

UDC 633.15; 577.214.5; 575.22; 575.113

Біоінформатичний аналіз вторинної структури транскриптів інтрона 1 гена *whp1* кукурудзи

Г. І. Сліщук, Н. Е. Волкова, Ю. М. Сиволап

Південний біотехнологічний центр в рослинництві НААН України
Овідіопольська дорога, 3, Одеса, Україна, 65036

natavolk@rambler.ru

Мета. Аналіз вторинної структури транскриптів інтрона 1 гена *whp1* кукурудзи. **Методи.** Вирівнювання, фолдинг *in silico*. **Результати.** Знайдено складні мікросателіти. Досліджено вторинну структуру 74 транскриптів інтрона 1 гена *whp1*. **Висновки.** Зроблено припущення щодо впливу транскриптів інтрона 1 гена *whp1* на відновлення фертильності кукурудзи з S-типом цитоплазматичної чоловічої стерильності.

Ключові слова: біоінформатика, ген *whp1*, ЦЧС, кукурудза, інтрон, вторинна структура.

Вступ. Цитоплазматична чоловіча стерильність (ЦЧС) у кукурудзи має декілька форм, які різняться за характером мутації мітохондріону, що спричиняє вади розвитку андроцею і, як наслідок, виникнення ЦЧС, а також за складом ядерних локусів, асоційованих з відновленням фертильності. Так, для відновлення фертильності у кукурудзи з S-типом ЦЧС потрібна наявність домінантного алеля гена *Rf3* (*restorer of fertility*) [1]. З трьох найвивченіших типів ЦЧС у кукурудзи інтерес викликає саме S-тип (молдавський), оскільки він позбавлений недоліків T-типу, зокрема, надчутливості до фітопатогенних грибів *Bipolaris maydis*.

Відомо, що ділянка гена *whp1* (*white pollen 1*) ядерного геному кукурудзи пов'язана з відновленням фертильності з молдавським типом ЦЧС у мексиканських ліній кукурудзи та теосінте [2], тому цей ген є цікавим з точки зору вивчення впливу його функціонування на відновлення фертильності кукурудзи. Ген *whp1* розташований в 3,5–5,5 сМ від гена *Rf3* [3], містить два екзони і один інтрон [4]; промоторна область цього гена виявляє високий рівень гомології з *pollen*-специфічним промотором.

Саме через комплексний характер експресії такої складної ознаки, як ЦЧС, використання лише класичних підходів молекулярної генетики недостатньо для повного розуміння механізмів її формування, тому доцільним є застосування методів біоінформатики.

Мета дослідження – біоінформатичний аналіз гена *whp1* та вивчення вторинної структури транскриптів інтрона 1 гена *whp1*.

Незважаючи на те, що інтрони (на відміну від екзонів) формально не беруть участі у формуванні білка, оскільки вирізаються під час сплайсингу, вони відіграють суттєву функціональну роль у клітині. Так, відомо про існування класу малих ядерних РНК (м'я-РНК, small nuclear RNA, snRNA), які формуються саме з транскриптів вирізаних інтронів [5]. Серед рослин малі РНК є досить розповсюдженими та виконують важливі функції, які залежать від вторинної структури вирізаних інтронів. Зокрема, дослідження геному *Arabidopsis thaliana* виявило близько 200 нових малих РНК з дволанцюговою структурою, функції яких обумовлені вторинною структурою вирізаних інтронів [6].

Матеріали і методи. Досліджували 74 нуклеотидні послідовності гена *whp1* кукурудзи з бази да-

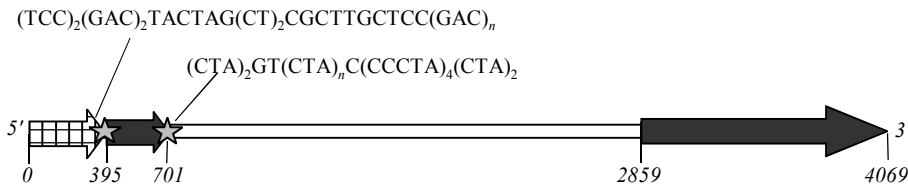


Рис. 1. Схема гена *whp1*. Ділянку промотору заштриховано клітинками, чорним кольором позначено екзони, зірочками – локалізацію мікросателітів, цифри показують позиції певних ділянок у парах нуклеотидів

