

АНДРЕЄВ І. О., КОНВАЛЮК І. І.[✉], КУНАХ В. А.Інститут молекулярної біології і генетики НАН України,
Україна, 03680, м. Київ, вул. Академіка Заболотного, 150[✉] konvalyuk.I.I@gmail.comБІОІНФОРМАТИЧНИЙ ПОШУК МОБІЛЬНИХ ГЕНЕТИЧНИХ ЕЛЕМЕНТІВ
НАДРОДИНИ *Ty1/COPIA* У *DESCHAMPSIA ANTARCTICA* E. DESV

Мета. Пошук *in silico* послідовностей ретротранспозонів надроддини *Ty1/Copia* в геномі *Deschampsia antarctica* E Desv. **Методи.** Біоінформатичний аналіз архівів коротких прочитань геному та транскриптому, представлених у базі даних GenBank. Як референтну послідовність для пошуку було використано МГЕ рису Tos 17 з надроддини *Ty1/Copia*. **Результати.** Проведений пошук виявив у геномі *D. antarctica* послідовності ретротранспозонів надроддини *Ty1/Copia*, які мають високий рівень подібності (до 75 % у районі гомологічних ділянок) до Tos 17. Нерівномірність розподілу знайдених прочитань уздовж референтної послідовності свідчить про існування в геномі групи послідовностей, які містять у своєму складі елементи, типові для надроддини, що і характеризуються різним ступенем подібності до Tos 17, причому більш консервативні з них представлені серед знайдених прочитань відповідно у більшій кількості. Присутність подібних до Tos 17 прочитань у транскриптомі вказує на те, що виявлені в геномі послідовності МГЕ характеризуються певним фоновим рівнем транскрипційної активності. **Висновки.** Методами біоінформатичного аналізу *in silico* встановлено наявність в геномі *D. antarctica* транскрипційно-активних послідовностей транспозонів надроддини *Ty1/Copia*. Знайдені послідовності можуть бути використані в подальших дослідженнях для створення праймерів та аналізу активності МГЕ цієї надроддини у *D. antarctica*.

Ключові слова: біоінформатичний пошук, *Deschampsia antarctica* E Desv., мобільні генетичні елементи, *Ty1/Copia*.

Мобільні генетичні елементи (МГЕ), які відкрито Барбарою МакКлінток [1], у більшості вищих рослин складають понад половину (у деяких злаків – до 90 % [2]) ядерного геному. На сьогоднішній день накопичується все більше

даних про те, що транспозиції МГЕ разом з іншими механізмами генетичної мінливості відіграють важливу роль у забезпеченні адаптивності та еволюції рослин [3–7]. Вони можуть індукувати зміни геному як на структурному, так і на функціональному рівнях. Питання про регуляцію активності МГЕ та роль спричиненої цим генетичної мінливості в адаптації рослин наразі викликає підвищену увагу дослідників.

МГЕ відрізняються за структурою і особливостями транспозиції, що дозволило поділити їх на два класи: ретротранспозони (механізм «копіювання і вставки») і ДНК-транспозони (механізм «вирізання і вставки») [4, 6, 8]. Реплікативний спосіб переміщення ретротранспозону дозволяє швидко накопичувати велику кількість його копій, що веде до збільшення розміру геному рослини [9–11]. Мутації, які виникають внаслідок інсерції ретротранспозону, є стабільними (на відміну від мутацій, зумовлених активністю ДНК-транспозонів), оскільки останні під час переміщення вирізають свою вихідну копію з геному, а потім вбудовуються в іншу його ділянку, тоді як копія ретротранспозону після вставки в геном залишається там перманентно [4]. До цього класу відносять LTR-вмісні ретротранспозони, особливістю будови яких є наявність довгих кінцевих повторів (long terminal repeats, LTR), які не кодують білки, але містять промотори та термінатори, що регулюють транскрипцію генів цих елементів. У складі LTR-ретротранспозонів виділяють надроддину *Ty1/Copia*-подібних елементів. Останнім часом у різних видів рослин і особливо у Triticaceae добре вивчено низку представників цієї надроддини ретротранспозонів [12–15].

