COMPLETE SEQUENCE OF LANDOMYCIN E BIOSYNTHETIC GENE CLUSTER FROM STREPTOMYCES GLOBISPORUS 1912

Streptomyces globisporus 1912 and their derivatives 1912-4Crt and 1912-2 are the producers of the landomycin E, carotenoids and regulatory diketopiperazine, respectively. The genome DNA of two mutant strains, 1912-2, the more effective producer of the landomycin E and regulator, and 1912-4Crt, the producer of beta-carotene and lycopene, was sequenced by Illumina. Comparative analysis of the DNA sequences of two neighboring contigs of 1912-2 and one contig of 1912-4Crt using GenBank data allowed localization of 36 landomycin E biosynthetic genes Ind of S. globisporus 1912 in one cluster. Twenty of these Ind genes have been sequenced for the first time. The new regulatory response gene IndRR and the sensor kinase gene IndY1 were proposed as the members of the putative two-component system. A high identity (94-95%) was found for the Ind genes of the 1912 strain and those of the metagenomic clone AZ97, and the lower similarity (80-85%) of Ind and Lan genes of S. cyanogenus S136 encoding landomycin A biosynthesis. Two direct repeats of 21 bp were shown in the crtY gene coding lycopene cyclase. A deletion in the IndRR gene renders the 1912-4Crt strain deficient in landomycin E production.

Keywords: Streptomyces globisporus, landomycin E, biosynthetic gene cluster, sequences.

Aromatic polyketide antibiotics are known to present a major and important group of antimicrobial and anticancer medicinal drugs [4, 18]. The family of angucycline antibiotics occupies a prominent place among polyketides [11, 16, 17]. The unique peculiarity of the angucycline structures is the angularly condensed ring A in the benz[a]antraquinone type tetracyclic aglycone moiety. Natural landomycins differ among themselves by a varying phenol-glycosidically linked oligosaccharide chain. To date, three landomycin (La) producing microorganisms have been discovered. Streptomyces cyanogenus S136 produces LaA, B, C and D [7], Streptomyces globisporus 1912 accumulates LaE and D [12], and Streptomyces albus synthesizes LaE as a result of heterologous expression of soil DNA from the metagenomic clone AZ97 [6]. LaE contains a trisaccharide chain: \( \alpha-L\text{-}rhodinose-(1-3)-\beta-D\text{-}olivose-(1-4)-\beta-D\text{-}olivose \) (Fig. 1). The deoxysugar chain of LaA is represented by hexasaccharide chain of four D-olivose and two L-rhodinose residues. Landomycin E is one of the more potent antitumour antibiotics of the angucycline family. It induces the apoptosis of the tumor cells in a G0/G1 phase including those resistant to doxorubicin [10, 16, 17]. LaA (lan) and LaE (Ind) biosynthetic genes have been cloned and their similar cluster organization was described [5, 6, 17, 19]. The complete sequences of 14 Ind genes of S. globisporus 1912 (prx, IndI, F, M, GT1, J, Z1, Z3, GT4, Z5, Y, YR, W2, W) were established earlier [5, 16, 17] and registered in the GenBank (access numbers: AY443343-45, AM258820, AM608714-15, AM640377, AM659997-8, DQ139409, DQ275159, HM204451). Information about genetic organization and sequence of the landomycin biosynthetic gene cluster may be used for the construction of hybrid enzymes by the combination of aromatic polyketidases from different producers of angucyclines [8, 15]. The subject of this paper was the determination of the sequences of the all Ind genes of the landomycin E cluster from S. globisporus 1912, including 20 previously unsequenced genes, and investigating the cause of the lack of expression of the pathway in mutant strain 1912-4Crt.

