

ІДЕНТИФІКАЦІЯ ГОМОЛОГІВ ГЕНІВ *Sr33* І *Sr35* У КУЛЬТУРНИХ ЗЛАКІВ, СПОРІДНЕНИХ З *AEGILOPS TAUSCHII* ТА *TRITICUM MONOCOCCUM*

Б.В. ІВАЩУК, Я.В. ПІРКО, А.П. ГАЛКІН, Я.Б. БЛЮМ

Інститут харчової біотехнології та геноміки НАН України, Київ
E-mail: antutilo@gmail.com

В геномах *Triticum aestivum* і *Triticum monococtum* ідентифіковано гомологи генів *Sr33* і *Sr35*, що забезпечують стійкість пшениці до високпатогенних рас стеблової іржі (*Ug99*). Здійснено їх порівняння з вихідними послідовностями для визначення важливих для стійкості амінокислотних сайтів. З'ясовано, що послідовності *S5DMA6* і *E9P785* є найближчими гомологами білка *RGA1e* – продукту гена *Sr33*, а послідовності *M7YFA9* (*CNL-C*) та *F2E9R2* гомологами білка *CNL9* – продукту гена *Sr35*. Передбачається, що гомологи генів *Sr33* і *Sr35*, отримані з диких родичів пшениці та ячменю, здатні надавати стійкість до різних форм стеблової іржі і можуть бути використані у подальших селекційних програмах, націлених на покращення вітчизняних сортів пшениці.

Ключові слова: стеблова іржа, гени стійкості, гомологи, дикі злаки.

Вступ. Серед хвороб, які вражають пшеницю, стеблова іржа (*Puccinia graminis* f. sp. *tritici*) посідає одне із перших місць за втратами врожаю. Після появи її нової раси *Ug99* в Уганді, а далі в Африці, на Близькому Сході і в Південній Азії ці втрати склали близько 3 млрд доларів США [1]. Одним з головних методів боротьби зі стебловою іржею є пошук генів стійкості до цього захворювання в диких родичах пшениці та їх подальше перенесення до нових сортів пшениці. На даний час відомо близько 50 генів стійкості до стеблової іржі пшениці [2], та лише нещодавно два з них – *Sr33* і *Sr35* – клоновано та секвеновано [3, 4].

В попередній роботі нами проаналізовано наявність гомологів генів стійкості до стеблової іржі *Rpg5*, *RGA1* та *Adf3* ячменю у пшениці та її близьких родичів, оскільки відомо, що ці гени лише разом забезпечують стійкість до стеблової іржі типу *Ug99* [5]. У переважній більшості випадків стійкість є результатом кооперативної взаємодії кількох генів, а іноді їх білкові продукти можуть взаємодіяти між со-

бою. Зокрема встановлено, що білкові продукти генів *Rpg5* та *RGA1* дуже подібні за структурою у двох доменах. Тому логічним є пошук гомологів генів *Sr33* і *Sr35* в геномах рослин з триби пшеничних, а також з'ясування можливості взаємодії білкових продуктів знайдених гомологів між собою, спираючись на відомості про такі взаємодії білків генів *MLA*.

Встановлено, що ген *Sr33* з *Aegilops tauschii* надає стійкості до багатьох патотипів стеблової іржі, зокрема раси *Ug99* [3]. Ген має 7 екзонів, і його білковий продукт містить мотиви, які є консервативними у багатьох білків *CNL* (*CC-NB-LRR*) (coiled-coil nucleotide-binding leucine rich protein). *Sr33* виявив найбільшу схожість до *MLA*-подібного гена з *T. monococtum*. Серед генів *MLA* ячменю найбільш подібним до *Sr33* є *MLA34* (80 % схожості). Підтверджено, що суперспіралізований домен білка гена *Sr33* не гомодимеризується і не взаємодіє з факторами транскрипції з родини *WRKY*. У рослин з цим геном при зараженні не спостерігається гіперчутливої загибелі клітин під час імунної відповіді. Трансформація чутливих рослин призводить до появи фенотипу, характерного для культур – носіїв гена *Sr33* [3].

Ген *Sr35* з *T. monococtum*, як і *Sr33*, забезпечує стійкість до стеблової іржі. Скринінг *T. urartu* і *T. turgidum* ssp. *dicoccoides* не виявив його ортологів, також не знайдено його гомологів в поліплоїдних видах пшениці. Експериментально доведено, що на відміну від гена *Sr33* одного гена *Sr35* вистачає для забезпечення повної стійкості до різних підтипів раси *Ug99*, і для забезпечення стійкості у цьому випадку не потребується додаткових генів специфічної стійкості. При зараженні патогеном рослин з інтрогресованим *Sr35* виявлено сильну гіперчутливу загибель клітин хазяїна [4].

Білкові продукти генів *Sr33* і *Sr35* належать до класу рецепторів *CC-NB-LRR*, і наявна до-

статня кількість інформації саме щодо цього класу білків або їм подібних. Виявлено, що два гени *Arabidopsis thaliana* – *RPS4* і *RRS1*, продукти яких належать до родини білків NB-LRR (з TIR N-кінцевим доменом), при перенесенні в парі до рослин з родин *Brassicaceae*, *Solanaceae* і *Cucurbitaceae* забезпечують стійкість до різних грибкових та бактеріальних патогенів. Це вказує на консервативність внутрішньоклітинних мішеней для патогенів у цих рослин, хоч і невідомо, якими саме є ці білки [6].

Для іншої пари білкових продуктів генів *RGA4* і *RGA5*, які належать до класу білків CC-NB-LRR, показано теж, що лише білок RGA4 опосередковує гіперчутливу клітинну загибель, а RGA5 працює як його репресор за відсутності активації ефектором. У цьому випадку RGA5 можна вважати пасткою для молекули ефектора, яка охороняється RGA4. Виявлено, що білкові продукти цих двох генів утворюють гетерокомплекси через їхні суперспіралізовані домени, а RGA4 при активації або повністю, або частково вивільняється від взаємодії з RGA5 і таким чином може виконувати свої сигнальні функції [7].

Одним із найбільш повно охарактеризованих генів стійкості є ген *MLA10* з ячменю. Його білковий продукт також належить до родини CC-NB-LRR. Показано, що одного суперспіралізованого домена вистачає для спричинення гіперчутливої загибелі клітин. Цей домен містить консервативний в цьому субкласі рецепторів EDVID-мотив, і разом з мотивом Р-петлі в нуклеотид-зв'язуючому домені є важливим для сигналіngu. Передбачається, що саме гомодимер суперспіралізованих доменів може викликати індуковану клітинну загибель. Домен, багатий на лейцинові повтори, виступає репресором активності білка MLA10. За відсутності цього домена навіть мутантні варіанти із замінами в EDVID-мотиві та Р-петлі можуть викликати клітинну загибель на противагу повноцінному білку, в якому мутації за згаданими мотивами спричиняють блокування низхідних сигналів [8].

