УДК 579.873.71 : 577.113.4 : 004.9

ПОЛІЩУК Л.В.
Інститут мікробіології і вірусології ім. Д.К. Заболотного НАН України,
Україна, ДО3680, Київ МСП, вул. Академіка Заболотного, 154, e-mail: LVPolishchuk@ukr.net

IN SILICO ПОШУК gvp-КЛАСТЕРІВ *STREPTOMYCES GLOBISPORUS* 1912-2

Газові вакуолі були здебільш знайдені у водних прокаріот як фото-, так і хемотрофних (пурпурних, сірчаних, зелених бактерій, деяких ціанобактерій і клостридій та ряду інших). Плавучість їх регулюється шляхом зміни числа та розміру аеросом. Протеїни, що утворюють мембрани газової вакуолі кодуються кластером gvp-генів. Наприклад, у штаму *Halobacterium* sp. NRC-1 gvp-кластер містить 14 генів, а gvp-кластер штаму *Pseudoanabaena* sp. PCC6901 – 4 гени [1, 2].

Неочікуванно у багатьох ґрунтових мікроорганізмів (у тому числі і актиноміцетів) за допомогою *in silico* аналізу первинної будови їх хромосом було виявлено кластери gvp-генів. Однак у представників данної родини відсутні ендоцитоплазматичні включення такого різновиду як аеросоми. Функції данних протеїнів у стрептоміцетів становлять великий інтерес для науковців. Встановлено, gvp-кластери актиноміцетів, наприклад, стрептоміцетів включають до 10 генів [1, 3].

Матеріали і методи

Досліджували бібліотеку контигів, що містять інформацію про первинну будову 1438 фрагментів тотальної ДНК штаму *S. globisporus* 1912-2. Визначення нуклетидних послідовностей фрагментів тотальної ДНК стрептоміцету було проведено акредитованою компанією «BaseClear» (Лейден, Голандія) у 2013 році. В роботі наводяться результати *in silico* аналізу контигів 264 (7468 пн), 756 (5821 пн) та 781 (2520 пн). gvp-Гени локалізовані на їх фрагментах: Contig 264 – 1582–7132 пн (5551 пн), Contig 756 – 2889–5821 пн (3084 пн), Contig 781 – 303–2519 пн (2218 пн).

Використана інформація з Інтернет баз серверу NCBI «Nucleotide» та «Genome» [4]. *In silico* аналіз (метод попарного вирівнювання) нуклеотидної будови ДНК стрептоміцетів проводили з допомогою пакету програм BLASTN 2.2.31+: megablast та discontiguous megablast [5]. Використали налаштування програми BLASTN за замовчуванням. В *in silico* аналізі бібліотеки контигів *S. globisporus* 1912-2 у якості реперних використовували послідовності 21 gvp-кластеру 11 стрептоміцетів (табл. 1).

Результати та обговорення

Наявність аеросом встановлено майже у 50 родів 5 класів царства *The Bacteria*. Була встановлена будова їх gvp-кластерів та функції більшості gvp-генів [1, 2].

Однак, завдяки прогресу у царині комп'ютерних технологій та методів визначення нуклеотидної будови ДНК, було виявлено кластери gvp-генів у ґрунтових мікроорганізмів, у яких аеросоми не виявлені, наприклад, у стрептомицетів, бацил, кишкових паличок. Крім того, встановлено, що мутанти *S. coelicolor* A3(2), які втратили обидва gvp-кластери мають здатність до флотації у рідких середовищах [1].

Як повідомлялося раніше, для вивчення первинної будови хромосоми *S. globisporus* 1912-2 було створено бібліотеку контигів, що включала 1438 фрагментів, молекулярний розмів яких становив від 358 пн (Contig 1421) до 75588 пн (Contig 2). Сумарний молекулярний розмір усіх контигів становить 7,125 мпн [6].

При аналізі *in silico* бібліотеки контигів *S. globisporus* 1912-2 було виявлено наявність послідовностей, гомологічних реперним послідовностям gvp-кластерів (табл. 1) на 3 контигах *S. globisporus* 1912-2 (Contigs 264, 781 та 756) (табл. 2). У таблиці 2 здебільшого представлені послідовності, для яких поріг відсіву (E-value) становть 0. Виняток – *S. scabiei* 87.22, для якого E-value = 2e-122.

Встановлено, що фрагменти послідовностей контигів 756 та 781 гомологічні послідовностям одних і тих же реперних gvp-кластерів (табл. 2). Однак, наприклад, послідовність контигу 756 гомологічна фрагменту gvp3-кластеру *S. avermitilis* MA-4680 2861440 пн – 2864626 пн (гени gvpK,S,L,J,Z,Y), а послідовність контигу 781 гомологічна фрагменту 2864925 пн – 2867124 пн (гени gvpY,G,F,A,O) (рис. 1). У той час, як послідовність (1585 – 7130 пн) Contig 264 гомологічна послідовності усіх 10 генів gvp2-кластеру S. albus DSM 41398 (7687341 – 7693045 пн).

