ISSN 0206-5657. Вісник Львівського університету. Серія біологічна. 2021 Випуск 85. С. 26–34
Visnyk of the Lviv University. Series Biology. 2021. Issue 85. P. 26–34

ГЕНЕТИКА

УДК 579.873.1:577.181.4 HTTPS://DOI.ORG/10.30970/VLUBS.2021.85.03

A DATABASE OF SEQUENCED GENOMES OF DIFFERENT *STREPTOMYCES ALBUS* J1074 STRAINS AND USES THEREOF

B. Dolya[1], T. Busche[2], T. S. Jørgensen[4], K. Ochi[3], T. Gren[4], T. Weber[4], J. Kalinowski[2], A. Luzhetskyy[5], B. Ostash[1,*]

[1]Ivan Franko National University of Lviv
4, Hrushevskyi St., Lviv 79005, Ukraine
**e-mail: bohdan.ostash@lnu.edu.ua*
[2]CeBiTec, Bielefeld University
25, Universitatstrasse, Bielefeld 33615, Germany
[3]Department of Life Sciences, Hiroshima Institute of Technology
Miyake 2-1-1, Saeki-ku, Hiroshima 731-5193, Japan
[4]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark Kemitorvet, Build. 220, 2800 Kgs. Lyngby, Denmark
[5]Department of Pharmacy, Saarland University
Campus C2.3, Saarbrucken 66123, Germany

Streptomyces albus J1074 has recently emerged as a powerful chassis strain for natural product research and discovery, as well as model to investigate various aspects of actinobacterial biology. A number of genetic tools have been developed to facilitate the use of this strain for the aforementioned purposes. One of the promising approaches is to introduce into J1074 genome mutations that would improve the antibiotic-producing capacity of J1074. Particularly, we reported a collection of spontaneous and genomically engineered J1074 mutants carrying mutation in genes for ribosomal protein S12, RNA polymerase beta subunit etc. We took advantage of this collection to build an in-house database which would host (both current and future) the primary and curated sequencing data for these mutants. The database is available at https://biotools.online/media/. The main benefits of the database lie in the known pedigree of the strains, which allows deep interpretation of the data. For example, there is lively – and unresolved – debate on the origins and consequences of the GC composition in actinobacteria. A better understanding of this issue should improve our knowledge of genome evolution in bacteria and will have a number of biotechnological ramifications. We used our *Streptomyces albus* J1074 dataset as an experimental model to reveal genome-wide spectrum of mutation, which appears to be biased towards elevated GC content. We also included the high-quality genomes of the other streptomycetes into our databse for comparative purposes. The genomic GC content in streptomycetes varies from 75 % to 66.5 %, with median value being 72 %. The GC content of secondary metabolic genes of *S. coelicolor* is less variable than that of primary metabolic genes, an indicative of different selection pressure on these gene groups. Along with selective constraints, the peculiarities of *Streptomyces* DNA mismatch repair might contribute to the skewed GC content of their genomes. Further uses of the database may include the development of a more precise knowledge of the mutation rate as well as population genetic processes within this species and genus.

Keywords: *Streptomyces albus*, genome sequencing, mutation spectra

Streptomycetes are Gram-positive bacteria of the phylum *Actinobacteria* known in first place for their ability to produce an impressive array of bioactive small molecules [1]. Linear

