ISSN 0206-5657. Вісник Львівського університету. Серія біологічна. 2022. Випуск 87. С. 45–53
Visnyk of the Lviv University. Series Biology. 2022. Issue 87. P. 45–53

УДК 579.873.1:577.181.4 HTTPS://DOI.ORG/10.30970/VLUBS.2022.87.04

GENOMIC POTENTIAL OF *STREPTOMYCES ROSEOCHROMOGENES* NRRL 3504 FOR THE PRODUCTION OF SPECIALIZED METABOLITES: ANALYSIS *IN SILICO*

S. Melnyk, P. Hrab, B. Ostash*

Ivan Franko National University of Lviv
4, Hrushevskyi St., Lviv 79005, Ukraine
**e-mail: bohdan.ostash@lnu.edu.ua*

Streptomyces roseochromogenes NRRL 3504 is the only known producer of aminocoumarin antibiotic clorobiocin – an inhibitor of bacterial DNA-gyrase and topoisomerase IV. Sequencing of NRRL 3504 genome revealed a plethora of specialized metabolite biosynthetic gene clusters (BGCs) within the latter, attesting to the significant potential of this species for the production of various as-yet-unknown bioactive compounds. Here we report bioinformatic analysis of NRRL 3504 genome aimed to better understand what kind of small molecules this strain could produce and genetic mechanisms that may limit their production. In addition to the most commonly applied bioinformatic service for BGCs detection antiSMASH, we turned to alternative tools for secondary metabolome *in silico* analysis such as PRISM, DeepBGC, ARTS, SEMPI and GECCO. While different genome mining applications pointed to a common core set of BGCs within the NRRL 3504, each tool having its own algorithm of BGCs detection also discovered certain number of non-overlapping clusters. This was especially true for machine learning tool DeepBGC that unearthed the biggest number of BGCs. To summarize the obtained results we used BGCViz tool, which visualizes and integrates BGC annotations from various sources by using genomic coordinates. We discuss the genetic and structural diversity of the BGCs and outline the most interesting, in opinion, targets for further investigations. Most of the described BGCs are most likely silent due to very low or zero transcription. Therefore, it might be needed to find the ways to activate the transcription of the BGCs of interest. To this end, we mined NRRL 3504 genome for the orthologs of global regulatory genes known to be involved in regulation of specialized metabolism of *S. coelicolor* A3(2). We were able to identify almost all plausible global regulators of interest in NRRL 3504, implying that overall scheme of regulation of specialized metabolism in A3(2) and NRRL 3504 might be similar. Results of our work set the stage for a more detailed experimental scrutiny of silent specialized metabolome of NRRL 3504.

Keywords: *Streptomyces roseochromogenes*, genome mining, bioinformatics

Streptomyces roseochromogenes NRRL 3504 was first described in 1970s as a producer of gyrase inhibitor clorobiocin [1]. Together with novobiocin, clorobiocin and their precursors (Fig. 1) constitute a small family of aminocoumarin antibiotics in actinobacteria [2]. For several decades this strain was known to produce only the aforementioned antibiotic, until the sequencing of its genome in 2014 revealed true potential of this species for biosynthesis of bioactive compounds [3].

Particularly, over 40 biosynthetic gene clusters (BGCs) were revealed which direct the production of various nonribosomal peptides, polyketides, terpenes etc [3]. However, the prioritization of the discovered BGCs with regard to the putative chemical uniqueness of the encoded small molecules or feasibility of their production in NRRL 3504 were not attempted. We also note that the description of BGCs in NRRL 3504 was based on *in silico* analysis with the help of server antiSMASH [4], which was the only available tool for BGC mining at the time of NRRL 3504 genome publication. More tools for BGC discovery have been developed over the last 10 years,

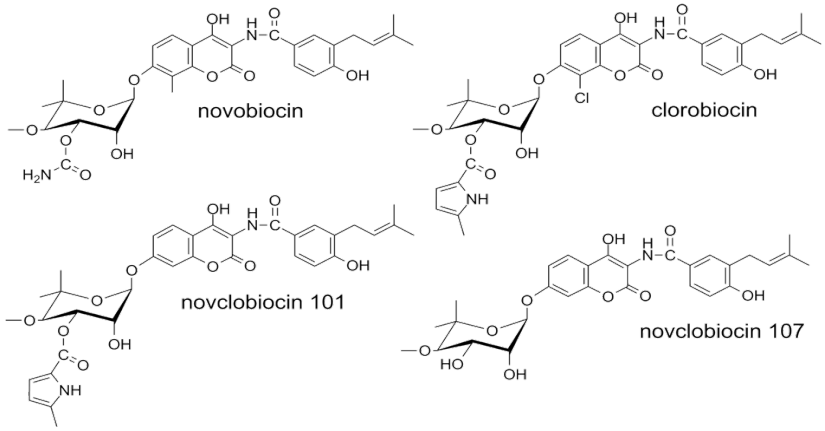

Fig. 1 Structural formulae of aminocoumarin anti-biotics mentioned in the main text

