

НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ  
ДЕРЖАВНА УСТАНОВА  
«ІНСТИТУТ ХАРЧОВОЇ БІОТЕХНОЛОГІЇ ТА ГЕНОМІКИ  
НАЦІОНАЛЬНОЇ АКАДЕМІЇ НАУК УКРАЇНИ»

Кваліфікаційна наукова  
праця на правах рукопису

**ПИДЮРА МИКОЛА ОЛЕКСАНДРОВИЧ**

УДК 575.113; 577.216

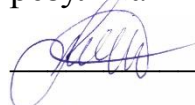
**ДИСЕРТАЦІЯ**  
**АНАЛІЗ СТРУКТУРНОЇ ОРГАНІЗАЦІЇ РОДИНИ ГЕНІВ ТУБУЛІНІВ І**  
**ЦЕЛЮЛОЗОСИНТАЗ ЛЬОНУ ТА ВЗАЄМОЗВ'ЯЗКУ ЇХ ЕКСПРЕСІЇ У**  
**ХОДІ ФОРМУВАННЯ ЦЕЛЮЛОЗНОГО ВОЛОКНА**

03.00.22 – молекулярна генетика

091– біологічні науки

Подається на здобуття наукового ступеня кандидата біологічних наук

Дисертація містить результати власних досліджень. Використання ідей,  
результатів і текстів інших авторів мають посилання на відповідне джерело.

 М. О. Пидюра

Науковий керівник: доктор біологічних наук, професор, академік НАН  
України Блюм Ярослав Борисович

Київ – 2021

## АНОТАЦІЯ

**Пидюра М.О. Аналіз структурної організації родини генів тубулінів і целюлозосинтаз льону та взаємозв'язку їх експресії у ході формування целюлозного волокна. – Кваліфікаційна наукова праця на правах рукопису.**

Дисертація на здобуття наукового ступеня кандидата біологічних наук за спеціальністю 03.00.22 – молекулярна генетика. – Державна установа «Інститут харчової біотехнології та геноміки Національної академії наук України», Київ, 2021.

У роботі визначено та охарактеризовано гени цитоскелетних білків ( $\alpha$ -,  $\beta$ -,  $\gamma$ -тубулінів і актину) та целюлозосинтаз у геномі льону. На основі структурного аналізу та філогенетичної подібності різні гени даних сімейств білків були віднесені до відповідних функціональних ізотипів. Окрім того, описано варіації довжини інтронів  $\alpha$ - і  $\beta$ -тубулінів та актину як важливі мішені для дослідження поліморфізму довжини інтронів різних видів *Linum* (метод LLP). Надалі, було виміряно рівні експресії окремих генів *CesA*,  $\alpha$ -тубуліну та актину у проростках та у стеблі на стадії швидкого росту у чотирьох сортів льону різних підвидів і зроблено висновки про функціональну роль різних ізотипів, кодованих цими генами білків. На основі отриманих результатів запропоновано стратегічні підходи для застосування методів редагування генів цитоскелетних білків, перш за все різних генів тубуліну, з метою посилення їх функціональної ролі для покращення целюлозних волокон льону.

Вперше було охарактеризовано родину генів  $\alpha$ -тубуліну льону. Показано, що вона складається з 6 генів (трьох паралогічних пар), організованих на геномному рівні, у вигляді екзонів, розділених у консервативних положеннях трьома або чотирма інтронами, залежно від відповідного класу  $\alpha$ -тубуліну – клас I і клас II, відповідно. Відповідно, дві паралогічні пари генів  $\alpha$ -тубуліну містять чотири інтрони у положеннях 37-2, 109-1, 176-0, 345-2, відносно кодону ініціації ATG, як це характерно для представників класу II тоді як третя пара містить три інтрони в положеннях 31-0, 109-1 та 233-0, як це характерно для класу I. Всі

інтрони генів  $\alpha$ -тубуліну льону мають ідентичні канонічні послідовності GT-AG у донор-акцепторних сайтах сплайсингу.

Вперше охарактеризовано родину генів  $\beta$ -тубулінів льону. Всі 14 ідентифіковані гени  $\beta$ -тубуліну льону мають три екзони, розділені двома інтронами у консервативних положеннях 131-1 та 221-1, відносно кодону ініціації ATG, які розмежовані консенсусними консервативними послідовностями GT-AG у донор-акцепторних сайтах сплайсингу. Так само, як це було показано для  $\alpha$ -тубуліну, 14 локусів  $\beta$ -тубуліну є попарно дуже гомологічними, і їх можна поєднати у п'ять пар генів-паралогів, а також *LusTub5*, який є синглетом, і триплет *LusTub 7a, b і c*. Загалом, 14  $\beta$ -тубулінових генів продукують лише 10 різних білків, довжиною від 442 до 450 амінокислот, із загальною ідентичністю послідовності 85,8%.

Було показано, що більша частина варіабельності, виявленої в амінокислотних послідовностях  $\beta$ -тубулінів, була зосереджена у невідповідному варіабельному С-кінцевому фрагменті – а.з. 431-450. У N-кінцевому домені, зокрема в межах N-петлі, між а.з. 33 і 44 *LusTub2* і *LusTub3* мають два додаткові а.з. так само, як це виявлено у більшості представників класу III у інших видів рослин, а додаткові амінокислотні залишки в цій області  $\beta$ -тубуліну сприяють вищій стабільності мікротрубочок.

Аналіз родини  $\gamma$ -тубулінів показав як і у багатьох інших рослин, наявність пари паралогічних генів, один з яких, ймовірно, втратив свою функцію. Наразі, лише локус *Lus10007851* слід розглядати як дійсний функціональний ген  $\gamma$ -тубуліну *Linum*. Ідентичність послідовності пептидного продукту *Lus10007851* та ізотипів *A. thaliana*  $\gamma 1$  та  $\gamma 2$  становить 94,26% та 92,98%, відповідно. У *Lus10007851* інтрони знаходяться в позиціях 43-0, 110-0, 147-1, 281-0, 299-0, 334-0, 359-0, 374-0, 417-0 та 452-0, відносно кодону ініціації ATG. Це збігається з положенням інтронів генів  $\gamma$ -тубуліну у *A. thaliana*, *Physcomitrella patens*, *Haplomitrium mnioides* та *Lupinus albus*. Всі інтрони генів  $\gamma$ -тубуліну льону мають ідентичні канонічні послідовності GT-AG у донор-акцепторних сайтах сплайсингу.

Вперше було визначено гени родини актинів льону та віднесено їх до різних ізотипів. Так, філогенетичний аналіз об'єднує 15 ідентифікованих генів актину льону у 7 пар генів-паралогів – сім ізотипів актину та послідовну дуплікацію одного з генів одного із ізотипів. Аналіз екзон-інтронної структури генів актину льону показав як мінливість кількості інтронів, яка може складати 2 або 3, так і поліморфізм довжини інтронів. Інтронний поліморфізм генів актину *Linum* дає можливість висловити припущення про те, що пов'язані з інтронами регуляторні елементи можуть регулювати експресію даних генів.

Показано, що поліморфізм довжини інтронів генів тубуліну і актину *Linum* свідчить про те, що пов'язані з інтронами регуляторні елементи можуть впливати на експресію цих генів, а варіації довжини інтрону 3  $\alpha$ -тубуліну, інтронів 1 і 2  $\beta$ -тубулінів та інтрону 2 актину позиціонують їх як важливі мішені для дослідження поліморфізму довжини інтронів різних видів *Linum* (метод ILP).

Під час виконання роботи було вдосконалено методики комп'ютерних високопродуктивних біоінформатичних досліджень на обчислювальних кластерах та у Грід, що сприяло значному покращенню продуктивності таких обчислень, як аналіз послідовностей та моделювання білкових молекул і їх комплексів, молекулярна динаміка, віртуальний скринінг і докінг лігандів. Також, було створено унікальну систему менеджменту структурно-біологічних даних, яка дозволила автоматизувати обробку, забезпечити накопичення, анотацію і повторне використання даних, отриманих у ході виконання досліджень.

На наступному етапі роботи використовуючи передбачені амінокислотні послідовності було проведено дизайн праймерів, які в подальшому були використані для оцінки експресії даних генів, що кодують  $\alpha$ - і  $\beta$ -тубуліни, актини і целюлозосинтази льону.

Показано, що пари генів онологів  $\alpha$ -тубуліну льону *LusTua2* переважно експресуються в апексі та флоемі, тоді як пара *LusTua3* переважно експресується у листі. Спостерігається певний рівень дисбалансу в експресії  $\alpha$ - та  $\beta$ -тубуліну, що призводить до загальної більшої кількості ізотипів *LusTua2* та *LusTua3* у апексі, листках та флоемі.

В свою чергу, для  $\beta$ -тубулінів показано, що 14 генів  $\beta$ -тубуліну льону експресуються, хоча і у різній мірі, у різних тканинах або на різних стадіях росту. Однак, обидва онологи кожної пари виявляють подібний рівень експресії у всіх аналізованих тканинах, за винятком пари генів ізотипу *LusTub4*, де є помітні відмінності між двома членами пари. У верхівці стебла, де більшість клітин активно ділиться, був виявлений високий рівень експресії для пар генів ізотипів *LusTub3* та *LusTub7*. І навпаки, рівень експресії пар генів ізотипів *LusTub2* і *LusTub5* був дуже низьким. Подібна закономірність диференціальної експресії спостерігалась також у корневих кінчиках, де гени ізотипу *LusTub7* мали найвищий рівень експресії, тоді як транскрипти ізотипів *LusTub2* були ледь помітними. Спостерігається підвищення регуляції експресії більшості генів у верхній частині стебла, а далі вниз по стеблу, експресія деяких генів  $\beta$ -тубуліну, *LusTub3a* та *b*, *LusTub5a* та *b* та *LusTub6*, була чітко регуляторно зниженою. І навпаки, регуляторне підвищення експресії *LusTub1*, 2 і 4 у верхній частині стебла зберігалось по всій довжині стебла, навіть у самих базальних його областях, де потовщення клітинної стінки було вже практично завершено. Паралоги *LusTub7* продемонстрували протилежну картину експресії – вони експресувались переважно у апексі, а у всіх інших частинах стебла експресія була регуляційно зниженою. У квіток льону було помічено, що повне дозрівання пилку під час цвітіння викликає зміни рівня експресії різних ізотипів  $\beta$ -тубуліну при сталій загальній кількості транскриптів, зі значним регуляторним підвищенням експресії пар генів *LusTub3* та 4, разом з паралельним зниженням кількості мРНК *LusTub2* та 5.