Для стійкості до листової іржі пшениці, опосередкованої *Lr10*, потрібно два гени класу CC-NB-LRR – *Lr10* і *RGA2*. Для алейних варіантів гена *Lr10* виявлено консервативність в частинах, які кодуєть нуклеотид-зв'язуючий

і збагачений лейциновими повторами домену. Проте в частині, яка кодує суперспіралізований N-кінцевий домен, знайдено велику кількість одонуклеотидних поліморфізмів та показано, що на неї діє дизруптивний добір. Ген *RGA2*, що також потрібний для *Lr10*-опосередкованої стійкості, має менше алелів, і як було продемонстровано раніше, поліморфізм цього гена набагато нижчий, ніж у *Lr10* чи інших подібних R-генах. Це можна пояснити тим, що білок RGA2 не бере участь у взаємодіях з ефекторами і в сигнальних шляхах розташований нижче від *Lr10*. Але не підтверджено, що білки *Lr10* і *RGA2* можуть взаємодіяти, хоча, можливо, така взаємодія і буде відбуватись за присутності ефектора патогена [9].

Два гени рису класу CC-NB-LRR – *Pi5-1* і *Pi5-2* – необхідні для набуття стійкості до *Magnaporthe oryzae*. Результати аналізу рівнів експресії цих генів свідчать про те, що транскрипти *Pi5-1* накопичуються після зараження патогеном, тоді як ген *Pi5-2* експресується конститутивно, тобто на однакових рівнях до і після зараження. Як показують результати філогенетичного аналізу, обидва гени відносяться до однієї класи і відрізняються від інших R-генів [10].

Продукти генів *Prf* і *Pto* – NB-LRR білок *Prf* і протеїнкіназа *Pto* – забезпечують впізнання ефекторів *AvrPro* і *AvrProB* патогена *Pseudomonas syringae*. *AvrPro* і *AvrProB* хоч і є різними структурно, взаємодіють з протеїнкіназою *Pto* через одні й ті ж амінокислотні залишки, а ця взаємодія в свою чергу викликає імунну відповідь. Показано, що *Prf* утворює олігомерні комплекси, які можуть взаємодіяти чи з протеїнкіназою *Pto*, чи з білками цієї ж родини *Fen*, *Pth2*, *Pth3* або *Pth5*. Комплекс білків *Prf* і *Pto* забезпечує слабшу стійкість, ніж комплекс білків *Prf* і *Pth*. Білок *Prf* не накопичується і зазнає деградації за відсутності протеїнкінази *Pto* чи білків цієї ж родини [11].

Таким чином, стає зрозуміло, що подібні пари генів відіграють ключову роль в забезпеченні стійкості до різних патогенів. До того ж підтверджено, що такі пари генів часто є зчепленими, і їх промотори можуть перекриватись для більш точної регуляції. У більшості випадків для набуття стійкості необхідні обидва гени, але іноді вони можуть забезпечувати стійкість окремо один від одного [12].

Цілком зрозуміло, що з появою нових високовірулентних рас стеблової іржі з'являється необхідність пошуку нових генів стійкості. Досліди з міжвидового перенесення *Sr33* і *Sr35* в *T. aestivum* підтвердили свою ефективність. Спираючись на той факт, що у більшості випадків для стійкості потрібно два гени, один з яких є консервативним, а інший поліморфним, і швидше за все *Sr33* і *Sr35* є поліморфними, можливо їх гомологи з диких злаків будуть ефективними проти стеблової іржі. Тому метою даної роботи був біоінформаційний пошук гомологів цих генів та з'ясування питання про їхню потенційну можливість забезпечувати стійкість до стеблової іржі.

Матеріали і методи. Повні амінокислотні послідовності білкових продуктів генів *Sr33*, *Sr35* та їх гомологів отримані за допомогою бази даних білкових послідовностей UniProtKB (www.uniprot.org) [13]. BLASTp-сканування (BLASTP 2.2.29+) бази виконували з дотриманням стандартних параметрів [14]. Відбір гомологів здійснювали на основі показника ідентичності (не нижче 30 %) та повноти амінокислотних послідовностей. Доменну архітектуру білків досліджували за допомогою online інструменту SMART (<http://smart.embl-heidelberg.de/>) [15]. Множинне вирівнювання виконували за допомогою сервера T-Coffee (<http://www.tcoffee.org/>) з використанням методів clustalw_msa і t_coffee_msa [16]. Редагування послідовностей множинного вирівнювання та вилучення гепів здійснювали за допомогою програми Jalview [17]. Філогенетичний аналіз проводили, використовуючи програмний пакет MEGA6.06 [18]. Для перевірки надійності розміщення відповідних таксономічних одиниць OTUs (Operational Taxonomic Units) застосовували Bootstrap-тест (1000 реплік) [19]. Еволюційні відстані розраховували за методом Джона-Тейлора Торнтонна (JTT) [20, 21]. Відмінності в швидкості заміну в однакових сайтах різних послідовностей змодельовані на основі гамма-розподілу. Послідовності для подальшого аналізу відбирали серед білкових продуктів відповідних генів та їх найближчих гомологів, з яких не вилучали пробіли після множинного вирівнювання. Для аналізу функціональних сайтів у домені, збагаченому на лейцинові повтори (leucine reach repeat) у

послідовностях білків RGA1e (кодується геном *Sr33*) і CNL9 (кодується геном *Sr35*), реконструйовано їх просторову структуру за допомогою Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>) [22] та LOMETS (<http://zhanglab.ccmb.med.umich.edu/LOMETS/>) [23].

Пошук можливих функціональних сайтів здійснювали з використанням програми Phyre Investigator (<http://www.sbg.bio.ic.ac.uk/phyre2/html/help.cgi?id=help/investigator>) [22] та COACH [24].

Результати досліджень та їх обговорення.
Пошук гомологів генів Sr33 і Sr35. Для пошуку гомологів білка RGA1e, який є продуктом гена *Sr33*, обрано послідовність S5DIH1 (ідентифікатор бази даних UniProtKB). Найближчими гомологами цієї повної послідовності, як і для всіх інших послідовностей, виявились D2KPB6 (MLA23, 80 % ідентичності), D2KPC4 (MLA34, 80 %), Q6WWJ4 (MLA10, 80 %), Q8GSK4 (MLA13, 81 %), M7YIN9 (Disease resistance protein RPP13, 80 %), E9P785 (MLA1, 88 %), S5DMA6 (RGA1e-like protein, 93 %), M8A3U1 (Disease resistance protein RPP13, 79 %), M7ZCM7 (Disease resistance protein RPP13, 80 %), W5A7S9 (78 %).

Для перевірки еволюційної спорідненості наведених послідовностей проведено кладистичний аналіз за допомогою методу об'єднання сусідів [19]. Встановлено, що послідовність S5DMA6 з *T. aestivum* можливо є ортологом білка RGA1e, а ген, що кодує послідовність E9P785 з *T. monococcum*, з високою ймовірністю є його паралогом (рис. 1, а). Незважаючи на високий рівень ідентичності, білкові продукти генів *MLA* ячменю розташовані на окремій гілці. Такий результат може вказувати на значну відмінність у функціях пшеничних і ячмінних гомологів гена *Sr33*.

Шляхом аналізу доменної архітектури послідовності білка RGA1e за допомогою інструменту SMART виявлено три домени: суперспіралізований, нуклеотид-зв'язуючий і домен, збагачений на лейцинові повтори. Для виявлення додаткових гомологів у подальшому виконували пошук окремо по кожному домену.