and each of these tools provides somewhat different picture of BGC suites in bacterial genomes [5]. For example, PRISM better predicts the chemical structures of secondary metabolites [6], DeepBGC calculates prediction scores of BGCs similarity and scores of their antimicrobial activity by using deep learning models [7]; ARTS prioritizes BGCs by the presence of resistance genes [8]; SEMPI predicts PKS and NRPS BGCs by analyzing available natural compound databases [9]; finally, GECCO identifies novel BGCs based on conditional random fields approach [10]. From practical point of view, it means that we might be able to identify more BGCs in any given genome on combining the powers of different search algorithms.

Except for the clorobiocin BGCs, all or most of the BGCs within NRRL 3504 genome are most likely silent due to very low or zero transcription. Therefore, it might be needed to find the ways to activate the transcription of the BGCs of interest. We recently described a set of genetic tools for NRRL 3504 [11] which in principle can be applied to study any silent BGC in this strain. Although manipulations of cluster-situated regulatory genes, likely present within the silent BGCs, might the simplest way to activate the latter, it is often possible to use higher-order, or pleiotropic (global) transcriptional factors to upregulate silent BGCs. As a first step towards the exploitation of global regulators in NRRL 3504, we searched orthologs of known global transcriptional factors in model strain *Streptomyces coelicolor* [12]. The resulting regulatory network in NRRL 30504 is outlined in this work.

Materials and Methods

NRRL 3504 genome was accessed through NCBI (accession number NZ_CM002285.1). Following BGC mining engines were used: antiSMASH, PRISM, DeepBGC, ARTS, SEMPI, GECCO (mentioned in the Introduction). These applications were used with default parameters. Results were summarized and visualized using BGCViz software (https://github.com/ostash-group/BGCViz).

NRRL 3504 orthologs of known *S. coelicolor* transcriptional factors involved in antibiotic production were identified as a reciprocal BLASTP hits (https://blast.ncbi.nlm.nih.gov) [12].

Results and Discussion

The antiSMASH-centered description of BGCs within NRRL 3504 genome. We re-analyzed NRRL 3504 genome with antiSMASH given vast improvements to the algorithm [4] since the time of its original application to NRRL 3504 genome. The antiSMASH data were further amended with some of the results of the other services. In total our analysis yielded 52 reliably identified BGCs, as shown in Table 1.

