Journal of Biotechnology 247 (2017) 6–10

Contents lists available at ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

Short communication

Complete genome sequence of *Amycolatopsis orientalis* CPCC200066, the producer of norvancomycin

Xuan Lei[a,1], Cong Zhang[a,1], Zhibo Jiang[a], Xingxing Li[a], Yuanyuan Shi[a], Ming Liu[b], Yunying Xie[a], Lifei Wang[a,*], Bin Hong[a,*]

[a] *Key Laboratory of Biotechnology of Antibiotics, The National Health and Family Planning Commission (NHFPC), Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050, China*
[b] *Immunology Program Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA*

ARTICLE INFO

Article history:
Received 9 September 2016
Received in revised form 13 February 2017
Accepted 13 February 2017
Available online 16 February 2017

Keywords:
Amycolatopsis orientalis CPCC200066
Complete genome sequence
Norvancomycin
Integrative vector pIMBT1

ABSTRACT

Amycolatopsis orientalis CPCC200066 is an actinomycete exploited commercially in China for the production of norvancomycin, an important glycopeptide antibiotic structurally close to the well-known vancomycin. The availability of the complete genome sequence of CPCC200066 would greatly strengthen our understanding of the regulation pattern of norvancomycin biosynthesis and ultimately improve its production, as well as potentiate discoveries of novel bioactive compounds. Here we report the complete genome sequence of *A. orientalis* CPCC200066, a circular chromosome consisting of 9,490,992 bp. Forty putative secondary metabolite biosynthetic gene clusters, including norvancomycin, were predicted, covering 20.3% of the whole genome. To facilitate genetic manipulation of this strain, an efficient transformation system was established by constructing a novel integrative vector pIMBT1, which could be transferred into CPCC200066 by electroporation with high efficiency. ΦBT1 *attB* sites were also identified in other known *Amycolatopsis* genomes, indicating pIMBT1's prospect to be a novel vector for genus *Amycolatopsis*.

Amycolatopsis orientalis CPCC200066 (originally named wan-23, deposited at the China Pharmaceutical Culture Collection) is a Gram-positive bacterium, isolated from a soil sample collected in the Guizhou Province, China in 1959. It was first discovered for its ability to produce an antibiotic that resembles the glycopeptide antibiotic vancomycin, and the antibiotic was confirmed to be norvancomycin in 1983. The strain was commercially developed by the North China Pharmaceutical Company for the production of norvancomycin, which lacks an *N*-methyl group at the *N*-terminus of the polypeptide of vancomycin. Norvancomycin resembles vancomycin in terms of antibacterial activity, exhibiting a wide range of inhibitory activity against Gram-positive bacteria, especially MRSA (Methicillin-resistant *Staphylococcus aureus*) and MRSE (methicillin-resistant *Staphylococcus epidermidis*), by non-covalent binding to the *C*-terminus of the Lys-D-Ala-D-Ala-peptide of bacterial cell wall precursors (Nicolaou et al., 1999). Norvancomycin has been widely used in China to treat severe infections such as endocarditis and osteomyelitis (Wu et al., 2012) for over 3 decades. Complete genome sequence information of CPCC200066 will enable us not only to better understand the biosynthesis of norvancomycin and its detailed regulatory mechanism, but also to genetically manipulate its production and potentiate discoveries of novel compounds.

The strain was grown at 28 °C in tryptic soy broth liquid medium (Lei et al., 2015), and genomic DNA was extracted using a DNA extraction kit (TANBead, China). The genome sequence of *A. orientalis* CPCC200066 was sequenced on a second generation sequencing platform, Illumina Hiseq2000, and a third generation sequencing platform, Pacbio RSII, resulting in 1093 Mb data (14,846,050 reads with 500 bp average insert size and 140-fold average coverage) and 732 Mb data (108,128 subreads with 6771 bp average length and the N_{50} length of the subreads is 8347 bp) separately. The genome was directly assembled into one contig (9,501,279 bp) by SOAPdenovo v2.04 (Li et al., 2010) based on the subreads obtained from Pacbio RSII platform. The linear contig was then checked for any overlap between the beginning and end sequences. The sequence of 2–10310 bp was identified to match 9,490,993–9,501,279 bp in the contig, confirmed a 9,490,992 bp chromosome with circular topology. SOAPsnp (Li et al., 2009), SOAPindel (Li et al., 2013) and GATK (McKenna et al., 2010)