них Національного центру біотехнологічної інформації (www.ncbi.nlm.nih.gov). Пошук нуклеотидних послідовностей гена *whp1* проводили, локально вирівнюючи нуклеотидні послідовності за допомогою програми *blastn*. Вивчали області промотору, екзона 1 і проксимальну частину інтрона, оскільки саме вони, по-перше, проявляють поліморфізм і, по-друге, нуклеотидні послідовності проксимальної частини інтрона здатні утворювати стабільні дволанцюгові структури. Глобальне вирівнювання нуклеотидних послідовностей за алгоритмом Нідлмана-Вунша [7] здійснювали з використанням підпрограми *AlignX* програми *VectorNTI*. *In silico* полімеразну ланцюгову реакцію проводили з використанням підпрограми *Alignent PCR* програми *VectorNTI*; умови ПЛР підбирали згідно з термодинамічними характеристиками матриці і праймерів, обчислених у програмі. Філогенетичну дендрограму реконструювали за результатами вирівнювання методом UPGMA [8] та будували із застосуванням програми *MEGA*.

Передбачення вторинної структури нуклеотидних послідовностей (фолдинг) ділянки інтрона 1 – як оптимальної структури (з найнижчим рівнем вільної енергії), так і субоптимальних структур (з рівнем вільної енергії, який можна порівняти з таким в оптимальній структурі) – здійснювали, використовуючи алгоритм Цукера за параметром найближчого сусіда (*nearest neighbor parameter*) [9, 10] та програму *RNAstructure* [11].

Вірогідність існування побудованих структур обчислювали із застосуванням статистичної суми для будовання матриці вірогідності термодинамічних характеристик системи. Точність передбачених структур за допомогою даної програми становить 73 % коректно знайдених пар нуклеотидів для послідовностей розміром менше 700 п. н.

Результати і обговорення. Пошук і вирівнювання нуклеотидних послідовностей гена *whp1*. Проведено пошук алелів гена *whp1* і знайдено 74 нуклеотидні послідовності, анотовані як *whp1*. Ген *whp1*

Праймери, які фланкують різні ділянки гена *whp1* кукурудзи

| Праймер | | |
|-----------|-------------------------|-------------|
| Назва | Послідовність (5'-3') | Локалізація |
| VSGI1 SP | aagaagagaaacaccataggccg | Промотор |
| VSGI1 ASP | tagtagttagttgggtcgcg | |
| VSGI2 SP | tggacgatgtgaggaagggc | Екзон 1 |
| VSGI2 ASP | ttgagtcggtgaggtggctc | |
| VSGI3 SP | tcttactaccctacccta | Інtron 1 |
| VSGI3 ASP | agaagaacaagagagcgag | |

довжиною 4069 п. н. містить промотор розміром 394 п. н., два екзони – 306 і 1210 п. н., а також інtron 2156 п. н. [4]. За результатами вирівнювання знайдено ділянки, що містять мікросателітні послідовності $(TCC)_2(GAC)_2TACTAG(CT)_2CGCTTGCTCC(GAC)_n$ (у промоторній області) і $(CTA)_2GT(CTA)_nC(CCCTA)_4(CTA)_2$ (в області інтрона 1) (рис. 1). Мікросателітам притаманні значний рівень поліморфізму і складна будова.

Внаслідок вирівнювання зроблено дизайн праймерів для промоторної області, екзона 1 та інтрона 1 (таблиця). Маркери, які генерувалися в полімеразній ланцюговій реакції (ПЛР) *in silico*, дозволяють диференціювати лінії кукурудзи.

Побудовано філогенетичну дендрограму (рис. 2). Виділено сім кластерів. Нуклеотидні послідовності за номерами AY731322 і AY731301 не належать до жодного кластеру та утворюють окремі дочірні гілки кластерів IV і VII, відповідно. Кластери II і III, у свою чергу, чітко розподіляються на два підкластери кожний.