Table 1

BGCs in NRRL 3504 genome as portrayed by BGC mining engines

BGC	Genomic interval[1]	Type[2]	Most similar BGC (similarity %)
1	132,030 – 165,370	NAPAA	
2	223,343 – 251,706	Betalactone	Platencin; terpene (6 %) *
3	251,738 – 303,426	NRPS,RiPP-like	Leinamycin; NRPS-PKS I (4 %) *
4	417,115 – 425,644	Siderophore	Natamycin; polyketide (9 %) *
5	454,533 – 532,064	T1PKS	Sceliphrolactam; poliketide (92 %)
6	903,971 – 919,646	Lanthipeptide-class-III	Informatipeptin, RiPP: lanthipeptide (100 %)
7	1,152,522 – 1,185,781	Aminocoumarins	Clorobiocin; saccharide (100 %)
8	1,198,728 – 1,255,895	NRPS-like;betalactone	Syringomycin; NRP (29 %) *
9	1,368,661 – 1,381,464	Terpene	Hopene; terpene (92 %)
10	1,531,132 – 1,574,167	NRPS-like	
11	1,580,006 – 1,621,908	NAPAA,NRPS	Stenothricin: cyclic depsipeptide (13 %) *
12	1,834,628 – 1,840,717	Siderophore	
13	1,841,787 – 1,846,610	RiPP-like	
14	1,876,591 – 1,920,303	T1PKS,siderophore	Paulomycin; other (13 %) *
15	2,073,191 – 2,075,353	Terpene	Geosmin; terpene (100 %)
16	2,105,451 – 2,116,289	RIPP-like	
17	2,160,971 – 2,180,930	Terpene	Meilingmycin; polyketide (5 %) *
18	2,361,368 – 2,372,564	Siderophore	
19	2,439,160 – 2,480,253	Ladderane, terpene	Merochlorins; terpene + PKSIII (7 %) *
20	2,820,651 – 2,846,507	Betalactone	Divergolides; PKSI (6 %) *
21	2,998,644 – 3,028,573	T1PKS, T3PKS	Venemycin; polyketide (100 %) [3]
22	3,028,645 – 3,058,382	NRPS	Thiazostatin: NRP (100 %) [3]
23	3,155,507 – 3,157,959	Terpene	Albaflavenone; terpene (100 %)
24	3,317,553 – 3,335,041	RRE-containing	Naphthomycin; polyketide (9 %) *
25	3,713,984 – 3,767,955	NRPS	Salinamides; NRPS-PKS(92 %)
26	4,202,052 – 4,222,838	Lanthipeptide-class-I	
27	4,549,129 – 4,576,738	Linaridin	Legonaridin; RIPP (16 %)
28	4,563,543 – 4,568,970	Terpene	
29	4,675,333 – 4,701,181	Polyketide	Colabomycin E; polyketide type II (20 %)
30	4,702,838 – 4,715,987	Lassopeptide	
31	5,334,727 – 5,339,264	Lassopeptide	Citrullasin D; RIPP (100 %)
32	5,358,233 – 5,439,812	NRPS-PKS, RRE	LL-D49194α1; polyketide (39 %)
33	5,656,729 – 5,667,757	Butyrolactone	Methylenomycin A (14 %)
34	6,239,049 – 6,245,818	Siderophore	Desferrioxamin (83 %)
35	6,350,666 – 6,352,022	Melanin	Melanin (100 %)
36	6,561,106 – 6,597,703	Thioamitides, RiPP	
37	6,679,165 – 6,701,768	Cyanobactin	
38	6,787,544 – 6,794,342	Terpene	2-methylisoborneol; terpene (75 %)
39	6,957,682 – 7,000,235	PKS	
40	7,271,875 – 7,325,011	NRPS, NRPS-like	Ecumicin; NRP (10 %) *
41	7,489,337 – 7,492,487	Ectoine	Ectoine (100 %)
42	7,719,515 – 7,751,356	NAPAA	
43	8,174,477 – 8,219,344	hglE-KS	Cinnamycin; RIPP: lanthipeptide (9 %)
44	8,223,946 – 8,245,107	CDPS	Purincyclamide; other (40 %)
45	8,279,264 – 8,342,793	NAPAA, T3PKS	Herboxidiene; polyketide (8 %)
46	8,498.110 – 8,503,471	NRPS	
47	8,584,205 – 8,603,224	PKS	
48	8,779,773 – 8,800,663	Melanin, terpene	Melanin; other (57 %)
49	8,882,064 – 8,902,774	CDPS	Peptide
50	9,386,952 – 9,409,720	Lanthipeptide-class-IV	Blasticidin S (10 %)
51	9,479,135 – 9,521,672	NRPS	
52	9,603,846 – 9,625,045	Terpene	

Notes: [1]Position in NRRL 3504 chromosome (accession number NZ_CM002285.1)

[2]Abbrevviated names of biosynthetic types: NAPAA – non-alpha poly-amino acids like e-Polylysin; NRPS – non-ribosomal peptide synthetase cluster; RiPP – ribosomally synthesised and post-translationally modified peptide product; RiPP-like – other RiPPs; T1PKS - type I PKS (polyketide synthase); T3PKS – type III PKS; RRE-containing – RRE-element containing cluster; PKS-like – other types of PKS cluster; hglE-KS – heterocyst glycolipid synthase-like PKS; CDPS – tRNA-dependent cyclodipeptide synthases.

[3]Clusters are connected.

*Non-significant hits, no evidence of large-scale similarity

Different BGC mining tools yield different sets of BGCs within NRRL 3504 genome. Next we run NRRL 3504 genome through the applications, that use different search logics (see the Introduction): DeepBGC, PRISM (+PRISM-Supp which identifies regulatory, resistance and transporter genes within clusters), ARTS, SEMPI and GECCO. We used BGCViz (application which integrates and visualizes results from various BGC annotation sources) to summarize our findings. The antiSMASH results were used as a reference. This led to a scheme of annotations comparison to the reference which reflected the location of identified BGCs within the genome of NRRL 3504. Fig. 2 shows the number of BGCs identified by each application, BGC location rela-