Встановлено, що активація та переміщення певних МГЕ можуть відбуватися за впливу деяких стресових чинників та інших зовнішніх стимулів. Наслідком цього можуть бути зміни профілю експресії генів і зумовлені цим фенотипові зміни, оскільки відомо, що вставка МГЕ

інколи надає розташованим поруч генам здатність до активації транскрипції в умовах, які зазвичай викликають активацію самого МГЕ [15, 16]. Загалом підвищення мінливості геному рослин внаслідок активації МГЕ разом із прискоренням темпів еволюції може забезпечувати ефективний механізм адаптації рослин до нових умов довкілля. Привабливою моделлю для перевірки цієї гіпотези є рослини-екстремофіли, які пристосувалися до існування в умовах, непридатних для більшості інших видів. Одна з таких рослин – *Deschampsia antarctica* E Desv. (Poaceae), єдиний злак, що зростає у Морській Антарктиці [17]. Геном цього виду поки що залишається практично не дослідженим, а про МГЕ у його складі нічого не відомо.

Метою роботи був пошук *in silico* методами біоінформатичного аналізу послідовностей ретротранспозонів надродини *Ty1/Copia* в геномі *D. antarctica*.

Матеріали і методи

Для пошуку було використано архіви коротких прочитань (sequence read archive, SRA) із бази даних GenBank: SRX465632, яка містить дані розшифрування геномної ДНК (153,3 млн парних прочитань), та SRX465633, який містить транскриптомні послідовності виду (14,9 млн парних прочитань) [18]. Пошук послідовностей, гомологічних до ретротранспозонів надродини *Ty1/Copia* у геномі *D. antarctica*, проводили за допомогою програми BLASTN [19] із використанням стандартного алгоритму пошуку blastn. Максимальна кількість послідовностей для пошуку становила 10000. Як референтну послідовність для пошуку використали повну послідовність МГЕ рису (*Oryza sativa*) Tos 17 [12], яка розташована на ділянці [26694762-26698875] 7-ї хромосоми із бази даних Rice Genome Research Program [20]. У результаті пошуку отримували файл у форматі .sam, який у текстовому форматі містить знайдені нуклеотидні послідовності, вирівняні до референтної послідовності. Для аналізу результатів пошуку використовували програму Unipro UGENE [21].

Результати та обговорення

Дані розшифрування геному та транскриптому *D. antarctica* наразі представлені у вигляді архівів коротких парних прочитань довжиною 101 пн. Збирання з цих послідовностей довгих контігів та їх анотування поки що не проводилося. Це перешкоджає безпосередньому

пошуку послідовностей МГЕ за їхніми характерними консервативними мотивами і потребує застосування іншої стратегії пошуку. Тому в цьому дослідженні ми проводили пошук прочитань, гомологічних до референтної послідовності. Як таку послідовність використали повну послідовність МГЕ рису Tos 17 (4114 пн), яка містить довгі кінцеві повтори (LTR) довжиною 105 пн та розташовану між ними відкриту рамку зчитування довжиною 3336 пн. На початку послідовності МГЕ між 173 та 190 нуклеотидами розташована ділянка зв'язування праймера (PBS). В межах транскрибованої ділянки Tos 17 ідентифіковано мотиви, споріднені з доменом *GAG-pre-integrase*, ферментами інтегразою та зворотною транскриптазою, тобто всі основні елементи, типові для представників надродини *Ty1/Copia*, за винятком специфічного для кожної групи антигенного білка GAG.

