

УДК 577.21+57.052

Михайлик С. Ю., Антонюк М. З.

## БІОІНФОРМАТИЧНИЙ АНАЛІЗ ПРОМОТОРНОЇ ДІЛЯНКИ ГЕНА *TaMSH7*

За результатом зіставлення послідовності κДНК гена *TaMSH7* та послідовності 3В хромосоми *Triticum aestivum* встановлено, що ген *TaMSH7* має 15 інтронів і на 3В хромосомі локалізований у положенні від 119371926 до 119373011. Дослідження 5'UTR ділянок генів *MSH7* дев'яти рослин, в тому числі злаків *Oryza sativa* та *Zea mays*, вказує на те, що послідовність κДНК гена *TaMSH7* ймовірно є вкороченою з 5'-кінця. Враховуючи ці дані, кандидатом на точку ініціації трансляції виступає кодон АТГ у положенні 433-435 нуклеотидів перед геном. На основі аналізу ймовірної промоторної ділянки гена *TaMSH7* встановлено велику кількість CpG точок метилювання, рівень метилювання яких може вказати на особливості регуляції експресії гена у геномах інтрогресивних ліній. Послідовність довжиною 816 нуклеотидів перед імовірним сайтом ініціації трансляції гена *TaMSH7* містить 75 CpG сайтів метилювання і є мішенню для розробки праймерів з метою бісульфітного секвенування та встановлення рівня метилювання ймовірної промоторної ділянки. З використанням програмного забезпечення *Bisulfite Primer Seeker*, розробленого фірмою *Zymo (Irvine, CA, США)*, вдалося отримати послідовності праймерів, що охоплюють 31 сайт метилювання у ймовірній промоторній ділянці гена *TaMSH7*. Перевірка рівня метилювання CpG сайтів цієї ділянки дасть змогу вказати на особливості експресії гена *TaMSH7*. Порівняння рівня метилювання ймовірної промоторної ділянки гена *TaMSH7* інтрогресивних ліній пшениці та їхніх батьківських форм дозволить встановити, чи впливає на рівень метилювання промоторного регіону досліджуваного гена наявність або відсутність чужинного генетичного матеріалу у геномі.

**Ключові слова:** *TaMSH7*, *MSH7*, промоторна ділянка гена, 5'UTR, сайт ініціації трансляції, бісульфітне секвенування, CpG точка метилювання.

### Вступ

Система виправлення помилок спарювання основ MMR (MisMatch Repair) є базовою та високонсервативною системою репарації геному, необхідною для передання коректної генетичної інформації нащадкам [1]. Ця система широко відома через свою роль у репарації некомплементарних пар основ і неспарених нуклеотидів, що виникають унаслідок помилок при реплікації, дезамінуванні 5-метилцитозина та рекомбінації між відмінними послідовностями. Система MMR також бере участь у обробці та/або передачі сигналів, що виникають унаслідок різних типів пошкоджень ДНК [1]. Репарація відбувається вна-

слідок складної взаємодії між численними білками, які є консервативними у бактерій, тварин і рослин [2]. У еукаріотів початкове розпізнавання пошкоджень ДНК здійснюють гетеродимерні комплекси, відомі як MutSα (MSH2-MSH6) та MutSβ (MSH2-MSH3). MutSα розпізнає некомплементарні пари основ та поодинокі індели (indel – insertion-deletion loop), в той час як MutSβ в основному взаємодіє з IDL довжиною 2-12 основ [3]. Рослини мають унікальний MutS білок – MutSγ, який складається з комплексу (MSH2-MSH7) [4]. Нині MutSγ досліджений значно менше, ніж MutSα та MutSβ. Попередні дослідження демонструють, що MutSγ переважно розпізнає певні

некомплементарні пари основ [5]. Функціональні дослідження репараційної системи на дріжджах показали, що MutSγ відіграє важливу роль у розпізнаванні послідовності, що містить T та/або G/T, A/C, T/C, G/A, T/T, або A/A, некомплементарні пари, які виникли внаслідок дезамінування цитозину та 5-метилцитозину, а також внаслідок дії УФ-світла чи окисного мутагенезу [6]. Крім того, існують дані стосовно суттєвої ролі MSH7 у процесі мейозу: у пшениці ген *TaMSH7* експресується на високому рівні у молодих квітках, також цей ген було картовано в *Ph2* локусі, що, як відомо, впливає на гомеологічну рекомбінацію [7]. У трансгенному *Hordeum vulgare* експресія гена *MSH7* виявилася необхідною для фертильності рослини [8]. У томата (*Solanum lycopersicum*) супресія гена *MSH7* була пов'язана з помірним підвищенням рівня гомеологічної рекомбінації між хромосомами томата та спорідненого дикорослого виду *Solanum lycopersicoides* [9].