Materials and methods. Strains and media. Two strains, the derivatives of the initial wild-type strain S. globisporus 1912, were used. The strain 1912 was isolated from a soil sample from Armenia and is stored in the Ukrainian Collection of Microorganisms at the Institute of Microbiology and Virology of the National Academy of Sciences of Ukraine as the strain Ac-2098 (http://www.inv.kiev.ua/images/doc/catalog/UCM_catalog.pdf). The strain 1912-2 is the more effective producer of landomycin E (200 mg/l) and a new regulatory diketopiperazine, generated by the action of nitrosoguanidine on the spores of the strain 1912 (14). The strain 1912-4Crt is the spontaneous mutant of 1912 producing beta-carotene (7.0 mg/l) and lycopene (3.24 mg/l), and not producing landomycin E [13]. Cultures were grown at 28°C in 20 ml liquid medium S [9] in 250 ml shake flask at speed © B.P. Matselyukh, L.V. Polishchuk, V.V. Lukyanchuk, 2015
240 rot/min during 48 h. Glycine was added to the medium at concentration 1% in order to increase the sensitivity of the cell wall to lytic action of the lysozime.

Isolation of chromosomal DNA. Chromosomal DNA of *S. globisporus* 1912-2 and 1912-4Crt was prepared according to the standard protocol of the Kirby method, dissolved in TE buffer, pH 8.0 and stored at 4°C (9). The A$_{260}$/A$_{280}$ index of the DNA preparations was equal to 2.1 – 2.3, and agarose gel electrophoresis showed one compact and localized band of DNA.

**DNA sequencing.** DNA sequencing was carried out in BaseClear B.V. (Leiden, Netherlands) using the following procedures. The FASTQ sequence reads were generated according to Illumina Casava pipeline version 1.8.3. Initial quality assessment was based on data passing the Illumina Chastity filtering. Subsequently, reads containing adapters and/or PhiX control signal were removed using an in-house filtering protocol. The second quality assessment was based on the remaining reads using the FASTQ quality control tool version 0.10.0. The final quality scores per sample are provided as Enclosure. The quality of the FASTQ sequences was enhanced by trimming off low-quality bases using the “Trim sequences” option of the CLC Genomics Workbench version 6.0.4. The quality-filtered sequence reads were assembled into a number of contig sequences. The analysis has been performed using the “De novo assembly” option of the CL C Genomics Workbench version 6.0.4. The optimal k-mer size was automatically determined using KmerGenie [3]. The contigs were linked and placed into scaffolds or supercontigs. The orientation, order and distance between the contigs was estimated using the insert size between the paired-end and/or matepair reads. The analysis has been performed using the SSPACE Premium scaffolder version 2.3 [1]. The gapped regions within the scaffolds were (partially) closed in an automated manner using GapFiller version 1.10 [2].

**Results and Discussion.** Sequence analysis of the genome of *S. globisporus* 1912. The genomes of the mutants 1912-2 and 1912-4Crt, according to Illumina paired end sequencing data, were represented by 1438 and 917 sequences of different length, summary sizes of which put together 7,125 kb and 7,368 kb, respectively. These numbers were around 90% of the length of a mid-sized streptomycete genome.