Передбачення вторинної структури нуклеотидних послідовностей (фолдинг). Проведено комп'ютерне моделювання транскриптів вторинної структури нуклеотидних послідовностей ділянки інтрона 1 гена *whp1*. Структури (фолди) можна розподілити на шість класів (рис. 3–5). Для 1-го класу (рис. 3, а) характерні термодинамічно стабільні (вірогідність

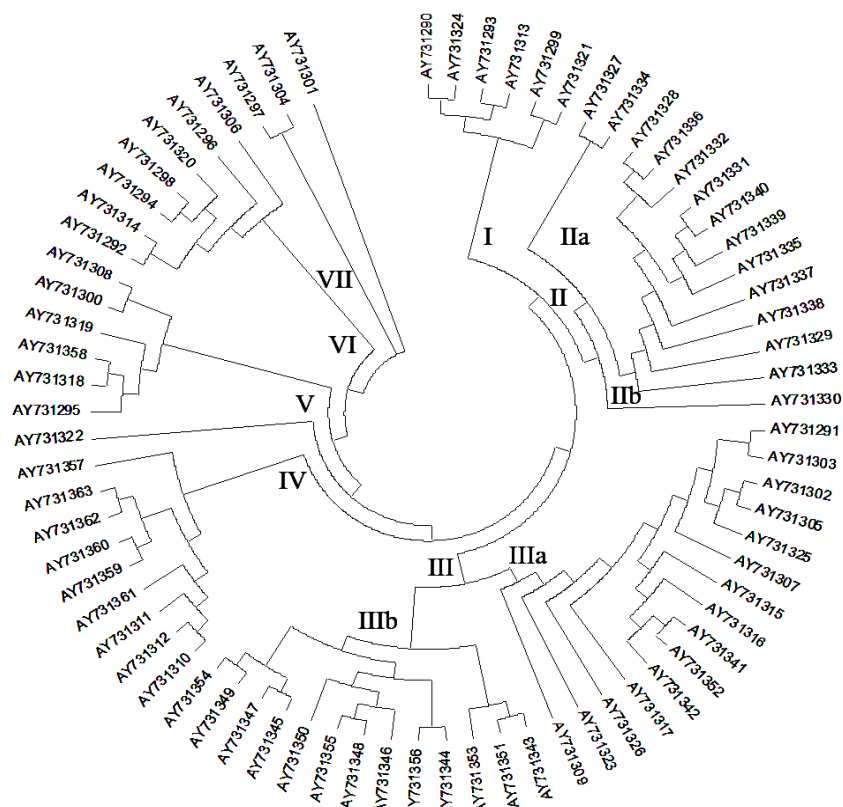


Рис. 2. Філогенетична дендрограма за результатами вирівнювання 74 послідовностей гена *whp1* кукурудзи. I–VII – кластери

утворення більше 80 %) вторинні структури, сформовані паліндромами класу 1-го типу (шпилька I) GCTCGCTCTACGCATCAAAACCGCAAGAGCTTGC і класу типу 2 (шпилька II) CGTTCGTCGTCATCGTCGTCG(N)₄₀CGTTCGACGATCGACCACGCACG.

Структурам класу 2 (рис. 3, б) також притаманна наявність шпильки I, але шпилька II у них на 20 нуклеотидів довша. Обидві шпильки є термодинамічно стабільнішими. Вірогідність утворення ними двоспіральної структури є вищою порівняно зі структурами класу 1, однак розташування обох шпильок викликає більше напруження ланцюга.

Для структур класу 3 (рис. 4, а), як і двох попередніх класів, характерною є шпилька I. Шпилька II має проміжну будову відносно структур класів 1 і 2. За результатами обчислення статистичної суми для цих структур відмічено високу стабільність дволанцюгових структур (як у класу 2) та незначне напруження між ланцюгами (як у класу 1), що свідчить про переважання утворення саме таких структур у порівнянні з вірогідністю формування структур класів 1 і 2.

У структур класу 4 (рис. 4, б) будова шпильки II є складнішою, шпилька I при цьому залишається незмінною. Довжина шпильки II проміжна між структурами класів 1 і 2. За стабільністю двоспіральної структури вона наближається до структур класу 2.