genomes and high GC content (around 73 %) are the defining traits of *Streptomyces* genomics. The reasons for GC-rich DNA in streptomycetes are unclear. In a larger context of bacterial genomics, the GC nucleotide content is within 25 %–75 % range [2]; this puts *Streptomyces* at one extreme of variation of GC content within entire *Bacteria* kingdom. Given that actinobacteria are among the most deeply branching lineages in bacterial phylogenetic tree [3, 4], elucidation of the reasons for and potential advantages of high GC content in *Streptomyces* may help better understand the factors that shape nucleotide composition of bacterial genomes. This issue is far from being settled. Biased mutation process was an initial assumption about differential GC content, although more recent studies questioned this view by providing the evidence for universal tendency of accumulation of AT bases [2]. The other studies pointed to relationship between GC content and aerobic versus anaerobic lifestyles [5]. As most of a bacterial genome is allocated to protein-coding sequences, the GC content could be tightly linked to selective processes acting on codon level [6, 7]. All above mentioned works based their theories about GC content from comparison of limited datasets collected across different species. Both factors undermine the validity of respective conjectures for following reasons. First, one cannot guarantee that trends observed for selected sets of genes will be true for the entire genomes. Second, when inferring mutational biases or rates from different species, it is impossible to know all factors that shaped their genomes; these factors could be different for different species and even different populations of the same species. High-throughput sequencing approaches offer potential remedy for these issues through the analysis of genomes of many lineages of the same species that were cultivated under fully controlled conditions over a number of generations [8, 9]. Through comparison of an ancestral genome and of evolved lineages, one will arrive at an evidence-based model of mutational rate and biases for this species. This kind of experimental setup so far provides the strongest possible evidence for the absence of unaccounted factors that would undermine the proposed mechanisms. Indeed, application of such an approach to genomes of *Burkholderia cenopacia* and *Mycobacterium smegmatis* portrayed a compelling mechanistic picture that explains why these organisms possess a particular GC content [10, 11]. We find these works especially interesting in the context of *Streptomyces* genomics, because both aforementioned species have GC-rich genomes and they were shown to have biased accumulation of spontaneous mutations. The predominant accumulation of GC nucleotides in *M. smegmatis* is thought to be associated with a deficient DNA repair system, although the exact mechanism remains debatable [12]. We decided to re-visit the issue of GC content of *Streptomyces* genomes. To this end, we identified types of mutations accumulated by a set of *S. albus* J1074 derivatives, taking advantage of nine in-house sequenced genomes of this species. We also compared the GC content of *Streptomyces* genomes on a larger dataset than in previous studies. Results of our findings are given below.

Materials and Methods

All strains analyzed in this work are derivatives of *S. albus* SAM2, which is a derivative of J1074 with deletion of pseudo-*attB*$^{\varphi C31}$ [13]. The strain pedigree is summarized in Fig. 1. Briefly, 10^9 spores/200 µL of SAM2 were plated onto GYM agar [14] supplemented with 100 µg/mL streptomycin to select for spontaneous streptomycin-resistant (Str^r) mutants KO-1296, KO-1297, KO-1298 and KO-1300. Strain R94G is a genomically engineered *rpsL* mutant described in [15]. This strain served as a platform for sequential introduction of spontaneous mutations conferring the resistance to streptomycin (KO-1295), lincomycin (KO-1304), erythromycin (KO-1305) and rifampicin (KO-1408). All aforementioned spontaneous mutants were generated in a single selection campaign and underwent no more than five passages prior to genome sequencing. Description of all aforementioned KO strains will be subject of separate publications.

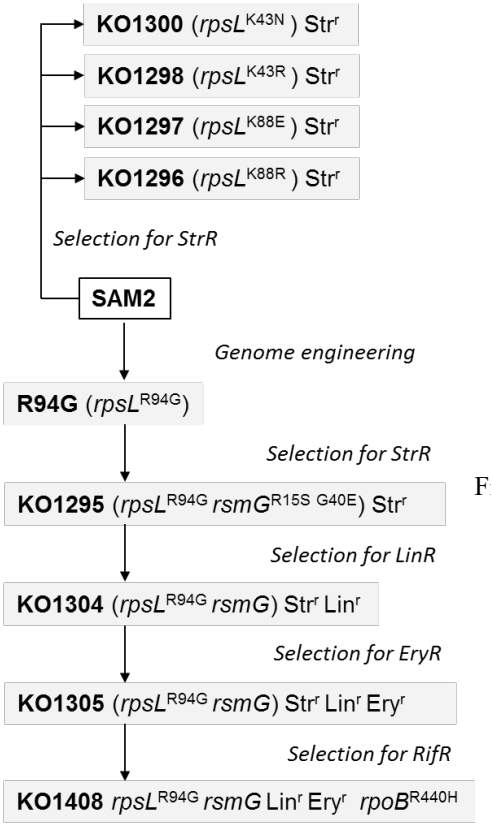

Fig. 1. Scheme that summarizes the generation of *S. albus* strains used in this work. Strain names are shown in bold; type of their generation is written in italic. SAM2 (boxed) is the initial strain, all derivatives are shown in grey background. Antibiotic resistance phenotype (Str^r, Lin^r, Ery^r, Rif^r – resistance to streptomycin, linkomycin, erythromycin and rifampicin, respectively) and mutations known prior to genome sequencing follow the strain name. See main text for more details