У роботі вперше визначено та охарактеризовано гени целюлозосинтаз у геномі льону. Проведений аналіз дозволив ідентифікувати 32 гени-кандидати целюлозосинтаз льону, 16 з яких кодують целюлозосинтази, а решта 16 кодують білки, подібні до целюлозосинтаз. Кількість екзонів у гені може служити критерієм точного віднесення генів льону до генів целюлозосинтаз – 11–14 екзонів та гени білків, подібних до целюлозосинтаз – 2–5 екзонів. Філогенетичний аналіз білкових продуктів ідентифікованих генів

целюлозосинтаз льону дозволив визначити шість груп, що охоплюють целюлозосинтази різних класів, а саме *CesA1/10*, *CesA3*, *CesA4*, *CesA5/6/2/9*, *CesA7* та *CesA8*. Гени, що входять до кластерів *CesA1/10* і *CesA5/6/2/9*, пов'язані з утворенням первинної клітинної стінки, тоді як гени, що відповідають за біосинтез целюлози вторинної клітинної стінки, утворюють окремі від них клади – *CesA4*, *CesA7* і *CesA8*. Аналіз 16 відібраних генів-кандидатів целюлозосинтаз льону показав, що існує принаймні 12 різних варіантів генів целюлозосинтаз у геномі льону, що представляють усі шість класів генів *CesA*.

Надалі, було виміряно рівні експресії окремих генів *CesA*,  $\alpha$ -тубуліну та актину у проростках та у стеблі на стадії швидкого росту у чотирьох сортів льону різних підвидів. Рівні експресії генів *LusCesA1*, *LusCesA4*, *LusCesA7*, *LusCesA6*, актину та TUA були виявлені у всіх чотирьох різновидах на обох стадіях. Рівні експресії досліджуваних генів значно варіювали. Загалом, рівні експресії генів *CesA* були нижчими порівняно з еталонними генами, тоді як експресія генів актину та тубуліну була сильнішою – приблизно на рівні еталонних генів.

Було показано, що рівні транскриптів генів *CesA* залежать від 1) функціонального підкласу продукованого білка, 2) стадії розвитку рослини та 3) підвиду льону, що використовується як рослинний матеріал. Гени, відповідальні за синтез вторинної клітинної стінки – *LusCesA4* та *LusCesA7*, були досить сильно експресовані в проростках, і їх експресія зростала у стеблах на стадії швидкого росту. У той же час, експресія генів, пов'язаних з синтезом первинної клітинної стінки – *LusCesA1*, *LusCesA6*, зменшилась у стеблах на стадії швидкого росту. Отже, експресія *LusCesA4*, *LusCesA7* була у 200-250 разів вища, порівняно із середньою експресією *LusCesA1* і *LusCesA6*, у стеблах на стадії швидкого росту. Також, була виявлена суттєва різниця у експресії гена *LusCesA1* між різними підвидами льону. Так, у стеблах підвидів *elongatum* і *mediterraneum* експресія гена *LusCesA1* знижується у 5,2-10,7 рази, і навпаки, експресія цього ж гена у підвиді *crepitans* не змінилася при порівнянні стадій швидкого росту і проростка. У свою чергу, середня експресія генів  $\alpha$ -тубуліну та актину є вищою

у порівнянні з середньою експресією генів целюлозосинтаз, і була стабільною між онтогенетичними стадіями та сортами льону.

Наукова новизна дисертаційної роботи полягає в тому, що вперше було ідентифіковано та охарактеризовано родини генів  $\alpha$ -,  $\beta$ - та  $\gamma$ -тубуліну, актину та целюлозосинтаз льону. Вперше проаналізовано експресію і взаємодію різних ізотипів генів даних родин льону у ході формування целюлозного волокна. Ідентифіковані послідовності на основі структури та філогенії було віднесено до різних функціональних ізотипів і досліджено їх експресію у різних тканинах на різних стадіях росту у чотирьох сортів льону різних підвидів. Продемонстровано, що у льону відкладання первинної та вторинної клітинної стінки і якість целюлозних волокон корелює з експресією певних ізотипів  $\alpha$ - і  $\beta$ -тубулінів і целюлозосинтаз.

На основі отриманих результатів запропоновано стратегічні підходи для застосування методів редагування генів цитоскелетних білків, перш за все різних генів тубуліну, з метою посилення їх функціональної ролі для покращення целюлозних волокон льону.

**Ключові слова:** *Linum usitatissimum*, льон, синтез целюлози, целюлозосинтази, тубулін, актин, цитоскелет, експресія генів, редагування геному, менеджмент даних, моделювання просторової структури.

## Список публікацій за темою дисертаційної роботи

1. Morello L., **Pydiura N.**, Galinousky D., Blume Ya., Breviario D. Flax tubulin and CesA superfamilies represent attractive and challenging targets for a variety of genome- and base-editing applications. *Funct. Integrat. Genomics*. 2020, 20(1), p. 163–176. DOI: <https://doi.org/10.1007/s10142-019-00667-2> (Здобувачем разом зі співавторами проведено дослідження, опрацьовано дані, написано статтю). **Q2**
2. **Pydiura N.**, Pirko Ya., Galinousky D., Postovoitova A., Yemets A., Kilchevsky A., Blume Ya. Genome-wide identification, phylogenetic classification, and exon-intron structure characterisation of the tubulin and actin genes in flax (*Linum usitatissimum*). *Cell Biol. Int.* 2019, 43(9), p. 1010–1019. DOI: <https://doi.org/10.1002/cbin.11001> (Здобувачем особисто проведено дослідження, опрацьовано отримані дані та написано статтю). **Q2**
3. Galinousky D., Padvitski Ts., Bayer G., Pirko Ya., **Pydiura N.**, Anisimova N., Nikitinskaya T., Khotyleva L., Yemets A., Kilchevsky A. Expression analysis of cellulose synthase and main cytoskeletal protein genes in flax (*Linum usitatissimum* L.). *Cell Biol. Int.* 2019, 43(9), p. 1065–1071. DOI: <https://doi.org/10.1002/cbin.10837> (Здобувачем разом зі співавторами проведено дослідження, опрацьовано отримані дані та написано статтю). **Q2**
4. Блюм, Я.Б., Ємець А.І., Пірко Я.В., Баєр Г.Я., Рабоконь А.М., Буй Д.Д., **Пидюра М.О.**, Галіновський Д.В., Анісімова Н.В., Хотильова Л.В., Кільчевський А.В. Оцінка експресії різних генів β-тубуліну і хітиназо-подібних генів при біогенезі волокон льону. *Анотований збірник проектів спільного конкурсу ДФФД-БРФФД*, 2017, Академперіодика: К., с. 66–69. (Здобувачем разом зі співавторами проведено дослідження, опрацьовано отримані дані та написано статтю).
5. Демчук О.М., Карпов П.А., Ожерєдов С.П., Співак С.І., **Пидюра М.О.**, Самофалова Д.О., Блюм Я.Б. Створення репозиторія моделей просторової



- структури тубулінів як реалізація одного із стратегічних завдань віртуальної організації CSLabGrid. *Фактори експериментальної еволюції організмів*. 2016, 19, с. 208–211. (Здобувачем разом зі співавторами проведено дослідження, опрацьовано отримані дані та написано статтю).
6. Karpov P.A., Brytsun V.M., Rayevsky A.V., Demchuk O.M., **Pydiura N.O.**, Ozheredov S.P., Samofalova D.A., Spivak S.I. , Yemets A.I., Kalchenko V.I., Blume Ya.B. High-throughput screening of new antimetabolic compounds based on CSLabGrid virtual organization. *Sci. Innov.* 2015, 11(1), p. 85–93. DOI: <https://doi.org/10.15407/scine11.01.085>) (Здобувачем разом зі співавторами проведено дослідження, опрацьовано отримані дані та написано статтю).
  7. **Pydiura N.A.**, Bayer G.Ya., Galinovsky D.V., Yemets A.I., Pirko Ya.V., Padvitski T.A., Anisimova N.V., Khotyleva L.V., Kilchevsky A.V., Blume Ya.B. Bioinformatic search for cellulose synthase genes in flax (*Linum usitatissimum*) and their phylogenetic analysis. *Cytol. Genetics.* 2015, 49(5), p. 279–287. DOI: <https://doi.org/10.3103/S0095452715050084> (Здобувачем разом зі співавторами проведено дослідження, опрацьовано отримані дані та написано статтю).
  8. Баєр Г.Я., **Пидюра М.О.**, Пірко Я.В., Ємець А.І., Блюм Я. Б. Біоінформаційний пошук послідовностей генів, що кодують тубуліни у геномі льону *Фактори експериментальної еволюції організмів*. 2014, 14, с. 14–17. (Здобувачем разом зі співавторами проведено дослідження, опрацьовано дані та написано статтю).
  9. **Pydiura N.**, Karpov P., Blume Ya. On the efficiency of CPU and hybrid CPU-GPU systems in computational biology tasks. *Comput. Sci. Appl.* 2014, 1(1), p. 48–59. (Здобувачем разом зі співавторами проведено дослідження, опрацьовано отримані дані та написано статтю).
  10. **Pydiura N.**, Karpov P., Blume Ya. Hybrid CPU-GPU calculations – a promising future for computational biology. Third Int. Conf. "High Performance Computing" (Kyiv, Ukraine, October 7–11, 2013), Nat. Technical Univ. of Ukraine «Kyiv Polytechnic Institute», p. 330–335. DOI: <https://doi.org/10.13140/2.1.1709.5366>

*(Здобувачем разом зі співавторами проведено дослідження, опрацьовано отримано дані та написано статтю).*

11. **Pydiura N.**, Karpov P., Blume Ya. Hardware environment for CSLabGrid: Reaching maximum efficacy of computations in structural biology and bioinformatics. In: Cluster Computing: Proc. II Int. Conf. (Lviv, Ukraine, June 3–5, 2013), Lviv Polytechnic Nat. Univ., p. 191–194. *(Здобувачем разом зі співавторами проведено дослідження, опрацьовано отримано дані та написано статтю).*
12. Галиновский Д.В., Подвицкий Т.А., Анисимова Н.В., **Пыдюра Н.А.**, Баер Г.Я., Емец А.И., Пирко Я.В., Блюм Я.Б., Хотылева Л.В., Кильчевский А.В. Экспрессия генов цитоскелета и целлюлозосинтезирующего комплекса клетки при формировании волокон льна. Технология органических веществ. Тезисы докл. 79-й научно-техн. конф-ции профессорско-преподавательского состава, научных сотрудников и аспирантов (Минск, Беларусь, 2–6 февраля 2015), Белорусский гос. технологический университет: Минск, 2015; с. 6.
13. **Pydiura N.**, Karpov P., Blume Y. COMODORE: a web system for high-throughput computational investigations in Grid environment. On-line Proc. of the 9th Int. Symposium on Integrative Bioinformatics, Gatersleben, Germany (March 18–20, 2013), Leibniz Inst. Plant Genetics and Crop Plant Research (IPK), 2013; pp 190–191.
14. **Pydiura N.A.**, Karpov P.A., Blume Y.B. The comparison of colchicine, vinblastine and taxol binding sites on animal and plant tubulins. III Int. Symposium «Intracellular Signaling and Bioactive Molecules Design» (Lviv, Ukraine, September 17–23, 2012), Ivan Franko Nat. University of Lviv, 2012, p. 129.
15. **Pydiura N.A.**, Karpov P.A., Blume Y.B. Bioinformatic search and phylogenetic analysis of the plant-specific MAPs in genomes of monocots and dicots. 8th Int. Conference on the Bioinformatics of Genome Regulation and Structure/Systems Biology (Novosibirsk, Russia, June 25–29, 2012), Institute of Cytology and Genetics, Novosibirsk, p. 260.