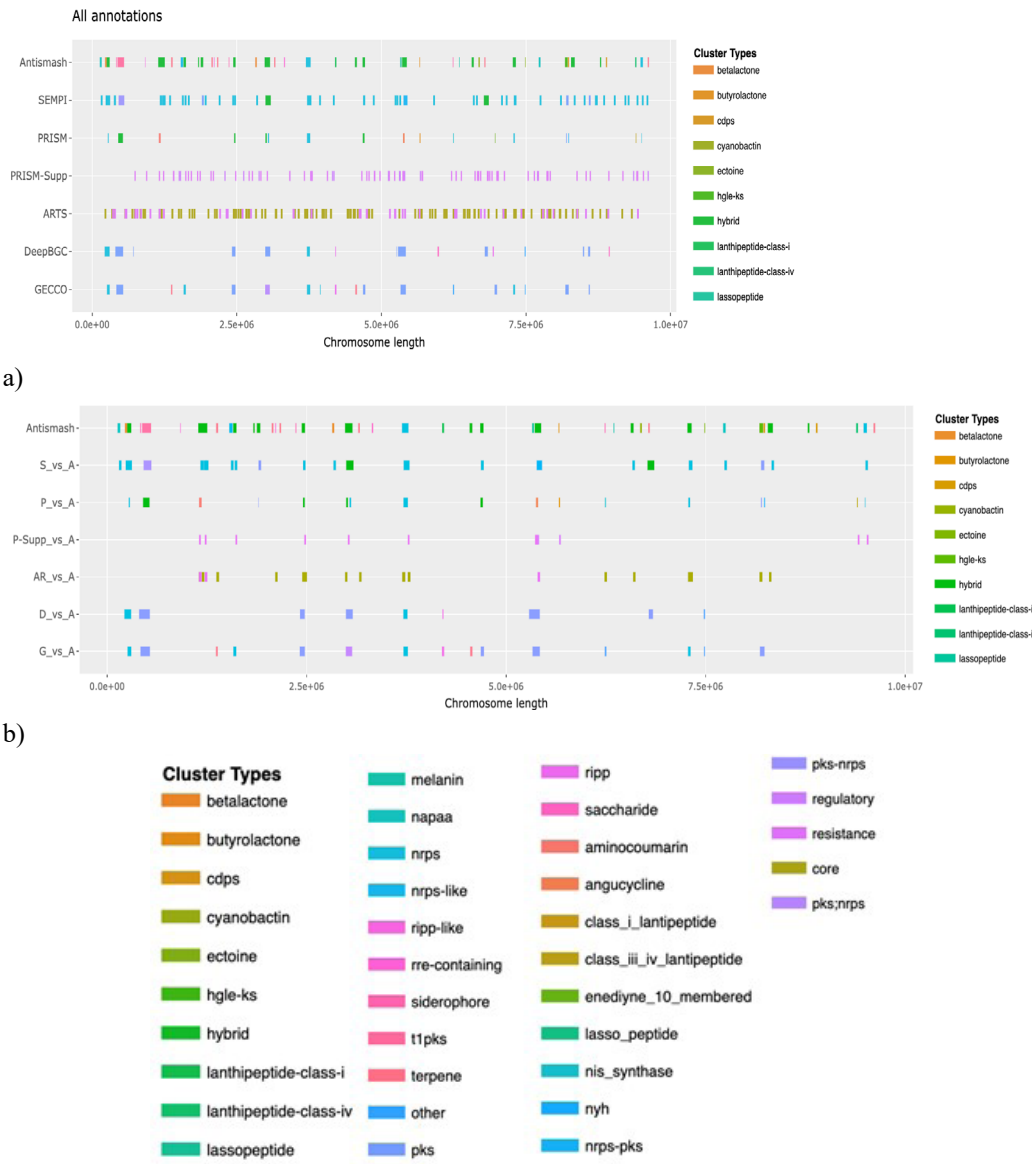

Fig. 2. BGCs annotations' all (a) and in comparison to the reference – antiSMASH (b). Abbreviations of rows: first letters of BGC applications (e.g. S_vs_A means SEMPI versus antiSMASH)

tive to each other, edges of each BGC, and the expected class of compounds encoded by the BGC. As we can see, each tool pinpointed a different number of gene clusters (fig. 2a). Particularly, antiSMASH has identified 44 BGCs, SEMPI – 48 BGCs, PRISM – 19 BGCs, PRISM-Supp – 84 regulatory, resistance and transporter genes within BGCs, ARTS – 186 core and resistance genes within BGCs. Machine learning engine DeepBGC revealed the highest number of BGCs (108), and the most recent addition to BGC-finding algorithm, GECCO, unearthed 30 BGCs. The aforementioned tools reveal overlapping set of BGCs (most of them listed in Table 1), although each service also leads to unique findings.