Genomic DNA isolation and Illumina sequencing were performed according to standard protocols. All identified mutations were verified via Sanger sequencing of PCR-amplified fragments. Raw data of *S. albus* strains, reference sequence of J1074 as well Excel table with called variants and supplementary materials can be found at the in-house *S. albus* genomics database maintained by Lviv University research group: https://biotools.online/media/. At the quality control stage, the sequence reads were examined for overall quality and presence of Illumina adapters with FastQC [18]. In order to omit poor quality data from further analysis we trimmed low quality read ends and filtered low quality reads by using Trimmomatic version 0.36 [19]. Sequencing reads were aligned to reference J1074 genome (accession number CP004370) with Bowtie2 version 2.2.5 [20]. SNP and DIP detection was performed with ReadXplorer [21]. Illumina coverage was 55-165X for all strains (detailed list with average coverage for assembled genomes one can find at the aforementioned webpage: Supplementary_data_(avg_coverage).xlsx? as well as in xlsx files in SNP folder for coverage in variant calling). In order to identify putative MutS and MutL orthologs within *Streptomyces* proteomes we used reciprocal best BLASTP hit strategy [22] and in-house scripts based on NCBI Datasets tools. JavaScript application Mutations Needle Plot v0.8.0 was used for visualization of mutation distributions along the genome [23].

Results and Discussion

Genomes of *Streptomyces albus* J1074 strains reveal biased accumulation of point mutations. Our collection of J1074 strains has simple and traceable genealogy; mutants carry certain mutations they were selected for [15], and also might carry additional spontaneous

mutations. We reasoned that this collection would be suitable to better understand mutational processes within *Streptomyces*. Nine strains and their parent SAM2 were selected for Illumina sequencing. The same approach was used to pre-process raw data and map the mutations in order to avoid artifacts arising from using different tools. Some portions (less than 1%) of *S. albus* genomes remained ambiguous due to low coverage or other sequencing artifacts. However, these regions (of, for instance, genome of KO-1305) were correctly represented in their derivatives (KO-1408), which ruled out the presence of unaccounted mutations. Our final estimates of the mutations accumulated in nine *S. albus* strains since their immediate ancestor (e.g., SAM2 → R94G; KO-1305 → KO-1408 etc) are summarized in Table.

Mutation spectra revealed by whole genome sequencing in nine *S. albus* SAM2 mutants

Strains / **Mutations**		**1300**	**1298**	**1297**	**1296**	**R94G**	**1295**	**1304**	**1305**	**1408**	**Total**
Transi-tions	A:T -> G:C	1	7	3	1	1	1		1		**15**
	G:C -> A:T		1	1			1			1	**4**
Trans-versions	A:T -> T:A										**0**
	G:C -> T:A	1									**1**
	A:T -> C:G					1					**1**
	G:C -> C:G			1			1				**2**
Coding		1	4	4	3	2	3	1		1	**19**
Intergenic		1	5	2			1		1		**10**
Synonymous			1								**1**
Nonsynonymous		1	2	4	1	2	2			1	**13**
Insertion					1						**1**
Deletion			1	1	1		1	1			**5**

Strain R94G carried two mutations in addition to the engineered *rpsL* substitution [15]. Hence, unanticipated rearrangements within genetically modified strain may contribute to the observed phenotypes [15]. The other eight strains (except for KO-1304; *vide infra*) that underwent selection for antibiotic resistance, carried mutations in expected targets (e.g., *rpsL*, *rpoB*) as well as the other genomic loci. The mutant KO-1304 was the only strain that carried a single deletion (within *xnr_2147*) and no single nucleotide variants (SNVs). Out of 29 mutations detected in total, there were six single-nucleotide indels (5 deletions and 1 insertion) and 23 SNVs. Two/thirds of the mutations (19) are located within coding sequences. Out of 23 transitions and transversions 16 (70 %) lead to replacement of AT with GC nucleotides. Almost all mutations were clustered within the core genome region spanning 2.0 – 6.2 Mbp segment of 6.8-Mbp *S. albus* genome (Fig. 2).