16. **Pydiura N.**, Karpov P., Blume Ya. Design of specific cytoskeleton related biological database and data management environment for bioinformatical cytoskeleton investigation and collaboration within virtual grid-organisation. Proc. Int. Moscow Conf. on Comput. Mol. Biol. (21–24 July, 2011, Moscow, Russia), НИИ Физико-химической биологии им. А. Н. Белозерского МГУ; p. 297–298.
17. **Pydiura N.A.**, Karpov P.A., Blume Ya.B. Modeling of the regulation of microtubule dynamic instability by row of associated proteins. 2nd IRB Barcelona PhD Student Symposium: Life in Motion: Dynamics of Molecules and Systems (Barcelona, Spain, November 17–18, 2011) Inst. Res. Biomedicine, Barcelona, p. 82–83.
18. **Пидюра М.О.**, Карпов П.А. Біоінформаційне дослідження білків, асоційованих з мікротрубочками вищих рослин та аналіз їх подібності. Збірник тез: Перша конф-ція мол. учених «Біологія рослин та біотехнологія» (5–7 жовтня 2011, Біла Церква), 2011, Київ: Ін-т харчової біотехнології та геноміки, с. 69.

## SUMMARY

**Pydiura M.O. Analysis of the structural organization of the family of tubulin and cellulose synthase genes and the relationship between their expression during the formation of cellulose fiber.**

Thesis for a Candidate degree in Biological Sciences (Ph.D) in speciality 03.00.22 – Molecular genetics. – Institute of Food Biotechnology and Genomics of the National Academy of Sciences of Ukraine, Kyiv, 2021.

The genes of cytoskeletal proteins ( $\alpha$ -,  $\beta$ -,  $\gamma$ -tubulins and actin) and cellulose synthases in the flax genome were determined and characterized. Based on structural analysis and phylogenetic similarity, different genes of these protein families were assigned to the corresponding functional isotypes. In addition, variations in the length of introns of  $\alpha$ - and  $\beta$ -tubulins and actin are described as important targets for the study of intron length polymorphism of different *Linum* species (ILP method). Furthermore, the expression levels of individual genes of CesA,  $\alpha$ - and  $\beta$ -tubulin and actin in seedlings and stems at the stage of rapid growth in four varieties of flax of different subspecies were measured and conclusions were made about the functional role of different isotypes encoded by these genes. Based on the obtained results, strategic approaches for the application of methods of editing genes of cytoskeletal proteins, especially various tubulin genes, are proposed in order to strengthen their functional role for the improvement of flax cellulose fibers.

The flax  $\alpha$ -tubulin gene family was characterized for the first time. It is shown that it consists of 6 genes (three paralogic pairs), organized at the genomic level, in the form of exons separated in conservative positions by three or four introns, depending on the corresponding class of  $\alpha$ -tubulin – class I and class II, respectively. Accordingly, two paralogic pairs of  $\alpha$ -tubulin genes contain four introns at positions 37-2, 109-1, 176-0, 345-2, relative to the codon of initiation of ATG, as is typical for representatives of class II while the third pair contains three introns in positions 31-0, 109-1 and 233-0, as is typical for class I. All introns of flax  $\alpha$ -tubulin genes have identical canonical GT-AG sequences at the donor-splice acceptor sites.

The flax  $\beta$ -tubulin gene family was characterized for the first time. All 14 identified flax  $\beta$ -tubulin genes have three exons separated by two introns at conserved positions 131-1 and 221-1, relative to the ATG initiation codon, which are delimited by consensus conserved GT-AG sequences at donor-acceptor splice sites. As shown for  $\alpha$ -tubulin, the 14  $\beta$ -tubulin loci are pairwise very homologous and can be combined into five pairs of paralog genes, as well as *LusTub5*, which is a singlet, and the *LusTub7a*, *b* and *c* triplet. In total, 14  $\beta$ -tubulin genes produce only 10 different proteins, ranging in length from 442 to 450 a. a., with an overall sequence identity of 85.8%.

It was shown that most of the variability found in the amino acid sequences of  $\beta$ -tubulins was concentrated in the disordered variable C-terminal fragment – a.a. 431-450. In the N-terminal domain, in particular within the N-loop, between residues 33 and 44 *LusTub2* and *LusTub3* have two additional a.a. as found in most representatives of class III in other plant species, and additional amino acid residues in this area of  $\beta$ -tubulin contribute to higher stability of microtubules.

Analysis of the  $\gamma$ -tubulin family showed as in many other plants the presence of a pair of paralogous genes, one of which probably lost its function. Currently, only the *Lus10007851* locus should be considered as a true functional gene of  $\gamma$ -tubulin *Linum*. The sequence identity of the peptide product *Lus10007851* and the isotypes of *A. thaliana*  $\gamma 1$  and  $\gamma 2$  is 94.26% and 92.98%, respectively. In *Lus10007851*, the introns are at positions 43-0, 110-0, 147-1, 281-0, 299-0, 334-0, 359-0, 374-0, 417-0, and 452-0 relative to the ATG initiation codon. This coincides with the position of the introns of the  $\gamma$ -tubulin genes in *A. thaliana*, *Physcomitrella patens*, *Haplomitrium mnioides* and *Lupinus albus*. All introns of flax  $\gamma$ -tubulin genes have identical canonical GT-AG sequences at donor-acceptor splice sites.

For the first time, the genes of the flax actin family were identified and assigned to different isotypes. Thus, phylogenetic analysis combines 15 identified flax actin genes into 7 pairs of paralog genes – seven actin isotypes and sequential duplication of one of the genes of one of the isotypes. Analysis of the exon-intron structure of flax actin genes showed both variability in the number of introns, which may be 2 or 3, and a polymorphism of intron lengths. The intron polymorphism of *Linum* actin genes

suggests that intron-related regulatory elements may regulate the expression of these genes.

It is shown that the polymorphism of the intron length of tubulin and actin *Linin* genes indicates that intron-related regulatory elements can affect the expression of these genes, and variations in the length of intron 3 of  $\alpha$ -tubulin, introns 1 and 2 of  $\beta$ -tubulin and intron 2 of actin position them as important targets for the study of intron length polymorphism of different species of *Linum* (ILP method).

During the work, the methods of computer high-performance bioinformatics research on computational clusters and in Grid were improved, which significantly improved the performance of such calculations as sequence analysis and modeling of protein molecules and their complexes, molecular dynamics, virtual screening and ligand docking. Also, a unique management system of structural biological data was created, which allowed to automate the processing, ensure the accumulation, annotation and reuse of data obtained during the research.

In the next step, using the predicted amino acid sequences, primers were designed, which were then used to assess the expression of these genes encoding  $\alpha$ - and  $\beta$ -tubulins, actins and cellulose cellulose synthase.

It has been shown that *LusTua2* flax  $\alpha$ -tubulin gene pairs are predominantly expressed in the apex and phloem, whereas the *LusTua3* pair is predominantly expressed in the leaf. There is a certain level of imbalance in the expression of  $\alpha$ - and  $\beta$ -tubulin, which leads to an overall greater number of isotypes of *LusTua2* and *LusTua3* in the apex, leaves and phloem.

In turn, for  $\beta$ -tubulins, it has been shown that 14 flax  $\beta$ -tubulin genes are expressed, albeit to different degrees, in different tissues or at different stages of growth. However, both ohnologs of each pair show a similar level of expression in all analyzed tissues, except for a pair of genes of the *LusTub4* isotype, where there are noticeable differences between the two members of the pair. At the apex of the stem, where most cells are actively dividing, a high level of expression was found for pairs of *LusTub3* and *LusTub7* isotype genes. Conversely, the expression level of *LusTub2* and *LusTub5* isotype gene pairs was very low. A similar pattern of differential

expression was also observed in the root tips, where the genes of the *LusTub7* isotype had the highest level of expression, while the transcripts of the *LusTub2* isotypes were barely noticeable. There is an increase in the regulation of the expression of most genes in the upper part of the stem, and further down the stem, the expression of some genes of  $\beta$ -tubulin, *LusTub3a* and *b*, *LusTub5a* and *b* and *LusTub6*, was clearly regulatory reduced. Conversely, the regulatory increase in *LusTub1*, 2, and 4 expression in the upper part of the stem persisted along the entire length of the stem, even in its most basal areas, where the thickening of the cell wall was almost complete. *LusTub7* paralogs showed the opposite pattern of expression – they were expressed mainly in the apex, and in all other parts of the stem the expression was regulatory reduced. In flax flowers, full maturation of pollen during flowering was observed to cause changes in the expression levels of different  $\beta$ -tubulin isotypes with a constant total number of transcripts, with a significant regulatory increase in *LusTub3* and 4 gene pair expression, along with a parallel decrease in *LusTub2* and 5 mRNA.

For the first time, cellulose synthase genes in the flax genome were identified and characterized. The analysis allowed us to identify 32 candidate genes for flax cellulose synthase, 16 of which encode cellulose synthases, and the remaining 16 encode proteins similar to cellulose synthases. The number of exons in the gene can serve as a criterion for the exact assignment of flax genes to cellulose synthase genes – 11–14 exons and genes of proteins similar to cellulose synthases – 2–5 exons. Phylogenetic analysis of protein products of identified flax cellulose synthase genes revealed six groups covering cellulose synthases of different classes, namely *CesA1/10*, *CesA3*, *CesA4*, *CesA5/6/2/9*, *CesA7* and *CesA8*. Genes belonging to the clusters *CesA1/10* and *CesA5/6/2/9* are associated with the formation of the primary cell wall, while the genes responsible for the biosynthesis of cellulose of the secondary cell wall, form separate clades – *CesA4*, *CesA7* and *CesA8*. Analysis of 16 selected flax cellulose synthase candidate genes showed that there are at least 12 different variants of cellulose synthase genes in the flax genome, representing all six classes of *CesA* genes.

Subsequently, the expression levels of individual Cesa,  $\alpha$ -tubulin and actin genes in seedlings and in the stem at the stage of rapid growth in four varieties of flax of different subspecies were measured. Levels of *LusCesA1*, *LusCesA4*, *LusCesA7*, *LusCesA6*, actin and *TUA* genes were detected in all four species at both stages. The expression levels of the studied genes varied significantly. In general, Cesa gene expression levels were lower compared to reference genes, while actin and tubulin gene expression was stronger at about the level of reference genes.