Вплив метилювання на рівень експресії, а отже і роботу системи MMR, на рослинах не досліджено. Інша ситуація спостерігається з дослідженням впливу метилювання промоторних ділянок генів *MSH2* та *MSH6* на їхню експресію у ракових клітинах [10; 11]. Ці дослідження вказують на те, що метилювання промоторних ділянок генів MMR системи репарації геному призводить до зменшення рівня експресії відповідних генів і, як наслідок, збільшення кількості помилок у геномі.

Інтрогресивні лінії м'якої пшениці під час свого створення неодноразово перебувають у стані геномного стресу, і цей стан може зберігатися навіть у цитологічно стабільних лініях [12]. Такі лінії є зручною моделлю для дослідження процесів, що відбуваються у геномах внаслідок привнесення чужинного генетичного матеріалу. Брак інформації стосовно роботи генів MMR системи у інтрогресивних геномах унеможливило розуміння повної ролі генів MMR у підтримці таких геномів. Дослідження рівня метилювання і відповідно особливостей експресії гена *TaMSH7* у геномі інтрогресивних ліній м'якої пшениці може вказати на особливості роботи MMR системи репарації в умовах стресу, спричиненого привнесенням чужинного генетичного матеріалу.

### Матеріали та методи

У дослідженні використовували послідовності таких генів і хромосом: кДНК гена *TaMSH7* *Triticum aestivum* (GenBank AF354709.1), хромосома 3В *Triticum aestivum* (GenBank HG670306.1),

ген *AtMSH7* (Gene ID 822040), ген *MSH7* *Vitis vinifera* (Gene ID 100260983), ген *MSH7* *Solanum lycopersicum* (Gene ID 100462656), ген *MSH7* *Oryza sativa* (Gene ID 4325353), ген *MSH7* *Glycine max* (Gene ID 100776965), ген *MSH7* *Sesamum indicum* (Gene ID 105162107), ген *MSH7* *Brassica napus* (Gene ID 106391373), ген *MSH7* *Gossypium raimondii* (Gene ID 105773605), кДНК гена *MUS2* *Zea mays* (GenBank AJ238787.1), ген *MUS2* *Zea mays* (GenBank AJ238786.1), ген *MSH7* *Oryza sativa* (GenBank XM\_015766983.1).

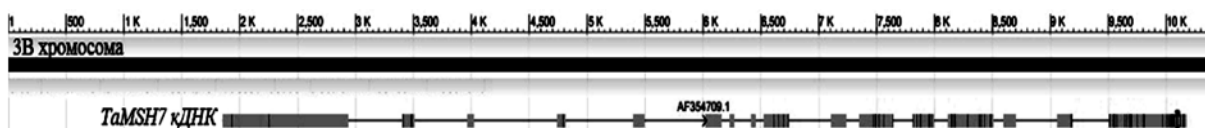
Для встановлення промоторної ділянки гена *TaMSH7* використовували такі програми, алгоритми та сервери: сервер ViroBLAST [13], Mega BLAST алгоритм NCBI [14; 15], NCBI Eukaryotic Genome Annotation Pipeline (з використанням miRBase [16], RefSeq [17] та Splign [18]), Sequence Viewer 3.12.2 (NCBI), база даних CpG-острівців та аналітичних додатків DBCAT (Database of CpG islands and Analytical Tool) [19], Bisulfite Primer Seeker розроблений Zymo (Irvine, CA, США).