Sequence analysis of the landomycin E biosynthetic gene cluster. Localization of the *Ind* genes in the assembled genome was identified with the BLAST program (www.ncbi.nlm.nih.gov/blast) using information about sequences of *Ind* and *lan* genes from the GenBank data. The identity of 14 *Ind* genes, described earlier, and corresponding *Ind* genes of the contig 220/23 (length 37416 bp) and contig 478 (length 4708 bp) of the strain 1912-2 was found to be 99-100%. The sequences of the last 20 *Ind* genes of the landomycin E biosynthetic gene cluster from *S. globisporus* 1912 were identified and localized in the contig 220/23 on the basis of high identity (94-95%) with the *Ind* genes from the metagenomic clone AZ97 (GenBank HQ828984). Comparison of the above mentioned *Ind* genes and *lan* genes, encoding biosynthesis of the landomycin A in *S. cyanogenus* S136 (GenBank AF080235), showed a lower degree of homology (identity 80-85%). Characteristics of the *Ind* and *lan* genes from the three different clusters are presented in the Table 1. The genome of the strain S136 contains additional three *lan* genes involved in the biosynthesis of the hexasaccharide chain (*lanGT3, lanZ2*) and regulation (*lank*). The genetic organization of the landomycin E biosynthetic gene cluster from *S. globisporus* 1912 is presented in the Fig. 2. The direction of transcription of the *prx* and *IndI* genes was reversed in comparison with earlier published data [17]. The sequence
of 780 bp downstream of the \textit{IndZ6} gene was identified as the new response regulatory gene \textit{IndRR}, and the sequence of the \textit{IndY} gene was divided into two probable genes, \textit{IndY1} and \textit{IndY2}. The genes \textit{IndRR} and \textit{IndY1} belong to a putative two-component system involved in the regulation of the landomycin E biosynthesis. The evidence for this assumption lies in a mutational change of the \textit{IndRR} sequence. This gene in the 1912-4Crt strain has deletion of 86 bp, from 31702 to 31787 bp (GenBank KJ645792) resulting in a strain deficient in landomycin E biosynthesis. The sequence of this deletion is underlined:

\begin{verbatim}
31681  TTCTGGCGGG TGGTTCCGCA ACGGCGCGGG GGCAGCTGCC GTGATCTCAC
31741  CCCGCGCCGC GGTTCCCACG ATCTGACGCG GGGATTCAGC CAGGCCGTCG
\end{verbatim}

Table 1

\begin{table}
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
Gene & Strain and gene size, bp & Gene localization & Gene product & \\
\hline
\textit{prx} & - & 1499 & " & endopeptidase \\
\textit{lndI*} & - & 786 & 786 & two-component regulatory system \\
\textit{lndE} & 1476 & 1485 & 1485 & oxygenase \\
\textit{lndF} & 330 & 330 & 330 & cyclase \\
\textit{lndA} & 1275 & 1236 & 1236 & ketosynthase \\
\textit{lndB} & 1221 & 1218 & 1218 & chain length factor \\
\textit{lndC} & 270 & 273 & 273 & acyl carrier protein \\
\textit{lndD} & 786 & 786 & 786 & polyketide ketoreductase \\
\textit{lndL} & 960 & 936 & 936 & cyclase \\
\textit{lndM} & 2250 & 2238 & 2247 & oxygenase-reductase \\
\textit{lndO} & 582 & 582 & 582 & reductase \\
\textit{lndP} & 1575 & 1575 & 1575 & decarboxylase \\
\textit{lndG} & 1068 & 1068 & 1068 & NDP-hexose synthetase \\
\textit{lndH} & 981 & 987 & 936 & NDP-hexose 4,6-dehydrogenase \\
\textit{lndQ} & 1305 & 1305 & 1305 & NDP-hexose 3,4-dehydratase \\
\textit{lndR} & 762 & 762 & 762 & 4-keto-reductase \\
\textit{lndS} & 1410 & 1407 & 1407 & NDP-hexose 2,3-dehydratase \\
\textit{lndT} & 966 & 969 & 969 & oxidoreductase \\
\textit{lndU} & 648 & 975 & 975 & unknown \\
\textit{lndV} & 762 & 762 & 762 & unknown \\
\textit{lndGT2} & 1122 & 1176 & 1176 & glycosyl transferase \\
\textit{lndX} & 444 & 417 & 417 & isobutyryl-CoA mutase \\
\textit{lndGT1} & 1173 & 1169 & 1169 & glycosyl transferase \\
\textit{lndJ} & 1554 & 1563 & 1566 & transporter \\
\textit{lndZ1} & 579 & 570 & 570 & glycosyl transferase \\
\textit{lndZ3} & 981 & 951 & 927 & glycosyl transferase \\
\textit{lndGT4} & 1254 & 1251 & 1251 & glycosyl transferase \\
\textit{lndZ4} & 603 & 585 & 585 & putative glycosyltransferase \\
\textit{lndZ5} & 1194 & 1158 & 1176 & oxygenase \\
\textit{lndZ6} & 717 & 1500 & 1445 & oxygenase \\
\textit{lndRR} & - & - & 780 & two-component regulatory system \\
\textit{lndY1} & - & - & 723 & sensor kinase \\
\textit{lndY2} & - & - & 762 & kinase \\
\textit{lndYR} & - & - & 426 & GntR-like regulator \\
\textit{lndW2} & - & - & 996 & ABC transporter \\
\textit{lndW} & - & - & 954 & ATP-ase of ABC transporter \\
\hline
\end{tabular}
\caption{Characterization of landomycin A and E biosynthetic genes}
\end{table}

