

**ПЛИГУН В. В.**, **АНТОНЮК М. З.**

Національний університет «Києво-Могилянська академія»,

Україна, 04070, м. Київ, вул. Григорія Сковороди, 2, ORCID: 0000-0001-5987-3293, 0000-0002-5877-969X

✉ [v.plyhun@ukma.edu.ua](mailto:v.plyhun@ukma.edu.ua)**МІНЛИВІСТЬ ПОСЛІДОВНОСТЕЙ НИЗКИ ГЕНІВ *Pm* У ПРЕДСТАВНИКІВ *TRITICEAE***

**Мета.** Співставлення послідовностей відібраних генів стійкості, розміщених у GenBank, для пошуку типів відмінностей у їхньому нуклеотидному складі та визначення значущості цих відмінностей при створенні праймерів. **Методу.** Пошук послідовностей проводили у біологічній базі даних GenBank. Порівняння різних сиквенсів одного і того ж гену здійснювали у DnaSP6 через розрахунок критерію Тадзіми для *Pm2*, *Pm3* та критерію Колмогорова-Смірнова у GeneDoc для *Pm21*. Множинне вирівнювання зробили у CLUSTALW перед тим як працювати у зазначених програмах. Розробляли праймери у Primer3. **Результати.** Варіабельність порівнюваних послідовностей є у нуклеотидних замінах, інсерціях / делеціях, наявності транспозону в одній з послідовностей *Pm3*. Найбільш варіабельними є ділянки, які кодують повтори, багаті на лейцин. Послідовності, ізольовані з *Triticum dicoccoides*, мають дві такі ділянки, в межах кожної деякі зі змін у нуклеотидній послідовності відображатимуться на амінокислотному складі білків. **Висновки.** Відмінності між порівнюваними послідовностями *Pm2*, *Pm3*, *Pm21* (у межах кожного з генів) є випадковими ( $p < 0,05$ ). За результатами вивчення сиквенсів створені комбінації праймерів для ідентифікації послідовностей, характерних для генів стійкості, на доступному рослинному матеріалі.

**Ключові слова:** гени стійкості, генні продукти, аналіз послідовностей генів.

Гени стійкості до борошністої роси (*Pm*) мають послідовності, які є консервативними у декількох генів, тому їхні генні продукти мають подібну будову. За продуктами гени класифікують на ті, що кодують нуклеотид-зв'язувальну ділянку (NBS), багаті на лейцин повтори (LRR), трансмембранні та протеїнаказні домени, Toll/інтерлейкін-1 рецептори (TIR). Білки на N-кінці можуть мати структуру типу спіраль-зміювик. Декілька доменів можуть поєднуватися у складі одного білка [1]. Наприклад, продукт

локусу *Pm3*, у якого 10 алелів, має структуру зміювик-петля, NBS та LRR. Серед цього кластеру є представники родини подібних до генів стійкості (resistance gene-like – RGL) послідовностей (*Pm3b*) на короткому плечі хромосоми 1A. RGL є і на гомеологічних хромосомах 1B та 1D пшениці. Паралоги *Pm3b* експресуються та потенційно можуть кодувати функціональні білки стійкості. Гени стійкості мають висококонсервативні послідовності, наприклад алелі *Pm3a*, *Pm3b*, *Pm3d* та *Pm3f* мають 97 % ідентичності, включаючи 5'-некодуєчу ділянку [2]. Зміни у послідовностях можуть відбуватися через інсерцію ретроелементів поблизу старт-кодону, що призводитиме до утворення різних алелів (функціональних або ні), що описано для локусу *Mla* ячменю. *Mla* теж пов'язують зі стійкістю до борошністої роси. Різниця між алелями може бути і на рівні інтронів [3, 4]. У межах кодувальних послідовностей можливе їх варіювання за нуклеотидним складом через заміни нуклеотидів, інсерції / делеції, обмін послідовностями шляхом міжгенної / міжалельної рекомбінації або конверсії генів [4]. Вивчення структури генів за еволюційними змінами їхніх нуклеотидних послідовностей є важливим для їхньої ідентифікації у геномах рослин, що контрастні за ознакою стійкості до борошністої роси, через створення олігонуклеотидних праймерів до певних регіонів цих генів.