## РЕЗУЛЬТАТИ ТА ОБГОВОРЕННЯ

### Встановлення структури гена *TaMSH7*

Ген *TaMSH7* (*Triticum aestivum* *MSH7*) пшениці було вперше ізольовано у 2002 р. Донгом та співавторами з використанням зонду до консервативного регіону гена *MSH6* дріжджів [7]. Послідовність гена було встановлено за допомогою 3'RACE та 5'RACE методики, однак ген демонстрував значно більшу гомологію до гена *MSH7* кукурудзи та *MSH7* арабідопсису, ніж до *MSH6*, отже це був новий ген *MSH7* пшениці – *TaMSH7*. 3'RACE методика дала змогу встановити кодуєчу послідовність гена до полі-А хвоста, в той час як 5'RACE методика – лише частково встановити послідовність початку гена. Отже, наявна у базі даних послідовність (GenBank AF354709.1) характеризує кДНК гена *TaMSH7*, однак лише часткову, без кодуєчої послідовності початку гена та без промоторної ділянки гена. З використанням нуллі-тетрасомних та дителосомних ліній пшениці ген *TaMSH7* було локалізовано на короткому плечі 3А, 3В та 3Д хромосоми пшениці [7]. Секвенування трьох *MSH7* генів, розташованих на короткому плечі 3 гомеогрупи хромосом пшениці, не виявило значних розбіжностей у послідовності нуклеотидів у цих генів [8].

У 2014 р. було секвеновано хромосому 3В м'якої пшениці [20]. Оскільки *TaMSH7* міститься на короткому плечі 3А, 3В та 3Д хромосом, порівняння сиквенсу 3В хромосоми з послідовністю гена *TaMSH7* дало змогу встановити положення



**Рис. 1.** Порівняння послідовності 3В хромосоми (нуклеотида від 119370077 до 119381076) з послідовністю кДНК гена *TaMSH7*, отримане у Sequence Viewer (NCBI). Сірі прямокутники – екзони, лінії між ними – інтрони. Вертикальні лінії у послідовності кДНК *TaMSH7* – SNP

гена на хромосомі, отримати дані стосовно його структури, а також вийти на фланкуючі послідовності досліджуваного гена (рис. 1). Ймовірно через вкрай велику послідовність 3В хромосоми (близько 770 мільйонів пар основ) BLAST із використанням програмного забезпечення серверів NCBI не зміг опрацювати порівняння послідовностей [15]. Використання сервера ViroBLAST дало змогу вирішити цю проблему [13]. Ген *TaMSH7* (без його початкової послідовності) розташований на 3В хромосомі пшениці у положенні від 119371926 до 119373011 і містить щонайменше 15 інтронів у своїй послідовності.

Окремі труднощі викликало встановлення промоторної ділянки гена та точки ініціації трансляції, що є необхідним для встановлення повної структури гена. Сучасне, наявне у доступі, біоінформатичне програмне забезпечення не дає змоги анотувати промоторні ділянки гена, його точки ініціації трансляції та інші дані силами окремого користувача. Для цього потрібна робота великих дослідних груп і проектів, таких як NCBI Eukaryotic Genome Annotation Pipeline, що використовує бази даних NCBI, алгоритми miRBase [16], RefSeq [17], Splign [18] та інше програмне забезпечення для анотації геномів. Наразі у цій базі даних відсутній анотований геном пшениці, але наявні геноми 39 інших рослинних видів. Порівняння структури генів *MSH7* цих видів із геном *TaMSH7* може допомогти встановленню промоторного регіону досліджуваного гена.

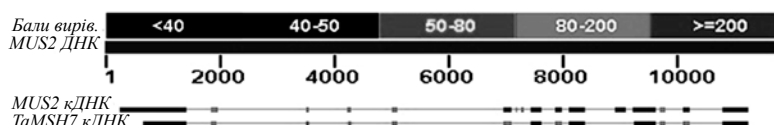
### Порівняння структури та положення гена *MSH7* на прикладі різних рослин

Якщо встановити промоторну та 5'UTR ділянки гена виключно біоінформатичним шляхом неможливо, можна проаналізувати вже встановлені 5'UTR ділянки генів *MSH7* інших видів рослин та вказати на приблизний її розмір у гена *TaMSH7* пшениці м'якої. У *Arabidopsis thaliana* ділянка від початку транскрипції гена до початку трансляції (кодону ATG) містить 60 нуклеотидів (Gene ID: 822040), сам ген *AtMSH7* розташова-