У результаті пошуку в геномі *D. antarctica* послідовностей, гомологічних до МГЕ рису Tos 17, було знайдено 7611 коротких прочитань. Вирівнювання знайдених прочитань до референтної послідовності показало, що вони покривають лише частину геномної ДНК (рис.). Гомологічні фрагменти концентрувалися переважно в транскрибованій ділянці, виходячи за її межі і частково захоплюючи 5'-LTR. У межах транскрибованої ділянки вони розташовувалися майже по всій її довжині, за винятком послідовності, яка охоплює мотив *gag-pre*, а також ділянки довжиною 402 п. н. у центральній частині транскрибованого регіону. Найбільшу кількість гомологічних фрагментів було знайдено до ділянок 1565–1660, 1848–1981 та 2560–2720. Перша і остання з цих ділянок збігаються з послідовностями, які кодують інтегразу та зворотню транскриптазу відповідно. Порівняння гомологічних ділянок консенсусної послідовності, зібраної на основі знайдених прочитань, із відповідними ділянками референтної послідовності показало, що рівень подібності між ними становив 74 %.

Пошук гомологічних до Tos 17 послідовностей, проведений в архіві транскриптомних послідовностей *D. antarctica*, дав аналогічні результати, за винятком дещо більшого покриття референтної послідовності. Загалом у транскриптомі вдалося знайти 2310 гомологічних прочитань. Усі знайдені фрагменти за результатами вирівнювання так само розташовувалися в межах відкритої рамки зчитування і трохи виходили за її межі в ділянці 5'-LTR.

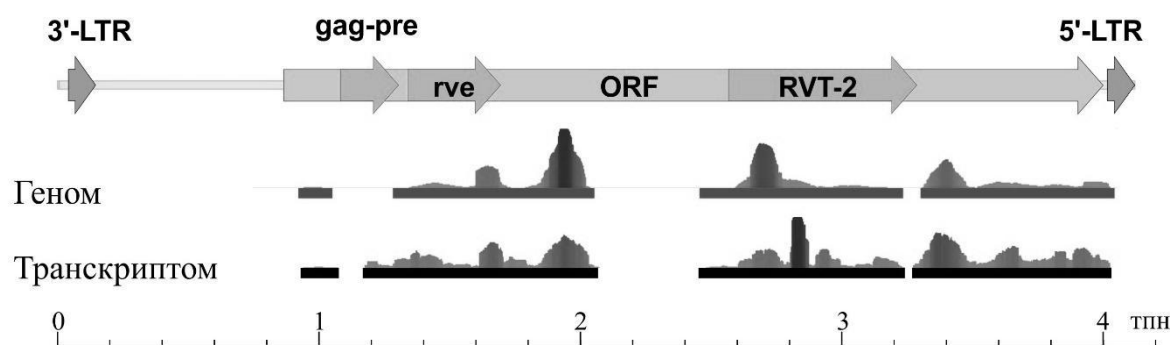


Рис. Схематична будова послідовності ретротранспозону Tos 17 за [12] та розташування знайдених у геномі (Геном) та транскриптомі (Транскриптом) *D. antarctica* гомологічних прочитань. Гістограма показує розподіл знайдених фрагментів уздовж референтної послідовності. ORF – відкрита рамка зчитування; LTR – довгі кінцеві повтори. Консервативні мотиви в транскрибованому регіоні: gag-pre – домен *GAG-pre-integrase*; rve – інтеграза; RVT-2 – зворотна транскриптаза.

Водночас розподіл знайдених фрагментів уздовж референтної послідовності дещо відрізнявся, що особливо помітно в кодувальній ділянці зворотної транскриптази. Подібність гомологічних ділянок зібраної на основі знайдених прочитань консенсусної послідовності і референтної послідовності Tos 17 становила 75 %.

Отже, у результаті проведеного пошуку в геномі *D. antarctica* виявлено послідовності ретротранспозонів надродини *Ty1/Copia*, які мають високий рівень подібності (до 75 % в районі гомологічних ділянок) до МГЕ рису Tos 17. Нерівномірність розподілу знайдених прочитань уздовж референтної послідовності свідчить про існування в геномі групи послідовностей, які містять у своєму складі елементи, типові для надродини, і характеризуються різним ступенем подібності до Tos 17, причому більш консервативні з них представлені серед знайдених прочитань відповідно у більшій кількості. Присутність у транскриптомі подібних до Tos 17 прочитань вказує на те, що наявні в геномі послідовності МГЕ характеризуються певним фоновим рівнем транскрипційної активності. Крім того, відмінності в характері розподілу гомологічних прочитань, знайдених у геномі та транскриптомі, можуть вказувати на те, що транскрипційно активною є лише певна частина із виявлених у геномі послідовностей МГЕ.