**Матеріали і методи**

Вивчення мінливості послідовностей генів стійкості здійснювали за сиквенсами, розміщеними у біологічній базі даних GenBank за допомогою програмного забезпечення GeneDoc 2.7, DnaSP6 (версія 6.12.03). Для використання зазначених програм потрібно мати вирівняні послідовності, що зробили у CLUSTALW, після чого зберігали файли з сиквенсами у форматі .msf для роботи з GeneDoc 2.7, для DnaSP6 прийнятним є формат .fasta. За допомогою GeneDoc 2.7 визначали які саме є відмінності між послі-

© ПЛИГУН В. В., АНТОНЮК М. З.

довностями, для різних сиквенсів *Pm3*. У DnaSP6 для генів, які мали низку секвенсованих послідовностей (більше чотирьох) у GenBank (*Pm2*, *Pm3*) визначали чи є статистично значущою різниця між цими послідовностями через розрахунок критерію Тадзіми. Для *Pm21* з цією ж метою послуговувалися критерієм Колмогорова-Смірнова у GeneDoc. Такі перевірки проводили, оскільки подальшою метою було створення олігонуклеотидних праймерів на основі знайдених послідовностей. Розробляли послідовності олігонуклеотидних праймерів у Primer3 (версія 0.4.0). Серед запропонованих ресурсом пар праймерів віддавали перевагу тим парам, які мали на 5'- та / або 3'-кінцях гуанін або цитозин, не були самокомплементарними та мали малу різницю у температурі гібридизації з матрицею.

### Результати та обговорення

Для дослідження структури генів стійкості використовують праймери до консервативних послідовностей *Pm*, серед яких є послідовності, які кодують повтори, багаті на лейцин, кіназні домени, домени для зв'язування нуклеотидів. Для створення праймерів здійснювали пошук сиквенсів *Pm* у базі даних GenBank. Умовами для відбору були: доступний сиквенс кодуючої послідовності, наявність інформації про продукт, що його кодує ген, та який тип стійкості забезпечує. Нами відібрано 7 генів: *Pm2*, *Pm3*, *Pm4b*, *Pm8*, *Pm21*, *Pm24*, *Pm41*, серед яких *Pm2* (2*a*-2*c*), *Pm24* (24*a*-24*b*) мають декілька алелів та *Pm3* алельний ряд (a-j, l-r, k). Гени походять від: *Triticum sp.* – *Pm3*, *Pm4b* (*T. carthlicum*), *Pm24* (китайський місцевий генотип Hulutou), *Pm41* (*T. turgidum* var. *dicoccoides*); жита посівного – *Pm8*; *Aegilops sp.* – *Pm2*; *Haunaldia villosa* (*Dasyphyrum villosum* (L.) Borbás) – *Pm21*. У гексаплоїдній пшениці перенесені гени можуть міститися у складі транслокацій, наприклад 6VS-6AL для *Pm21* (табл. 1). Мультирезистентність забезпечують *Pm2* [5], *Pm21* [6], *Pm24* [7]; расоспецифічними є гени *Pm3* [8], *Pm4b* [9], *Pm8* [8]. Для *Pm41* описана стійкість на різних онтогенетичних стадіях розвитку рослин [9].

У GenBank можуть бути розміщені послідовності одного і того ж гену, отримані з різних рослинних зразків / генотипів чи різними дослідницькими групами, чи послідовності алелів, що спостерігали для *Pm2*, *Pm3*, *Pm21*. Тому виникла потреба перевіряти такі послідовності на відсутність статистично значущої різниці. Для

гена *Pm24* у літературних джерелах [7] зазначають два алелі, проте у базі даних депонована одна послідовність. Для інтрогредованого від *Ae. taushii*, *Pm2* у GenBank є сім послідовностей, які секвенсовані з генотипів пшениці м'якої та *Ae. taushii* (алелі h, g, i, j, f, e, d). Для *Pm3* у базі даних розміщено послідовності генів, ізольовані з *Triticum aestivum*, *Triticum monococcum*, *Triticum dicoccoides*, *Triticum turgidum* subsp. *durum*. Під час порівняння до уваги не брали псевдогени та неповні сиквенси. Для послідовності з кодом доступу AY146587.2, отриманої з *Triticum turgidum* subsp. *durum*, розміщений нуклеотидний склад локусу *Pm3*, включно з мобільними генетичними елементами. При співставленні сиквенсів наявність мобільних генетичних елементів не буде впливати на збільшення рівня розбіжностей, оскільки послідовності вирівняні та програми не враховують пробіли (gaps). Для *Pm21* у GenBank розміщено тільки три послідовності, ізольовані з *Triticum aestivum*, які ми і використовували.