* Corresponding authors.
E-mail addresses: lifeiwang2002@hotmail.com, wanglifei@imb.pumc.edu.cn (L. Wang), binhong69@hotmail.com, hongbin@imb.pumc.edu.cn (B. Hong).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jbiotec.2017.02.013
0168-1656/

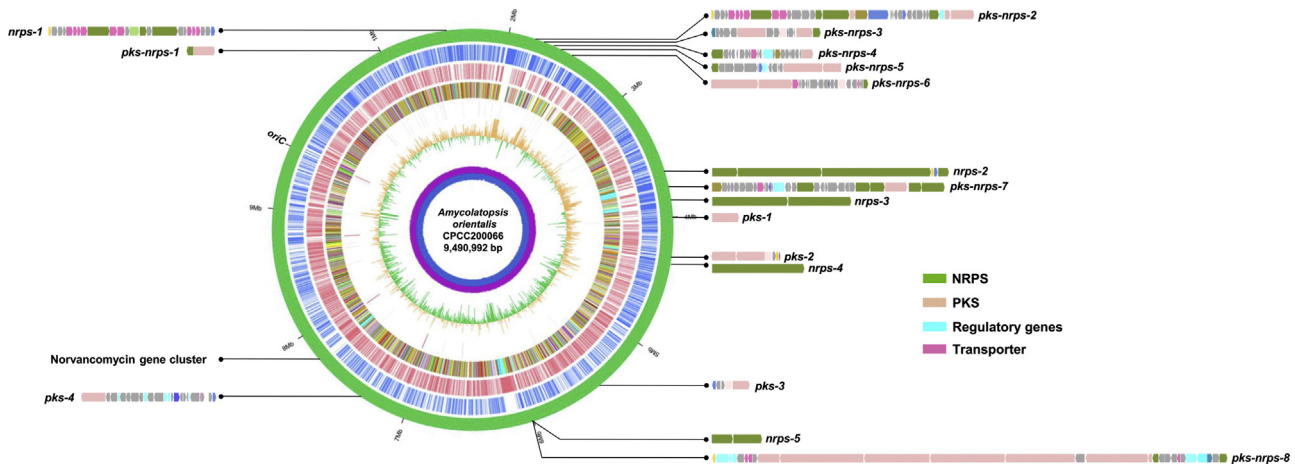

Fig. 1. Circular genome map of *A. orientalis* CPCC200066. The circles are numbered from the outside in and the outer scale is numbered in megabases. Circle 1 represents the genome. Predicted modular PKS and NRPS are illustrated outside the circle and the norvancomycin biosynthetic gene cluster is further illustrated in Fig. 2; circle 2 and 3 are genes in forward and backward strand respectively; circle 4 indicates the annotated CDSs according to COG functional categories; circle 5 denotes tRNA and rRNA; circle 6 illustrates the GC skew (G – C/G + C); circle 7 and 8 demonstrates G and C content respectively.

Table 1
Genome features of *Amycolatopsis orientalis* CPCC200066.

Feature	Chromosome
Length (bp)	9,490,992
G +C content (%)	68.8
Total genes	8356
Protein coding genes	8211
tRNA genes	50
rRNA genes	12
ncRNA genes	3
Pseudo genes	80
Hypothetical protein with unknown function	2582
Protein with function assigned	5629
GenBank accession	**CP016174**

were used to correct the errors found in the assembled sequence based on the data obtained from Illumina Hiseq2000 platform. Protein-coding sequences (CDSs) were predicted using Glimmer 3.0 (Delcher et al., 2007). Gene functional annotation was performed using BLASTP with KEGG, COG, Swiss-Port, TrEMBL, NR and GO databases. tRNA-encoding genes and rRNA operons were found by software tRNAscan and rRNAmmer.

A. orientalis CPCC200066 has a circular chromosome of 9,490,992 bp with a G + C content of 68.8% (Fig. 1). The genome of CPCC200066 totally contains 8356 genes, which include 8211 protein coding genes, 80 pseudo genes and 65 RNA genes (12 rRNA genes organized into 4 rRNA operons, 50 tRNA genes and 3 non-coding RNA (ncRNA) genes) (Table 1). Circular chromosome were also identified in the other seven complete genomes of *Amycolatopsis* strains currently available, entailing that this is a common feature for the genus *Amycolapotsis* (Table S1).