*The break between contigs 478 (\textit{lndI}) and 220/23 (\textit{lndE}) consists of 6 bp (5'-CGTCCG-3').

The \textit{Ind} gene cluster in DNA of the strain 1912-4Crt in comparison with the 1912-2 strain was localized in one contig N10 (length 71750 bp). Comparative analysis of both \textit{Ind} clusters from 1912-2 and 1912-4Crt showed 100% identity of the \textit{Ind} genes, except for \textit{IndRR} and \textit{crtY}. The break
between the 478 (GenBank KJ701191) and 220/23 (GenBank KJ45792) contigs of 1912-2 consists of 6 bp (downstream from \textit{lndI} to \textit{lndE}: 5’-CGTCCG-3’). Despite of the presence of the carotenoid biosynthetic gene cluster in the genome, the strain 1912-2 does not produce beta-carotene and lycopene under normal conditions, or after blue-light illumination, in quantities detectable by standard methods. Gene \textit{crtY} of 1912-2 contains two non-punctual direct repeats (NPDRs) of 21 bp each flanking a sequence of 96 bp. It was proposed that site-specific recombination between these NPDRs formed the deletion of 117 bp leading to the activation of the silent \textit{crt} gene cluster in the 1912-4Crt strain (GenBank KM349312):

\begin{verbatim}
5’-GGGGCGTGCGGAAGTCCATCA....
....GGGGCGGGCGGGAGTCGAACA – 3’
\end{verbatim}

Similar NPDRs of 21 bp were found in the \textit{crtY} genes of many other representatives of \textit{Streptomyces} and \textit{Bacteria} except for 100% homology of the first 6 bp 5’-GGGGCG-3’ which does not permit a site-specific recombination leading to next deletion.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{Organization of the landomycin E biosynthetic gene cluster. The \textit{lnd} genes, directions of their transcription and sizes (base pair) are indicated by letters, arrows and numbers, respectively.}
\end{figure}

The organization and sequence analysis of the carotenoid gene cluster from 1912-4Crt strain will be present in another publication.

Start and stop codons can be found at the beginning and end of the \textit{lnd} genes. The termination codon, TGA, of the genes \textit{lndA}, G, Q, T, GT2, Z4 and W2 overlapped the initiation codons, ATG or GTG, of the genes \textit{lndB}, H, R, U, X, Z5 and W, correspondingly. The rarest codon TTA, recognized in mRNA by a developmentally important tRNA, was found in the sequences of the regulatory gene \textit{lndI} and glycosyl transferase gene \textit{GT1}. The transcription of the majority of \textit{lnd} genes take place from (+) DNA strand and only the marginal genes \textit{prx}, \textit{lndI} in the left part of the cluster, and \textit{lndYR}, W2 and W in the right part, are transcribed from (–) strand (Fig. 2).

So, the following new results were obtained in this paper: the sequences of 20 early described landomycin E biosynthetic genes \textit{lnd} were identified and the sizes and the direction of transcription of some of the last 14 \textit{lnd} genes were defined more precisely; the response regulatory gene \textit{lndRR} from putative two-component system was proposed; two non-punctual direct repeats of 21 bp were shown in the \textit{crtY} gene coding lycopene cyclase; the explanation of the loss of the landomycin E production and activation of the \textit{crt} gene cluster in the strain 1912-4Crt was given; the overlapped startstop codon ATGA (more rarely GTGA) in the structures of the seven \textit{lnd} genes was identified and the rarest codon TTA was found in the sequences of two genes.