За допомогою програмного забезпечення GeneDoc у межах послідовностей кожного з генів визначали, які є відмінності між сиквенсами. Для *Pm2*, який кодує білок з N-термінальним кінцем (50–420 нуклеотид), NB-ARC (691–1466 нуклеотид) та протеїнкіназоподібний домен (3176–3743) (PKC-like). Перша кодуюча ділянка найменш варіабельна та має одну транзицію й одну трансверсію. У другій, крім зазначених заміни є одно- (T, G, C), дво- (TT, TG) інсерції / делеції. На відміну від попередніх двох, у регіоні, який кодує PKC-like домен, є транзиції та трансверсії, розташовані поруч. Для прикладу на рис. 1 наведено тільки послідовності, які кодують функціональні домени білка, пов'язаного зі стійкістю, для *Pm2*.

У генах *Pm3* та *Pm21* спостерігали ті самі типи відмінностей у нуклеотидних послідовностях. З сиквенсами *Pm3* найбільша кількість змін згрупована у послідовностях, які кодують повтори, багаті на лейцин, особливо у другій ділянці (з близько 2600 до 3700 нуклеотиду). Зміни у нуклеотидах, які кодують LRR будуть впливати і на амінокислотний склад білків. Вирівнювали послідовності гена *Pm3* у NCBI Multiple Sequence Alignment Viewer, версія 1.25.0, оскільки мали 106 сиквенсів і таке вирівнювання більш наочно показує, в яких кодуючих ділянках зосереджені відмінності. Найбільш варіабельними є послідовності, які кодують LRR, в їхній C-термінальній частині. Такі зміни були в усіх

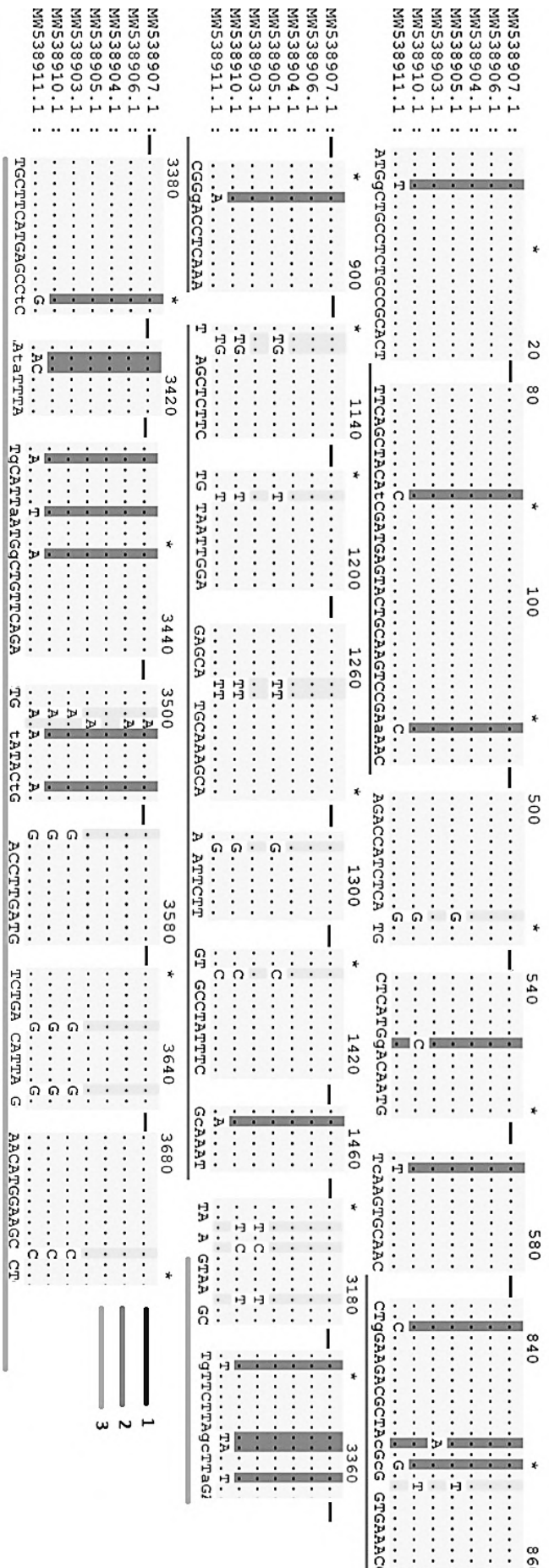


Рис. 1. Вирівняні послідовності гена *Rm2* у GenBank. Нуклеотидні послідовності, які кодують: 1 – Rх N-термінальну ділянку білка; 2 – NV-ARC; 3 – домен, подібний до протейнінази.