В результаті пошуку гомологів за суперспіралізованим доменом білка RGA1e знайдено такі послідовності: E9P785 (95,33 % ідентичності), S5DMA6 (100 %), Q8GSK4 (82 %),

Ідентифікація гомологів генів *Sr33* і *Sr35* у культурних злаків

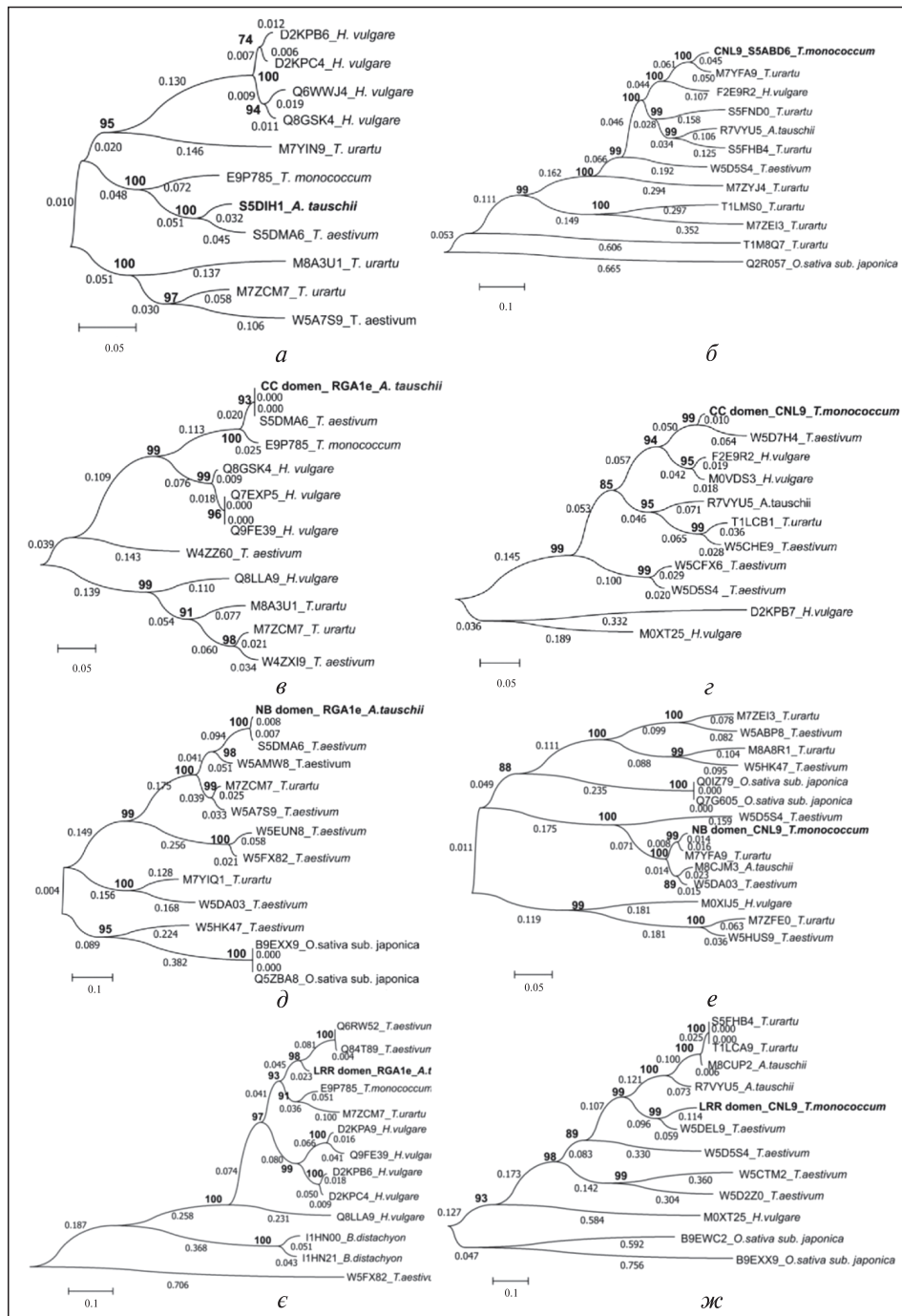


Рис. 1. Результати філогенетичного аналізу послідовності білка RGA1e (S5DIH1), білка CNL9 (S5ABD6) та їх гомологів. Зображено кладограми з гомологами, пошук яких проводився з використанням таких послідовностей: *а* – повної послідовності RGA1e; *б* – повної послідовності CNL9; *в* – суперспіралізованого домена RGA1e; *г* – суперспіралізованого домена CNL9; *д* – нуклеотид-зв’язуючого домена RGA1e; *е* – нуклеотид-зв’язуючого домена CNL9; *е* – багатого на лейцинові повтори домена RGA1e; *ж* – багатого на лейцинові повтори домена CNL9. Жирним позначено результати bootstrap-тесту. Генетична відстань вказана біля відповідних гілок

Q7EXP5 (MLA1, 81 %), Q9FE39 (81 %), W4ZZ60 (72 %), Q8LLA9 (RGH1a, 64 %), M8A3U1 (64 %), M7ZCM7 (62 %), W4ZXI9 (61 %). Як і в попередньому випадку, результати кладистичного аналізу підтверджують, що найближчими гомологами за послідовністю суперспіралізованого домена білка RGA1e є послідовності E9P785 і S5DMA6 (рис. 1, б).

За допомогою blastp пошуку знайдено такі гомологи за послідовністю нуклеотид-зв'язуючого (nucleotide-binding – NB) домена: S5DMA6 (99 %), W5AMW8 (87 %), M7ZCM7 (83 %), W5A7S9 (83 %), W5EUN8 (63 %), W5FX82 (64 %), M7YIQ1 (Disease resistance protein RPP13, 62 %), W5DA03 (60 %), W5HK47 (59 %), B9EXX9 (55 %), Q5ZBA8 (MLA6 protein-like, 55 %). Результати філогенетичного аналізу підтвердили спорідненість цих гомологів (рис. 1, д). Послідовність E9P785, хоча і є високоідентичною за нуклеотид-зв'язуючим доменом до відповідної послідовності білка RGA1e (82 % ідентичності), була виключена з дерева після проведеного bootstrap-тесту. Не вдалось також отримати кладограму, в якій одночасно були б представлені гомологи за цим доменом з ячменю та пшениці. При врахуванні при множинному вирівнюванні додаткових послідовностей з *T. aestivum* і *T. urartu* подальші результати побудови кладограми та bootstrap-тесту були незадовільні. Першою причиною такого результату може бути недостатня кількість послідовностей для виявлення гомологів ячменю в цих видах пшениці, другою причиною – різне еволюційне походження цього домена в родах *Hordeum* і *Triticum* саме в цій групі рецепторів.

За допомогою збагаченого на лейцинові повтори домена білка RGA1e в базі UniProt знайдено такі гомологи: Q6RW52 (Mla-like protein, 92 %), Q84T89 (Mla-like protein, 92 %), E9P785 (87 %), M7ZCM7 (85 %), D2KPA9 (MLA8, 82 %), Q9FE39 (82 %), D2KPB6 (82 %), D2KPC4 (83 %), Q8LLA9 (75 %), I1HN00 (62 %), I1HN21 (63 %), W5FX82 (54 %). Виявлено, що найбільш подібними до збагаченого на лейцинові повтори домена білка RGA1e є послідовності Q6RW52 і Q84T89 з *T. aestivum* (рис. 1, е). Слід зазначити, що послідовність збагаченого на лейцинові повтори домена білкової послідовності S5DMA6, яка ідентична на 93 % до

білка RGA1e, була виключена з кладограми через низьке значення bootstrap-тесту. Саме за цим доменом послідовність RGA1e-подібного білка найбільше відрізняється від послідовності білка RGA1e.