Структурам класу 5 (рис. 5, а) властиве виникнення подвійної нестабільності на кінці шпильки II. Лише невелика ділянка розміром 10 п. н. на дистальній частині шпильки II може утворюватися з високою вірогідністю (більше 80 %) і стабільністю, будова проксимальної частини подібна до такої структур класу 1. Шпилька I залишається незмінною і спільною для інших структур.

Для структур класу 6 (рис. 5, б) характерна аналогічна з іншими структурами будова шпильки I та низька можливість формування шпильки II.

Інші 19 транскриптів не виявили здатності до утворення стабільних двоспіральних структур. Даний факт дає можливість припустити нефункціональність цих типів транскриптів через їхнє перебування у клітині в одноланцюговому стані (за результатами передбачення транскриптів *in silico*).

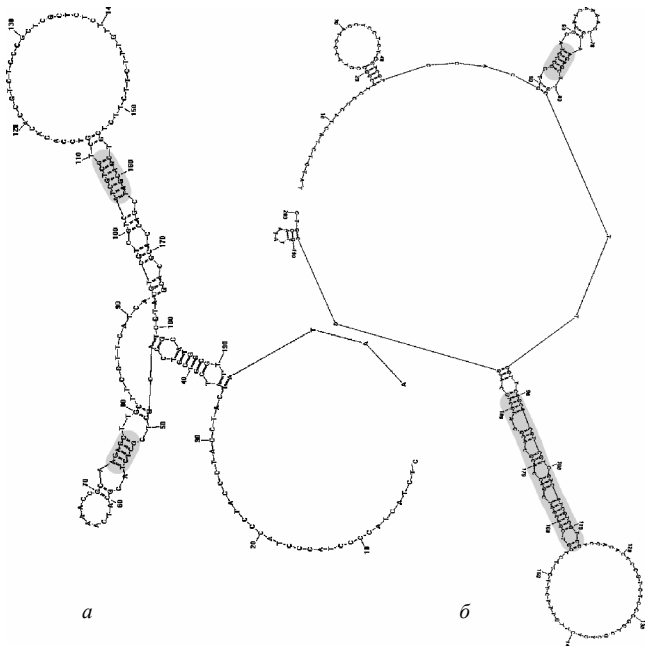


Рис. 3. Вторинні структури нуклеотидних послідовностей ділянки інтрона 1 гена *whp1* класів 1 (а) і 2 (б). Сірим кольором позначено ділянки, для яких вірогідність утворення двоспиральної структури становить більше 80 %. Розгалужена структура 1 (домен) містить шпильки I і II; структура 2 є лінійною, окрім шпильок I і II, вона містить дві мінорні шпильки з низькою вірогідністю утворення

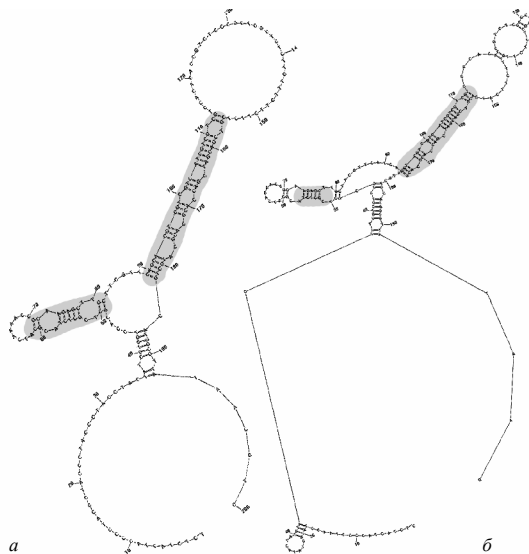


Рис. 4. Вторинні структури нуклеотидних послідовностей ділянки інтрона 1 гена *whp1* класів 3 (а) і 4 (б). Сірим кольором позначено ділянки, для яких вірогідність утворення двоспиральної структури становить більше 80 %. Розгалужена структура 4 (домен) містить шпильки I і II

Змодельовані структури демонструють певний ступінь структурної подібності до вже описаних мяРНК, які можуть змінювати рівень експресії ге-