Таблиця 1. Перелік відібраних генів стійкості та їхні характеристики

Ген	Алелі (за наявності)	Хромосомна локалізація у пшениці м'якої	Джерело	Ген	Алелі (за наявності)	Хромосомна локалізація у пшениці м'якої	Джерело
1	<i>Rm2b</i> , <i>Rm2c</i> , <i>RmLX66</i> , <i>RmW14</i>	SDS	[5]	<i>Rm21</i>	секвеновані послідовності зразків <i>D. villosum</i> , відсутня номенклатура	6VS-6AL	[6]
2	<i>Rm3a-j</i> та <i>Rm3l-r</i> ( <i>T. aestivum</i> ), <i>Rm3k</i> ( <i>T. turgidum subsp. discosum</i> )	IAS	[8]	<i>Rm24</i>	<i>Rm24a, b</i>	IDS	[7]
3	<i>Rm4b</i>	відсутні	[9]		відсутні	ZVL	[9]
4	<i>Rm8</i>	IBL-IRS	[8]	<i>Rm41</i>			

порівнюваних послідовностей. У *Pm3* зазначених ділянок дві. Нуклеотидні зміни в цих ділянках призводять до заміни амінокислот. У першій кодуючій LRR послідовності зміни зосереджені у частині послідовностей, отриманих з *Triticum dicoccoides*. LRR відповідає за розпізнавання патогена, що потрібно для розвитку відповіді на різні раси патогенів.

Крім того, що є різні відмінності між послідовностями важливо знати чи є статистично достовірною ця різниця, щоб далі, для розробки праймерів до сиквенсу конкретного гену, використовувати лише одну послідовність. За нульову гіпотезу взяли припущення, що розбіжності між послідовностями є випадковими. Якщо припущення підтверджувалося, то в подальшому праймери створювали лише до одного з сиквенсів у межах кожного з генів. Мірою відмінності між послідовностями слугувало значення D критерію Ф. Тадзіми [11], яке розраховували за допомогою програмного забезпечення DnaSP6, попередньо провівши множинне вирівнювання послідовностей, які співставляли, у CLUSTALW. У DnaSP6 порівняння послідовностей відбувається методом ковзного вікна (sliding window), яке поступово переміщується по нуклеотидній послідовності. Вікно – це ділянка ДНК певного розміру, в межах якої по чергово з певним кроком порівнюються нуклеотиди. У нашому випадку розмір вікна – 100 нуклеотидів, крок – 25 нуклеотидів. Місця з прогалинами не враховуються при визначенні довжини вікна, всі вікна мають однакову кількість вирівняних («чистих») нуклеотидів. Показники, які визначаються, розраховуються у кожному вікні [12]. Для послідовностей ДНК визначали кількість поліморфних сайтів (segregating sites, S) та середню кількість нуклеотидних відмінностей, які отримують через попарне порівняння сиквенсів. Кількість поліморфних сайтів не є сама по собі хорошим показником відмін-

ностей між послідовностями, оскільки залежить від розміру вибірки. Якщо розмір вибірки малий, тоді значення показників великі, а зі збільшенням розміру – показники зменшуються. За допомогою зазначених показників розраховують критерій D, розроблений Тадзімою. Під час розрахунків припускають, що всі поліморфізми є випадковими, нейтральними, тобто не піддаються добору. D визначають за формулою:

$$D = \frac{k - \frac{S}{a_n}}{\sqrt{\text{var}(k - \frac{S}{a_n})}}, \text{ де}$$

$k$  – середнє значення кількості відмінних нуклеотидів на послідовність,  $S$  – кількість поліморфних сайтів,  $a_n$  – кількість пар порівнянь і та  $j$  послідовностей й розраховується як  $a_n = \sum_{i=1}^{n-1} \frac{1}{i}$ , де  $n$  – розмір вибірки (кількість послідовностей, яка порівнюється),  $i$  – індекс сумачі.

