ISSN 0206-5657. Вісник Львівського університету. Серія біологічна. 2024 Випуск 91. С. 14–21
Visnyk of the Lviv University. Series Biology. 2024. Issue 91. P. 14–21

ГЕНЕТИКА

УДК 579.873.1:577.181.4 HTTPS://DOI.ORG/10.30970/VLUBS.2024.91.02

MINING THE CRYPTIC SPECIALIZED METABOLOME OF *STREPTOMYCES CYANOGENUS* S136

I. Ostash[1,2], M. Deneka[1], M. Lopatniuk[3], T. Busche[4], J. Kalinowski[4], A. Luzhetskyy[3], V. Fedorenko[1], B. Ostash[1,*]

[1]Ivan Franko National University of Lviv
4, Hrushevskyi St., Lviv 79005, Ukraine
**e-mail: bohdan.ostash@lnu.edu.ua*
[2]Explogen LLC
20, Zelena St., Lviv 79005, Ukraine
[3]Department of Pharmacy, Saarland University
Campus C2.3, Saarbrucken 66123, Germany
[4]CeBiTec, Bielefeld University
25, Universitatstrasse, Bielefeld 33615, Germany

Streptomyces cyanogenus S136 is known to produce landomycin family antibiotics, particularly its largest congener, landomycin A. Except for landomycins and polyene lucensomycin, no other specialized metabolites were sourced from S136. Nevertheless, S136 genome sequencing revealed over 40 biosynthetic gene clusters (BGCs), implying underappreciated potential of this strain for the production of novel bioactive natural compounds. We set out to gain deeper insight into the specialized metabolome of this strain. First, transcriptomic analysis of S136 grown under landomycin production conditions has been carried out, revealing that most of them are expressed at a basal level. This, likely, leads to a phenotypic silence of most of the BGCs. Nevertheless, several notable exceptions have been spotted. First of all, landomycin BGC is expressed at high level (at least 100 Transcripts Per Million mapped reads (TPM); and around 1000 TPM for minimal polyketide synthase genes *lanFABC*). Similarly, high levels of expression showed BGCs # 2, 4, 7 and 33, of which #2, encoding unknown saccharide, is the most dissimilar to the described precedents. RNAseq data also allowed us to delineate better the borders of several presumed BGCs. In the next phase of the work we singled out a few BGCs within S136 that appeared to be promising. First, these BGCs exhibited low similarity to the other gene clusters directing the production of known natural products. Second, the BGCs harbored cluster-situated regulatory genes that can be employed in the attempts to activate the expression of cryptic pathways. For one such BGC we constructed two plasmids for expression of several such regulatory genes and introduced them into S136 and its derivative deficient in production of landomycin A. Bioassays showed no differences in bioactivity of the recombinant strains as compared to the initial strains. Liquid chromatography coupled to mass spectrometry (LC-MS) analysis of several *S. cyanogenus* samples revealed the effects of genotype, growth conditions and extraction on specialized metabolome of this species, setting reference point for further studies.

Keywords: *Streptomyces cyanogenus* S136, genes, physiology, cryptic specialized metabolome

Streptomyces cyanogenus S136 was first described in 1990s as a producer of angucycline polyketides landomycins that exhibit potent antiproliferative properties [1]. No other natural products were known to be produced by S136 until recent report on identification of lucensomy-

cin through manipulation of global transcriptional factor AdpA [2]. However, the aforementioned two classes of small molecules (Fig. 1) are perhaps a tip of an iceberg of natural products encrypted in S136 genome reported to harbour at least 40 BGCs [3]. This situation is typical for streptomycetes, whose genome sequencing routinely uncovers 30-40 BGCs, most of which remain a "white spot" in terms of the structure of encoded small molecule. BGCs, not expressed under laboratory conditions commonly employed to grow streptomycetes, are referred in literature to as dormant, silent or cryptic ones [4], although exact reason for disparate levels of phenotypic expression of a BGCs are not always known.