It has been shown that the levels of Cesa gene transcripts depend on 1) the functional subclass of the produced protein, 2) the stage of plant development and 3) the flax subspecies used as plant material. The genes responsible for the synthesis of the secondary cell wall – *LusCesA4* and *LusCesA7*, were quite strongly expressed in seedlings, and their expression increased in the stems at the stage of rapid growth. At the same time, the expression of genes associated with the synthesis of the primary cell wall – *LusCesA1*, *LusCesA6*, decreased in the stems at the stage of rapid growth. Thus, the expression of *LusCesA4*, *LusCesA7* was 200-250 times higher, compared to the average expression of *LusCesA1* and *LusCesA6*, in the stems at the stage of rapid growth. Also, a significant difference in *LusCesA1* gene expression was found between different subspecies of flax. Thus, in the stems of the subspecies *elongatum* and *mediterraneum*, the expression of the *LusCesA1* gene is reduced by 5.2-10.7 times, and conversely, the expression of the same gene in the subspecies *crepitans* did not change when comparing the stages of rapid growth and seedling. In turn, the mean expression of  $\alpha$ -tubulin and actin genes was higher than the mean expression of cellulose synthase genes, and was stable between ontogenetic stages and flax varieties.

The scientific novelty of the dissertation is that for the first time the families of genes  $\alpha$ -,  $\beta$ - and  $\gamma$ -tubulin, actin and cellulose synthase of flax were identified and characterized. For the first time, the expression and interaction of different isotypes of genes of these flax families during the formation of cellulose fiber were analyzed. Identified sequences based on structure and phylogeny were assigned to different functional isotypes and their expression in different tissues at different stages of growth in four varieties of flax of different subspecies was studied. It has been shown that in



flax the deposition of primary and secondary cell walls and the quality of cellulose fibers correlate with the expression of certain isotypes of  $\alpha$ - and  $\beta$ -tubulins and cellulose synthases.

Based on the obtained results, strategic approaches for the application of methods of editing genes of cytoskeletal proteins, especially various tubulin genes, are proposed in order to strengthen their functional role for the improvement of flax cellulose fibers.

**Key words:** *Linum usitatissimum*, flax, cellulose synthesis, cellulose synthase, tubulin, actin, cytoskeleton, gene expression, genome editing, data management, structural modelling.

## ЗМІСТ

СПИСОК УМОВНИХ СКОРОЧЕНЬ .....	20
ВСТУП.....	23
РОЗДІЛ 1 .....	30
ОГЛЯД ЛІТЕРАТУРИ .....	30
1.1 Льон як джерело високоякісних текстильних та композитних волокон....	30
1.2 Геномні ресурси, присвячені льону .....	32
1.3 Цитоскелетні білки .....	33
1.3.1 Родина генів тубулінів.....	33
1.3.2 Родина генів актинів .....	37
1.3.3 Білки асоційовані з мікротрубочками.....	39
1.4 Родина генів целюлозосинтаз та целюлозосинтазо подібних білків льону	40
1.5 Синтез целюлози. Тубуліни і CesA: дві сімейства генів, що сприяють відкладенню клітинної стінки.....	45
1.6 Менеджмент біологічних даних .....	48
1.6.1 Менеджмент структурно-біологічних та генетичних даних генів білків пов'язаних з цитоскелетом та генів целюлозосинтаз.....	52
РОЗДІЛ 2 .....	55
МАТЕРІАЛИ ТА МЕТОДИ ДОСЛІДЖЕННЯ .....	55
2.1 Біоінформатичні методи досліджень .....	55
2.2 COMODORE – система менеджменту структурно-біологічних даних.....	57
2.3 Менеджмент структурно-біологічних та генетичних даних генів білків пов'язаних з цитоскелетом та генів целюлозосинтаз.....	61
2.4 Аналіз експресії генів .....	66
РОЗДІЛ 3 .....	70
ПОШУК ТА АНАЛІЗ ГЕНІВ ЦИТОСКЕЛЕТНИХ БІЛКІВ У ГЕНОМІ ЛЬОНУ .....	70
3.1 Пошук генів тубулінів у геномі льону .....	70
3.2 Пошук генів актинів у геномі льону .....	73

3.3 Біоінформаційне дослідження білків, асоційованих з мікротрубочками вищих рослин та аналіз їх подібності .....	75
3.4 Аналіз генів цитоскелетних білків у геномі льону .....	77
3.4.1 $\alpha$ -тубуліни .....	77
3.4.2 $\beta$ -тубуліни .....	80
3.4.3 $\gamma$ -тубуліни .....	86
3.4.4 Актини .....	87
3.4.5 Висновки .....	92
3.5 Моделювання рослинних тубулінів за гомологією .....	93
3.6 Структурні особливості цитоскелетних білків льону як основа їх функцій .....	101
3.6.1 Тубуліни $\alpha$ , $\beta$ та $\gamma$ .....	102
3.6.2 МАРи які стабілізують мікротрубочки .....	106
3.6.3 МАРи, які дестабілізують мікротрубочки .....	108
3.7 Аналіз експресії цитоскелетних білків льону .....	109
3.7.1 $\alpha$ -тубуліни .....	109
3.7.2 $\beta$ -тубуліни .....	111
РОЗДІЛ 4 .....	113
ПОШУК ТА АНАЛІЗ ГЕНІВ ЦЕЛЮЛОЗОСИНТАЗ У ГЕНОМІ ЛЬОНУ .....	113
4.1 Пошук генів целюлозосинтаз у геномі льону .....	113
4.2 Аналіз генів целюлозосинтаз льону .....	117
4.3 Аналіз експресії генів целюлозосинтаз льону .....	125
4.4 Гени суперсімейств тубулінів та целюлозосинтаз льону як перспективні мішені для покращення кількості та якості целюлозних волокон біотехнологічними методами .....	131
УЗАГАЛЬНЕННЯ .....	140
ВИСНОВКИ .....	147
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ .....	150
Список публікацій здобувача .....	177

## СПИСОК УМОВНИХ СКОРОЧЕНЬ

CesA	– целюлозосинтаза (Cellulose Synthase)
CSC	– целюлозосинтезуючий комплекс (Cellulose Synthase Complex)
CSL	– целюлозосинтазо подібні білки (Cellulose Synthaselike)
CW	– клітинна стінка (Cell Wall)
PCW	– первинної клітинної стінки (Primary Cell Wall)
SCW	– вторинної клітинної стінки (Secondary Cell Wall)
MYA	– мільйонів років тому (Million Years Ago)
WGD	– повногеномна дуплікація генома (Whole-Genome Duplication)
MT	– мікротрубочки (Microtubule)
c-MT	– кортикальні мікротрубочки (c-MT)
TUA	– тубулін $\alpha$ (tubulin $\alpha$ )
TUB	– тубулін $\beta$ (tubulin $\beta$ )
TUG	– тубулін $\gamma$ (tubulin $\gamma$ )
ACT	– актин (actin)
БAM / MAP	– білок, асоційований з мікротрубочками (Microtubule-Associated Protein)
н. п.	– нуклеотидна пара (bp, base pair)
а. з. / а. а.	– амінокислотний залишок (amino acid)
СТТ	– С-кінцевий хвіст (C-Terminal Tail)
PTM	– посттрансляційна модифікація (Post-Translational Modification)
UTR	– нетранслюємі області (Untranslated Regions)
IME	– інтрон-медійоване посилення експресії (Intron-Mediated Enhancement)
ILP	– поліморфізм довжин інтронів (Intron Length Polymorphisms)
HVR	– гіперваріабельний регіон або домен (Hypervariable Region)
CCD	– Центральний цитоплазматичний домен (Central Cytoplasmic Domain)
CSR	– клас-специфічний регіон (Class-Specific Region)

P-CRs	– рослинно-консервативні регіони (Plant-Conserved Regions)
VO / VO	– віртуальна організація (Virtual Organisation)
CDS	– кодуюча послідовність (Coding Sequence)
PDB	– Protein Data Bank
GCP	– білки гама-тубулінового комплексу (Gamma-Tubulin Complex Proteins)
SAP	– одиничний амінокислотний поліморфізм (Single Amino-Acid Polymorphism)
SNP	– поліморфізм одного нуклеотиду (Single Nucleotide Polymorphism)
NCBI	– National Center for Biotechnology Information
CG	– Conjugate Gradient
RMSD	– середньоквадратичне відхилення (Root-Mean-Square Deviation)
MTOC	– центри організації мікротрубочок (Microtubule Organizing Centers)
CFP	– синій флуоресцентний білок (Cyan Fluorescent Protein)
ssODN	– одноланцюговий олігодезоксинуклеотид (Single-Stranded Oligodeoxynucleotides)
CRISPR	– короткі паліндромічні кластерні повтори (Clustered Regularly Interspaced Short Palindromic Repeats)
Cas	– білок, асоційований з CRISPR (CRISPR Associated Protein)
HDR	– репарація на основі гомології (Homology Directed Repair)
EPSPS	– Enolpyruvylshikimate-3-phosphate Synthase
sgRNA	– одинарна направляюча РНК (Single Guide RNA)
ZFN	– ендонуклеаза цинкових пальців (Zinc Finger Nuclease)
TALEN	– транскрипційно-активаторна ефекторна нуклеаза (Transcription Activator-like Effector Nuclease)
PAM	– мотив, дотичний до протоспейсеру (Protospacer adjacent motif)

- NHEJ – негомологічне з'єднання кінців (Non-Homologous End Joining)
- SV – секреторний везикул (secretory vesicle)
- CSI – білок-посередник целюлозосинтаз (Cellulose Synthase Interactive Protein)

## ВСТУП

**Обґрунтування вибору теми дослідження.** Льон, як одна з найдавніших луб'яних культур, є унікальною модельною системою для вивчення процесів росту та диференціації рослинних клітин, а також формування клітинної стінки. Мікрофібрили целюлози синтезуються мультиферментними целюлозосинтезуючими комплексами (CSC), які складаються з 18-36 субодиниць *CesA* (Doblin et al., 2002; Fernandes et al., 2011; Newman et al., 2013), що після доставки до плазматичної мембрани за допомогою везикул, які рухаються по тяжках F-актину, під час синтезу целюлози рухаються вже вздовж кортикальних мікротрубочок (Wightman and Turner, 2008; Watanabe et al., 2015). Синтез целюлози відбувається на плазматичній мембрані, а комплекси CSC приєднуються і переміщуються вздовж кортикальних мікротрубочок за допомогою проміжних білків, які взаємодіють з тубулінами та актином (Bringmann et al., 2012), що підтверджує важливість їх взаємодії у процесі біосинтезу клітинної стінки. Відповідно, вивчення експресії цих генів у різних сортів льону які продукують волокна різної якості може допомогти виявити фактори, що визначають якість лляного волокна та можливі механізми біотехнологічних модифікацій.