The antiSMASH annotation was compared against DeepBGC and GECCO results in more details. The plots (Fig. 3, a, b) show the number of BGCs detected by applications in accordance with different activity scores. antiSMASH independently annotated more than 40 clusters with the highest activity scores, while DeepBGC annotated only 6 BGCs with a score of 50 %. At the same time, DeepBGC and antiSMASH together identified 14 BGCs with activity score of 50 %. The situation is similar when comparing antiSMASH and GECCO results: GECCO independently didn't identify clusters with a high activity score, while in combination with antiSMASH it annotated 19 clusters with a score of 50 %.

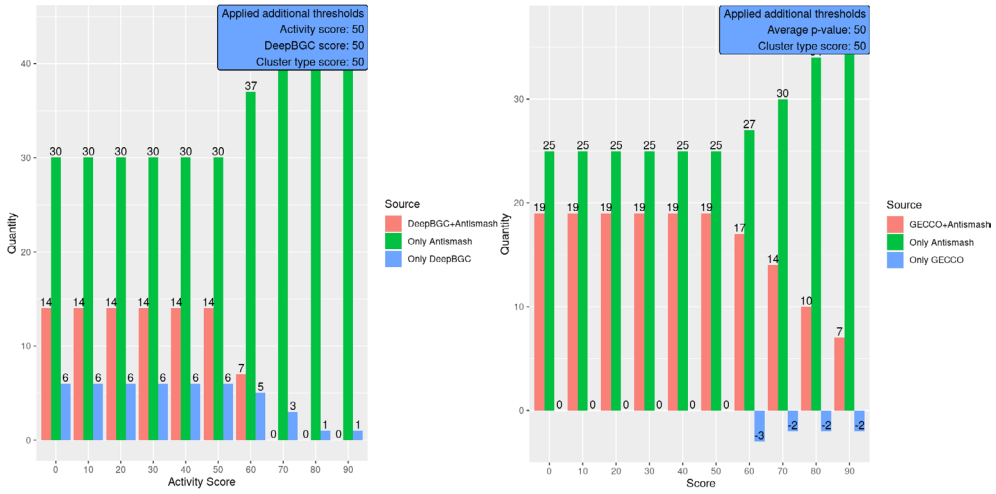

Fig. 3. a) Comparison of antiSMASH and DeepBGC annotations at a given score threshold; b) comparison of antiSMASH and GECCO annotations at a given score threshold

We summarized all BGC mining results in a combined Biocircos plot where arcs connect the same genomic intervals identified as a BGC by different tools (outer face of the plot, Fig. 4). Different arc colors stand for biogenetically different types of specialized metabolism. The reference column-based mode of coloring links was used, as shown to the right of Biocircos plot.

For example, a cluster of genes for the biosynthesis of a compound similar to manumycin-type antibiotic colabomycin was annotated by antiSMASH as arylpolyene-ladderane-lassopeptide-type BGC №25 in a genomic interval from 4 675 333 to 4 717 422. Analyzing Biocircos plot, we can see this BGC was also identified by the other bioinformatic tools, however, the genomic coordinates and biosynthetic type of the BGC slightly differ from the antiSMASH results. GECCO annotated colabomycin BGC as pks-type cluster in a region ranging from 4 682 664 to 4 722 714 bp; SEMPI – as nrps-type cluster (from 4 683 996 to 4 720 559 bp), PRISM – as lassopeptide-nrps-type cluster (from 4 678 456 to 4 708 906 bp). DeepBGC did not detect the aforementioned BGC, while ARTS and PRISM-Supp tools did not identify resistance genes within this BGC.

Global specialized metabolism regulatory network of NRRL 3504. We attempted to reconstruct the regulatory system in *S. roseochromogenes* by searching for orthologs of global regulators of model organism *S. coelicolor*. The proteins of the Bld cascade are key to the transition from substrate to aerial mycelium and the main protein of this cascade, BldD, represses the expression of the other cascade components [13]. Orthologs of genes encoding the Bld-cascade were found: the search revealed orthologs with a high percentage of amino acid sequence identity for BldD (99 %), BldM (100 %), BldN (90 %). The central place in the Bld-cascade is given to the pleiotropic regulator AdpA involved in regulation of a number of other proteins. We revealed ortholog of AdpA of *S. coelicolor* in the genome of *S. roseochromogenes* with the identity of amino acid sequences of 89 %. Analysis of the functional regions (dimerization and DNA-binding domains) showed that the level of sequence identity in them is close to 100 %.