Для пошуку гомологів білка CNL9, який є продуктом гена *Sr35*, обрано послідовність S5ABD6 (ідентифікатор UniProtKB). Найближчими гомологами цього білка виявились M7YFA9 (білок CNL-C, 92 % ідентичності), F2E9R2 (84 %), S5FND0 (білок CNL-D, 75 %), R7VYU5 (RPP8-like protein 3, 77 %), S5FHB4 (білок CNL-A, 77 %), W5D5S4 (73 %), M7ZYJ4 (RPP13-like protein, 66 %), T1LMS0 (58 %), M7ZEI3 (RPP13-like protein, 57 %), T1M8Q7 (52 %), Q2R057 (49 %). На основі цих послідовностей було проведено кладистичний аналіз і виявлено, що найближче до повної послідовності білка CNL9 розташовані гомологи з *T. urartu* (рис. 1, б).

Гомологічними за послідовністю суперспіралізованого домена білка CNL9 є W5D7H4 (93 % ідентичності), F2E9R2 (88 %), M0VDS3 (89 %), R7VYU5 (79 %), T1LCB1 (75 %), W5CHE9 (76 %), W5CFX6 (75 %), W5D5S4 (75 %), D2KPB7 (MLA25-1, 57 %) та M0XT25 (60 %). На відміну від повної послідовності, найближчими гомологами за суперспіралізованим доменом цього білка виявились послідовності з *T. aestivum* та *H. vulgare* (рис. 1, з).

За допомогою послідовності нуклеотид-зв'язуючого домена білка CNL9 знайдено такі гомологи: M7ZEI3 (64 % ідентичності), W5ABP8 (65 %), M8A8R1 (Disease resistance protein RPP13, 62 %), W5HK47 (64 %), Q0IZ79 (65 %), Q7G605 (65 %), W5D5S4 (79 %), M7YFA9 (98 %), M8CJM3 (Disease resistance protein RPM1, 94 %), W5DA03 (94 %), M0XIJ5 (64 %), M7ZFE0 (62 %), W5HUS9 (63 %). За даним доменом найбільший рівень спорідненості виявлений для послідовності M7YFA9 (рис. 1, е).

Застосувавши для пошуку як запит послідовність домена збагаченого на лейцин білка CNL9, знайдено такі гомологи: S5FHB4 (71 %), T1LCA9 (71 %), M8CUP2 (72 %), R7VYU5 (75 %), W5DEL9 (86 %), W5D5S4 (64 %), W5CTM2 (55 %), W5D2Z0 (59 %), M0XT25 (52 %), B9EWC2 (47 %), B9EXX9 (42 %). На відміну від нуклеотид-зв'язуючого домена CNL9 найближчими гомологами за

Ідентифікація гомологів генів *Sr33* і *Sr35* у культурних злаків

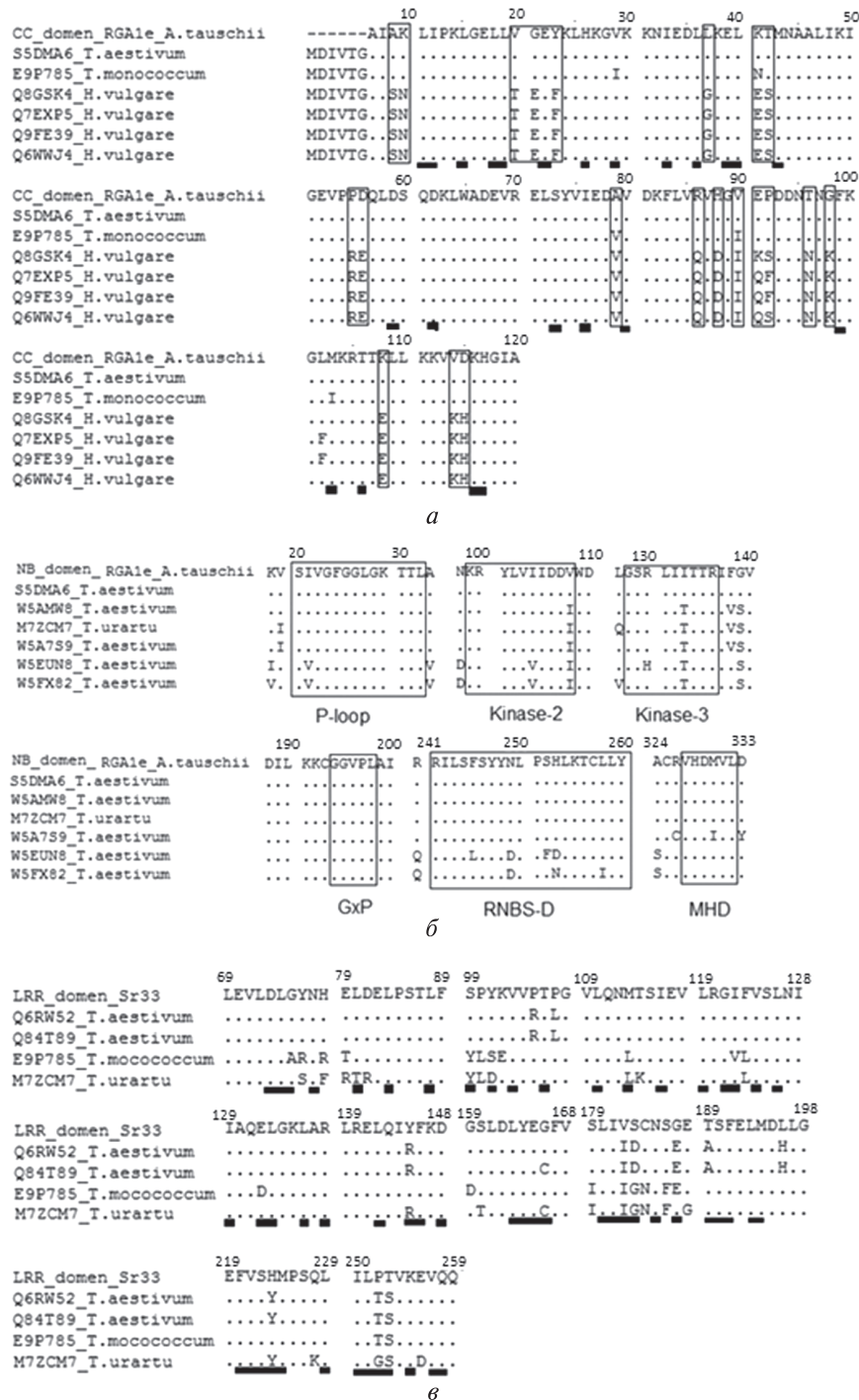


Рис. 2. Фрагменти множинних вирівнювань послідовностей доменів гомологів RGA1e: *a* – вирівнювання гомологів за послідовністю суперспіралізованого домена, рамками зазначено місця, де гомологи відрізняються між собою; *б* – вирівнювання гомологів за послідовністю нуклеотид-зв’язуючого домена, рамками зазначено головні мотиви домена; *в* – вирівнювання гомологів за послідовністю домена, багатого на лейцинові повори. Чорні прямокутники під вирівнюваннями вказують на передбачені важливі функціональні сайти цих послідовностей

послідовністю домена, багатого на лейцинові повтори, виявились послідовності з *T. aestivum* і *A. tauschii* (рис. 1, ж).