We are interested in uncovering true potential of S136 for the production of specialized metabolites. On one hand, this may lead to identification of novel natural products desperately needed to combat various diseases, first of all infectious ones. On the other hand, by studying the genetic control of biosynthesis of natural products in S136, we may arrive at novel solutions for landomycin A overproduction (e.g. by elimination of competing metabolic pathways). We applied transcriptomic and metabolomic approaches to S136 to understand which specialized pathways are likely to be expressed concomitantly with landomycin A accumulation, and to lay the ground for further genetic manipulations of selected pathways. Our studies agree that insufficient level of BGC transcription is the key reason for its crypticity. For several cases in S136 genome RNAseq data provide opportunity to better solve the problem of precise mapping of BGC borders, an issue that currently plagues all BGC identification pipelines [4]. Two plasmids for expression of putative cluster-situated regulatory genes have been generated, laying the groundwork for studies of promising naphthomycin-type BGC.

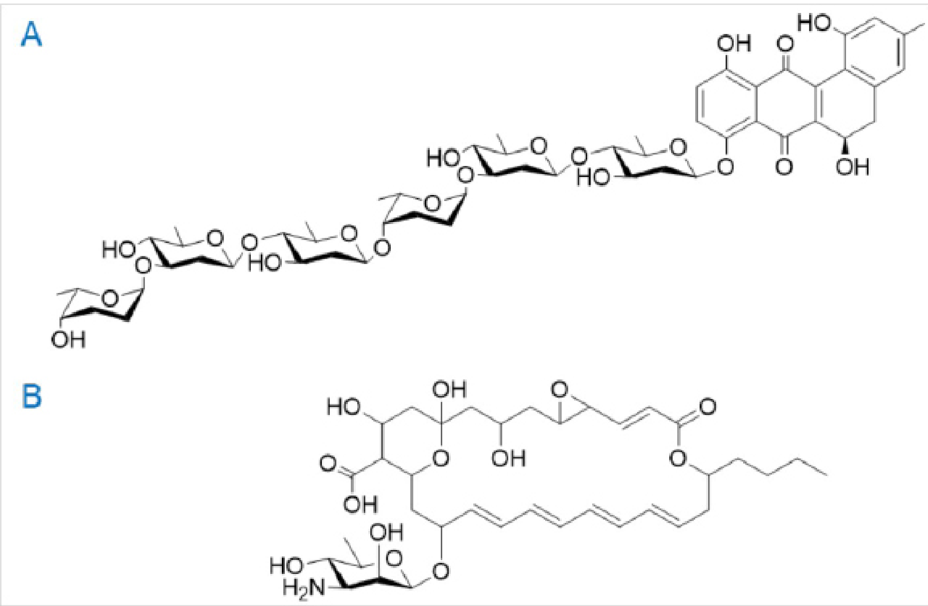

Fig. 1. Structural formulae of landomycin A (A) and lucensomycin (B), the only classes of natural products known to be accumulated by S136

Materials and Methods

Genomes and strains. *S. cyanogenus* S136 genome was accessed through NCBI (accession number CP071839). We used wild type S136 and its landomycin-deficient mutant ΔlanI7 [2] in experimental work, as described below. *E. coli* DH5α was used in routine cloning while

ET12567 (pUZ8002) [2] was employed to transfer plasmids into *S. cyanogenus*. *Bacillus cereus* ATCC19637 and *Debaryomyces hansenii* VKM-Y9 were used as bioassay strains.

Media. Biomass for RNAseq experiments was grown at 30 °C for 48 hours in medium SG [5] without calcium carbonate, which interferes with mRNA isolation procedures. Solid media OM, TSA, SFM, GYM, R5, GA, SMMS, MM, NL5 [6] were used to grow *S. cyanogenus* strains for agar plug bioassays. Liquid media YMPG, SG were also used to grow recombinant *S. cyanogenus* strains for LC-MS analysis. LB, its solid version LA and tryptic soy agar (TSA) were used to grow *E. coli* strains. To select strains apramycin sulfate (50 mcg/mL) was used.

Methods. For RNAseq analysis mRNA was isolated and purified from frozen mycelial samples and then sequenced (Illumina) essentially as described in [7]. Gene transcription values were calculated as "transcripts per million of mapped reads" (TPM) using Geneious Prime software. Values given in the Table 1 represent averages out of three biological repeats; deviations constitute no more than 20 % of the mean value. Orthologs were identified as a reciprocal BLASTP hits (https://blast.ncbi.nlm.nih.gov) [8]. Routine recombinant DNA techniques (DNA isolation, PCR, other enzymatic DNA treatments) were used to construct plasmids. Conditions of LC-MS experiments are described in [2]. Bioassays were carried out as described in [6]. Raw RNAseq and LC-MS data are available from authors upon request.