ний на 3-й хромосомі. У *Vitis vinifera* 5'UTR регіон гена *MSH7* (Gene ID: 100260983) має 173 нуклеотиди, ген розташований на 11-й хромосомі. У *Solanum lycopersicum* ген *MSH7* (Gene ID: 100462656) розташований на 7-й хромосомі, його 5'UTR ділянка має два інтрони, загалом послідовність має довжину від 10347436 (локація гена на хромосомі) до 10347987, тобто 551 нуклеотид. Перший екзон має довжину 240 нуклеотидів, другий – 241–297, третій з 10347945 по 10347987 – 42 нуклеотиди [4]. *Oryza sativa* має ген *MSH7* (Gene ID: 4325353) на хромосомі 1, 5'UTR ділянка представлена 124 нуклеотидами. У *Glycine max* ген *MSH7* (Gene ID: 100776965) локалізований на хромосомі 7, його 5'UTR ділянка розташована від 8112544 до 8112619 нуклеотиду хромосоми – має у довжину 75 нуклеотидів. *Sesamum indicum* має 5'UTR ділянка гена *MSH7* (Gene ID: 105162107) з 5642578 по 5642446 нуклеотид на LG5-й хромосомі, 132 нуклеотиди. 5'UTR регіон гена *MSH7* (Gene ID: 106391373) ріпаку *Brassica napus* розташований на хромосомі C4 у положенні 23871332–23871414 – довжина становить усього 82 нуклеотиди. 5'UTR регіон досліджуваного гена бавовника *Gossypium raimondii* (Gene ID: 105773605) розташований на хромосомі 10 і складається з 161 нуклеотиду.

Порівняння восьми 5'UTR ділянок генів *MSH7* різних рослин вказує на те, що її довжина коливається у межах від 551 нуклеотида у *Solanum lycopersicum* до 60 нуклеотидів у *Arabidopsis thaliana*. Можливо, 5'UTR ділянка гена *TaMSH7* міститься у таких межах.

Донт і співавтори у дослідженні гена *TaMSH7* порівнювали його з кДНК гена кукурудзи *MUS2* (GenBank AJ238787.1) [7]. Структура *MUS2* схожа зі структурою гена *TaMSH7*, зокрема послідовність *MUS2* є неповною, в ній також відсутній початок гена. Однак у 2006 р. Хорват і Кунз ізолювали та охарактеризували повну геномну ДНК гена *MUS2* (GenBank AJ238786.1) (неопубліковані дані). Цей ген також містить 15 інтронів, 5'UTR ділянка має довжину 283 нуклеотиди. Порівняння геномної ДНК гена *MUS2* з його кДНК та геном *TaMSH7* дає змогу ідентифікувати можливу



**Рис. 2.** Вирівнювання послідовностей ДНК гена *MUS2*, κДНК гена *MUS2* та κДНК гена *TaMSH7*.  
Бали вирів. – бали вирівнювання, що характеризують ідентичність послідовностей

5'-ділянку останнього. Перший екзон *TaMSH7* демонструє високу ідентичність до κДНК гена *MUS2* (більше ніж 200 балів вирівнювання), однак є коротшим (рис. 2).

Відстань від початку 5'UTR ділянки ДНК *MUS2* до першого екзону *TaMSH7* становить 680 нуклеотидів. Перший екзон гена *MUS2* на 396 нуклеотидів довший за аналогічний екзон *TaMSH7* (рис. 2). Ймовірно, ця відстань у 396 нуклеотиди і є тією частиною екзону, яка була відсутня у секвенуванні послідовності з дослідження Донга.

Порівняння κДНК гена *TaMSH7* м'якої пшениці з геном *MSH7* (GenBank XM\_015766983.1) *Oryza sativa* демонструє високий рівень ідентичності цих генів – 83 %. Ген рису анотований, і це може допомогти в анотації *TaMSH7*. Порівняння κДНК гена *TaMSH7* з послідовністю гена *MSH7* рису вказує на те, що ген пшениці на 340 нуклеотидів коротший з 5'-кінця. 5'UTR гена *MSH7 Oryza sativa* становить 124 нуклеотиди (рис. 3). Загалом κДНК гена *TaMSH7* коротша за відповідну κДНК аналогічного гена у рису на 216 нуклеотидів.