Аналіз функціональних сайтів та головних мотивів суперспіралізованого (CC), нуклеотид-зв'язуючого (NB) і збагаченого на лейцинові повтори (LRR) доменів білкових продуктів генів *Sr33* і *Sr35*. Для аналізу суперспіралізованого домена білка RGA1e використовували його амінокислотну послідовність та послідовності цього ж домена найближчих гомологів – E9P785, S5DMA6, Q8GSK4, Q7EXP5, Q9FE39, а також послідовність Q6WWJ4 (білок MLA10), для якої є відомості про важливі для функціонування суперспіралізованого домена амінокислотні залишки [7] (рис. 2, а). В попередніх дослідженнях виявлено, що для гомодимеризації суперспіралізованих доменів двох білків MLA10 важливими є близько 20 амінокислотних залишків (позначені чорними прямокутниками на рис. 2, а), що стабілізують кластер альфа-спіралей, з яких складається цей домен. Нині відомо також, що суперспіралізовані домени білків RGA1e не димеризуються [3]. Проте нами не виявлено відмінностей в цих сайтах між послідовністю суперспіралізованого домена білка RGA1e та його гомологами і послідовністю суперспіралізованого домена білка MLA10. Але з 21 амінокислотного залишка, за якими послідовності суперспіралізованого домена білків RGA1e, E9P785, S5DMA6 відрізняються від їхніх ячмінних гомологів, у п'яти сайтах спостерігається зміна полярності залишків (A9S, V20T, G21E, G98K, V114K), а в дев'яти сайтах є відмінності в заряді амінокислотних залишків (K10N, G21E, K41E, P55R, R96Q, E91K/Q, G98K, K108E, V114K) (рис. 2, а). Ймовірно, саме ці відмінності не дозволяють відбуватись димеризації, якщо така взагалі можлива і потрібна для функціонування білка RGA1e.

Для аналізу послідовності нуклеотид-зв'язуючого домена продукта гена *Sr33* обрано його найближчі гомологи – S5DMA6, W5AMW8, M7ZCM7, W5A7S9, W5EUN8, W5FX82. Розглядалися основні мотиви, які характерні для білків класу NB-ARC [25, 26]. При цьому виявлено такі відмінності між гомологами: в мотиві Р-петлі послідовності W5EUN8 і W5FX82 відрізнялись від решти (сайт I22V). В мотиві

RILSFSYYNLPShLKTCLLY (мотив позначається як Kinase-2) відрізнялись ці ж послідовності (сайт I104V), а також послідовності S5DMA6 і RGA1e (V108I). В мотиві Kinase-3 від решти відрізнялась W5EUN8 (H130R), а також SD5MA6 і RGA1e (I134T). В мотиві GxP різниці не виявлено, а в мотиві RNBS-D відрізнялись W5EUN8 (L245F, N249D, F252S, D253H) і W5FX82 (D249N, N253H, I258L). В мотиві MHD від гомологів відрізнялась лише W5A7S9 (I330M) (рис. 2, б). Оскільки ці мотиви є доволі консервативними для рецепторів NB-ARC класу у багатьох видів, навіть незначні заміни можуть призводити до зміни активності білків. У даному випадку абсолютну ідентичність до послідовності білка RGA1e має лише послідовність S5DMA6.

За допомогою програм також була відтворена просторова структура домена збагаченого лейциновими повторами білка RGA1e та визначені потенційні сайти його міжмолекулярних взаємодій. Для порівняння обрали гомологи Q6RW52, Q84T89, E9P785 та M7ZCM7. В результаті виявлено 16 позицій, в яких спостерігались відмінності між послідовностями: послідовність E9P785 – A75G, Y99S, S101Y, L113M, V122I, D132E, I183V, G184S, E187G, T252P, S253T; послідовність M7ZCM7 – T80L, Y99S, D101Y, L113M, R145Y, C166G, I183V, G184S, Y223H, T252P, S253T; послідовність Q84T89 – R145Y, C166G, I183V, D184S, E187G, A189T, Y223H, T252P, S253T; послідовність Q6RW52 – R145Y, I183V, D184S, E187G, A189T, Y223H, T252P, S253T). З них в порівнянні з білком RGA1e сукупно у восьми позиціях змінювалась полярність і в трьох – заряд залишків.

Результати аналізу послідовності суперспіралізованого домена білка CNL9 свідчать про те, що ця послідовність та послідовність суперспіралізованого домена білка MLA10 в сайтах, важливих для гомодимеризації, відрізняються в трьох позиціях (L12I, V60D, M76I), а його гомологи (W5D7H4, F2E9R2, M0VDS3, R7VYU5, T1LCB1, W5CHE9), крім вже зазначених позицій, ще в кількох для кожного окремо (разом в сайтах 22, 25, 29, 39, 59, 76) (рис. 3, а). Відомо, що при перенесенні до пшениці гена, який кодує білок CNL9, вона набуває стійкості до стеблової іржі [4]. І хоч ортологів білка CNL9