Plasmid pTESypdB. Gene *ypdB* (*S1361_33920*) located on the left fringe of BGC #37 (putative naphthomycin BGC, see Table 1) was amplified with primers ypdB_29_up (AAATCTAGA CCGTCGTCCGTGATCTCGAC) and ypdB_29_rp (AAAGAATTCACCGGTTCGGTCG CCATAC). The resulting 1.1-kb amplicon was digested with restriction endonucleases XbaI and EcoRI and cloned into respective sites of vector pTES [2], yielding plasmid pTESypdB where the cloned gene is located under control of strong constitutive promoter *ermEp**.

Plasmid pTESalkS. Gene *alkS* (*S1361_33925*) located near *ypdB* was amplified with primers alkS_29_up (AAATCTAGAGTCTTCCGAGCCAGGCCGTTC) and alkS_29_rp (AAAGAATTCGGCATCCCTCAGGTTGGTAG). The resulting 1.5-kb amplicon was digested with restriction endonucleases XbaI and EcoRI and cloned into respective sites of vector pTES, yielding plasmid pTESalkS.

Results and Discussion

RNAseq data portray complex picture of transcription of BGCs across S136 genome under landomycin A production conditions. We were interested to find out which specialized metabolic pathways are expressed under conditions that support high level of landomycin A production. For this purpose, we grew wild type S136 strain for 48 h in medium SG favorable for landomycin A biosynthesis and subjected to RNAseq analysis. Here we mostly focused on transcriptional levels of BGC scattered across S136 chromosome. All relevant data are summarized in Table 1. As a reference we used numbered list of S136 BGCs published in [3]. We have to note here two caveats in analysis of RNAseq data described below. First, BGCs are usually multi-gene units that are not expressed uniformly; for the sake of clarity and brevity (and concealing, to some extent, a true pattern of highly heterogeneous gene expression) we focus therefore on two extreme values – minimal and maximal TPM ones that can be found across each BGC. Second, in some instances in Table 1, e.g. BGC # 6 and 16, we show TPM values for the core biosynthetic genes, which are not necessarily extreme ones in terms of expression, as the latter was for genes with hypothetical function. In this case we prefer to show TPM values for the genes directly relevant for the production of a small molecule.

TPM values of no less than 100 revealed for landomycin biosynthesis set a threshold level of how a BGC is to be expressed to yield a natural product (Table 1). Only a handful of BGC displayed this kind of expression level (#2, 4, 27 and 33; these entries in Table 1 are highlighted in grey). We suggest tentatively that these BGCs are most likely to be actively expressed in parallel to landomycin BGC, and their corresponding product could in principle be identified in the extracts.