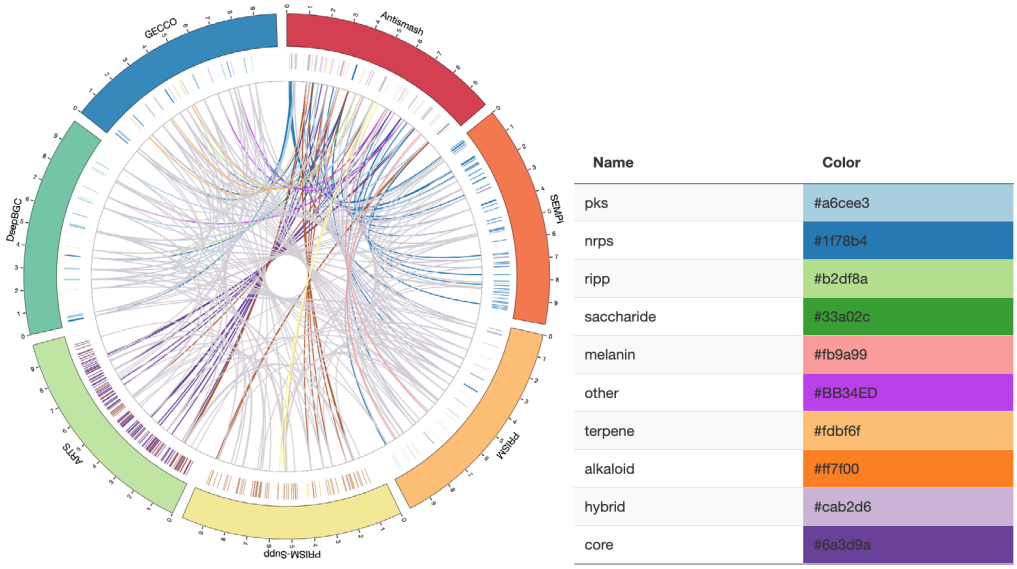

Fig. 4. Biocircos visualizing plot and its coloring scheme; see main text above for more details

Streptomyces genomes of contain genes of chaplins and rodlins, which encode small proteins hydrophobins, necessary for the formation of the water-repellent surface of the air mycelium and involved in the processes of spore formation. Our analysis revealed a low level of identity of the amino acid sequence of AmfR of *S. coelicolor* to its ortholog in *S. roseochromogenes* genome. AmfR controls the synthesis of the morphogenetic surfactant peptide SapB, involved in the straightening of hyphae in the air. It is likely that the low level of sequence similarity or the complete absence of orthologs of genes encoding these proteins are the causes of the defective sporulation of NRRL 3504.

The next step was to find orthologs of genes that encode components of the Whi-cascade involved in the maturation of spores (pigmentation, septation). The search identified orthologs with a high percentage of amino acid sequence identity for WhiA (95 %), WhiB (100 %), WhiG (94 %), WhiH (89 %), WhiI (97 %). An orthologs with a low percentage of identity (32 %) was identified for WhiE involved in the synthesis of polyketide spore pigment during the late stages of sporulation.

In addition to analogs of genes encoding global regulators, shown in Fig. 5, we found orthologs for a number of other transcriptional regulators of secondary metabolism, such as: ArgR

(92 %), Crp (97 %), DasR (96 %), RelA (93 %), SlbB (91 %), WblA (96 %). No significant differences were found in functionally important areas (DNA or effector molecule binding regions).