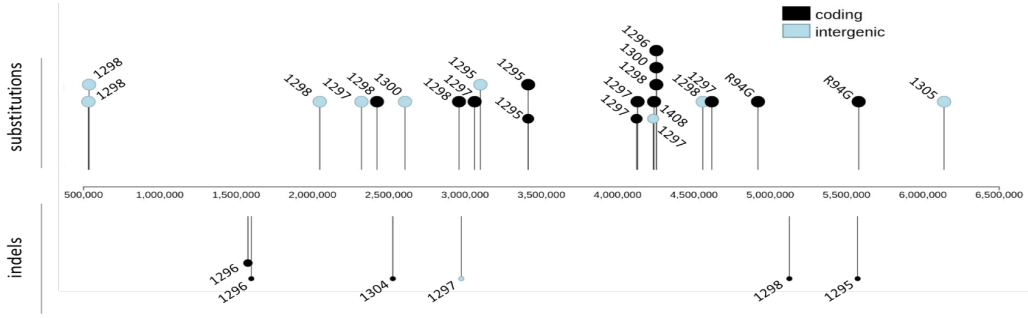

Fig. 2. Distribution of the identified mutations along *S. albus* chromosome. Strains are marked with numbers, e.g. *S. albus* KO-1296 is denoted as 1296

GC content across Streptomyces genomes. What would be the plausible reasons for the biased accumulation of mutations in *S. albus* J1074? We decided to approach this issue by taking

a closer look on the GC content of the available *Streptomyces* genomes. Although the general notion of high GC content for this genus is valid, there is no detailed analysis of this genomic parameter, despite significant growth of current databases. We undertook such analysis for several different datasets. The first dataset consisted of 200 high-quality complete *Streptomyces* genomes selected from the NCBI Genome database. The second one featured 1500 Streptomyces genomes, including the complete genomes as well as draft chromosomes having no more than 5 % ambiguities (see lists of the genomes at https://biotools.online/media/). The mean GC% value for both datasets was near 72 % (Fig. 3), although a number of notable outliers was revealed. Particularly, the genome of the strain *Streptomyces sp. NP1*60 showed the highest GC content (74.93 %), while *Streptomyces sp. SID10244* – at 66.53 % had the lowest GC content. The chromosome of the latter strain is in the draft stage, but the assembly quality permits to conclude that the computed GC% was not caused by sequencing artifacts. Hence, it can be concluded that noticeable variation in GC content can be observed for this genus. Finally, we explored the possibility that GC content of *Streptomyces* genes may depend on their essentiality. Namely, primary metabolism genes could be under more stringent selection against the biased accumulation of GC base pairs. At the same time, secondary metabolism genes, being less constrained that essential genes imposed on essential genes, could accumulate more AT → GC replacements. Using extensively validated annotation of the genome of model species *S. coelicolor* A3(2) [24], we compared the GC% for the primary and secondary metabolic datasets of this species. The distribution of GC% values for primary metabolic genes had a mean value similar to that observed for the entire genome, with a long tail into low GC part of the plot. The GC% values for secondary metabolic genes were more clustered, and the mean value for the entire dataset was slightly above that for primary metabolic genes or the entire genome (data not shown).

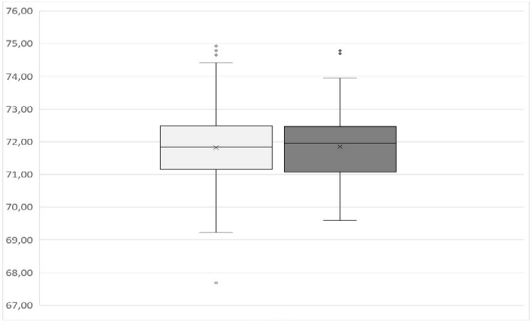

Fig. 3. The range of GC content in *Streptomyces* genomes. Box plot charts summarize the distribution of GC% values for 200 *Streptomyces* complete genomes (light grey boxplot), and complete plus draft genomes (dark grey one). Error bars indicate confidence interval for the mean value (95 %)

Genomic portrait of DNA mismatch repair systems in *Streptomyces*. Recent work [12] has demonstrated that actinobacterial genomes invariably lack orthologues of MutLS DNA mismatch repair system, featuring instead NucS homologues, likely of archaeal origin. The conclusion about the universal absence of MutLS-encoding genes within Actinobacteria has been supported by analysis of 300 genomes encompassing different genera (including several dozens of *Streptomyces*) of this huge taxon. Availability of the larger *Streptomyces* datasets (see above) have prompted us to re-visit the distribution of NucS orthologues within this single actinobacterial genus. Our analysis supports the conclusion made by Castaneda-Garcia et al. and shows little level of diversity of gene content around NucS homologues. Only at the fourth and fifth rightmost position did we observe a presence of different genes in different species (Fig. 4). We have not found any *Streptomyces* genome encoding MutL or MutS orthologues, despite the use of the query sequences of different origin and different search strategies.