Отримані значення критерію Тадзіми D є статистично незначущими для генів *Pm2*, *Pm3*, тобто розбіжності між послідовностями є випадковими на прийнятому рівні значущості (табл. 2). Тому для створення праймерів до послідовностей генів брали лише одну послідовність (у межах кожного гена) і далі її використовували.

Для розрахунку критерію Тадзіми потрібно мати чотири або більше послідовностей, тому розрахувати його для *Pm21* не вдалося. Для визначення достовірності відмінностей послідовувалися критерієм Колмогорова-Смірнова у програмі GeneDoc. Як і в попередньому випадку нульовою гіпотезою було припущення, що відмінності у нуклеотидному складі послідовностей є випадковими. Використовували вирівняні сиквенси у CLUSTALW. Критерій забезпечує сувору оцінку відмінностей двох розподілів [13].

Таблиця 2. Результати порівняння різних послідовностей генів стійкості

Показник, який розраховували	<i>Pm2</i>	<i>Pm3</i>	<i>Pm21</i>
Довжина порівнюваних послідовностей (Number of sites)	4425	9177	2588
Загальна кількість нуклеотидів (сайтів) порівняння	4402	4016	2377
Кількість поліморфних (тих, що розщеплюються) сайтів, S	104	622	88
Середнє значення відмінних нуклеотидів, k	43,143	62,390	59,000
Нуклеотидна різноманітність, π	0,009	-1,633	0,0248
Критерій Тадзіми, D	0,095**	0,10*	#

Примітки: \* P < 0,05, \*\* P < 0,01; # – для розрахунку критерію потрібно мати 4 або більше послідовностей.



Рис. 2. Порівняння послідовностей гена *Pm21*, ізольованих з *Triticum aestivum* сорту Yang Mai 158: JF439308.1 – *serine/threonine protein kinase Stpk-B (Stpk-B)*; JF439307.1 – *serine/threonine protein kinase Stpk-D (Stpk-D)*; JF439306.1 – *serine/threonine protein kinase Stpk-A (Stpk-A)*. Цифри на графіку з правого боку вказують на пари порівнянь послідовностей.

Результат порівнянь зображений у вигляді графіка, де червона лінія вказує на відсоток ідентичності між послідовностями. На осі ординат нанесені оцінки порівняння послідовностей. Ці оцінки розподіляються за відсотком ідентичності (вісь абсцис) між парою порівнюваних послідовностей. У нашому випадку було три пари порівнянь (рис. 2). На графіку пари розташовані у порядку зростання відсотка подібності: перша – JF439308.1 та JF439306.1 (89 %), JF439308.1 та JF439307.1 (92 %), JF439307.1 та JF439306.1 – 94 %. Чим ближче червона лінія до правої вертикальної осі, тим більша гомологія. На рівні значущості 0,01 прийняли нульову гіпотезу, оскільки емпіричні значення критерію менші, ніж критичне 1,63, хоч між послідовностями є розбіжності, але вони не є статистично значущими. Емпіричні значення критерію знаходяться в межах від нуля до одиниці. Тобто є ділянки, які систематично представлені у послідовностях.

У результаті проведених порівнянь були розроблені послідовності праймерів за допомогою програми Primer3, які обмежують різні ділянки відібраних генів стійкості (рис. 3), які кодують: N-термінальні ділянки білків, подібну до протеїнкіназного домену (ПКс), домен NB-ARC з трьома субдоменами NB, ARC1 та ARC2, ділянки, багаті на лейцин (LRR), каталітичний

домен серин / треонінових кіназ, рецептор-асоційованих кіназ інтерлейкіну-1 і споріднених серин / треонінових кіназ (STKс – IRAK) (рис. 3).

До регіонів, які кодують домени, пов'язані зі стійкістю, за допомогою Primer3 розробили 20 пар праймерів.