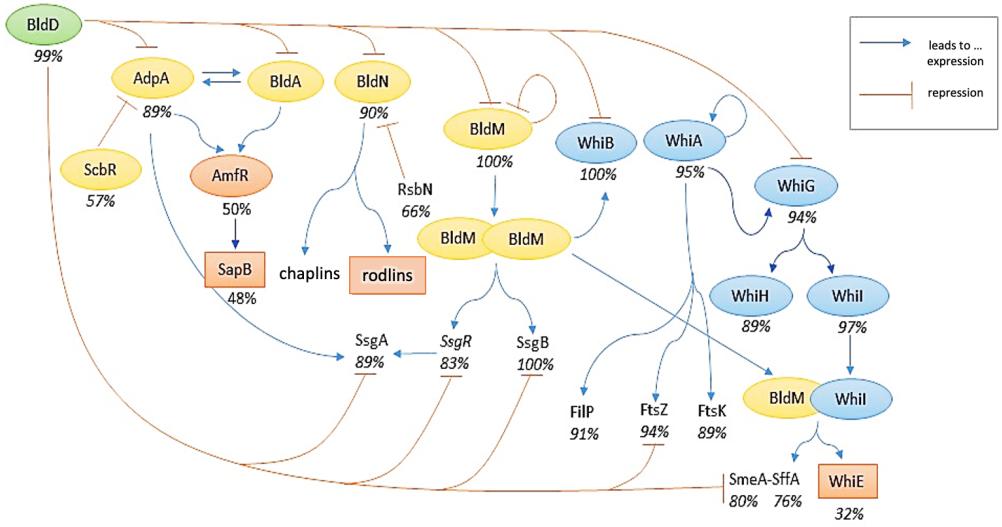

Fig. 5. Scheme of the putative regulatory network of NRRL 3504 (based on the network of *S. coelicolor*). The percentages indicate the identity of the amino acid sequences of *S. roseochromogenes* proteins with the corresponding regulators in *S. coelicolor*. Components absent in NRRL 3504 or having a low degree of similarity to the corresponding *S. coelicolor* proteins are highlighted with red

Streptomyces roseochromogenes possesses an exceptional biosynthetic potential with 52 BGCs identified in our analysis. BGCs № 8, 19, 32, 45 deserve special attention as they harbor one or several resistance markers according to PRISM and ARTS. For example, BGC №32 consists of 5 candidate BGCs, identified by antiSMASH tp carry 5 transport genes, 4 regulatory genes; 1 resistance target was identified by ARTS. Further decomposition of such a complex region is not feasible in the absence of experimental results. One annotated unusual genomic interval contained two BGCs (№ 21,22), that are identical to venemycin and thiazostatin BGCs, which we could separate into corresponding regions.

NRRL 3504 is rich in novel RiPPs, which take up to 16,3 % of all identified BGCs. As RiPPMiner-Genome results show, BGCs № 26, 30, 37, 50 are particularly interesting, as they have both: biosynthetic machinery and precursor peptides for natural product synthesis. Given short length of RiPP BGCs, they could be an interesting target for heterologous expression.

Clusters № 5, 11 and 25 can be also considered targets for further analysis. The first one is similar to sceliphrolactam, with rearranged tailroing enzymes and different PKS domains. Hence its product is likely different from known ones, but the BGC is similar enough to be functional. BGC №11 displays interesting D-ala-D-ala ligase signature, identified by PRISM, responsible for vancomycin resistance. Therefore, the product could target peptidoglycan cell wall of Gram-positive bacteria. On the other hand, cluster 25 is similar to salinamides with different NRPS domains and several tailoring enzymes. Also, it encodes EF-Tu isoform, which could shed the light on the potential target of its product. This incomplete homology to salinamides could mean that the cluster direct the biosynthesis of a complete product different from the known one.

We found that antiSMASH data are the most informative ones when used in combination with ARTS and PRISM results. Collectively these tools allowed to refine cluster boundaries

and prioritize targets for further investigation. GECCO results were mostly congruent with antiSMASH results, yielding also 2 unique BGCs (№39 and №47). We included them into the Table 1 as BGC №39 was also identified by PRISM as a PKS-based BGC harboring an efflux transporter and a core gene within it's boundaries. BGC №47 was identified by SEMPI as a BGC encoding PKS along with MFS transporter within its boundaries. We added NRPS cluster from PRISM (BGC №46) as well, given its simultaneous identification with SEMPI.

Overall, we believe that combination of different genome mining approaches provides the most exhaustive annotation of BGCs in genomes, and so the use of non-antiSMASH tools should not be neglected.