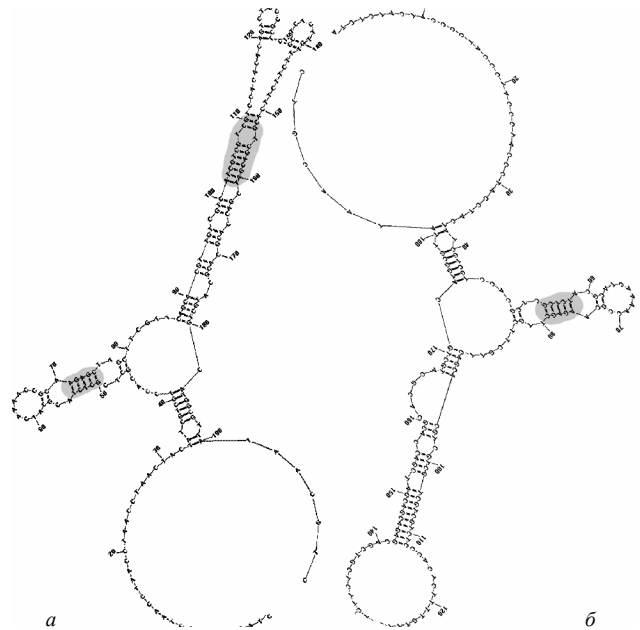


Рис. 5. Вторинні структури нуклеотидних послідовностей ділянки інтрона 1 гена *whp1* класів 5 (а) і 6 (б). Сірим кольором позначено ділянки, для яких вірогідність утворення двоспиральної структури становить більше 80 %. Структура 5 в апікальній частині шпильки II містить короткі (10 і 11 п. н.) шпильки

нів. Наприклад, 7SK РНК людини і деяких тварин здатні інгібувати активність комплексів CDK9/циклін Т [12], а U1 мяРНК може інтенсифікувати експресію генів [13].

Враховуючи те, що саме ділянка гена *whp1* є джерелом генів – відновників фертильності у кукурудзи з молдавським типом ЦЧС [2], а також результатами вивчення вторинної структури транскриптів ділянки інтрона 1 гена *whp1* можна зробити припущення щодо впливу цих транскриптів на процес відновлення фертильності. Так, відомо, що відновлення фертильності у кукурудзи з молдавським типом ЦЧС відбувається внаслідок зміни рівня експресії мітохондріальних генів. Однак для підтвердження гіпотези і з'ясування можливого механізму дії знайдених структур необхідною є експериментальна верифікація.

Таким чином, проведено біоінформатичний аналіз гена *whp1*; за допомогою методу вирівнювання досліджено поліморфізм гена серед 74 нуклеотидних послідовностей; побудовано філогенетичну дендрограму, за якою генотипи розподіляються на сім клад.

Знайдено області геному, що містять раніше невизначені та невивчені мікросателітні послідовності

складної будови; розроблено дизайн праймерів, які дозволяють диференціювати лінії кукурудзи. Встановлені за допомогою біоінформатичних методів структури теоретично здатні впливати на рівень експресії мітохондріальних генів.

Висновки. За результатами біоінформатичного аналізу 74 нуклеотидних послідовностей гена *whp1* визначено структуру гена, знайдено складні мікросателітні повтори. Обрано праймери, які фланкують найполіморфніші області, використання яких при ПЛР-аналізі *in silico* дозволяє диференціювати лінії кукурудзи. Досліджено вторинну структуру транскриптів інтрона 1 гена *whp1*. За результатами комп'ютерного моделювання вторинної структури нуклеотидних послідовностей зроблено припущення щодо можливого впливу транскриптів гена *whp1* на відновлення фертильності у ліній кукурудзи з молдавським типом ЦЧС.

G. I. Shishchuk, N. E. Volkova, Yu. M. Sivolap

Bioinformatics analysis of the secondary structure of maize *whp1* gene intron 1 transcripts

South Plant Biotechnology Center, NAAS of Ukraine
3, Ovidiopol'ska doroga, Odesa, Ukraine, 65036

Summary

Aim. Analysis of the secondary structure of maize *whp1* gene intron 1 transcripts. **Methods.** Alignment, *in silico* folding. **Results.** Novel complex microsatellites of the *whp1* gene were discovered. 74 *whp1* intron 1 transcripts secondary structures were predicted. **Conclusions.** The influence of *whp1* gene intron 1 transcripts on restoration of CMS-S sterile maize fertility has been proposed.