Transcriptional activity of specialized metabolic BGCs within S136 genome

BGC	BGC edges, S1361_nnnnn	Type (size, kb)	TPM values, 48 h		Notes[1]
			Min.	Max.	
1	01085 → 01155	PKS I (42)	4 (PKS)	39	
2	01415 → 01435	Xenobiotic transformation? (6)	159	1168	Chlorohexane?
3	03240 →03255	Melanin (4)	10	15	Tyrosinase-based
4	04560	PKS III (1)	875		Germicidin
5	04600 → 04665	Other + glycoside (24)	12	52	
	S1361_04950	NRPS TycC	20		
6	04955 → 05125	PKS I (91)	18 (PKS)	44 (Reg)	Chlorothricin
7	05240 → 05245	NRPS (36)	290	850	Acyldepsipeptide
8	06195 → 06205	PKS III (3)	9	18	Flaviolin
9	07145 → 07340	PKS I (46)	16 (PKS)	160	
10	10235 → 10250	Amino acid (3)	28	52	Ectoine
11	12040 → 12085	PKS II (9)	17 (KS)	42 (KS)	Curamycin
12	14265 → 14275	Terpene (4)	12	38	Caryolan?
13	14845 → 14850	Melanin (1)	7	21	Melanin
14	15220 → 15240	Hydroxamate, NRPS (6)	11	31	Desferrioxamine
15	18455 → 18670	RiPP? (47)	8	500	Linaridin?
16	20280 → 20335	NRPS (26)	15 (PKS)	479 (AT)	
17	20420 → 20455	Other, lincomycin (9)	21	111	
18	20930 → 20965	Peptide, RiPP (8)	15	26664*	Cold shock prot.
19	22435 → 22460	RiPP (6) + TcmN?	5	42	
20	25895 → 25900	Terpene (2)	16	20	Albaflavenone
21	28760 →28785	Other (10; + LexA?)	22	109	
22	29070 → 29315	NRPS (67)	8 (PS)	163 (Reg)	
23	29325 → 29345	RiPP (5)	15	242	Lagmysin
24	29465 → 29500	Terpene (8)	8	15	Isorenieratene
25	29610 → 29645	Bacteriocin (9)	19	36	
26	29685 → 29780	NRPS (25)	19 (PS)	31	
27	29845 → 29995	PKS II (32)	128	5292	Landomycin A
28	30205 → 30255	Bacteriocin (10)	12	40	
29	30400	Terpene (2)	21		Geosmin
30	31160 → 31185	Siderophore (7), LucA	17	40	
31	31295 → 31320	Butyrolactone (6)	20	98	
32	31440 → 31475	Betalactone	11	33	
33	32700 → 32720	RiPP (4)	102	4599	Citrulassin
34	32725 → 32750	RiPP, lasso (9)	124	417	
35	33270 → 33330	Terpene (14)	28	103	Hopene
36	33530 → 33555	RiPP, lasso (5)	35	88	
37	33925 → 34055	PKS I (94)	10	27	Naphthomycin
38	34235 → 34320	PKS I (87)	12 (PKS)	33(PKS)	Lucensomycin
39	35330 → 35495	RiPP + NRPS (77)	8	38	
40	36990 → 37065	PKS I (25)	8	20 (PKS)	
41	37150 → 37285	Saccharide (30)	8	184 (GT)	
	hrdB (17215)	Highly expressed control	270		Sigma factor
	rpoB (23335)	Highly expressed control	342		RNAP
	rpsL (23355)	Highly expressed control	1980		S12 protein

Note: [1]The BGC is likely to direct the production of a small molecule identical or similar to the one mentioned in this entry. Grey background marks BGC that is actively expressed or has minimal TPM value over 100, an indicative of active expression (medium SG)

Of course, low TPM value cannot exclude that some of the BGCs are indeed expressed, and lead to compound accumulation. This especially is true for small BGCs often encoded by a single or a few genes (#4, 20, 29). For large BGCs, such as BGC #37 for presumed polyketide of naphthomycin family (94 kb, 26 genes) on the other hand, low TPM values (no more than 27) are most likely a sign of *de facto* transcriptional and phenotypic silence. Overall, we believe that most of the BGCs across S136 genome are not expressed under conditions that support high-titer production of landomycin A.

A few more findings from analysis of RNAseq data analysis are worth noting. In several cases we stumbled upon abrupt changes in gene transcription right on the edge of annotated BGC. For example, BGC #15 for putative ribosomally produced protein (RiPP) was originally annotated as BGC starting on the "left" side from gene *S1361_18430* [3]. The latter gene as well as two genes flanking *S1361_18430* displayed TPM values over 2000 and most likely belong to an operon encoding signal transduction system consisting from single histidine kinase, three apparently paralogous sensory transduction proteins and phosphate transport accessory protein. The rest of the genes from BGC #15 display significantly lower expression levels. We therefore suggest that *S1361_18430* and three downstream genes are not part of the BGC #15, and accordingly revised "left" edge of the latter (see Table 1). Next, we spotted a gene for nonribosomal peptide synthetase (NRPS) *S1361_04950* which was not included in any of the S136 BGCs [3]. This gene forms a separate entry in the Table 1, between BGCs #5 and #6, and most likely is part of the latter BGC; their expression level is within the same range. Then, annotation of BGC #6 would have to be changed from polyketide synthase (PKS) type I to hybrid NRPS-PKS. These examples demonstrate how functional genomic data can be used to better delineate the boundaries of BGCs within bacterial genomes [9].