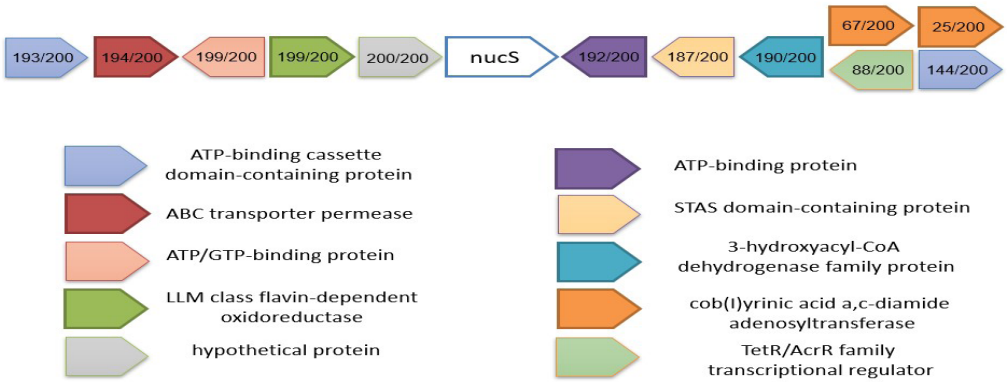

Fig. 4. Summary of genetic organization of homologous segments of 201 *Streptomyces* genomes carrying NucS orthologue (*SCO5388* in *S. coelicolor* M145). Numbers on gene symbols (arrows) denote the number of genomes carrying certain gene in that position. The leftmost gene is found in 193 out of 201 analyzed genomes (193/201). See graphical legend for predicted gene functions.

Streptomyces bacteria are notable for large linear genomes possessing high proportion of GC base pairs. Numerous explanations have been put forward about the mechanisms behind variable GC content across bacterial taxa (see Introduction). These explanations most often take the shape of revealing new correlations between the GC content and various genomic and/or physiological aspects [25]. In this work we took advantage of in-house S. albus J1074 genomic database to find the evidence of GC-biased mutagenesis in *Streptomyces*, which will serve as a stepping stone to understand the mechanisms behind this phenomenon. We show, on a limited experimental dataset, that in *S. albus* J1074 genome mutations towards GC nucleotides are prevalent. This biased accumulation of GC bases may be caused by the function of NucS-based DNA mismatch repair in all streptomycetes. In related actinobacterial genus *Corynebacterium* NucS showed preference for certain mismatch types, thus mitigating the asymmetric accumulation of replication errors [26]; the specificity of a *Streptomyces* NucS protein awaits experimental scrutiny. We believe that analysis of more expertly annotated *Streptomyces* genomes will yield valuable insights into the role of selection in shaping the GC content of this exciting group of bacteria.

***Acknowledgements**. B.O. thanks for grant support of the Ministry of Education and Science of Ukraine (BG-80F). T.W., T.G, T.S.J thank for grant support of the Novo Nordisk Foundation (NNF16OC0021746, NNF20CC0035580).*

REFERENCES

1. *Barka E., Vatsa P., Sanchez L.* et al. Taxonomy, physiology, and natural products of Actinobacteria // Microbiol. Mol. Biol. Rev. 2015. Vol. 80. P. 1–43. doi.org/10.1128/MMBR.00019-15.
2. *Hershberg R., Petrov D.* Evidence that mutation is universally biased towards AT in bacteria // PLoS Genet. 2010. Vol. 6(9). P. e1001115. doi.org/10.1371/journal.pgen.1001115.
3. *Segata N., Börnigen D., Morgan X., Huttenhower C.* PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes // Nat. Commun. 2013. Vol. 4. P. 2304. doi.org/10.1038/ncomms3304.
4. *Hug L., Baker B., Anantharaman K.,* et al. A new view of the tree of life // Nat. Microbiol. 2016. Vol. 1. P. 16048. doi.org/10.1038/nmicrobiol.2016.48.