**Рис. 3.** Порівняння послідовностей гена *MSH7 Oryza sativa* з послідовністю κДНК гена *MSH7 Oryza sativa* та геном *TaMSH7*, отримане у Sequence Viewer (NCBI). Вертикальні лінії у послідовності κДНК *TaMSH7* – SNP

### Ідентифікація 5'-фланкуючої послідовності гена *TaMSH7*

Порівняння гена *TaMSH7* пшениці з аналогічними генами *MSH7* двох інших злаків, кукурудзи та рису, дає змогу встановити можливі

приблизні межі невідомого фрагмента першого екзону, а також ймовірні розміри 5'UTR ділянки (рис. 4–5).

TCA .GCT .GCC .GCG .GTA .TCT .AGC .TAA .TGG .TTC .CGC .CCG .CGG .ATT .CGT .TTA .GGT .GGT .TGC .CGC .TTG .ATA .GTG .GAT .GCC .GAT .TCG .GTC .GAT .TTA .TCG .TTG .GAT .TGG .TCT .CCG .CGT .GTA .TGC .TGT .TGG .AAT .TGC .TCG .AAC .AAA .TTC .TGG .CGT .GGC .TTC .GTT .TGT .TGT .GGG .TAG .CGC .CGT .GGC .GGT .ATT .GTG .TGG .CCA .GTA .GCA .GCA .GGG .**ATG** .CGA .TGA .**ATG** .CGA .GTT .TGC .AGT .TGC .GAT .GTC .TTA .GCT .CCT .GAC .CGC .GCC .TGG .TGT .TTG .GCC .ACT .ACA .TGT .GTC .GAG .TGT .TGT .GCC .TGA .TAG .TTT .ACG .GGC .CGC .ATT .GTC .ATC .TTC .AGT .TAG .TCA .GCT .TGT .TTT .TGC .TAT .TGC .ATT .GTC .CTC .ATC .AGT .CGC .CAT .GCT .TGC .TGT .TAA .ATC .CGT .TTC .TCT .TAA .CTC .CAA .TTT .GCT .TAC .TGT .CCG .TGT .CTC .TTG .GCT .GTT .GGA .TGG .TGT .TTT .**ATG** .TGA .TCT .ATC .GAT .GTT .TTG .ATT .CAC .**ATG** .TTT .GTA .GTT .TTT .GAG .CAT .GGA .TAG .**ATG** .ACA .CAC .CAT .AAT .TTT .GCG .AAC .CAA .AGT .AAA .CTT .CAG .ATT .GGA .AGC .TGT .TCC .ATA .CAT .TCG .AAA .GGA .AGA .ATC .TGG .GCT .AAT .ACA .CAT .TAT .CAG .CCA .TTG .CCG .ATT .CGT .AAT .TTG .TAT .GTG .GTT .CCC .TCC .ACC .TAC .ATT .GCA .ACT .AGC .CTT .AAG .ATC .ATT .GGA .TTT .TGG .GTC .CAG .TGT .TTA .ACT .GTG .СТА .ATT .GTC .CTT .CAC .TTT .ATT .GCA .AGC .TTG .GAT .TGC .TTT .AAT .GGC .СТА .TGG .AAT .GGT .TGG .GTT .GGA .TGC .ATT .TGT .TTC .TGT .GTT .AAT .GGG .TGT .**ATG** .ACA .GGT .GTA .TGT .TAT .TGC .CTT .CTT .GGT .CCC .TAA .AGT .ATC .TTG .TTT .GTT .GGT .TTG .CAT .CGG .CAG .ATT .TTC .TGC .TTA .TTT .TGG .AAG .ACA .CTT .TTT .TTT .TGC .TGT .GCC .ATC .**ATG** .GTG .**ATG** .CTC .TCT .GTG .TTT .GTT .GCA .GAA .ACA .AAG .**ATG** .CCG → початок κДНК *TaMSH7*

**Рис. 4.** Фланкуюча 5'-послідовність ділянки κДНК гена *TaMSH7* довжиною 933 нуклеотиди. Сірим кольором виділено початок 5'UTR аналогічного гена кукурудзи. Підкреслено початок κДНК гена кукурудзи. **ATG** – сайт ініціації трансляції