Також, на даний момент немає даних про конкретні білок-білкові взаємодії між відповідними ізо типами тубуліну, актину та целюлозосинтаз. Вирішення цього питання є особливо цікавим, оскільки сімейства даних генів складаються з декількох членів, що групуються у різні класи, де паралоги можуть демонструвати досить різний рівень експресії у різних тканинах та на різних стадіях росту (Breviario et al., 2013; Ermawar et al., 2015; Dhugga, 2001). Слід було б справедливо розраховувати, що спираючись на доступну геномну інформацію, можна ефективно використати технології рекомбінантних ДНК для того, щоб отримати хоча б частину відповідей на ці питання.

Льон має довгу історію генетичної трансформації, починаючи з кінця 80-х років минулого століття. Вміст жирних кислот і склад насіння, вміст лігніну,

якість волокон, толерантність до гербіцидів та стійкість до хвороб належать, серед інших, до властивостей льону, які вже були модифіковані за допомогою генної інженерії (Ludvíková and Griga, 2015). Незважаючи на це, до цього часу в Канаді зареєстровано лише один трансгенний сорт льону (McHughen, 1989; McHughen et al., 1997), що характеризується парадигматичною історією, яка призвела до його зняття з реєстрації в 2001 р. та тривалого зниження загального експорту льону Канадою (Ryan and Smyth, 2012). Звичайно, ці події слугували однією з причин зменшення зусиль, що могли би бути прикладені генетичними інженерами для поліпшення процесів формування целюлозного волокна у льону.

Цей негативний досвід, швидше за все, може бути усунутий завдяки впровадженню та використанню технологій редагування геномів, які дозволяють модифікувати ендogenous гени, спрямовані на сайт, без введення чужорідних генів. Модифіковані рослини, створені у такий спосіб, швидше за все, будуть прийняті тими країнами, які негативно ставляться до трансгенних культур, і це має відкрити нові перспективи для поліпшення агрономічних ознак у культурних рослин, в тому числі і льону. Окрім свого потенціалу для підвищення врожайності сільськогосподарських культур, редагування геному в даний час є одним з найкращих інструментів для проведення зворотної генетики, і воно виступає як особливий універсальний інструмент для вивчення базової біології рослин.

Тому виходячи з актуальності цієї проблеми необхідно було провести повний аналіз генів, що кодують основні білки цитоскелету (тубуліну та актин), та генів *CesA*, оцінити закономірності їх експресії у різних тканинах та підвидах льону з метою з'ясування можливих взаємозв'язків білкових продуктів цих генів у процесах формування целюлозного волокна і запропонувати нові підходи для їх регулювання шляхом редагування геному.

**Зв'язок роботи з науковими програмами, планами, темами, грантами.** Дисертаційна робота виконувалась в рамках бюджетних НДР відділу геноміки та молекулярної біотехнології ДУ «Інститут харчової біотехнології та геноміки НАН України» «Використання Грід для високопропускового скринінгу біологічно



активних речовин з підвищеною спорідненістю до тубулінів як базової складової реалізації програмних засад віртуальної організації CSLabGrid» (№ ДР 0113U004047, 2011–2013рр.); «Вивчення цитоскелету як критичної мішені для розробки нових агробіотехнологій та пошуку біологічно активних речовин за допомогою засобів геноміки та біоінформатики» (№ ДР 0110U001224, 2010–2014 рр.); «Геноміка та клітинна біологія цитоскелету рослин як інструмент для вивчення його структури і функцій та розвитку нових біотехнологій» (№ ДР 0115U002084, 2015–2019 рр.); «Створення молекулярно-генетичних маркерів для диференціації різних генотипів рослин на основі вивчення поліморфізму інтронів генів їх цитоскелетних білків» (№ ДР 0115U005025, 2017–2019 рр.).

**Мета та завдання дослідження.** Мета дисертаційної роботи – визначити можливі взаємодії між генами суперсімейств цитоскелетних білків –  $\alpha$ -,  $\beta$ - і  $\gamma$ -тубулінів та актину, а також целюлозосинтаз льону в процесі синтезу клітинної стінки як перспективні мішені для покращення кількості та якості целюлозних волокон за допомогою біотехнологічних методів, запропонувавши на основі цих даних стратегії редагування геному з метою покращення якості сировини (целюлозного волокна) цієї технічної культури.

Для досягнення мети поставлено такі завдання:

1. Визначити гени  $\alpha$ -,  $\beta$ - і  $\gamma$ -тубуліну та актинів, а також целюлозосинтаз у геномі льону;
2. Охарактеризувати особливості структурної організації генів  $\alpha$ -,  $\beta$ - і  $\gamma$ -тубуліну та актину, а також целюлозосинтаз льону;
3. Описати досліджувані гени (їх білкові продукти) льону до різних функціональних ізотипів по аналогії до ізотипів генів цитоскелетних білків та целюлозосинтаз у геномі *A. thaliana* та провести філогенетичний аналіз;
4. Провести диференційований аналіз рівнів експресії генів  $\alpha$ - і  $\beta$ -тубуліну та актину у різних підвидах та на різних стадіях росту рослин льону;
5. Провести диференційований аналіз рівнів експресії різних ізотипів генів *CesA* у різних підвидах та на різних стадіях росту рослин льону;

6. На основі отриманих генетичних даних та даних по експресії запропонувати стратегії редагування геному, які дозволили би поглибити знання про синтез целюлози і покращити якість продукції целюлозного волокна льону.

*Об'єктом* дослідження є генетичні механізми, що регулюють механізм синтезу целюлози на різних стадіях росту рослини за рахунок модулювання експресії різних ізотипів цитоскелетних білків та білків *CesA*.

*Предмет* дослідження – гени суперсімейств тубулінів  $\alpha$ ,  $\beta$  і  $\gamma$ , актинів та целюлозосинтаз льону, особливості та механізми регуляції експресії різних ізотипів даних білків у різних підвидах на різних стадіях розвитку рослини.

**Методи дослідження:** Пошук генів  $\alpha$ -,  $\beta$ -,  $\gamma$ -тубулінів, актину, целюлозосинтаз льону і білків, подібних до целюлозосинтаз у геномі льону, множинне вирівнювання послідовностей, аналіз екзон-інтронної структури, визначення ступеня філогенетичної спорідненості, здійснювали за допомогою біоінформатичних методів. Побудова структурних моделей білків цитоскелету та їх аналіз здійснювались структурно-біологічними методами. Для аналізу експресії генів тубулінів, актину і целюлозосинтаз використовувались молекулярно-генетичні методи: екстракція та очищення РНК та синтез кДНК, полімеразна ланцюгова реакція в режимі реального часу, аналіз експресії за допомогою методу  $\Delta\Delta Ct$ .

**Наукова новизна отриманих результатів.** Вперше описано родини генів  $\alpha$ -,  $\beta$ -,  $\gamma$ -тубулінів, актину, целюлозосинтаз льону і білків, подібних до целюлозосинтаз і віднесено їх до різних функціональних ізотипів. Побудовано структурні моделі рослинних тубулінів та проаналізовано центри взаємодії з агентами, що впливають на (де)полімеризацію мікротрубочок. Показано, що варіації довжини інтрону 3  $\alpha$ -тубуліну, інтронів 1 і 2  $\beta$ -тубуліну та інтрону 2 актину льону можуть зробити їх життєздатними мішенями для дослідження поліморфізмів довжини інтронів різних видів *Linum* за допомогою оцінки поліморфізму довжини їх інтронів (ILP). Визначено рівні експресії генів *CesA*, тубуліну та актину на двох стадіях росту у чотирьох сортів культивованого льону, що належать до різних підвидів. Було показано, що рівні експресії генів

*CesA* залежать від 1) функціонального підкласу продукованого білка, 2) стадії розвитку рослини та 3) підвиду льону, що використовується як рослинний матеріал.

**Практичне значення отриманих результатів.** Досліджено гени суперсімейств цитоскелетних білків – тубулінів і актину, та целюлозосинтаз льону, що в подальшому дозволяє дослідити механізми відкладання целюлозних волокон у різних підвидах льону за рахунок експресії різних ізотипів білків даних сімейств. Для менеджменту біологічних даних, отриманих в результаті структурного моделювання, створено базу даних цитоскелетних білків CSMoDB (<http://csmodb.ifbg.org.ua/comodore/>), присвячену виключно білкам цитоскелету. Розроблено стратегії редагування геному для розкриття ролі та взаємодії між членами надродини тубуліну та *CesA*, які можуть сприяти відкладанню клітинної стінки у льону, та покращення кількості та якості целюлозного волокна за допомогою біотехнологічних методів.

**Особистий внесок здобувача.** Результати, викладені у дисертації, автор отримав особисто або за безпосередньої участі у виконанні експериментів. Здобувачем спільно з науковим керівником академіком НАН України, д. б. н., проф. Я.Б. Блюмом, обрано ідею та тему дослідження, сформульовано основну мету та завдання роботи, сплановано експерименти, та проведено обговорення отриманих результатів, а в подальшому – інтерпретовано отримані результати та опубліковано ряд наукових публікацій. Здобувачем особисто виконані основні експериментальні дослідження та опрацьовані літературні джерела за темою дисертаційної роботи. Аналіз експресії було проведено спільно з групою академіка НАН Білорусі О. Кільчевського з Інституту генетики і цитології НАН Білорусі (Мінськ). Розробку шляхів редагування геномів було опрацьовано спільно з групою доктора D. Breviario (Istituto di Biologia e Biotecnologia Agraria IBBA-CNR, Milan, Italy). Автором обґрунтовано основні положення та узагальнення дисертаційної роботи.

**Апробація результатів дисертації.** Матеріали дисертаційної роботи доповідалися і обговорювалися на конференціях: Первая конференция молодых

ученых «Биология растений и биотехнология» с международным участием (Біла Церква, 5-7 жовтня 2011 р.); 2nd IRB Barcelona PhD Student Symposium 2011 (Барселона, 17-18 листопада 2011 р.), Int. Moscow Conf. on Computational Mol. Biology» (Москва, РФ, 21-24 липня 2011 р.); 8th Int. Conf. on the Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS-SB-2012) (Новосибірськ, 25-29 червня 2012 р.); III Int. Symposium «Intracellular Signaling and Bioactive Molecules Design 2012» (Львів, 17-23 вересня 2012 р.); 9th Int. Symposium «Integrative Bioinformatics 2013» – Leibniz Institute of Plant Genetics and Crop Plant Research (Gatersleben, Germany, 18–20 березня 2013р.); 79 научно-техн. конференция профессорско-преподавательского состава, научных сотрудников и аспирантов (Мінськ, Білорусь, 2-6 лютого 2015р.).

**Публікації.** За результатами роботи опубліковано 18 наукових робіт, з яких 11 статей в українських та зарубіжних наукових фахових виданнях, зокрема, 3 – у виданнях Q2, та 7 тез доповідей у збірниках матеріалів всеукраїнських та міжнародних наукових конференцій.