5. *Aslam S., Lan X., Zhang B.* et al. Aerobic prokaryotes do not have higher GC contents than anaerobic prokaryotes, but obligate aerobic prokaryotes have // BMC Evol. Biol. 2019. Vol. 19. P. 35. doi.org/10.1186/s12862-019-1365-8.
6. *Vetsigian K., Goldenfeld N.* Genome rhetoric and the emergence of compositional bias // Proc. Nat. Acad. Sci. USA. 2009. Vol. 106. P. 215–220. doi.org/10.1073/pnas.0810122106.
7. *Rokytskyy I., Kulaha S., Mutenko H.* et al. Peculiarities of codon context and substitution within streptomycete genomes // Visn. Lviv Univ. Ser. Biol. 2017. Vol. 75. P. 66–74. dx.doi.org/10.30970/vlubs.2017.75.07.
8. *Good B., McDonald M., Barrick J.* et al. The dynamics of molecular evolution over 60,000 generations // Nature. 2017. Vol. 551. P. 45–50. doi.org/10.1038/nature24287.
9. *Lenski R.* Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations // ISME J. 2017. Vol. 11. P. 2181–2194. doi.org/10.1038/ismej.2017.69.
10. *Kucukyildirim S., Long H., Sung W.* et al. The rate and spectrum of spontaneous mutations in *Mycobacterium smegmatis*, a bacterium naturally devoid of the postreplicative mismatch repair pathway // G3 Genes|Genomes|Genetics. 2016. Vol. 6. P. 2157–2163. doi.org/10.1534/g3.116.030130.
11. *Dillon M., Sung W., Lynch M., Cooper V.* The Rate and molecular spectrum of spontaneous mutations in the GC-rich multichromosome genome of *Burkholderia cenocepacia* // Genetics. 2015. Vol. 200. P. 935–946. doi.org/10.1534/genetics.115.176834.
12. *Castañeda-García A., Prieto A., Rodríguez-Beltrán J.* et al. A non-canonical mismatch repair pathway in prokaryotes // Nat. Commun. 2017. Vol. 8. P. 14246. doi.org/10.1038/ncomms14246.
13. *Bilyk B., Luzhetskyy A.* Unusual site-specific DNA integration into the highly active pseudo-*attB* of the *Streptomyces albus* J1074 genome // Appl. Microbiol. Biotechnol. 2014. Vol. 98. P. 5095–104. doi: 10.1007/s00253-014-5605-y.
14. *Shima J., Hesketh A., Okamoto S.* et al. Induction of actinorhodin production by *rpsL* (encoding ribosomal protein S12) mutations that confer streptomycin resistance in *Streptomyces lividans* and *Streptomyces coelicolor* A3(2) // J. Bacteriol. 1996. Vol. 178. P. 7276–7284. doi.org/10.1128/jb.178.24.7276-7284.1996.
15. *Lopatniuk M., Myronovskyi M., Nottebrock A.* et al. Effect of «ribosome engineering» on the transcription level and production of *S. albus* indigenous secondary metabolites // Appl. Microbiol. Biotechnol. 2019. Vol. 103. P. 7097–7110. doi.org/10.1007/s00253-019-10005-y.
16. *Kieser T., Bibb M., Buttner M.* et al. Practical *Streptomyces* genetics. Norwich: John Innes Foundation. 2000. 634 p.
17. *Bekiesch P., Zehl M., Domingo-Contreras E., Martín J.* et al. Viennamycins: lipopeptides produced by a *Streptomyces* sp. // J. Nat. Prod. 2020. Vol. 83. P. 2381–2389. doi.org/10.1021/acs.jnatprod.0c00152.
18. *Andrews S.* FastQC: a quality control tool for high throughput sequence data / 2010. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
19. *Bolger A., Lohse M., Usadel B.* Trimmomatic: a flexible trimmer for Illumina sequence data // Bioinformatics. 2014. Vol. 30. P. 2114–2120. doi.org/10.1093/bioinformatics/btu170.
20. *Langmead B., Salzberg S.* Fast gapped-read alignment with Bowtie 2 // Nat. Methods. 2012. Vol. 9. P. 357–359. doi.org/10.1038/nmeth.1923.
21. *Hilker R., Stadermann K., Doppmeier D.* et al. ReadXplorer-visualization and analysis of mapped sequences // Bioinformatics. 2014. Vol. 30. P. 2247–2254. doi.org/10.1093/bioinformatics/btu205.