© ПОЛІЩУК Л.В.

Таблиця 1

gvp-Кластери стрептоміцетів, що були використані як реперні

Штами стрептоміцетів	Посилання в GenBank	Локалізація gvp-кластерів, пн
S. coelicolor A3(2) фрагм.25/29	AL939128.1, 296500 пн	23830-29188
S. coelicolor A3(2) фрагм. 3/29	AL939106.1, 314100 пн	101016-106218
S. lividans TK24	CP009124.1, 8345283 пн	1316696-1322048 7676311-7681459
S. avermitilis MA-4680 = NBRC 14893	BA000030.3, 9025608 пн	746752-750848 2307611-2312879 2861430-2867207
S. scabiei 87.22	FN554889, 10148695 пн	1952635-1958115 4318447-4323402
S. pratensis ATCC 33331	CP002475.1, 7337497 пн	300816-306378
Streptomyces sp. PAMC26508	CP003990.1, 7526197пн	7201966-7207300
Streptomyces sp. SirexAA-E	CP002993.1, 7414440 пн	7274383-7280090
S. albus DSM 41398	CP010519.1, 8384669 пн	4428226-4432827 7687287-7693052
S. albus J1074	CP004370.1, 6841649 пн	5038209-5043486
S. venezuelae ATCC 10712	FR845719.1, 8226158 пн	411040-416610 7502637-7506707
S. rapamycinicus NRRL 5491	CP006567.1, 12700734 пн	1909544-1913233 3056067-3059885 4928360-4931998 9184485-9187989

Таблиця 2

Гомологія нуклеотидної будови фрагментів ДНК *S. globisporus* 1912-2 та реперних послідовностей стрептоміцетів*

Показники гомології первинної будови послідовностей ДНК стрептоміцетів (Query)		
Contig 264 (1582-7132 пн)	Contig 781 (303-2519 пн)	Contig 756 (2389-5821 пн)
S. albus DSM 41398 CP010519.1 (7687341-7693045 пн) Id. - 72% Q.c. - 96%	*S. avermitilis* MA-4680 BA000030.3 (2864925 - 2867124 пн) Id. - 75% Q.c. - 86%	*S. avermitilis* MA-4680 BA000030.3 (28614400 -2864500 пн) Id. - 73% Q.c. - 98%
S. venezuelae ATCC 10712 FR845719.1, (411060-416553 пн) Id. - 71% Q.c. - 90%	*S. pratensis* ATCC 33331 CP002475.1 (304109 - 306275 пн) Id. - 74% Q.c. - 99%	*S. pratensis* ATCC 33331 CP002475.1 (300816 - 303869 пн) Id. - 73% Q.c. - 97%
S. scabiei 87.22 FN554889, (4318832-4323394 пн) Id. - 74% Q.c. - 55%	*S. sp.* PAMC26508 CP003990.1 (7201978 -7204055 пн) Id. - 74% Q.c. - 95%	*S. sp.* PAMC26508 CP003990.1 (7204289-7207300 пн) Id. - 73% Q.c. - 97%
	Streptomyces sp. SirexAA-E CP002993.1 (7274464 -7276662 пн) Id. - 72% Q.c. - 99%	*Streptomyces sp.* SirexAA-E CP002993.1 (7276987-7280090 пн) Id. - 72% Q.c. - 97%

П р и м і т к и: Id. – ідентичність нуклеотидної будови послідовностей ДНК стрептоміцетів, Q.c. (Query cover – покриття запросу), * – попарне вирівнювання проводили програмою bl2seq: discontiguous megablast.

Рис. 1. Локалізація гомологічних послідовностей контигів *S. globisporus 1912-2* (756 та 781) відносно генів gvp3-кластеру *S. avermitilis* MA-4680

За даними літератури, у більшості просіквенованих хромосом стрептоміцетів було виявлено gvp-кластери [1, 3, 7]. Встановлено, що стрептоміцетна хромосома може включати до 4 gvp-кластерів. Наприклад, у штаму *S. pratensis* ATCC 33331 виявлено 1 gvp-кластер (локалізований на фрагменті 2300816 пн – 306378 пн), а у штаму *S. rapamycinicus* NRRL 5491 – 4 gvp-кластер (1909544 пн – 1913233 пн, 3056067 пн – 3056378 пн, 4928360 пн – 4931998 пн, 9184485 пн – 9187989 пн). Показано, що у актиноміцетів кількість кластерів gvp-генів у хромосомі корелює з її довжиною. Так, довжина хромосоми штаму *S. rapamycinicus* NRRL 5491 (GenBank CP002475.1) становить 7337497 пн, а штаму *S. rapamycinicus* NRRL 5491 (CP006567.1) – 12700734 пн. Цікавим є те, що в хромосомі *S. griseus* subsp. *griseus* NBRC 13350 (AP009493, 8545929 пн) не було виявлено жодного gvp-кластеру [4]. Виявлений найбільший фрагмент *S. griseus* в 387 пн (6635477-6635863 пн), послідовність якого ідентична на 76 % послідовності gvpO-гена (4428226-4428651 пн) *S. albus* DSM 41398 (90 %). Протеїн BAG2248 (156 аа) детермінує стресовий протеїн ASP23 [4, 5].