**Структура та обсяг дисертації.** Дисертація складається з вступу, огляду літератури, матеріалів і методів досліджень, результатів експериментальних досліджень та їх обговорення, аналізу та узагальнення результатів, висновків, списку використаних джерел (240 найменувань) та додатку. Роботу викладено на 149 сторінках машинописного тексту і проілюстровано 29 рисунками та 15 таблицями.

\*\*\*

Дисертант висловлює глибоку подяку науковому керівнику академіку НАН України, професору, д.б.н. Блюму Ярославу Борисовичу за наукове керівництво та обрання актуального напрямку дисертаційного дослідження, допомогу при інтерпретації отриманих результатів та написанні дисертаційної роботи. Автор висловлює подяку співробітникам ДУ «Інститут харчової біотехнології та геноміки»: д.б.н. Карпову Павлу Андрійовичу, к.б.н. Співак Світлані Ігорівні, к.б.н. Баєр Галині Ярославівні, д.б.н. Пірку Ярославу Васильовичу, к.б.н. Демчуку Олегу Миколайовичу за допомогу в проведенні досліджень, та консультації при плануванні експериментів та обробці даних. Пошукач також висловлює подяку всім колегам та співробітникам ДУ «Інститут харчової біотехнології та геноміки НАН України» за допомогу та консультації при плануванні експериментів та обробці даних. Також, здобувач дякує колегам та співавторам у кого я багато чому навчився: Галіновському Дмитру Валентиновичу, Подвіцкому Тимофію Олександровичу – НАН Білорусі; Gavazzi Floriana, Morello Laura та Breviario Diego – Istituto di biologia e biotecnologia agraria (IBBA), Milan, Italy. Не в останню чергу автор дякує рідним та близьким за розуміння і підтримку під час виконання дисертаційної роботи.

## РОЗДІЛ 1

### ОГЛЯД ЛІТЕРАТУРИ

#### **1.1 Льон як джерело високоякісних текстильних та композитних волокон**

Вважається, що льон (*Linum usitatissimum* L., родина *Linaceae*) є однією з восьми сільськогосподарських культур, одомашнених людьми для отримання масла та клітковини понад 7–8000 років тому на Близькому Сході – van Zeist and Bakker-Heeres 1975. В цілому, 54 600 зразків зберігаються у банках насіння, а під його обробіток зайнято площі у 3 мільйони гектарів у всьому світі, забезпечуючи виробництво 0,8 мільйона тонн волокон на рік та 2,9 мільйона тонн на рік насіння льону в 2016 році (<http://www.fao.org/faostat>). Льон є як цінним ресурсом, так і цікавою модельною культурою.

**Стале виробництво.** Вирощування льону не вимагає зрошення під час росту і не потребує майже ніякої хімічної обробки. Поряд з цінними текстильними волокнами та багатою  $\Omega$ -3  $\Omega$ -6 олією, – Singh et al. 2011, льон також може бути джерелом інших корисних сполук із побічних продуктів: корм для домашніх тварин із льняного шроту; короткі технічні волокна, які будуть використовуватися як біоматеріали, – Zhu et al. 2013, слиз із оболонки насіння, фармацевтичні та нутрицевтичні сполуки, тобто полімери для доставки ліків, молекули антиоксидантів лігнани, фітостероли, PUFA, – Oomah 2001. Сировина з ринку матеріалів низької якості – залишкова волокниста біомаса нарешті використовується у виробництві енергії. Поєднання екстракції олії та виробництво енергії можуть збільшити доходи відповідно до так званого принципу біопереробки (biorefinery principle). Сприяння використанню натуральних волокон як ресурсу нейтрального щодо продукції CO<sub>2</sub> та екологічно чистого матеріалу, швидше за все, сприятиме збільшенню культивування льону.

**Льняне волокно.** – одне з найбільш міцних рослинних волокон, яке здавна використовується для виробництва текстилю та композиційних матеріалів, – Dissanayake et al., 2009; Kvavadze et al., 2009. З огляду на важливість цієї рослини, нещодавно отримано та проаналізовано послідовність цілого геному, – Wang et al., 2012, протеом клітковини, – Hotte and Deyholos, 2008 та транскриптом апікальної меристеми льону, – Zhang and Deyholos, 2016.

Подібно волокнам конопель та джуту, текстильні волокна льону (полотно) – це флоемні клітини. Льняні волокна класифікують як луб'яні; вони утворюються у зовнішніх тканинах стебла, між флоемою та епідермісом, і досягають довжини близько 7 см за 2–4 дні, в результаті, складаючи 20–30% від загальної маси стебла. Як правило, лляні волокна містять близько 70% целюлозних волокон, що відкладаються в клітинних стінках разом із геміцелюлозою, пектином та лігніном, – Titok et al. 2006. Відомо, що спрямований ріст рослинної клітини забезпечується відкладенням та модифікацією компонентів клітинної стінки під впливом тургорного тиску. Відповідно, ключовим регуляторним елементом анізотропного росту та форми клітин є спрямоване відкладення фібрил целюлози, – Bringmann et al., 2012b. Цінні сорти лляних волокон були розроблені традиційними методами селекції. У таких сортів врожайність волокна значно зросла без помітних змін у зовнішній архітектурі стебла або діаметрі волокна, – Goudenhooff et al. 2017.

Структура та склад клітинної стінки луб'яного волокна відповідають за питоми міцність текстильних волокон на розрив, розвиток яких детально досліджено в льоні за допомогою мікроскопії та біохімічних підходів, – Gorshkova et al. 2003, після чого, також, за допомогою досліджень експресії генів за допомогою мікрочипів, – Roach and Deyholos 2007, методами qPCR та протеоміки, – Hotte and Deyholos 2008. Подовження волокна супроводжується відкладенням специфічної вторинної клітинної стінки, яка після розвитку в драглистому шарі (G-layer) називається третинною клітинною стінкою, – Mellerowicz and Gorshkova 2012; Gorshkova et al. 2018. Третинна клітинна стінка може досягати товщини до 10 мкм і характеризується типовою кристалічною

морфологією завдяки осьовій орієнтації мікрофібрил целюлози, високому вмісту целюлози, – приблизно 70%, окрім геміцелюлози та пектину, а також низькому вмісту лігніну та ксилану, – Gorshkova et al. 2010.

## 1.2 Геномні ресурси, присвячені льону

Льон (*Linum usitatissimum* L., родина *Linaceae*) – цінна харчова та клітковинна культура, яка культивується завдяки своїй якісній клітковині та олії насіння; деякі форми виду також вирощують як декоративні рослини. Стебло льону є основним джерелом багатого целюлозою луб'яного волокна, що використовується в текстильній промисловості для виробництва лляної тканини. Вживання лляної олії, багаті омега-3 жирними кислотами, корисно для здоров'я людини. Лляне насіння також іноді додають до різних марок круп, хліба та інших зернових продуктів. Крім того, лляне масло використовується для приготування різних промислових розчинників. Через економічну важливість льону його геном нещодавно був сиквенований і опубліковано попередній драфт зборки геному, – Wang, 2012.

Перший проект збірки генома льону, який орієнтовно охоплює 81% геному, був опублікований у 2012 році після повно-геномного сиквенування методом шотган сорту CDC Bethune, – Wang et al. 2012. Хоча, геном льону і поводитьься як диплоїд ( $n = 15$ ), геном льону це результат кількох раундів повногеномних дуплікацій генома (WGD), – Sveinsson et al. 2014, останній з яких відбувся у спільного предка культивованого льону та його дикого родича *Linum bienne*, після його відокремлення від споріднених ліній, що містять такі види, як *L. grandiflorum*, число хромосом якого –  $n = 8$ . На підставі кількості синонімічних заміщень у ідентичних позиціях серед знайдених 9920 продубльованих генів, було зроблено висновок про час розбіжності даних видів складає 5–9 МґА, млн. років тому. Загалом було передбачено 43384 генів, що кодують білки, у 77% з яких визначено один або кілька охарактеризованих доменів сімейства білків (Pfam), – Wang et al. 2012.



Мікрочіпи, а також дані EST та транскриптомічні дані дають багато інформації про гени-кандидати, потенційно пов'язані з агрономічно важливими ознаками, такими як ріст волокон, – Gorshkov et al. 2017, розвиток насіння, – Venglat et al. 2011 та посухостійкість, – Dash et al. 2017.

### 1.3 Цитоскелетні білки

$\alpha$ -,  $\beta$ -,  $\gamma$ -тубуліни та актини є основними структурними білками цитоскелета; вони беруть участь у кількох ключових біологічних процесах, таких як поділ клітин, везикулярний транспорт та відкладення клітинної стінки. Чотирнадцять генів  $\beta$ -тубуліну, що належать до 7 ізотипів були передбачені у геномі льону в більш ранніх дослідженнях, – Gavazzi 2017. Гени  $\alpha$ - та  $\gamma$ -тубулінів досі не були ідентифіковані, так, як і гени актину. Сімейства  $\alpha$ - і  $\beta$ -тубулінів додатково поділяються на класи, які походять від двох та чотирьох генів-предків відповідно.

#### 1.3.1 Родина генів тубулінів

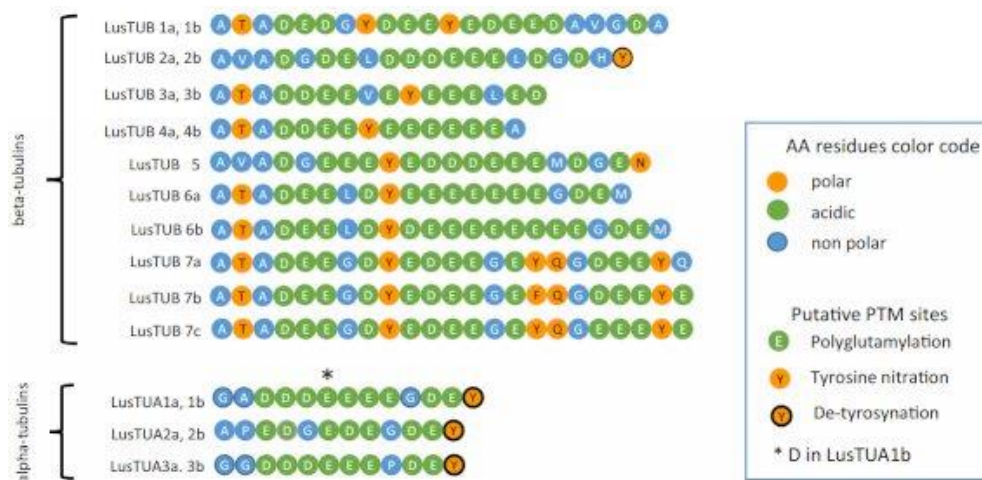
Геном *A. thaliana* містить шість генів  $\alpha$ -тубуліну, що кодують чотири різні ізотипи, – Korczak, 1992, дев'ять генів  $\beta$ -тубуліну, що кодують дев'ять різних ізотипів, – Snustad, 1992, та два ізотипу  $\gamma$ -тубуліну, – Liu, 1994.