References

1. McClintock B. The significance of responses of the genome to challenge. *Science*. 1984. Vol. 226. P. 792–801.
2. Devos K.M., Brown J.K., Bennetzen J.L. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res*. 2002. Vol. 12. P. 1075–1079.
3. Bennetzen J.L. Transposable element contributions to plant gene and genome evolution. *Plant Mol. Biol.* 2000. Vol. 42. P. 251–269.
4. Todorovska E. Retrotransposons and their role in plant – genome evolution. *Biotechnology & Biotechnological Equipment*. Vol. 21 (3). P. 294–305. doi: 10.1080/13102818.2007.10817464.

МГЕ надродини *Ty1/Copia* є одними із найрозповсюдженіших серед ретротранспозонів рослин і знайдені майже у всіх представників рослинного царства від мохів до покритонасінних [1, 4]. Здебільшого ці мобільні елементи знаходяться в неактивному стані внаслідок метилювання, однак за дії стресу або ж у процесі онтогенезу їх транскрипційна активність може зростати, як це, зокрема, було описано для МГЕ Tos 17 в культурі тканин рису [12]. Отримані нами результати свідчать про транскрипційну активність *Ty1/Copia* транспозонів у пагонах *D. antarctica*. Чим це зумовлено: особливостями регуляції в геномі рослини-екстремофіла чи ж умовами, в яких вирощували рослини для отримання сумарної РНК, наразі сказати однозначно не можна. Однак цей факт вказує на те, що *D. antarctica* може виявитися перспективним об'єктом для дослідження активності МГЕ.

Висновки

Методами біоінформатичного аналізу *in silico* встановлено наявність у геномі *D. antarctica* транскрипційно активних послідовностей транспозонів надродини *Ty1/Copia*. Знайдені послідовності можуть бути використані в подальших дослідженнях для створення праймерів і аналізу активності МГЕ цієї надродини у *D. antarctica*.