Нами встановлено наявність послідовності 2 gvp-кластерів на хромосомі *S. globisporus* 1912-2. Сумарний молекулярний розмір усіх 1438 контигів бібліотеки становив 7125 тпн. Однак ми вважаємо, що бібліотека містить інформацію відносно менше, ніж 90 % нуклеотидної послідовності хромосоми мікроорганізму, тому можлива наявність у хромосомі 3 gvp-кластеру.

За даними літератури gvp-кластери стрептоміцетів можуть містити до 10 генів (наприклад, обидва gvp-кластери *S. coelicolor* A3(2)) [1, 7] .

Показано, що послідовність генів у більшості кластерів стрептоміцетів мають подібний порядок: gvpO, A, F, G, Y, Z, J, L, S, K. У той же час зареєстровані випадки відсутності окремих gvp-генів та відмінної їх послідовності в кластері. Так, в gvp2-кластері (4318447 пн – 4323403 пн) *S. scabiei* 87.22 відсутній gvpL-ген, а в gvp2-кластері *S. venezuelae* ATCC 10712 (7502637 пн – 7506707 пн) gvpY- та Z-гени локалізовані перед gvpO-геном. Крім того, встановлено, що ряд генів у gvp-кластерах стрептоміцетів є паралогічними (gvpA-gvpJ-gvpS та gvpF-gvpL) [1, 2, 7].

Наш *in silico* аналіз показав, що обидва кластери gvp-генів *S. globisporus* 1912-2 мають класичну будову як по кількості генів, так і по їх порядку у кластері:

gvpO - A - F - G - Y - Z - J - L - S - K.

За літературними данними, *in silico* аналіз послідовностей gvp-кластерів одного і того ж штаму стрептоміцета, виявив, що, як загалом послідовність цих кластерів, так і окремих їх генів не є ідентичними за нуклеотидною будовою [1]. Так, ідентичность 2 gvp-кластерів *S. coelicolor* A3(2) становить 72 %, однак гомологічними є тільки 27 % довжини послідовності. Однак знайдена значна гомологія первинної будови gvp-кластерів різних штамів стрептоміцетів: так 98 % довжини gvp-кластера *S. coelicolor* A3(2) на фрагменті 3/29 (AL939106.1 – 101016-106218 пн) є ідентичними на 99 % gvp2-кластеру *S. lividans* TK24 (CP009124.1 – 7676311-7681459 пн).

Наш *in silico* аналіз показав, що окремі послідовності gvp-кластерів *S. globisporus* 1912-2 мають значну гомологію, однак довжина гомологічних фрагментів не перевищує 30 % загальної довжини кластеру (табл. 3).

In silico аналізом (megablast) показана ідентичність ряду окремих невеликих фрагментів 3 досліджуваних контигів *S. globisporus* 1912-2. Так, виявлена гомологія кількох фрагментів Contig 264 та Contig 756, найбільші з яких у (відповідно 533 пн та 773 пн) гомологічні gvpZ-генам. Довжина gvpZ-гена *S. venezuelae* ATCC 10712 становить 1140 пн, а gvpZ-ген *S. avermitilis* MA-4680 – 1131 пн. Встановлено, що ідентичні між собою фрагменти в послідовностей Contig

Таблиця 3

Гомологія послідовностей gvp-кластерів *S. globisporus* 1912-2*

Query \ Subject	Contig 264	Contig 756	Contig 781
Contig 264	100 %	Id. 69 %, Q.c. 16 %	Id. 80 %, Q.c. 6 %
Contig 756	Id. 69 %, Q.c. 37 %	100 %	0 %
Contig 781	Id. 80 %, Q.c. 16 %	0 %	100 %

П р и м і т к и: Id. – ідентичність нуклеотидної будови послідовностей ДНК стрептоміцетів, Q.c. (Query cover – покриття запросу), * – попарне вирівнювання проводили програмою bl2seq: megablast.

