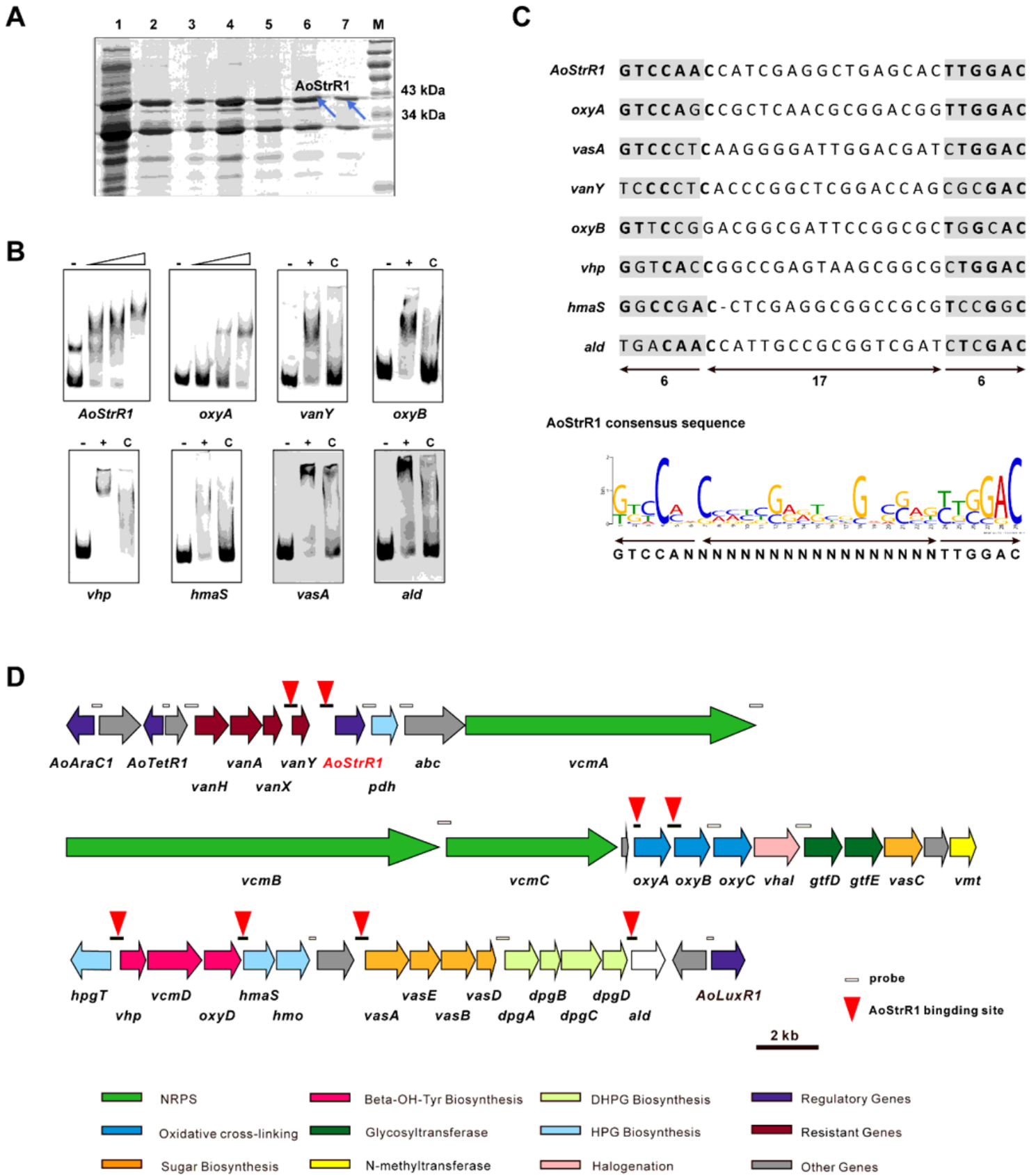


Figure 5

AoStrR1 binds to the promoter regions of the *nvcM* cluster. (A) Overexpression and purification of His10-tagged AoStrR1 in *E. coli*. SDS-PAGE showing purification of N-terminal His10-tagged AoStrR1 by nickel-

affinity chromatography. lanes 1- 5, elution with 500 mM imidazole; lanes 6-7, dialyzed by the PD10 Desalting Columns. M stands for protein size marker. (B) AoStrR1 binds to the promoter region of these eight genes identified by EMSA. Lane -, probe only; Lane +, 1000 nM AoStrR1 incubated with the probe; Lanes under the triangle, increasing concentrations of His10-AoStrR1 (0, 400, 800, 1,000 nM) incubated with the probe; Lane C, 1000 nM His10-AoStrR1 incubated with 100-fold excess unlabeled specific competitor DNA fragment. (C) Sequence alignment of the putative AoStrR1 binding sites. Consensus sequence was generated using MEME. Arrows indicate inverted repeats and the double-headed arrow denotes the spacer. The most highly conserved nucleotides are bold. (D) The genes present in norvancomycin biosynthesis gene cluster are indicated by arrows. The short bars are the probes in the nvcm cluster. Red triangles are the binding sites of His10-AoStrR1.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [NVCMmanuscriptV7supplhb.docx](#)