### Висновки

Гени стійкості до борошнистої роси ізольовані з різних представників *Triticeae* мають консервативні ділянки, які кодують продукти з сайтами зв'язування нуклеотидів, повторами, багаті на лейцин, кіназними домени. Представленість цих доменів різниться, залежно від білків стійкості. Для *Pm2*, *Pm3* та *Pm21*, які мають декілька сиквенсів у GenBank, відмінності, які є між послідовностями (заміни нуклеотидів, інсерції / делеції) не є статистично значущими в межах кожного з генів ( $p < 0,05$ ), що було перевірено через використання критеріїв Тадзіми (для *Pm2*, *Pm3*), Колмогорова-Смірнова для *Pm21*. З використанням ресурсу Primer3 розроблені праймери до сиквенсів, характерних для генів *Pm2*, *Pm3*, *Pm8*, *Pm21*, *Pm24*, *Pm41*, з метою подальшого їх використання для ідентифікації послідовностей, які вони фланкують, на доступному рослинному матеріалі для вивчення структури генів стійкості.

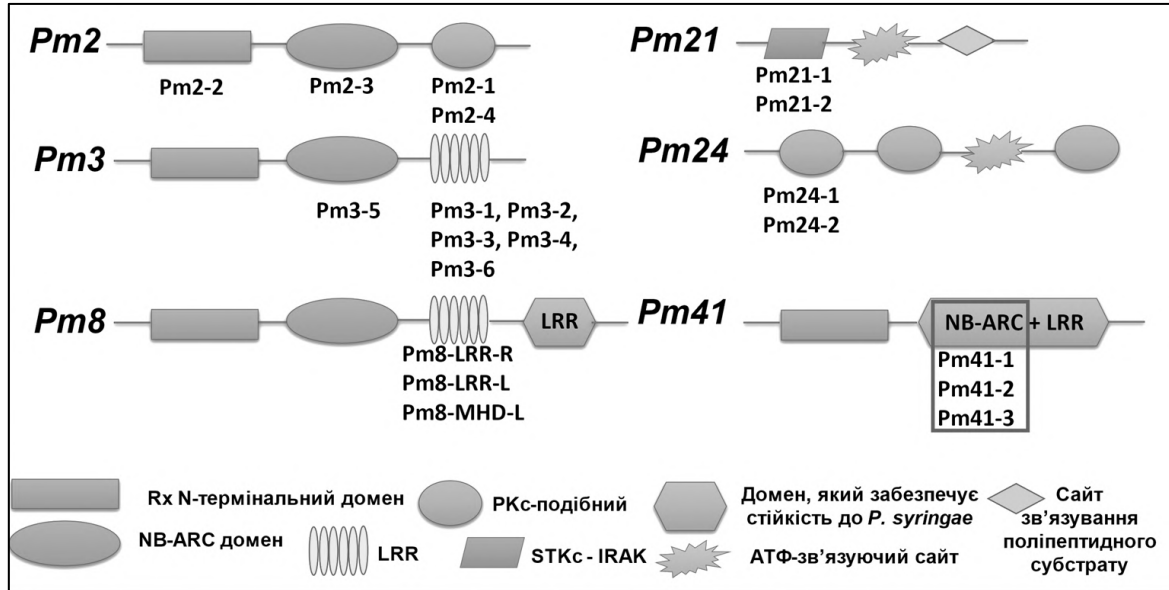


Рис. 3. Будова генів стійкості, за доменами білків, які вони кодують, і перелік назв праймерів до цих ділянок генів.