Ідентифікація гомологів генів *Sr33* і *Sr35* у культурних злаків

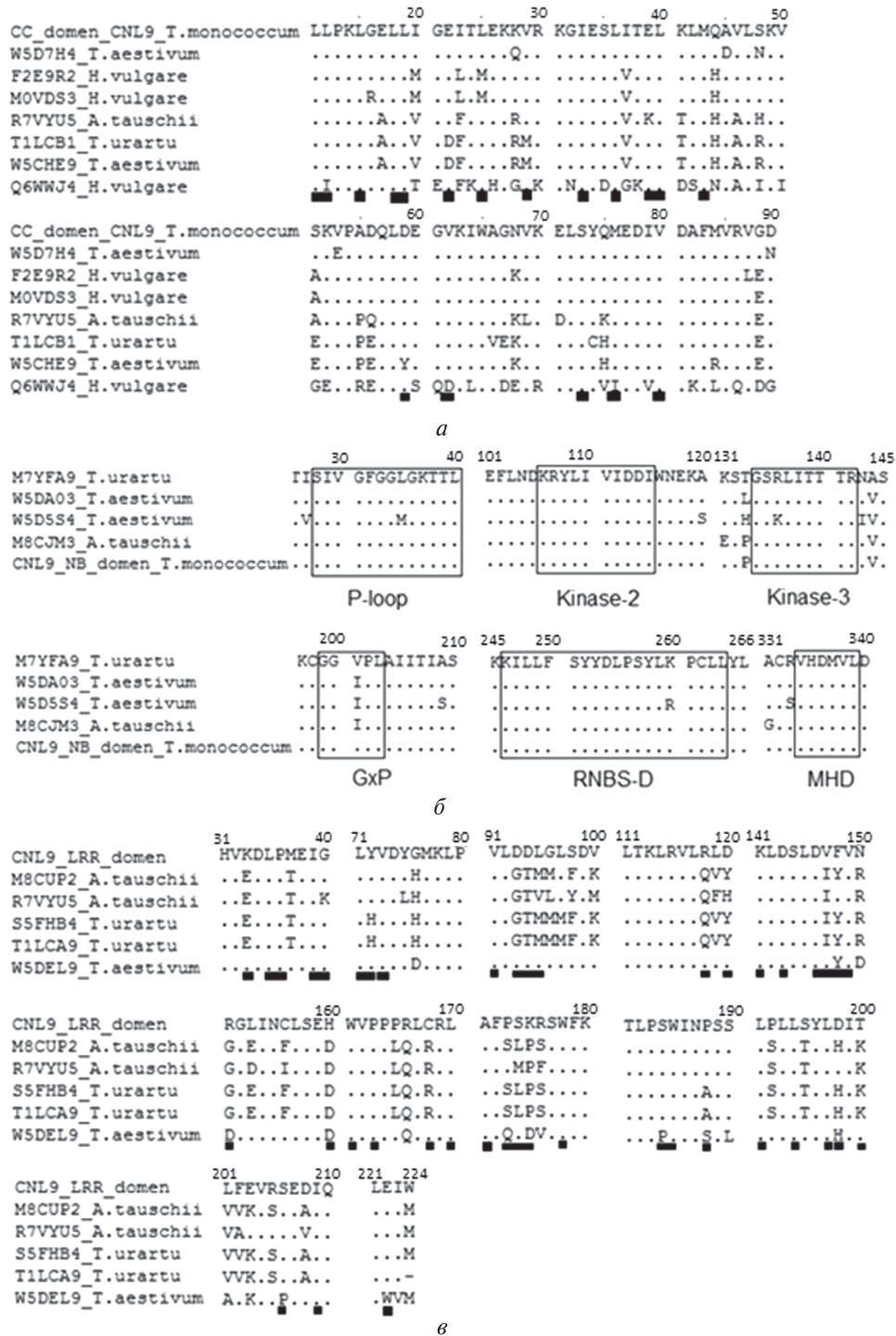


Рис. 3. Множинні вирівнювання гомологів послідовностей доменів CNL9: *a* – вирівнювання для послідовності суперспіралізованого домена; *б* – вирівнювання гомологів за послідовністю нуклеотид-зв'язуючого домена, рамками зазначено головні мотиви домена; *в* – вирівнювання гомологів за послідовністю багатого на лейцинові повтори домена. Чорні прямокутники під вирівнюваннями вказують на передбачені важливі функціональні сайти цих послідовностей

в *T. urartu* і *T. aestivum* не знайдено, схоже, що ген *Sr35* може виконувати свої функції, а отже низхідні посередники для нього та його гомологів мають бути однакові.

Дослідження відмінностей в послідовності нуклеотид-зв'язуючого домена білка CNL9 та його гомологів M7YFA9, W5DA03, W5D5S4, M9CJM3 здійснено так само, як і для нуклеотид-зв'язуючого домена білка RGA1e. Виявлено, що в мотивах Р-петлі, Kinase-3, RNBS-D від решти гомологів відрізняється лише білкова послідовність W5D5S4 (M35L, K136R, R260K відповідно). За мотивом GxP від своїх гомологів відрізняються послідовності M7YFA9 і CNL9 (V201I). За мотивом MHD всі послідовності-гомологи не відрізняються (рис. 3, б).

За допомогою комп'ютерних програм також відтворено просторову структуру домена білка CNL9, збагаченого на лейцинові повтори, та визначені потенційні сайти міжмолекулярних взаємодій. Для порівняння обрано гомологи W5DEL9, T1LCA9, S5FHB4, R7VYU5, M8CUP2. У результаті в сукупності виявлено 21 позицію, за якими спостерігались відмінності між послідовностями. З них у порівнянні з CNL9 найменшу кількість відмінностей мали послідовності W5DEL9 (Y148F, D151R, D160H, Q173P, D175K, P184S, S188P, H198D, P206S, W222E) та R7VYU5 (K40G, G83D, T84D, V85L, Y120H, I147V, G151R, D160H, M174S, P175K) – по 10 замін. Послідовності T1LCA9 і S5FHB4 мали по 17 замін (H42Y, G83D, T84D, M85L, Q118R, Y120D, I147V, Y148F, G151R, D160H, R178C, S173P, L174S, P175K, A188P, H198D, K200T), послідовність M8CUP2 – 15 (G83D, T84D, M85L, Q118R, Y120D, I147V, Y148F, G151R, D160H, R178C, S173P, L174S, P175K, H198D, K200T) (рис. 3, в). В середньому у порівнянні з білком CNL9 в 10 позиціях змінювалась полярність залишків гомологів та у 8 – заряд.

В результаті проведеної роботи з'ясовано, що найближчими гомологами генів *Sr33* і *Sr35* є послідовності, виявлені в *T. aestivum*. Для генів *Sr33* і *Sr35* не вдалось знайти ортологів. Знайдені гомологи до повних послідовностей білків цих генів відрізнялись від гомологів, виявлених при blast пошуку з використанням як запиту кожного з доменів окремо. Пояснити це можна дуплікаціями/делеціями кластерів

паралогічних генів, які спричиняють зсув ортологів один відносно одного, і наступною рекомбінацією між ортологами і паралогами. Результати деяких досліджень частково заперечують подібну гіпотезу [27], проте також є відомості, які її підтверджують: наявні дані про конверсію між кластерами генів паралогів різних субгеномів у кавовому дереві [28]. До того ж повідомляється, що штучна перестановка суперспіралізованих доменів між RPS5 і RPS2, які належать до класу білків CC-NB-LRR і забезпечують стійкість до *P. syringae*, не змінює активності цих білків, а лише помірно зменшує їхню здатність активувати програму клітинної загибелі [29]. Це додатково вказує на можливість урізноманітнення генів стійкості через кросинговер між паралогами. Існує також ймовірність, що швидка еволюція генів, білки яких прямо взаємодіють з ефекторами [30] і до яких можливо належать *Sr33* і *Sr35*, призводить до великої кількості змін в домені, збагаченому лейциновими повторами. Тому деякі послідовності паралогів різних видів можуть бути ближчими між собою, ніж послідовності ортологів, завдяки конвергентній еволюції під тиском одного й того ж патогена.

Більшість знайдених найближчих гомологів білка CNL9 та білка RGA1e (S5DMA6, Q6RW52 та Q84T89, W5D7H4 і W5DEL9) отримані з геному *T. aestivum*. Оскільки звичайна пшениця є донором найбільшої кількості генів стійкості до стеблової іржі (близько 20) [2] та в її поліплоїдному геномі частково представлені геноми *T. turgidum*, *A. tauschii*, *A. speloides*, а також *T. urartu* [31], це дає змогу передбачити виникнення великої кількості нових варіантів R-генів в результаті змін функцій ортологів при кожній спонтанній гібридизації, що призводило до розширення геному. Таким чином, геном звичайної пшениці *Triticum aestivum* може містити значну кількість паралогів в кластерах генів стійкості і тим самим містити гомологи як повних послідовностей, так і їх фрагменти, отримані від різних близьких родичів.

Результати детального аналізу функціональних мотивів гомологів за окремими доменами білків CNL9 та RGA1e виявили, що їхні найближчі гомологи за нуклеотид-зв'язуючим доменом (M7YFA9 та S5DMA6) не містять амінокислотних замін, які можуть призводити

до функціональних відмінностей. Проте домени, збагачені на лейцинові повтори, у найближчих гомологів містять значну кількість замін, що вказує на відсутність ортологів білків CNL9 та RGA1e.

На жаль сьогодні відсутні дані, які пояснюють, чим відрізняється помірна стійкість базового імунного захисту від помірної стійкості, яка забезпечується специфічними генами стійкості (наприклад, завдяки гену *Sr33*). До того ж в більшості досліджень не розрізняються типи гіперчутливої загибелі клітин хазяїна, хоча відомо кілька різних механізмів її виникнення [32].