TAGAGAAATTCGTTTTTTTTTCGTTCCGGCGGAGTTATTTCAATCGAATTCGGTTATCGCGATGTAGTCGCGGCGGTAGT  
 AGTAGTTTTTTTTTTTTTTTTTTTTTATTCGGTTAGAGTTCCGGCGTAGGAGGCGTTCCGGCGTACGTTTTTCGAGAGG  
 TTTTCGCGGTTTTTCGTCGCGTTCGTCGCGTTCGACGGTATTATGGAGAGGTTTCGTCGCGTTCGTTATAGGGGAGGTACGT  
 GCGTACGGACGGACGCGATTAATGTAGTTTTTCGTCGTTTTAGGGTTTTGTTGGGTTTGGGTGTGCGTTTTGATTGGGTTAG  
 TTGTCCGCGTATTAGTTAATGGTTTTCGTTCCGCGATTTCGTTAGGTGGTTGTGCGTTGATAGTGGATGTCCGATTCCGGTC  
 GATTTATCGTTGGATTGGTTTTCCGCGTGTATGTTGTTGGAATTGTTCCAATAAATTTGGCGTGGTTTTCGTTGTTGTGG  
 GTAGCGTCCGTTGGCGGTTATTGTGTGGTTAGTAGTAGTAGGGATCCGATGAATCCGAGTTTGTAGTTCCGATGTTTTAGTTT  
 TTGATCCGCGTTTTGGTGTGGTTATTATATGTGTCCAGTGTGTGTTTGATAGTTTACGGGTCCGATTGTTATTTTAGT  
 TAGTTAGTTGTTTTGTTATTGTTATGTTTTATTAGTCCGTTATGTTGTTGTTAAATTCGTTTTTTTTAAATTTAATT  
 TGTATTATGTTCCGTGTTTTTTGGTTGTTGGATGGTGTGTTTTATGTGATTTATCGATGTTTTGATTTATATGTTTGTAGTTT  
 TTGAGTATGGATAGATGATGATATATAAATTTGCGAATTAAGTAAATTTAGATTGGAAGTTGTTTTATATATTCGAAA  
 GGAAGAAATTTGGGTTAATATATATATTAGTTATTGTCGATTCGTAATTTGTATGTGGTTTTTTTTATTTATATTGTAAT  
 TAGTTTTAAGATTAATTGGATTTTGGGTTTGTGTTAATT

**Рис. 8.** Бісульфітно конвертована 5'-фланкуюча послідовність кДНК гена *TaMSH7*.  
 Сірим кольором виділено CpG сайти метилювання та послідовності праймерів.  
 ATG – ймовірний сайт ініціації трансляції

*Таблиця.* Послідовності праймерів, специфічних до ймовірної 5'UTR ділянки гена *TaMSH7*

Назва праймера	Довжина праймера, нт	Розмір продукту, нт	Кількість CpG	Tm	Послідовність (5'а)
TaMSH7-5'UTR/L	25	492	31	57.1	GTTTTAGGGTTTTGTTGGGTTTGG
TaMSH7-5'UTR/R	25			55.8	ATCACATAAAACACCATCCAACAAC

Використовуючи програмне забезпечення Bisulfite Primer Seeker, розроблене фірмою Zymo (Irvine, CA, США), вдалося отримати послідовності праймерів, що охоплюють 31 сайт метилювання у ймовірній промоторній ділянці гена *TaMSH7* (рис. 8, таблиця).

Використання ймовірної промоторної ділянки гена *TaMSH7* як мішені для бісульфітного секвенування з метою перевірки рівня метилювання промоторного регіону гена *TaMSH7* дасть змогу вказати на особливості регуляції експресії досліджуваного гена у геномах інтрогресивних ліній.