Дві пари з шести (TUA1–6)  $\alpha$ -тубулінів *A. thaliana*, анотованих у UniProt, мають однакові білкові послідовності: TUA2 – TUA4 (клас Ia) та TUA3 – TUA5 (клас IIa), відповідно. Крім того, TUA6 має лише дві амінокислотні заміни порівняно з TUA2/TUA4. TUA1 належить до класу IIb. Виходячи з їх переважної схеми експресії, гени  $\alpha$ - та  $\beta$ -тубуліну *Arabidopsis* можна позначити як вегетативні (або конститутивні) або репродуктивні, що означає, що вони згодом майже виключно експресуються в квіткових органах, – Meagher, 1999. Таким же чином два з дев'яти (TUB1–9) генів  $\beta$ -тубулінів кодують однакові білки TUB2 і TUB3, – Meagher and Fehheimer, 2003.

Рослинні тубуліни належать до трьох різних сімейств:  $\alpha$  та  $\beta$ . Гетеродимери  $\alpha$  та  $\beta$  тубулінів утворюючи основну одиницю МТ. В свою чергу, гамма-тубуліни необхідні для зародження МТ. Мономери тубулінів – це глобулярні білки, тривимірна структура яких є однією з найбільш консервативних у еукаріотів, аж до того, що тубуліни видів, які належать до різних царств, можуть збиратися в мікротрубочки *in vivo*, – Zhang et al. 1990; Yuan et al. 1994. Гетеродимери  $\alpha$ -/ $\beta$ -тубулінів здатні самостійно збиратись у ругулярну структуру, яка, врешті-решт, утворює мікротрубочку за допомогою асоціації 13 протофіламентів, біч-о-біч одна до одної. Нуклеація нових МТ головним чином починається в кортикальному шарі клітин з кільцевих комплексів  $\gamma$ -тубуліну ( $\gamma$ -TuRC), нуклеаційної структури, що містить гамма-тубулін. МТ є високодинамічними структурами, що піддаються циклам полімеризації-деполімеризації, що регулюються різними сигналами і великою кількістю білків асоційованих з мікротрубочками (MAP), – Oda 2015; Gardiner 2013.

У той час як гамма-тубуліни, як правило, кодуються одним або двома генами,  $\alpha$ - та  $\beta$ -тубуліни кодуються багатогенними сімействами у багатоклітинних еукаріотів. Кожного окремого члена сімействами часто називають ізотипом. Послідовності амінокислот, хоча і дуже консервативні (60% у різних царствах), демонструють незначні відмінності у ключових регіонах, що може спричинити важливі функціональні відмінності. Найбільш мінливою областю є С-кінцевий хвіст (СТТ), багатий на залишки кислот, які можуть відрізнятися за довжиною та послідовністю серед різних ізотипів тубуліну, – Рис. 1.1. Цей короткий домен, який відходить назовні від поверхні мікротрубочок, відіграє вирішальну роль у регулюванні збирання стабільності МТ, як це продемонстровано у досліджах збиранням *in vitro* тубулінів з видаленими СТТ фрагментами, – Serrano et al. 1984, та моделюванням молекулярної динаміки, – Freedman et al. 2011. СТТ також є сайтом для взаємодії з MAPами, – Hamada 2007, а також для важливих посттрансляційних модифікацій (PTM), таких як детирозинація, – Smertenko et al. 1997, нітрування тирозину, – Jovanović et al.

2010; Blume et al. 2013 та поліглутамілування, – Smertenko et al. 1997; Wang et al. 2004.



**Рис. 1.1.** Схематичне зображення С-кінцевого домену всіх ізотипів альфа- та бета-тубуліну льону із зазначенням передбачуваних сайтів посттрансляційних модифікацій.

Інші важливі варіабельні домени знаходяться в N-кінцевій області і включають сайти зв'язування антимітотичних препаратів, надаючи різні фармакологічні властивості мікротрубочкам рослин, грибів, тварин та протистів, – Dostál and Libusová 2014. У всіх досліджуваних рослин експресія ізотипів тубуліну регулюється тканинно-специфічно та у відповідь на гормони та зовнішні подразники. Шаблони експресії не є унікальними, а перекриваються досить сильно. Незважаючи на ці варіації, жодна особлива функція ніколи не приписувалася жодному конкретному ізотипу рослинних тубулінів, і питання про наявність кількох генів, що синтезують взаємозамінний структурний білок, залишається відкритим для різних пояснень, отриманих з моделі клітин тварин, включаючи переглянутий варіант так званої «мульти-тубулінової гіпотези» (“multi-tubulin hypothesis”), – Wilson and Borisy 1997, та модель «коду тубуліну» (“tubulin code”), – Gadadhar et al. 2017. Ймовірно, що фактичний склад пулу внутрішньоклітинного тубуліну, а також панель білків, асоційованих з мікротрубочками, може помірно впливати на кінетичні властивості масивів

мікротрубочок, – Freedman et al. 2011, впливаючи тим самим на різні фази деполімеризації та нуклеації МТ, які забезпечують переорієнтацію МТ.

Перекривання експресії та функціональних властивостей генів тубуліну завжди ускладнювало можливість зроби́ти висновок про конкретну функцію ізотипів на основі експериментів з нокаутуючими мутаціями або сайленсингом генів на основі RNAi, – Bao et al. 2001. Раніше нуль-мутанти тубулінів за певними окремими локусами не виявляли жодного фенотипового ефекту в морфології рослин через функціональну надмірність різних ізотипів тубулінів, – Thitamadee et al. 2002. І навпаки, регуляторне зниження експресії кількох генів тубуліну або генів тубуліну, які експресуються на основних стадіях розвитку, призвело до серйозних вад росту, – Mayer et al. 1999; Bao et al. 2001. Міссенс мутації у ключових позиціях можуть виступати головним чинником, який впливає на полімеризацію та/або стабільність МТ. Велика колекція мутантів тубулінів *Arabidopsis* дозволила визначити критичні залишки на інтрадимерному інтерфейсі  $\alpha$ - $\beta$  тубулінів, поздовжнього інтерфейсу між гетеродимерами та бічного контакту між протофіламентами, – Ishida and Hashimoto 2007. Ступінь вираженості спостережуваних фенотипів суворо пов'язана з відносним рівнем експресії мутованого гена та його тканиноспецифічністю, – Ishida and Hashimoto 2007. Рідко повідомляється, що зміна вмісту клітинного пулу тубулінів внаслідок надмірної експресії окремих генів проявляються у фенотипі, як це спостерігалось у *Eucalyptus*, коли надмірна експресія EgrTUB1 у соматичних трансгенних секторах деревини впливала на орієнтацію мікротрубочок у клітинних стінках ксилеми, – Spokevicius et al. 2007.

Вирівнювання амінокислотної послідовності всіх генів  $\alpha$ - і  $\beta$ -тубулінів кількох рослин виявляє наявність підродин або класів, що виникли в результаті дублювання генів-предків: їх принаймні два для  $\alpha$ - та чотири для  $\beta$ -тубулінів, – Breviario et al. 2013. Кількість генетичних локусів різна серед різних видів рослин, тому практично неможливо точно визначити гени-ортологи.

### 1.3.2 Родина генів актинів

У геномі *Arabidopsis thaliana* було виявлено десять генів актину, вісім з яких, як відомо, експресуються конститутивно, а два (ACT5 і ACT9) переважно експресуються в ендоспермі, – McDowell, 1996b, Šlajcherová, 2012. ACT2/8 та 7 експресуються переважно у вегетативних тканинах, а ACT1/3, 4/12 та 11 – у репродуктивних тканинах. Примітно, що ген *Arabidopsis* ACT7 містить вище рамки ТАТА різні унікальні гормоночутливі регуляторні елементи – ауксин, абсцисову кислоту, гіберелін та інші, і має надзвичайно високий рівень експресії в молодих швидко зростаючих тканинах, – McDowell, 1996a, Kandasamy, 2002. Ізотип ACT7 у *Arabidopsis*, також, необхідний для нормальної реакції на стрес – на рану та на ауксин, – Kandasamy, 2002.

Рослинні вегетативні та репродуктивні актини генетично розійшлися і згодом еволюціонували з єдиного загального рослинного гена-предка актину, – Meagher, 1999. У той час, як вегетативні актини є консервативною групою, яка зберегла древні актинові функції базальних еукаріот, а саме протистів, – Kandasamy, 2012, група репродуктивні актинів рзійшлась значно більше і набула нових функцій у нових репродуктивних структурах, що виникли в процесі еволюції, втрачаючи тим самим частину основних властивостей (оглянуто у Šlajcherová, 2012. Зокрема, репродуктивні актини нездатні взаємодіяти належним чином з вегетативною кладою актинозв'язуючих білків, – Kandasamy, 2007.

*Arabidopsis* ACT2 & ACT8, ACT1 & ACT3 та ACT4 & ACT12 визначають три пари тісно пов'язаних між собою паралогів, проте порівняння відповідних їм передбачуваних амінокислотних послідовностей показує наявність лише однієї амінокислотної заміни. Геномна організація генів актину *A. thaliana* досить консервативна. Транскрибована частина починається з одного некодуєчого екзона лідера, перерваного послідовністю лідера-інтрона, розташованого в 5' нетранслюємій області (UTR). Трансльована область містить чотири екзони та

три інтрони. Єдиним винятком є *Arabidopsis ACT2*, де відсутній інтрон 2 (McDowell, 1996b).

Експресія різних ізотипів актину модулюється за участю різних регуляторних елементів у відповідь на різні умови або у різних органах. У зв'язку з цим, лідируюча послідовність 5'UTR та інтрон 1 відіграють важливу роль у посттранскрипційній регуляції експресії актину (оглянуто у Šlajcheroová, 2012).

Білкові послідовності тубуліну та актину, є еволюційно висококонсервативними, – Demchuk, 2005, Dominguez, 2011. Це наслідок фундаментальної ролі, яку виконує цитоскелет у ключових клітинних процесах. Такий високий рівень еволюційного консерватизму робить аналіз філогенетичних та еволюційних зв'язків послідовностей білків тубуліну та актину складним завданням, особливо для генів того самого виду або близькоспоріднених видів.

З іншого боку, через менший еволюційний тиск, послідовності інтронів менш консервативні і можуть бути більш ефективно використані для еволюційних та філогенетичних досліджень. Наприклад, поліморфізми інтронів  $\beta$ -тубуліну раніше використовувались для генетичної характеристики роду винограду *Vitis L.*, – Gavazzi, 2016. Як було показано, конвенційні сплайсосомні рослинні інтрони, як і інтрони інших еукаріот, мають консенсусні послідовності у сайтах розпізнавання 5'-GT(/C) та 3'-AG, – Hanley and Schuler, 1988.