Despite the very different origins and relationship of the three \textit{Streptomyces} strains, there is a similarity between the landomycin A and landomycin E biosynthetic genes in respect of their size, sequence, order of localization and common function. An unexpected surprise was the very high similarity between the sequences and the length of the \textit{lnd} genes from \textit{S. globisporus} 1912 and metagenomic clone AZ97 isolated from the soil samples of Armenia and Arizona, respectively. This may be an example of the high degree of conservation of these genes in the nature or their recent evolutionary divergence.

Uncharacterized sequences between the \textit{lnd} genes may be further regulatory regions or have other functions and will be the subject of future investigations.
The differences between the landomycin biosynthetic gene clusters including the presence of extra genes in flanking regions and varied lengths, provides useful information for future studies into the biosynthetic mechanisms responsible for the accumulation of specific compounds and regulation of their biosynthesis.

The function and cluster organization of 14 lnd genes of S. globisporus 1912 were established earlier [5, 17]. The 100% identity of these sequences of 14 lnd genes identified earlier and the same genes reported in this paper may testify to correct identification of the last 20 lnd genes of the landomycin E cluster. Repeat Illumina analysis of the sequences of the lnd and crt gene clusters of two strains 1912-2 and 1912-4Crt showed their 100% identity except for two genes, lndRR and crtY, bearing the deletions. One can suppose that above mentioned direct repeats may be the recognition site for the unknown enzyme participating in the rearrangement of the genome DNA in S. globisporus 1912.

The complete sequences of landomycin E and carotenoid biosynthetic gene clusters of S. globisporus 1912 can be found in GenBank (accession numbers KJ645792, KJ701191 and KM349312).

Acknowledgement. This work was supported by the National Academy of Sciences of Ukraine (0111U000606).

B.P. Maceychuk, L.V. Poliuchuk, V.V. Luk’ianchuk
Institute of microbiology and virology, d.k. Zabolotnyi National Academy of Ukraine, 154, Kiev, 03143, Ukraine

THE COMPLETENESS OF THE LANDOMYCIN E CLUSTER

Streptomyces globisporus 1912 is the only isolate 1912-4Crt and 1912-2 is a producer of landomycin E, carotenoids and regulates a diketo-epiperoxidase of this nature, respectively. Genomic DNK of two strains: 1912-2 is a larger effective producer of landomycin E and regulates it, and 1912-4Crt - a producer of beta-carotin and lycopene. Sequencing was performed using Illumina. The comparative analysis of the last 20 genes of the genome DNA in S. globisporus 1912 and the sequences of the landomycin E cluster and carotenoid clusters of the same strains 1912-2 and 1912-4Crt showed their 100% identity except for two genes, lndRR and crtY, bearing the deletions. One can suppose that above mentioned direct repeats may be the recognition site for the unknown enzyme participating in the rearrangement of the genome DNA in S. globisporus 1912.

The complete sequences of landomycin E and carotenoid biosynthetic gene clusters of S. globisporus 1912 can be found in GenBank (accession numbers KJ645792, KJ701191 and KM349312).
геномного клона AZ97 и меньшую схожесть (80-85%) генов ind и lan S. cyanogenus S136, кодирующих биосинтез ландомицина А. Показано два не пунктуальные прямые повторы из 21 п.о. в гене crтY, кодирующего ликонин циклазу. Деления в гене indRR вызывают потерю штаммом 1912-4Cт способности продуцировать ландомицин Е.

Ключевые слова: Streptomyces globisporus, ландомицин Е, кластер биосинтетических генов, секвенирование.

Адрес автора: Мацелюх Б.П. Институт микробиологии и вирусологии им. Д.К. Заболотного НАН Украины, ул. Академика Заболотного, 154, Киев, 03143, Украина.


Отправлено 20.11.2014