### Висновки

Порівняння послідовності кДНК гена *TaMSH7* (GenBank AF354709.1) з послідовністю 3В хромосоми *Triticum aestivum* (GenBank HG670306.1) показало, що досліджуваний ген має 15 інтронів, а 5'-послідовність кДНК не є повною. Порівняння послідовностей гена *MSH7 Oryza sativa* (GenBank XM\_015766983.1) з послідовністю кДНК гена *MSH7* вказує на те, що останній має кДНК,

коротшу за відповідну кДНК аналогічного гена у рису на 216 нуклеотидів. Аналіз послідовностей генів *TaMSH7* та гена *MUS2 Zea mays* (GenBank AJ238786.1) демонструє, що перший екзон послідовності *TaMSH7* гена пшениці коротший на 396 нуклеотидів. З цих даних випливає, що ген пшениці *TaMSH7* ймовірно має точку ініціації трансляції у положенні перед відомою послідовністю гена. Аналіз 933 нуклеотидів вказує, що найбільш ймовірним кандидатом точки початку трансляції виступає кодон ATG у положенні 433-435, усі інші кодони розташовані значно ближче до початку послідовності кДНК гена.

Майже відразу після ймовірної точки ініціації трансляції у напрямку 3'-5' йде ділянка, багата на CpG точки метилювання, що є природнім для промоторних ділянок декотрих генів родини MMR [10; 21]. Детальне дослідження такої ділянки та використання бісульфітного секвенування для перевірки рівня метилювання промоторного регіону гена *TaMSH7* дасть змогу вказати на особливості експресії цього гена у геномах інтрогресивних ліній м'якої пшениці.

### Список літератури

- Li G.-M. Mechanisms and functions of DNA mismatch repair / G.-M. Li // Cell research. – 2008. – Vol. 18, No. 1. – P. 85–98.
- Spampinato C. P. From bacteria to plants: a compendium of mismatch repair assays / C. P. Spampinato, R. L. Gomez, C. Galles, L. D. Lario // Mutation research. – Vol. 682, No. 2–3. – P. 110–128.
- Acharya S. hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6 / S. Acharya, T. Wilson, S. Gradia [et al.] // Proceedings of the National Academy of Sciences of the United States of America. – 1996. – Vol. 93, No. 24. – P. 13629–13634.

4. Tam S. M. Characterization and comparative sequence analysis of the DNA mismatch repair MSH2 and MSH7 genes from tomato / S. M. Tam, S. Samipak, A. Britt, R. T. Chetelat // *Genetica*. – 2009. – Vol. 137, No. 3. – P. 341–354.
5. Culligan K. M. Arabidopsis MutS homologs-AtMSH2, AtMSH3, AtMSH6, and a novel AtMSH7-form three distinct protein heterodimers with different specificities for mismatched DNA / K. M. Culligan, J. B. Hays // *The Plant cell*. – 2000. – Vol. 12, No. 6. – P. 991–1002.
6. Gómez R. Mismatch recognition function of Arabidopsis thaliana MutSγ / R. Gómez, C. P. Spampinato // *DNA repair*. – 2013. – Vol. 12, No. 4. – P. 257–264.
7. Dong C. A DNA mismatch repair gene links to the ph2 locus in wheat / C. Dong, R. Whitford, P. Langridge // *Genome / National Research Council Canada = Génome / Conseil national de recherches Canada*. – 2002. – Vol. 45, No. 1. – P. 116–124.
8. Lloyd A. H. *TaMSH7*: a cereal mismatch repair gene that affects fertility in transgenic barley (*Hordeum vulgare* L.) / A. H. Lloyd, A. S. Milligan, P. Langridge, J. A. Able // *BMC plant biology*. – 2007. – Vol. 7. – P. 67.
9. Tam S. M. Effects of suppressing the DNA mismatch repair system on homeologous recombination in tomato / S. M. Tam, J. B. Hays, R. T. Chetelat // *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*. – 2011. – Vol. 123, No. 8. – P. 1445–1458.
10. Felsberg J. Promoter methylation and expression of mgmt and the dna mismatch repair genes MLH1, MSH2, MSH6 and PMS2 in paired primary and recurrent glioblastomas / J. Felsberg, N. Thon, S. Eigenbrod [et al.] // *International journal of cancer. Journal international du cancer*. – 2011. – Vol. 129, No. 3. – P. 659–670.
11. Gazzoli I. Regulation of the human MSH6 gene by the Sp1 transcription factor and alteration of promoter activity and expression by polymorphisms / I. Gazzoli, R. D. Kolodner // *Molecular and cellular biology*. – 2003. – Vol. 23, No. 22. – P. 7992–8007.
12. Antonyuk M. Z. Permanent genetic variability in the introgressive lines and amphidiploids of Triticeae / M. Z. Antonyuk, V. V. Shpylychyn, T. K. Ternovska // *Cytology and Genetics*. – 2013. – Vol. 47, No. 4. – P. 242–251.
13. Deng W. ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets / W. Deng, D. C. Nickle, G. H. Learn [et al.] // *Bioinformatics (Oxford, England)*. – 2007. – Vol. 23, No. 17. – P. 2334–2336.
14. Johnson M. NCBI BLAST: a better web interface / M. Johnson, I. Zaretskaya, Y. Raytselis [et al.] // *Nucleic acids research*. – 2008. – Vol. 36, No. Web Server issue. – P. W5–9.
15. Zhang Z. A greedy algorithm for aligning DNA sequences / Z. Zhang, S. Schwartz, L. Wagner, W. Miller // *Journal of computational biology : a journal of computational molecular cell biology*. – Vol. 7, No. 1–2. – P. 203–214.
16. Griffiths-Jones S. The microRNA registry / S. Griffiths-Jones // *Nucleic acids research*. – 2004. – Vol. 32, No. Database issue. – P. D109–111.
17. Pruitt K. D. NCBI reference sequences (RefSeq): current status, new features and genome annotation policy / K. D. Pruitt, T. Tatusova, G. R. Brown, D. R. Maglott // *Nucleic Acids Research*. – 2011. – Vol. 40, No. D1. – P. D130–D135.
18. Kapustin Y. Splign: algorithms for computing spliced alignments with identification of paralogs / Y. Kapustin, A. Souvorov, T. Tatusova, D. Lipman // *Biology direct*. – 2008. – Vol. 3. – P. 20.
19. Kuo H.-C. DBCAT: database of CpG islands and analytical tools for identifying comprehensive methylation profiles in cancer cells / H.-C. Kuo, P.-Y. Lin, T.-C. Chung [et al.] // *Journal of computational biology : a journal of computational molecular cell biology*. – 2011. – Vol. 18, No. 8. – P. 1013–1017.
20. Choulet F. Structural and functional partitioning of bread wheat chromosome 3B / F. Choulet, A. Alberti, S. Theil [et al.] // *Science (New York, N. Y.)*. – 2014. – Vol. 345, No. 6194. – P. 1249721-1–1249721-7.
21. Szadkowski M. Identification and functional characterization of the promoter region of the human MSH6 gene / M. Szadkowski, J. Jiricny // *Genes, chromosomes & cancer*. – 2002. – Vol. 33, No. 1. – P. 36–46.