Attempts to activate BGC #37. Out of several dozens of BGCs scattered over S136 genome we focused our further experimental efforts on BGC #37. According to antiSMASH predictions, this gene cluster directs the biosynthesis of naphthomycin family polyketide. Its overall expression level is low enough to assume that it is phenotypically silent (see Table 1). In the same time, one of the edges of the BGC harbors two genes for transcriptional factors, *S1361_33920* and *S1361_33925*, whose low transcription levels (18-20 TPM) might underlie the crypticity of BGC #37. As a first step to understand this BGC, we cloned the aforementioned genes individually into expression vector pTES and introduced the resulting plasmids (see Methods section) into wild type S136 and its landomycin-deficient mutant ΔlanI7. The initial and recombinant strains were cultivated on a number of solid media, such as GA, SG2, SFM, OM, TSA, SMMS, R5, GYM, MM and Czapek. Then agar plugs were cut off from 120-h-old lawns of the strains and tested against *B. cereus*, *E. coli* and *D. hansenii*. In no case did we detect the differences in the bioactivity of the initial and recombinant strains. Likewise, the tested strains did not differ in their morphology on agar plates, and their resistance to commonly used antibiotics was similar. It appears, at first sight, that our attempts to activate the production of specialized metabolite encoded by BGC #37 were largely unsuccessful. If true, it may be because we missed some important cluster-situated regulatory gene [10], or did not achieve a proper level of overexpression of the target genes. However, it might also be that the expression of BGC #37 was activated and led to a specialized metabolite that is not colored and/or has no activity against the test cultures that we used. Therefore, definitive conclusions will be drawn after rigorous analysis of the recombinant strains with the help of mass spectrometry methods.

Initial insights into specialized metabolism of S136 grown under landomycin A conditions. As a first step towards disentanglement of the complexity of specialized metabolism of S136, we grew the wild type and ΔlanI7strain under various conditions of submerged fermenta-

tion, and the extracts subjected to liquid chromatography coupled to mass spectrometry (LC-MS). The most notable results are summarized below. In liquid medium YMPG supporting both landomycin and lucensomycin production [2], S136 accumulated a number of unknown compounds as well as some landomycins. Intriguingly, knockout of cluster-situated regulatory gene *lanI* abrogated the production of not only landomycins, but also the other metabolites (Fig. 2), suggesting its cross-regulatory role.

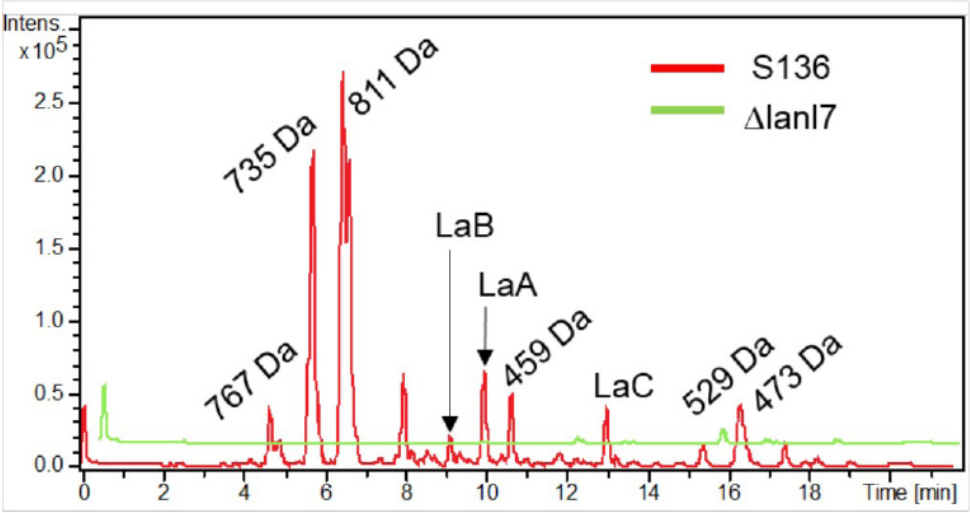

Fig. 2. Overlaid MS traces of ethyl acetate extracts of S136 and ΔlanI7 grown in YMPG for 48 h. Landomycins are marked as LaA, LaB, LaC. The other unknown mass peaks are denoted by the masses (Da) of their anions. Note almost flat trace line (green) for ΔlanI7 strain. Equal amounts of the biomass were used for analysis