Незважаючи на труднощі виявлення генів стійкості та їх аналізу, нещодавно за допомогою маркерних технологій та селекції перенесено ген *Sr43* стійкості до стеблової іржі з *Thinopyrum ponticum* до звичайної пшениці *T. aestivum*. Цей ген є ефективним проти раси Ug99. Як наслідок, було отримано дві лінії пшениці з вкороченою ділянкою чужорідного хроматину. Тепер ці лінії є джерелом важливої стійкості до стеблової іржі і можуть бути використані в подальших дослідженнях [33].

Крім цього повідомляється, що картовано ген *SrTm4* з *T. monococcum*, який теж є ефективним проти раси Ug99. З'ясовано, що цей ген розміщений на хромосомному плечі 2A^mL, а регіон, де знаходиться цей ген, є синтенічним до регіону п'ятої хромосоми *Brachypodium*. Підсумовуючи, можна зазначити тенденції продовження пошуку нових генів стійкості в диких родичах пшениці. Слід згадати, що на даний час, крім гена *SrTm4* з *T. monococcum*, до пшениці перенесено такі гени стійкості до стеблової іржі, як *Sr21*, *Sr22* і *Sr35* [34]. Проте найбільш вражаючою є нещодавня ідентифікація гена *Sr50*. Цей ген, як і широко поширений ген стійкості *Sr31*, а також нещодавно виявлений ген стійкості до стеблової іржі *SrR^{Amigo}*, знаходиться на короткому плечі першої хромосоми жита (*Secale cereale*). Встановлено, що локус з цими генами асоційований з кластером генів CC-NB-LRR, які є ортологічними до ячмінного локусу *Mla* генів стійкості до борошнистої роси. При використанні частини ячмінного гена *Mla1* як зонду для скринінгу бібліотек геному ячменю ідентифіковано кластер з семи генів CC-NB-LRR, серед яких згодом і було знайдено *Sr50* [35]. Схоже на те, що кластери таких

генів є надзвичайно важливими для стійкості проти різних патогенів, а наявність подібних локусів у пшениці та інших злаків відкриває нові можливості для досліджень стійкості та покращення зернових культур.

Очевидно, що подальші дослідження, спрямовані на уточнення молекулярної природи генів стійкості, та їх прискорення шляхом використання арсеналу методів біоінформатики дозволять уникнути трудомістких та затратних методів «сліпого» пошуку не лише генів стійкості до стеблової іржі, а й генів стійкості до інших шкідників сільськогосподарських культур.

Sr33 & *Sr35* GENE HOMOLOG IDENTIFICATION IN GENOMES OF CEREALS RELATED WITH *AEGILOPS TAUSCHII* & *TRITICUM MONOCOCCUM*

B.V. Ivaschuk, Ya.V. Pirko, A.P. Galkin, Ya.B. Blume

Institute of Food Biotechnology and Genomics, NAS of Ukraine, Kyiv

E-mail: antutilo@gmail.com, cellbio@cellbio.viaduk.net

Using bioinformatics analysis, the homologues of the genes *Sr33* and *Sr35* were identified in the genomes of *Triticum aestivum*, *Hordeum vulgare* and *Triticum urartu*. It is known that these genes provide resistance to highly virulent wheat stem rust races (Ug99). To identify important for resistance amino acid sites, the comparison of the founded homologues with the *Sr33* and *Sr35* protein sequences was performed. It was found that the sequences S5DMA6 and E9P785 are the closest homologues of RGA1e protein – a product of the *Sr33* gene, and the sequences M7YFA9 (CNL-C) and F2E9R2 are the homologues of CNL9 – a product of the gene *Sr35*. It is assumed that the homologues of the genes *Sr33* and *Sr35*, which derived from the wild relatives of wheat and barley, can provide resistance to various forms of a stem rust and can be used in the future breeding programs for wheat improvement.

ИДЕНТИФИКАЦИЯ ГОМОЛОГОВ ГЕНОВ *Sr33* И *Sr35* В КУЛЬТУРНЫХ ЗЛАКАХ, ЯВЛЯЮЩИХСЯ РОДСТВЕННИКАМИ *AEGILOPS TAUSCHII* И *TRITICUM MONOCOCCUM*

Б.В. Иващук, Я.В. Пирко, А.П. Галкин, Я.Б. Блюм

С помощью биоинформационного анализа в геномах *Triticum aestivum*, *Triticum urartu* и *Hordeum vulgare* идентифицированы гомологи генов *Sr33* и *Sr35*, ранее обнаруженных в *Aegilops tauschii* и *Triticum monococcum* и обеспечивающих устойчивость к высокопатогенной расе стеблевой ржавчины Ug99. Про-

ведено сравнение аминокислотных последовательностей белковых продуктов найденных гомологов с таковыми продуктов генов *Sr33* и *Sr35* с целью определения важных для проявления устойчивости аминокислотных участков. Установлено, что последовательности S5DMA6 и E9P785 являются ближайшими гомологами белка RGA1e – продукта гена *Sr33*, а последовательности M7YFA9 (CNL-C) и F2E9R2 гомологами белка CNL9 – продукта гена *Sr35*. Предполагается, что гомологи генов *Sr33* и *Sr35*, полученные из диких родственников пшеницы и ячменя, способны обеспечивать устойчивость к различным формам стеблевой ржавчины и могут быть использованы в дальнейших селекционных программах по улучшению пшеницы.