*S. Mykhailyk, M. Antonyuk*

## BIOINFORMATIC ANALYSIS OF *TAMSH7* PROMOTER REGION

*Comparison of cDNA sequences of TaMSH7 gene and Triticum aestivum 3B chromosome revealed that TaMSH7 contains 15 introns and is localized in the position from 119371926 to 119373011 on 3B chromosome. Analysis of nine MSH7 5'UTR derived from different plants species, including cereals Oryza sativa and Zea mays, indicates that the cDNA sequence of TaMSH7 is probably shortened from 5'-end. Considering these data, a possible translation initiation codon ATG is located at 433-435 nucleotides upstream to the available cDNA gene sequence. Based on the analysis of the predicted promoter region of TaMSH7 gene, a large number of CpG methylation points were determined. The methylation level of such points can reveal the features of gene expression regulation in the genomes of introgressive lines. An 816 bp sequence upstream to predicted TaMSH7 translation initiation site contains 75 CpG methylation sites and becomes a target for bisulfite sequencing to establish the predicted promoter region methylation level. The sequences of primers covering 31 site of methylation in the predicted promoter region of TaMSH7 gene were obtained by Bisulfite Primer Seeker software developed by Zymo (Irvine, CA, USA). Evaluating the level of CpG sites methylation within the predicted promoter region allows to specify features of TaMSH7 gene expression in Triticum aestivum. Possible differences in the methylation level of TaMSH7 predicted BY THE gene promoter region in genomes of introgressive wheat lines and in genomes of their parental forms may show if the presence of alien genetic material in genome affects the level of methylation within the promoter region of the studied gene.*

**Keywords:** *TaMSH7*, *MSH7*, gene promoter region, translation initiation site, bisulfite sequencing, CpG methylation site, 5'UTR.