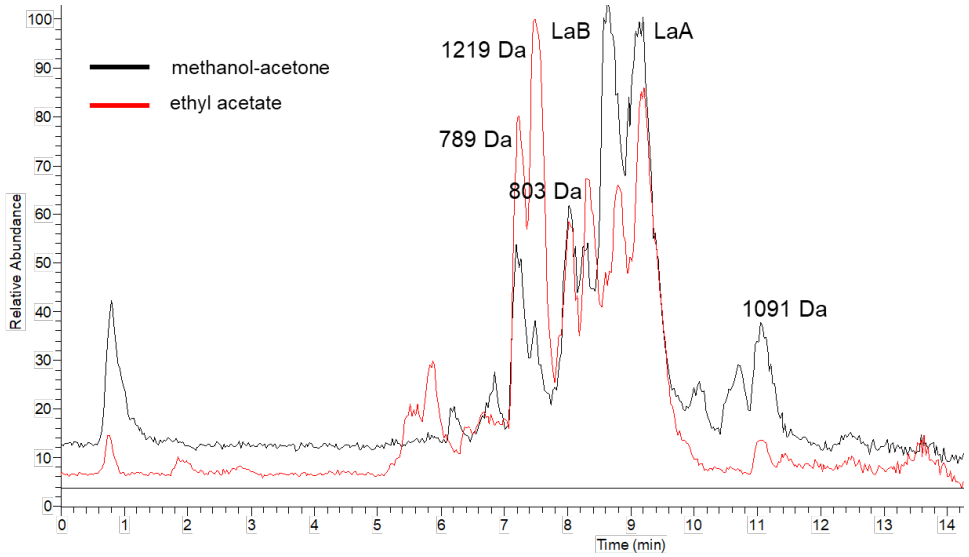

Fig. 3. Overlaid MS traces of ethyl acetate and methanol-acetone (1:1) extracts of S136 grown in SG for 48 h. Landomycins are marked as LaA, LaB. The other unknown mass peaks are denoted by the masses (Da) of their anions

We also prepared methanol-acetone (1:1) and ethyl acetate extracts from the 48-h-old cultures grown in optimal landomycin production medium SG. Choice of the organic solvent had pronounced effect on the spectrum of extracted metabolites, as evident from Fig. 3. Again we detected a number of unknown mass peaks that deserve further in-depth elucidation.

In this work we provide first evidence of disparate patterns of expression of different BGCs found in *S. cyanogenus* S136 genome under conditions conducive of landomycin production. Using landomycin BGC as a reference case of actively expressed pathway, we suggest the threshold transcription levels that define cryptic BGC in this strain, and, most likely, the other streptomycetes. We demonstrate, through LC-MS approach, that several media used to produce landomycins also support the production of a number of unknown compounds, whose identity awaits further experimental scrutiny. It appears therefore that *S. cyanogenus* S136 is a promising source of novel bioactive natural products.

Acknowledgements*. B. O. thanks for grant support of the Ministry of Education and Science of Ukraine (BG-21F).*

REFERENCES

1. *Ostash B., Korynevska A., Stoika R., Fedorenko V*. Chemistry and biology of landomycins, an expanding family of polyketide natural products // Mini Rev. Med. Chem. 2009. Vol. 9. P. 1040–1051. doi: 10.2174/138955709788922593.
2. *Yushchuk O., Ostash I., Mösker E.* et al. Eliciting the silent lucensomycin biosynthetic pathway in *Streptomyces cyanogenus* S136 via manipulation of the global regulatory gene *adpA* // Sci Rep. 2021. Vol. 11. P. 3507. doi: 10.1038/s41598-021-82934-6.
3. *Hrab P., Rückert C., Busche T.* et al. Complete genome sequence of *Streptomyces cyanogenus* S136, producer of anticancer angucycline landomycin A // 3 Biotech. 2021. Vol. 11. P. 282. doi: 10.1007/s13205-021-02834-4.
4. *Hemmerling F., Piel J.* Strategies to access biosynthetic novelty in bacterial genomes for drug discovery // Nat. Rev. Drug Discov. 2022. Vol. 21. P. 359–378. doi: 10.1038/s41573-022-00414-6.
5. *Rebets Y., Ostash B., Luzhetskyy A.* et al. Production of landomycins in *Streptomyces globisporus* 1912 and *S. cyanogenus* S136 is regulated by genes encoding putative transcriptional activators // FEMS Microbiol. Lett. 2003. Vol. 16. P. 149–153. doi: 10.1016/S0378-1097(03)00258-1.
6. *Koshla O., Lopatniuk M., Rokytskyy I.* et al. Properties of *Streptomyces albus* J1074 mutant deficient in $tRNA^{Leu}_{UAA}$ gene *bldA*. Arch. Microbiol. 2017. Vol. 199. P. 1175–1183. doi: 10.1007/s00203-017-1389-7.
7. *Koshla O., Rokytskyy I., Ostash I.* et al. Secondary metabolome and transcriptome of *Streptomyces albus* J1074 in liquid medium SG2 // Cyt. Genet. 2019. Vol. 53. P. 1–7. doi: 10.3103/S0095452719010080.
8. *Kuzniar A., van Ham R., Pongor S.* et al. The quest for orthologs: finding the corresponding gene across genomes // Trends Genet. 2008. Vol. 24. P. 539–551. doi.org/10.1016/j.tig.2008.08.009.
9. *Blin K., Shaw S., Kloosterman A.* et al. antiSMASH 6.0: improving cluster detection and comparison capabilities // Nucleic Acids Res. 2021. Vol. 49. P. 29-35. doi: 10.1093/nar/gkab335.
10. *Liu G, Chater KF, Chandra G, Niu G, Tan H.* Molecular regulation of antibiotic biosynthesis in streptomyces // Microbiol. Mol. Biol. Rev. 2013. Vol. 77. P. 112–143. doi: 10.1128/MMBR.00054-12.