СПИСОК ЛИТЕРАТУРЫ

1. Khan, M.H., Bukhari, A., Dar, Z.A., and Rezi, S.M., Status and strategies in breeding for rust resistance in wheat, *Agricult. Sci.*, 2013, vol. 4, no. 6, pp. 292–301.
2. McIntosh, R.A., Wellings, C.R., and Park, R.F., *Wheat Rusts: An Atlas of Resistance Genes*, Melbourne, CSIRO, 1995.
3. Periyannan, S., Moore, J., Ayliffe, M., Bansal, U., Wang, X., Huang, L., Deal, K., Luo, M., Kong, X., Bariana, H., Mago, R., McIntosh, R., Dodds, P., Dvorak, J., and Lagudah, E., The gene *Sr33*, an ortholog of barley *Mla* genes, encodes resistance to wheat stem rust race Ug99, *Science*, 2013, vol. 341, no. 6147, pp. 786–788.
4. Sainenac, C., Zhang, W., Salcedo, A., Rouse, M.N., Trick, H.N., Akhunov, E., and Dubcovsky, J., Identification of wheat gene *Sr35* that confers resistance to Ug99 stem rust race group, *Science*, 2013, vol. 341, no. 6147, pp. 783–786.
5. Ivashuk, B.V., Samofalova, D.O., Pirko, Ya.V., Fedak, G., and Blume, Ya.B., The homologous identification of the stem rust resistance genes *Rpg5*, *Adf3* and *RGA1* in the relatives of barley, *Cytol. Genet.*, 2016, vol. 50, no. 2, pp. 96–105.
6. Narusaka, M., Kubo, Y., Hatakeyama, K., Imamura, J., Ezura, H., Nanasato, Y., Tabei, Y., Takano, Y., Shirasu, K., and Narusaka, Y., Interfamily transfer of dual NB-LRR genes confers resistance to multiple pathogens, *PLoS One*, 2013, vol. 8, no. 2, e55954.
7. Césari, S., Kanzaki, H., Fujiwara, T., Bernoux, M., Chalvon, V., Kawano, Y., Shimamoto, K., Dodds, P., Terauchi, R., and Kroj, T., The NB-LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance, *EMBO J.*, 2014, vol. 33, no. 17, pp. 1941–1959.
8. Bai, S., Liu, J., Chang, C., Zhang, L., Maekawa, T., Wang, Q., Xiao, W., Liu, Y., Chai, J., Takken, F.L., Schulze-Lefert, P., and Shen, Q.H., Structure-function analysis of barley NLR immune receptor MLA10 reveals its cell compartment specific activity in cell death and disease resistance, *PLoS Pathog.*, 2012, vol. 8, no. 6, e1002752.
9. Loutre, C., Wicker, T., Travella, S., Galli, P., Scofield, S., Fahima, T., Feuillet, C., and Keller, B., Two different CC-NBS-LRR genes are required for Lr10-mediated leaf rust resistance in tetraploid and hexaploid wheat, *Plant J.*, 2009, vol. 60, no. 6, pp. 1043–1054.
10. Lee, S.K., Song, M.Y., Seo, Y.S., Kim, H.K., Ko, S., Cao, P.J., Suh, J.P., Yi, G., Roh, J.H., Lee, S., An, G., Hahn, T.R., Wang, G.L., Ronald, P., and Jeon, J.S., Rice Pi5-mediated resistance to *Magnaporthe oryzae* requires the presence of two coiled-coil-nucleotide-binding-leucine-rich repeat genes, *Genetics*, 2009, vol. 181, no. 4, pp. 1627–1638.
11. Gutierrez, J.R., Balmuth, A.L., Ntougakis, V., Mucyn, T.S., Gimenez-Ibanez, S., Jones, A.M., and Rathjen, J.P., Prf immune complexes of tomato are oligomeric and contain multiple Pto-like kinases that diversify effector recognition, *Plant J.*, 2010, vol. 61, no. 3, pp. 507–518.
12. Eitas, T.K., and Dangl, J.L., NB-LRR proteins: pairs, pieces, perception, partners and pathways, *Curr. Opin. Plant Biol.*, 2010, vol. 13, no. 4, pp. 472–477.
13. UniProt Consortium, The universal protein resource (UniProt), *Nucl. Acids Res.*, 2008, vol. 36, iss. suppl. 1, pp. D190–D195.
14. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L., BLAST+: architecture and applications, *BMC Bioinform.*, 2009, vol. 10, e421.
15. Letunic, I., Doerks, T., and Bork, P., Smart 7: recent updates to the protein domain annotation resource, *Nucl. Acids Res.*, 2012, vol. 40, pp. D302–D305.
16. Notredame, C., Higgins, D.G., and Heringa, J., T-Coffee: A novel method for fast and accurate multiple sequence alignment, *J. Mol. Biol.*, 2000, vol. 302, no. 1, pp. 205–217.
17. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J., Jalview version 2 – a multiple sequence alignment editor and analysis workbench, *Bioinformatics*, 2009, vol. 25, no. 9, pp. 1189–1191.
18. Tamura, K., Stecher, G., Peterson, D., Flipski, A., and Kumar, S., MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0, *Mol. Biol. Evol.*, 2013, vol. 30, no. 12, pp. 2725–2729.
19. Saitou, N., and Nei, M., The neighbor-joining method: A new method for reconstructing phylogenetic trees, *Mol. Biol. Evol.*, 1987, vol. 4, no. 4, pp. 406–425.
20. Felsenstein, J., Confidence limits on phylogenies:

- An approach using the bootstrap, *Evolution*, 1985, vol. 39, no. 4, pp. 783–791.
21. Jones, D.T., Taylor, W.R., and Thornton, J.M., The rapid generation of mutation data matrices from protein sequences, *Bioinformatics*, 1992, vol. 8, no. 3, pp. 275–282.
 22. Kelley, L.A., and Stenberg, M.J., Protein structure prediction on the Web: a case study using the Phyre server, *Nat. Protoc.*, 2009, vol. 4, no. 3, pp. 363–371.
 23. Wu, S., and Zhang, Y., LOMETS: A local meta-threading-server for protein structure prediction, *Nucl. Acids Res.*, 2007, vol. 35, no. 10, pp. 3375–3382.
 24. Yang, J., Roy, A., and Zhang, A., Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment, *Bioinformatics*, 2013, vol. 29, no. 20, pp. 2588–2595.
 25. Kar, B., Nanda, S., Nayak, P.K., Nayak, S., and Joshi, R.K., Molecular characterization and functional analysis of *CzRI*, a coiled-coil-nucleotide-binding-site-leucine-rich repeat R-gene from *Curcuma zedoaria* Loeb. that confers resistance to *Pythium aphanidermatum*, *Physiol. Mol. Plant Pathol.*, 2013, vol. 83, pp. 59–68.
 26. Pan, Q., Liu, Y. S., Budai-Hadrian, O., Sela, M., Carmel-Goren, L., Zamir, D., and Fluhr, R., Comparative genetics of nucleotide binding site-leucine rich repeat resistance gene homologues in the genomes of two dicotyledons: tomato and arabidopsis, *Genetics*, 2000, vol. 155, no. 1, pp. 309–322.
 27. Rairdan, G.J., and Moffett, P., Distinct domains in the ARC region of the potato resistance protein Rx mediate LRR binding and inhibition of activation, *Plant Cell*, 2006, vol. 18, no. 8, pp. 2082–2093.
 28. Ribas, A.F., Cenci, A., Combes, M.-C., Etienne, H., and Lashermes, P., Organization and molecular evolution of disease-resistance gene cluster in coffee trees, *BMC Genom.*, 2011, vol. 12, e240.
 29. Qi, D., de Young, B.J., and Innes, R.W., Structure-function analysis of the coiled-coil and leucine-rich repeat domains of the RPS5 disease resistance protein, *Plant Physiol.*, 2012, vol. 158, no. 4, pp. 1819–1832.
 30. McHale, L., Tan, X., Koehl, P., and Michelmore, R.W., Plant NBS-LRR proteins: adaptable guards, *Genome Biol.*, 2006, vol. 7, no. 4, e212.
 31. Gong, C., Cao, S., Fan, R., Wei, B., Chen, G., Wang, X., Li, Y., and Zhang, X., Identification and phylogenetic analysis of a CC-NBS-LRR encoding gene assigned on chromosome 7B of wheat, *Int. J. Mol. Sci.*, 2013, vol. 14, no. 8, pp. 15330–15347.
 32. Dickman, M.B., and Fluhr, R., Centrality of host cell death in plant-microbe interactions, *Ann. Rev. Phytopathol.*, 2013, vol. 51, pp. 543–570.
 33. Niu, Z., Klindworth, D.L., Yu, G., Friesen, T.L., Chao, S., Jin, Y., Cai, X., Ohm, J.B., Rasmussen, J.B., Xu, S.S., Development and characterization of wheat lines carrying stem rust resistance gene *Sr43* derived from *Thinopyrum ponticum*, *Theor. Appl. Genet.*, 2014, vol. 127, no. 4, pp. 969–980.
 34. Briggs, J., Chen, S., Zhang, W., Nelson, S., Dubcovsky, J., and Rouse, M.N., Mapping of *SrTm4*, a recessive stem rust resistance gene from diploid wheat effective to Ug99, *Phytopathology*, 2015, vol. 105, no. 10, pp. 1347–1354.
 35. Mago, R., Zhang, P., and Vautrin, S., The wheat *Sr50* gene reveals rich diversity at a cereal disease resistance locus, *Nature Plants*, 2015, vol. 1, art. no. 15186.

Надійшла 03.03.15