Рис. 2. Локалізація гомологічних послідовностей контигів 756 та 781 *S. globisporus 1912-2* відносно генів gvp-кластеру Contig 264

264 та Contig 781 (відповідно 132 пн та 164 пн) є частиною gvpA-генів. gvpA-ген *S. venezuelae* ATCC 10712 має довжину 438 пн, а gvpA-ген *S. avermitilis* MA-4680 – 471 пн (рис. 2).

Відсутність значної гомології нуклеотидної будови різних gvp-кластерів одного і того ж мікроорганізму (наприклад, *S. coelicolor* A3(2), *S. lividans* TK24 чи *S. globisporus* 1912-2) надає можливість припускати, що значні зміни їх первинної будови сталися після дуплікації в наслідок мутацій чи ортологічність їх походження.

Таким чином, у бібліотеці контигів тотальної ДНК *S. globisporus* 1912-2 було виявлено 3 контиги (264, 756 та 781), на яких містилися 2 gvp-кластери. Один з gvp-кластерів локалізований на фрагменті Contig 264 (1582–7132 пн), другий кластер – на фрагментах 2 контигів (Contig 756 – 2889-5821 пн та Contig 781 – 303-2519 пн). Обидва gvp-кластери *S. globisporus* 1912-2 містять по 10 генів і мають традиційний для стрептоміцетів порядок їх у кластері: gvp**O - A - F - G - Y - Z - J - L - S - K.** Показано, що тільки невеликі послідовності gvp-кластерів *S. globisporus* 1912-2 мають достатню гомологію.

Висновки

Встановлена наявність двох gvp-кластерів у хромосомі *S. globisporus* 1912-2, що містять по 10 генів і мають традиційний для стрептоміцетів порядок їх у кластері: gvp**O - A - F - G - Y - Z - J - L - S - K.** Кластери gvp-генів *S. globisporus* 1912-2 не мають значної взаємної гомології нуклеотидної будови послідовностей, тому можна вважати їх ортологічними чи припустити, що сталися значні зміни їх первинної будови в наслідок мутацій після дуплікації.

ЛІТЕРАТУРА

1. van Keulen G., Hopwood D.A., Dijkhuizen L., Sawers R.G. Gas vesicles in actinomycetes: old buoys in novel habitats? // Trends Microbiol. – 2005. – 13, N 8. – P. 350–354.
2. Pfeifer F. Distribution, formation and regulation of gas vesicles // Nat. Rev. Microbiol. – 2012. – 10, N 10. – P. 705–715.
3. Walsby A.E., Dunton P.G. Gas vesicles in actinomycetes? // Trends Microbiol. – 2006. – 14, N 3. – P. 399–400.
4. The National Center for Biotechnology Informatin [Електронний ресурс]. – Режим доступу: http://www.ncbi.nlm.nih.gov/nucleotide/ та http://www.ncbi.nlm.nih.gov/genome/.
5. The National Center for Biotechnology Informatin [Електронний ресурс]. – Режим доступу: http://blast.ncbi.nlm.nih.gov/Blast.cgi BLASTN (bl2seq: megablast, discontiguous megablast).
6. Полищук Л.В., Мацелюх Б.П. рРНК-гены актиномицетов, гомологичные генам рРНК-кластеров *Streptomyces globisporus* 1912-2 // Фактори генетичної еволюції організмів. – 2014. – 14. – С. 129–133.
7. Karoonuthaisiri N., Weaver D., Huang J., Cohen S.N., Kao C.M. Regional organization of gene expression in *Streptomyces coelicolor* // Gene. – 2005. – 353, N 1. – P. 53–66.

POLISHCHUK L.V.

Zabolotny Institute of Microbiology and Virology of Natl. Acad. Sci. of Ukraine,
Ukraine, 03680, Kyiv, Akademika Zabolotnogo str., 154, e-mail: LVPolishchk@ukr.net

IN SILICO SEACHING OF *STREPTOMYCES GLOBISPORUS 1912-2* GVP-CLASTERS

*The **aims*** – to find of gvp-clasters in the genome of *S. globisporus* 1912-2. ***Methods***. The programs BLASTN were used for *in silico* analysis of the *S. globisporus* 1912-2 contigs library. ***Results***. Two gvp-clusters in the library of *S. globisporus* 1912-2 chromosome fragments were found. One gvp-cluster was localized on Contig 264, the second cluster – on Contig 756 and Contig 781. Both gvp-clusters contain 10 genes in traditional order for streptomycetes cluster: gvpO - A - F - G - Y - Z - J - L - S - K. ***Conclusions***. *In silico* analysis of the primary structures of two *S. globisporus* 1912-2 gvp-clusters did not find significant homology in their nucleotide sequences, therefore we could suggest that they were orthologous or significant changes in their primary structures occurred as a results of mutations after duplication.

Keywords: Streptomyces globisporus 1912-2, gvp-claster, homology, chromosomal DNA.