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

REFERENCES

1. *Mancy D, Ninet L, Preud'Homme J*. Antibiotic 18631 RP. U.S. patent. 1974. 3(793):147.
2. *Heide L*. The aminocoumarins: biosynthesis and biology // Nat. Prod. Rep. 2009. Vol. 26. P. 1241–1250. https://doi.org/10.1039/b808333a.
3. *Rückert C., Kalinowski J., Heide L.* et al. Draft genome sequence of *Streptomyces roseochromogenes* subsp *oscitans* DS 12.976, producer of the aminocoumarin antibiotic clorobiocin // Genome Announc. 2014. 2(1):e01147-13. https://doi.org/10.1128/ GENOMEA.01147-13.
4. *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.
5. *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.
6. *Skinnider M., Johnston C., Gunabalasingam M*. et al. Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences // Nat Commun. 2020. Vol. 11. P. 6058. https://doi.org/10.1038/s41467-020-19986-1.
7. *Hannigan G., Prihoda D, Palicka A.* et al. A deep learning genome-mining strategy for biosynthetic gene cluster prediction // Nucleic Acids Res. 2019. Vol. 47: e110. doi:10.1093/nar/gkz654.
8. *Mungan M., Alanjary M., Blin K.* et al. ARTS 2.0: feature updates and expansion of the Antibiotic Resistant Target Seeker for comparative genome mining // Nucleic Acids Res. 2020. Vol. 48. P. 546–552. doi: 10.1093/nar/gkaa374.
9. *Zierep P., Ceci A., Dobrusin I.* et al. SeMPI 2.0-A web server for PKS and NRPS predictions combined with metabolite screening in natural product databases // Metabolites. 2020. Vol. 11. 13. doi: 10.3390/metabo11010013. PMID: 33383692; PMCID: PMC7823522.
10. *Carroll L., Larralde M., Fleck J.* et al. Accurate *de novo* identification of biosynthetic gene clusters with GECCO // 2021. doi:10.1101/2021.05.03.442509.
11. *Melnyk S., Stepanyshyn A., Yushchuk O*. et al. Genetic approaches to improve clorobiocin production in *Streptomyces roseochromogenes* NRRL 3504 // Appl. Microbiol. Biotechnol. 2022. Vol. 106. P. 1543–1556. doi:10.1007/s00253-022-11814-4.
12. *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.
13. *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 02.08.22

прийнята до друку 22.08.22

ГЕНОМНИЙ ПОТЕНЦІАЛ *STREPTOMYCES ROSEOCHROMOGENES* NRRL 3504 ЩОДО ПРОДУКЦІЇ СПЕЦІАЛІЗОВАНИХ МЕТАБОЛІТІВ: АНАЛІЗ *IN SILICO*

С. Мельник, П. Граб, Б. Осташ*

Львівський національний університет імені Івана Франка
вул. Грушевського, 4, Львів 79005, Україна
**e-mail: bohdan.ostash@lnu.edu.ua*

Streptomyces roseochromogenes NRRL 3504 – єдиний відомий продуцент амінокумаринового антибіотика хлоробіоцину, інгібітора бактерійних ферментів ДНК-гіраз і топоізомераз IV. Секвенування геному NRRL 3504 виявило багато кластерів генів біосинтезу спеціалізованих метаболітів (BGC), що свідчить про значний потенціал даного виду у продукції різноманітних, не відомих раніше біоактивних сполук. У даній статті ми представляємо біоінформативний аналіз геному NRRL 3504, спрямований на те, аби краще зрозуміти, які сполуки може синтезувати NRRL 3504, а також генетичні механізми, що потенційно обмежують цей синтез. Окрім найширше застосовуваного біоінформативного сервісу для виявлення BGC antiSMASH, ми звертаємось до альтернативних *in silico* інструментів дослідження вторинного метаболому, таких як PRISM, DeepBGC, ARTS, SEMPI та GECCO. Різні застосунки, маючи власний механізм детекції, виявили не лише спільний набір BGC у NRRL 3504, а й низку нетипових BGC. Особливо це стосується базованого на принципі машинного навчання застосунку DeepBGC, що ідентифікував найбільшу кількість BGC. Для узагальнення отриманих результатів нами використано біоінструмент BGCViz, що візуалізує та інтегрує геномні координати BGC, отримані з різних джерел. Ми обговорюємо генетичну та структурну різноманітність BGC і окреслюємо найцікавіші, на нашу думку, цілі для подальших досліджень. Більшість описаних BGC, найімовірніше, «мовчазні» через дуже низьку або нульову транскрипцію. Таким чином, може стати доцільним пошук способів активації транскрипції таких цільових BGC. З цією метою ми здійснили пошук у геномі NRRL 3504 ортологів глобальних регуляторних генів, які, як відомо, беруть участь у регуляції спеціалізованого метаболізму *S. coelicolor* A3(2). Нам вдалось ідентифікувати майже усі глобальні регулятори у NRRL 3504, а це може свідчити, що загальна схема регуляції спеціалізованого метаболізму в A3(2) та NRRL 3504 є подібною. Результати нашої роботи закладають основу для детальнішого експериментального вивчення «мовчазного» спеціалізованого метаболому NRRL 3504.

Ключові слова: Streptomyces roseochromogenes, аналіз геному, біоінформатика