Стаття надійшла до редакції 04.01.24

прийнята до друку 10.01.24

ДОСЛІДЖЕННЯ КРИПТИЧНОГО СПЕЦІАЛІЗОВАНОГО МЕТАБОЛОМУ *STREPTOMYCES CYANOGENUS* S136

І. Осташ[1,2], М. Денека[1], М. Лопатнюк[2], Т. Буше[3], Й. Каліновскі[3], А. Лужецький[2], В. Федоренко[1], Б. Осташ[1,*]

[1]Львівський національний університет імені Івана Франка
вул. Грушевського, 4, Львів 79005, Україна
**e-mail: bohdan.ostash@lnu.edu.ua*
[2]ТОВ «Експлоджен»
вул. Зелена, 20, Львів 79005, Україна
[3]Університет Саарланду
Кампус С2.3, Саарбрюкен 66123, Німеччина
[4]CeBiTec, Університет Білефельда
25, Університетштрассе, Білефельд 33615, Німеччина

Штам Streptomyces cyanogenus S136 відомий як продуцент ландоміцинової родини антибіотиків, серед яких найбільше синтезується ландоміцин А. За винятком ландоміцинів і люцензоміцину, зі штаму S136 не виділено жодних інших спеціалізованих метаболітів. Однак геном S136 містить понад 40 генних кластерів біосинтезу (ГКБ) спеціалізованих метаболітів, що робить цей штам недооціненим потенційним продуцентом нових біоактивних природних сполук. Ми вирішили глибше дослідити спеціалізований метаболом цього штаму. Ми проаналізували дані транскриптому штаму S136 за умов продукції ландоміцину та виявили, що більшість ГКБ транскрибується на базальному рівні. Це призводить до фенотипової криптичності більшості ГКБ, виявлених у геномі штаму S136. Втім, є кілька винятків. Кластер генів біосинтезу ландоміцинів експресується на високому рівні (100 транскриптів на мільйон картованих рідів (Transcripts Per Million mapped reads (TPM)) і близько тисячі TPM для генів мінімальної полікетидсинтази *lanFABC*. Подібний рівень експресії спостерігали у ГКБ № 2, 4, 7 і 33, з яких кластер генів № 2 кодує біосинтез невідомого сахариду, який відрізняється від усіх описаних на сьогодні. Дані РНК-секвенування дали можливість чіткіше окреслити межі передбачуваних ГКБ. На наступному етапі цієї роботи ми виокремили кілька цікавих для дослідження ГКБ у геномі S136. Ці кластери містять гени, які мають низьку подібність до відомих генів і, ймовірно, кодують синтез ще невідомих природних сполук. Також у цих кластерах розташовані регуляторні гени, які можна використати для активації криптичних кластерів. Нами сконструйовано плазміди для надекспресії кількох таких регуляторних генів і введено у штам S136 та у мутантний штам, який не здатен продукувати ландоміцини. Проте біотести не показали жодної відмінності між рекомбінантними штамами з плазмідами та вихідними штамами. Втім, аналіз кількох штамів *S. cyanogenus* за допомогою рідинної хроматографії, спряженої з мас-спектроскопією, вказує на вплив генотипу, умов культивування та екстракції на дані метаболому цих штамів. Дані метаболому штамів *S. cyanogenus* є референтними для подальших досліджень цього виду.

Ключові слова: Streptomyces cyanogenus S136, гени, фізіологія, криптичний спеціалізований метаболом