Amycolatopsis belongs to a rare actinomycetes genus, which contains important producers of bioactive secondary metabolites including clinically used antibiotics, such as vancomycin and rifampicin. As novel anti-multidrug-resistant Gram-negative bacteria antibiotics are urgently needed more than ever before, it has promoted us to explore the secondary metabolites of *A. orientalis* CPCC200066. In the fermentation extracts of *A. orientalis* CPCC200066, antibacterial activity against the highly refractory pathogen *Pseudomonas aeruginosa* as well as an efflux pump mutant strain *Escherichia coli ΔtolC* (Fralick, 1996) were identified in addition to *Bacillus subtilis*, which is ascribed to norvancomycin (Best and Durham, 1964) (Fig. S1). AntiSMASH (Weber et al., 2015) analysis of the complete genome of *A. orientalis* CPCC200066 revealed the presence of forty putative biosynthetic gene clusters, which is estimated at ~1926 kb, 20.3% of the whole genome. Compared with the antiSMASH results of the other seven known complete genomes of *Amycolatopsis* strains, *A. orientalis* CPCC200066 contains the most secondary metabolite biosynthetic gene clusters and the highest secondary metabolite biosynthetic gene cluster density (Table S1), holding great potential for producing diverse and novel antibiotics. At least four type I PKS, six NRPS and eight hybrid PKS-NRPS gene clusters were found in the genome (Fig. 1). In accordance to the assembly line mechanism (Fischbach and Walsh, 2006), chemical backbones that each cluster would synthesis were predicted based on the results of antiSMASH and NRPS-PKS knowledgebase (http://www.nii.ac.in/~pksdb/sbspks/master.html) (Table 2).

Norvancomycin is a heptapeptide synthesized by *nrps-6* (Table 2) and 36 genes (*SD37_33590* to *SD37_33765*) were identified in the gene cluster, which display a high resemblance with reported gene clusters of vancomycin (*vcm*) in *A. orientalis* HCCB10007 (Xu et al., 2014) and *A. orientalis* subsp. *orientalis* KCTC 9412^T (Jeong et al., 2013) (Table S2). A StrR-like regulator SD37_33745 was identified within the norvancomycin biosynthetic gene cluster (*nvcm*), and its homologous genes were also found in the reported glycopeptide antibiotic biosynthesis gene clusters including vancomycin, balhimycin, chloroeremomycin, teicoplanin, A40926 and A47934 (Donadio et al., 2005). In balhimycin gene cluster (*bal*) the StrR-like regulator Bbr was characterized to regulate some of the biosynthetic genes at a transcriptional level (Shawky et al., 2007). Four vancomycin resistant genes (*vanH*, *vanA*, *vanX* and *vanY*) were identified next to *strR* in the norvancomycin biosynthetic gene cluster (Fig. 2). Vancomycin resistant genes, such as *vanH*, *vanA*, *vanX* and *vanY* typically under the control of a sensory membrane-bound His-Kinase (*vanS*) and its cognate response regulator (*vanR*) (Yim et al., 2014), have been reported to be present in the known glycopeptide antibiotic biosynthesis gene clusters. For example, *vanHAX* are present in the vancomycin gene cluster in *A. orientalis* HCCB10007 (Xu et al., 2014). In balhimycin gene cluster, *vanRS* and *vanY* are present within the cluster while *vanHAX* locates at least 2 Mb away from the cluster (Schaberle et al., 2011) (Fig. 2). The vancomycin resistant genes may affect the production of glycopeptide antibiotics, e.g., as inactivation of *vanAst* resulted in a delay in A47934 production (Pootoolal et al., 2002). In the genome of *A. orientalis* CPCC200066, a two-component system (*SD37_04185* and *SD37_04180*) similar to the *vanRS* of *bal* was identified, but no homologues of *vanRS* were located immediately next to *vanHAXY*. Interestingly, three more regulators were present closely adjacent to the identified norvancomycin gene cluster, which are