## References

1. Pečenková T., Sabol P., Kulich I., Ortmannová J., Žárský V. Constitutive Negative Regulation of R Proteins in Arabidopsis also via Autophagy Related Pathway? *Front. Plant Sci.* 2016. Vol. 7. doi: 10.3389/fpls.2016.00260.
2. Srichumpa P., Brunner S., Keller B., Yahiaoui N. Allelic Series of Four Powdery Mildew Resistance Genes at the *Pm3* Locus in Hexaploid Bread Wheat. *Plant Phys.* Vol. 139. P. 885–895. doi: 10.1104/pp.105.062406.
3. Seeholzer S., Tsuchimatsu T., Jordan T., Bieri S., Pajonk S., Yang W., et al. Diversity at the *Mla* powdery mildew resistance locus from cultivated barley reveals sites of positive selection. *Mol Plant Microbe Interact.* 2010. Vol. 23. P. 497–509. doi:10.1094/mpmi-23-4-0497.
4. Kusch S., Qian J., Loos A., Kümmel F., Spanu P. D., Panstruga R. Long- term and rapid evolution in powdery mildew fungi. *Mol Ecol.* 2023. doi: 10.1111/mec.16909.
5. Qie Y., Sheng Y., Xu H., Jin Y., Ma F., Li L., et al. Identification of a new powdery mildew resistance gene *pmDHT* at or closely linked to the *Pm5* locus in the Chinese wheat landrace Dahongtou. *Plant Dis.* 2019. Vol. 103 (10). P. 2645–2651. doi: 10.1094/PDIS-02-19-0401-RE28.
6. He H., Zhu S., Zhao R., Jiang Z., Ji Y., Ji J., et al. *Pm21*, encoding a typical CC-NBS-LRR protein, confers broad-spectrum resistance to wheat powdery mildew disease. *Mol Plant.* 2018. Vol. 11 (6). P. 879–882. doi: 10.1016/j.molp.2018.03.004.
7. Lu P., Guo L., Wang Z., Li B., Li J., Li Y., et al. A rare gain of function mutation in a wheat tandem kinase confers resistance to powdery mildew. *Nat Com.* 2020. Vol. 11 (1). doi: 10.1038/s41467-020-14294-0.
8. Hurni S., Brunner S., Stirnweis D., Herren G., Peditto D., McIntosh R. A., et al. The powdery mildew resistance gene *Pm8* derived from rye is suppressed by its wheat ortholog *Pm3*. *Plant J.* 2014. Vol. 79 (6). P. 904–913. doi: 10.1111/tj.12593.
9. Sánchez-Martín J., Widrig V., Herren G., Wicker T., Zbinden H., Gronnier J., et al. Wheat *Pm4* resistance to powdery mildew is controlled by alternative splice variants encoding chimeric proteins. *Nat Plants.* 2021. Vol. 7 (3). P. 327–341. doi: 10.1038/s41477-021-00869-2.
10. Zhang Q., Li Y., Li Y., Fahima T., Shen Q., Xie C. Introgression of the powdery mildew resistance genes *Pm60* and *Pm60b* from *Triticum urartu* to common wheat using durum as a “bridge.” *Pathogens.* 2021. Vol. 11 (1). doi: 10.3390/pathogens11010025.
11. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* 1989. Vol. 123 (3). P. 585–595. doi: 10.1093/genetics/123.3.585.
12. Rozas J., Ferrer-Mata A., Sánchez-DelBarrio J. C., Guirao-Rico S., Librado P., Ramos-Onsins S., Sánchez-Gracia A. DnaSP v6 Documentation. Barcelona, 2018. 112 p. Retrieved from: [http://www.ub.edu/dnasp/DnaSP6\\_Documentation\\_6.12.pdf](http://www.ub.edu/dnasp/DnaSP6_Documentation_6.12.pdf).
13. Nicholas K. B., Nicholas H. B., Deerfield D. W. Genedoc: analysis and visualization of genetic variation. *EMBNEW News.* 1997. Vol. 4.

**PLYHUN V. V., ANTONYUK M. Z.**

*National University of Kyiv-Mohyla Academy,  
Ukraine, 04070, Kyiv, Skovorody, str., 2*

**VARIABILITY OF *Pm* GENE SEQUENCES IN REPRESENTATIVES OF *TRITICEAE***

**Aim.** Comparison of the sequences of selected resistance genes deposited in GenBank to find types of differences in their nucleotide composition and to determine the significance of these differences for primer design. **Methods.** Sequence searches were performed in the GenBank biological database. Comparison of different sequences of the same gene was carried out in DnaSP6 by calculating the Tajima test for *Pm2*, *Pm3* and the Kolmogorov-Smirnov test in GeneDoc for *Pm21*. Multiple alignment was performed in CLUSTALW before using the specified programs. Primers were designed in Primer3. **Results.** Variability of the compared sequences is represented by nucleotide substitutions, insertions/deletions, the presence of a transposon in one of the *Pm3* sequences. The regions coding for leucine-rich repeats are the most variable. The sequences isolated from *Triticum dicoccoides* have two such regions, within each of which some of the changes in the nucleotide sequence will be reflected in the amino acid composition of the proteins. **Conclusions.** The differences between the compared sequences of *Pm2*, *Pm3*, *Pm21* (within each of the genes) are random ( $p < 0.05$ ). According to the results of the sequences analysis, combinations of primers were developed for the identification of sequences characteristic to resistance genes on the available plant material.

**Keywords:** resistance genes, gene products, analysis of gene sequences.