22. *Kuzniar A., van Ham R.C., Pongor S., Leunissen J. A.* 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.
23. *Schroeder M. P.* muts-needle-plot: Mutations Needle Plot v0.8.0 / 2015. Available from: doi.org/10.5281/zenodo.14561.
24. *Jeong Y., Kim J. N., Kim M. W.* et al. The dynamic transcriptional and translational landscape of the model antibiotic producer *Streptomyces coelicolor* A3(2) // Nat. Commun. 2016. Vol. 7. P. 11605. doi.org/10.1038/ncomms11605.
25. *Almpanis A., Swain M., Gatherer D., McEwan N.* Correlation between bacterial G+C content, genome size and the G+C content of associated plasmids and bacteriophages // Microb. Genom. 2018. Vol. 4. P. e000168. doi.org/10.1099/mgen.0.000168.
26. *Takemoto N., Numata I., Su'etsugu M., Miyoshi-Akiyama T.* Bacterial EndoMS/NucS acts as a clamp-mediated mismatch endonuclease to prevent asymmetric accumulation of replication errors // Nucleic Acids Res. 2018. Vol. 46. P. 6152–6165. doi.org/10.1093/nar/gky481.

Стаття надійшла до редакції 11.10.21

прийнята до друку 19.10.21

БАЗА СЕКВЕНОВАНИХ ГЕНОМІВ РІЗНИХ ШТАМІВ *STREPTOMES ALBUS* J1074 ТА ЇЇ ВИКОРИСТАННЯ

Б. Доля[1], Т. Буше[2], Т. Ш. Йорґенсен[4], К. Очі[3], Т. Грень[4], Т. Вебер[4], Й. Каліновскі[2],А. Лужецький[5], Б. Осташ[1*]

[1]Львівський національний університет імені Івана Франка
вул. Грушевського, 4, Львів 79005, Україна
**e-mail: bohdan.ostash@lnu.edu.ua*
[2]CeBiTec (ЦеБіТек), Білефельдський університет
вул. Universitatstrasse, 25, Білефельд 33615, Німеччина
[3]Відділ природничих наук, Хіросімський технологічний інститут
Міяке 2-1-1, Саекі-ку, Хіросіма 731-5193, Японія
[4]Центр Біосталості компанії Ново Нордіск,
Технологічний університет Данії
Кеміторвет, буд. 220, Люнгбю 2800, Данія
[5]Відділ фармакології, Саарландський університет
Кампус С2.3, Саарбрюкен 66123, Німеччина

Streptomyces albus J1074 – модельний об'єкт-шасі для дослідження різних аспектів біології актинобактерій, який, хоч і з'явився відносно недавно, широко застосовують і для дослідження наявних, і для відкриття нових природних сполук. Створено велику кількість генетичних знарядь, покликаних полегшити використання штаму для названих потреб. Одним із перспективних підходів є введення в геном J1074 мутацій, які покращили би його здатність продукувати антибіотики. Зокрема, нами попередньо описано колекцію спонтанних і генно-інженерних мутантів J1074, які містять мутації в генах рибосомного білка S12, бета-субодиниці РНК-полімерази тощо. На основі цієї колекції ми створили власну базу даних, яка містить первинні та куровані дані геномних послідовностей спонтанних і генетично сконструйованих мутантів J1074. База даних доступна за адресою https://biotools.online/media/. Основні переваги бази даних полягають у відомому родоводі штамів, що дає змогу поглиблено

аналізувати й інтерпретувати отримані дані. Наприклад, ведуться жваві та досі не завершені дискусії щодо походження і впливу ГЦ складу в актинобактерій. Краще розуміння цього питання покращить наші знання про еволюцію геномів у бактерій, а також, як наслідок, матиме низку практичних застосувань у біотехнології. Ми використали наш набір даних *Streptomyces albus* J1074 як експериментальну модель для виявлення загальногеномного спектру мутацій, котрі, як бачимо, зміщені у бік підвищеного ГЦ вмісту. Для порівняння ми включили до нашої бази і високоякісні геноми інших стрептоміцетів. ГЦ відсоток у геномах стрептоміцетів коливається від 75 % до 66,5 %, зі середнім значенням 72 %. Вміст ГЦ у генах вторинного метаболізму *S. coelicolor* менш мінливий порівняно з генами первинного метаболізму, що може свідчити про різний тиск добору на ці групи генів. Поряд зі селективними обмеженнями, особливості системи репарації ДНК у *Streptomyces* можуть сприяти зміщенню ГЦ вмісту в їхніх геномах. Подальше використання бази даних може забезпечити розвиток більш точних знань про швидкість появи мутацій, а також про популяційні генетичні процеси у межах цього виду та роду загалом.

Ключові слова: Streptomyces albus, секвенування геному, спектр мутацій