5. Gerashchenkov G.A., Rozhnova N.A. Mobile genetic elements in plant sex evolution. *Russ. J. Genet.* 2010. Vol. 46 (11). P. 1271–1281. doi: 10.1134/S1022795410110013g.
6. Kunakh V.A. Mobile genetic elements and plant genome plasticity. Kyiv: Logos, 2013. 288 p. ISBN 978-966-171-721-2. [in Ukrainian] / Кунах В.А. Мобільні генетичні елементи і пластичність геному рослин. К.: Логос, 2013. 288 с.
7. Oliver K.R., McComb J.A., Greene W.K. Transposable elements: powerful contributors to angiosperm evolution and diversity. *Genome Biol. Evol.* 2013. Vol. 5(10). P. 1886–1901. doi: 10.1093/gbe/evt141.
8. Wicker T., Sabot F., Hua-Van A. et al. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 2007. Vol. 8. P. 973–982. doi: 10.1038/nrg2165.
9. Pearce S.R., Pich U., Harrison G. et al. The Ty1-copia group retrotransposons of *Allium cepa* are distributed throughout the chromosomes but are enriched in the terminal heterochromatin. *Chromosome Res.* 1996. Vol. 4. P. 357–364.
10. Kumar A., Bennetzen J.L. Plant retrotransposons. *Annu. Rev. Genet.* 1999. Vol. 33. P. 479–532.
11. Bennetzen J.L., Ma J., Devos K.M. Mechanisms of recent genome size variation in flowering plants. *Ann. Bot.* 2005. Vol. 95. P. 127–132. doi: 10.1093/aob/mci008.
12. Sabot F. Tos17 rice element: incomplete but effective. *Mobile DNA.* 2014. Vol. 5 (1). P. 10. doi: 10.1186/1759-8753-5-10.
13. Orłowska R., Machczyńska J., Oleszczuk S. et al. DNA methylation changes and TE activity induced in tissue cultures of barley (*Hordeum vulgare* L.). *J. Biol. Res.* 2016. Vol. 23 (1). P. 19. doi: 10.1186/s40709-016-0056-5.
14. Picault N., Chaparro C., Piegou B., Stenger W., Formey D., Llauro C., Descombin J., Sabot F., Lasserre E., Meynard D., Guiderdoni E., Panaud O. Identification of an active LTR retrotransposon in rice. *Plant J.* 2009. Vol. 58. P. 754–765. doi: 10.1111/j.1365-3113X.2009.03813.x.
15. Makarevitch I., Waters AJ, West PT, Stitzer M, Hirsch CN, et al. Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet.* 2015. Vol. 11 (1). e1004915. doi: 10.1371/journal.pgen.1004915.
16. Dubin M.J., Scheid O. M., Becker C. Transposons: a blessing curse. *Current Opinion in Plant Biology.* 2018. Vol. 42. P. 23–29. doi: 10.1016/j.pbi.2018.01.003.
17. Ozheredova I.P., Parnikoza I.Yu., Poronnik O.O., Kozeretska I.A., Demidov S.V., Kunakh V.A. Mechanisms of Antarctic vascular plant adaptation to abiotic environmental factors. *Cytol. Genet.* 2015. Vol. 49 (2). P. 139–145.
18. Lee J., Kang Y., Shin S.C., Park H., Lee H. Combined analysis of the chloroplast genome and transcriptome of the Antarctic vascular plant *Deschampsia antarctica* Desv. *Plos One.* 2014. Vol. 9 (6). e101100. doi: 10.1371/journal.pone.0092501.
19. Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T.L. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009. Vol. 10. P. 421. doi: 10.1186/1471-2105-10-421.
20. Rice genome: Os-Nipponbare-Reference-IRGSP-1.0 pseudomolecules. URL: http://rice.plantbiology.msu.edu/annotation_pseudo_current.shtml (Last accessed: 2019).
21. Okonechnikov K., Golosova O, Fursov M; UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics.* 2012. Vol. 28 (8). P. 1166–1167. doi: 10.1093/bioinformatics/bts091.

ANDREEV I. O., KONVALYUK I. I., KUNAKH V. A.

*Institute of molecular biology and genetics of NAS of Ukraine,
Ukraine, 03143, Kyiv, Akademika Zabolotnogo str., 150, e-mail: konvalyuk.I.I@gmail.com*

**BIOINFORMATIC IDENTIFICATION OF TY1/COPIA-LIKE TRANSPOSABLE ELEMENTS IN
DESCHAMPSIA ANTARCTICA E. DESV.**

Aim. Identification of *Ty1/Copia*-like transposable elements in *Deschampsia antarctica* E Desv. *in silico*.

Methods. Bioinformatic analysis of sequence read archives of *D. antarctica* genome and transcriptome contained in the GenBank database was conducted. The search was carried out using the rice *Ty1/Copia* TE Tos 17 as a reference sequence.

Results. The search revealed the sequences of *Ty1/Copia* retrotransposons in the *D. antarctica* genome to show a high level of identity (up to 75% in the homologous regions) to Tos 17. The uneven distribution of the found reads along the reference sequence indicates the existence of a group of the sequences in the genome to contain elements typical of the family and have varying degrees of identity to Tos 17, with more conservative ones being represented in correspondingly greater numbers among the found reads. The presence of the reads identical to Tos 17 in transcriptome indicates that the TE sequences identified in genome have a certain background level of transcriptional activity.

Conclusions. Transcriptionally active *Ty1/Copia*-like transposable elements were identified *in silico* in *D. antarctica* genome using methods of bioinformatics. The sequences found can be used to construct primers for PCR and to analyze the activity of the TE of this family in *D. antarctica* in further studies.

Keywords: bioinformatic analysis, *Deschampsia antarctica* E Desv., transposable elements, *Ty1/Copia*.