Keywords: bioinformatics, *whp1* gene, CMS, maize, intron, secondary structure.

Г. І. Слищук, Н. Э. Волкова, Ю. М. Сиволап

Биоинформатический анализ вторичной структуры транскриптов интрона 1 гена *whp1* кукурудзы

Резюме

Цель. Анализ вторичной структуры транскриптов интрона 1 гена *whp1* кукурудзы. **Методы.** Выравнивание, фолдинг *in silico*. **Результаты.** Найдены сложные микросателлиты. Исследована вторичная структура 74 транскриптов интрона 1 гена *whp1*. **Выводы.** Сделано предположение о влиянии транскриптов интрона 1

гена *whp1* на восстановление фертильности кукурудзы с S-типом цитоплазматической мужской стерильности.

Ключевые слова: биоинформатика, ген *whp1*, ЦМС, кукуруза, интрон, вторичная структура.

PERELIK LITERATURY

1. Zabala G., Gabay-Laughnan S., Laughnan J. R. The nuclear gene *Rf3* affects the expression of the mitochondrial chimeric sequence R implicated in S-type male sterility in maize // *Genetics*.—1997.—**147**, N 2.—P. 847–860.
2. Gabay-Laughnan S., Chase C. D., Ortega V. M., Zhao L. Molecular-genetic characterization of CMS-S restorer-of-fertility alleles // *Genetics*.—2004.—**166**, N 2.—P. 959–970.
3. Schaeffer M. L., Harper L. C., Gardiner J. M., Andorf C. M., Campbell D. A., Cannon E. K. S., Sen T. Z., Lawrence C. J. MaizeGDB: curation and outreach go hand-in-hand // *J. Biol. Databases and Curation*.—2011. doi: 10.1093/database/bar022.
4. Franken P., Niesbach-Klosgen U., Weydemann U., Marechal-Drouard L., Saedler H., Wienand U. The duplicated chalcone synthase genes *C2* and *Whp* (white pollen) of *Zea mays* are independently regulated; evidence for translational control of *Whp* expression by the anthocyanin intensifying gene in // *EMBO J.*—1991.—**10**, N 9.—P. 2605–2612.
5. Mount S. M., Gotea V., Lin C.-F., Hernandez K., Makalowski W. Spliceosomal small nuclear RNA genes in 11 insect genomes // *RNA*.—2007.—**13**, N 1.—P. 5–14.
6. Zheng Q., Ryvkin P., Li F., Dragomir I., Valladares O., Yang J., Cao K., Wang L.-S., Gregory B. D. Genome-wide double-stranded RNA sequencing reveals the functional significance of base-paired RNAs in *Arabidopsis* // *PLoS Genet.*—2010.—**6**, N 9.—e1001141.
7. Needleman S. B., Wunsch C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins // *J. Mol. Biol.*—1970.—**48**, N 3.—P. 443–453.
8. Sneath P. H. A., Sokal R. R. Numerical taxonomy: The principles and practices of numerical classification.—San-Francisco: Freeman, 1973.—573 p.
9. Zuker M., Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information // *Nucleic Acids Res.*—1981.—**9**, N 1.—P. 133–148.
10. SantaLucia J., Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics // *Proc. Natl Acad. Sci. USA.*—1998.—**95**, N 4.—P. 1460–1465.
11. Reuter J. S., Mathews D. H. RNAstructure: software for RNA secondary structure prediction and analysis // *BMC Bioinformatics*.—2010.—N 11.—P. 129.
12. Nguyen V. T., Kiss T., Michels A. A., Bensaude O. TSK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes // *Nature*.—2001.—**414**, N 6861.—P. 322–325.
13. Alexander M. R., Wheatley A. K., Center R. J., Purcell D. F. J. Efficient transcription through an intron requires the binding of an Sm-type U1 snRNP with intact stem loop II to the splice donor // *Nucleic Acids Res.*—2010.—**38**, N 9.—P. 3041–3053.

Received 18.11.11