Інтрони можуть впливати на експресію генів різними способами, впливаючи на кожен етап переробки та дозрівання мРНК, залежно від того, чи забезпечується регуляторна активність регуляторним елементом, який присутній у послідовності інтронів, або сплайсосомою, – Chorev and Carmel, 2012; Jo, 2015. Інтрон-медійоване посилення експресії (IME) є одним із найбільш помітних ефектів на транскрипцію генів, яке здійснюється регуляторними інтронами у рослин, – Mascarenhas, 1990; Luehrsen and Walbot, 1991; Akua, 2010. На відміну від енхансерів, інтрони повинні бути присутніми в транскрибованій області гена, щоб забезпечити IME, – Rose, 2008. Коли IME активність забезпечується інтронами, присутніми в послідовності лідера 5'UTR, вона зменшується зі

збільшенням відстані від сайту початку транскрипції, – Morello, 2002, Parra, 2011. ІМЕ може збільшити рівень транскрипційної експресії свого контрольованого гена до 1000 разів, – Gallegos, 2015. Проте, як саме інтрони можуть впливати на експресію генів у такій мірі, залишається відкритим питанням, – Parra, 2011, Morello, 2011. Було показано, що інтрони впливають на експресію  $\alpha$ - і  $\beta$ -тубуліну в рисі саме за допомогою механізму ІМЕ, – Fiume, 2004; Giani, 2009.

Предметом та методом дослідження у даній роботі, є, зокрема, пошук генів  $\alpha$ -,  $\beta$ - та  $\gamma$ -тубулінів та актину у геномі льону та аналізували філогенетичні зв'язки між льоном та генами тубуліну та актину *A. thaliana*. Через порівняно недавню 5–9 мільйонів років тому подію дублювання геному льону, – Wang, 2012 ми очікували знайти принаймні два паралоги кожного гена у геномі льону. З цих паралогів обидва можуть бути функціональними, або лише один, або жоден.

### 1.3.3 Білки асоційовані з мікротрубочками

Білки, асоційовані з мікротрубочками (БММ / MAP) – білки, що зв'язуються з субодиницями мікротрубочок і виконують функції, стабілізації або дестабілізації мікротрубочок, клітинного транспорту, зв'язування мікротрубочок між собою і є посередниками при взаємодії мікротрубочок з іншими клітинними білками.

Організація рослинних мікротрубочок (МТ) вимагає наявності широкого ряду рослинно-специфічних білків, асоційованих з мікротрубочками (MAP), оскільки у рослин нема центріолей, які організовують МТ у цитоплазмі клітин тварин. Чинячи вплив на поведінку мікротрубочок, MAPи є важливими об'єктами біотехнології. Однак, послідовності специфічних для рослин MAP ідентифіковані та анотовані лише в декількох геномах рослин. Раніше, – Karrov and Blume 2008, ми проводили біоінформаційний пошук рослинних гомологів структурних білків MAP тварин у геномі *A. thaliana*. У даній роботі ми проводимо біоінформатичний пошук та послідовне багаторазове вирівнювання

та філогенетичний аналіз рослинно специфічних MAP сімейств MAP60 та MAP70 та сімейства EB1 у геномах дводольних – *Arabidopsis*, *Brassica*, *Glycine*, *Medicago*, та однодольних – *Oryza*, *Hordeum*, *Triticum*, *Zea*.

#### **1.4 Родина генів целюлозосинтаз та целюлозосинтазо подібних білків льону**

Льон звичайний – одна з найдавніших технічних культур, що успішно вирощується в ряді країн світу, зокрема, в країнах Західної Європи. Льон – це унікальна модельна система для вивчення процесів росту та диференціації рослинних клітин, а також формування клітинної стінки. Останнє має особливе значення, оскільки звичайний льон цінується переважно за його стебло, яке містить 20–30% лігноцелюлозної клітковини. Льняне волокно називають лігноцелюлозним волокном, оскільки воно здебільшого складається з целюлозних волокон (приблизно 70%) з додаванням геміцелюлози, пектину та лігніну. Поширена думка, що мікрофібрили целюлози є основними механічними елементами клітинної стінки. Льняні волокна є класичним прикладом клітин, метаболізм яких налаштований на активне продукування целюлози, а целюлоза є найпоширенішим біополімером на Землі. Існують дані, що щорічно у світі виробляється понад  $10^{11}$  тонн целюлози Brown 2004.

Основними структурними компонентами фібрил є молекули целюлози, які відкладаються у вторинній клітинній стінці флоєми. Вони багато в чому визначають фізико-хімічні характеристики лляного волокна. Біосинтез целюлози здійснюється багатоферментним целюлозосинтезуючим комплексом, який зв'язаний з цитоплазматичною мембраною. У вищих рослин білки целюлозосинтази розташовуються у вигляді розеткоподібного комплексу діаметром 25 нм, локалізованого в плазматичній мембрані. На думку різних авторів, для того щоб комплекс целюлозосинтази був функціонально активним, до комплексу має входити 18 або 36 субодиниць білка целюлозосинтази Kimura et al. 1999; Scheible et al. 2001; Taylor et al. 2003; Sethaphong et al. 2013.



Гени, що кодують целюлозосинтази (*CesA*), вперше були ідентифіковані як частина бактеріального оперону синтази целюлози, – Ross et al. 1991. Пізніше гени *CesA* були виявлені у понад 170 видах рослин. Дослідження геному арабідопсису показало, що ця рослина містить гени синтезу целюлози, що належать до десяти різних класів, а також 30 генів, що кодують целюлозосинтазо подібні білки (*Csl*), – Hamann et al 2004.

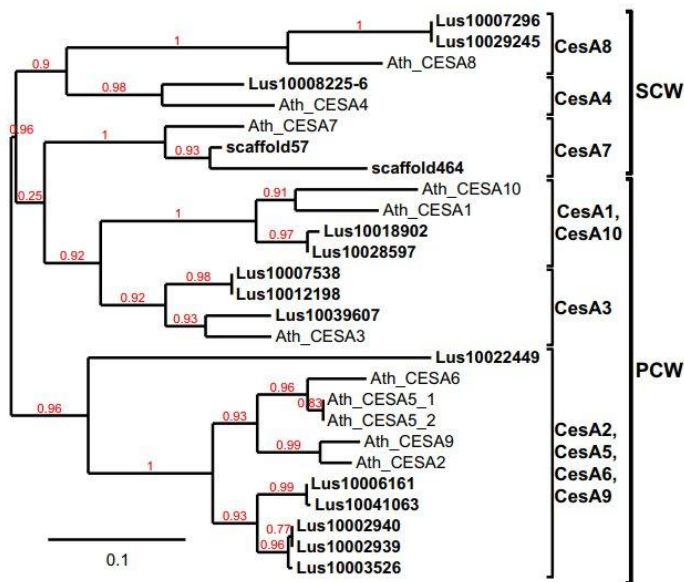
Філогенетичний аналіз відповідних генів сперматофітів показав наявність шести окремих клад генів *CesA*, кожен з яких відповідає одному з шести важливих компонентів первинного та вторинного комплексів синтази целюлози у *A. thaliana*, – Holland et al. 2000; Ranik et al 2006; Kumar et al 2009; Carroll et al. 2011. Експресія різних субодиниць целюлозосинтаз цього комплексу була експериментально продемонстрована в різних органах і тканинах волокна льону, – Galinousky et al., 2014; Mokshina et al., 2014.

Всі гени целюлозосинтаз можна розділити на дві групи генів, які регулюються диференційовано, – Burton et al., 2000. Гени з кожної групи експресуються протягом первинного або вторинного біогенезу клітинної стінки. У *Arabidopsis thaliana* білки *ATCESA1*, *ATCESA2*, *ATCESA3*, *ATCESA5*, *ATCESA6* та *ATCESA9* пов'язані з утворенням первинної клітинної стінки, тоді як *ATCESA4*, *ATCESA7* та *ATCESA8* необхідні для біогенезу товстої та міцної вторинної клітинної стінки, – Kumar et al., 2009.

У даній роботі ми, також, здійснили *in silico* пошук нуклеотидних послідовностей, що кодують целюлозосинтази у геномі звичайного льону, і порівняли їх з послідовностями генів-ортологів у дводольних рослин (*Populus tomentosa*, *Populus tremula / tremuloides*, *Eucalyptus grandis*, *Gossypium hirsutum*, *Brassica napus* та *Arabidopsis thaliana*).

**Родина генів *CesA* льону.** У геномі льону за різними даними ідентифіковано 15–16 генів *CesA*, – Mokshina et al. 2014; Pydiura et al. 2015; Chantreau et al. 2015, серед яких виявлено 12 різних ізоформ білка – Рис. 1.2. Вони об'єднуються у шість кластерів *CesA*: *LusCesA1*, *LusCesA3* та *LusCesA6* беруть участь у відкладанні первинної клітинної стінки (PCW), а *LusCesA4*, *LusCesA7* та

*LusCesA8* асоціюються з утворенням вторинної клітинної стінки (SCW), – Rudiura et al. 2015. Гени льону *CesA* містять характерну кількість екзонів, від 10 до 14, що відображає змінну довжину нуклеотидної послідовності різних ізотипів і відрізняє це сімейство від генів целюлозосинтаз подібних білків, що характеризуються наявністю лише 2–4 екзонів, – Rudiura et al. 2015.



**Рис. 1.2.** Філогенетичне дерево генів *Linum* та *A. thaliana* *CesA*. Ізотипи *AtCesSA1/10*, *AtCesA3* та *AtCesA6/2/5/9* переважно пов'язані з утворенням первинної клітинної стінки (PCW), яка відкладається на етапах клітинного поділу та розширення клітин. Три інші ізотипи, *AtCesA4*, *AtCesA7* і *AtCesA8*, необхідні для синтезу більш товстої вторинної клітинної стінки (SCW), що утворюється між ПКВ і плазматичною мембраною. Дерево було побудоване за допомогою веб-сервісу Phylogeny.fr.

Гени *CesA* кодують целюлозосинтази (код ферменту: EC 2.4.1.12) – фермент, який належить до сімейства глікозилтрансфераз 2 (GT-2, [www.cazy.org](http://www.cazy.org)). Рослинні білки *CesA* мають типову доменну організацію. У N-кінцевій області *CesA*-пептид містить потенційний пальцевий цинковий домен (zinc finger domain), який важливий для утворення димеру, – Holland et al. 2000. Далі за ним слідує гіперваріабельний домен (HVR), важливий для зв'язування UDPглюкози, – Pear et al. 1996, за яким слідує два трансмембранні домени.

Центральний цитоплазматичний домен (CCD) та шість термінальних трансмембранних доменів знаходять далі в послідовності даного білка, – Long et al. 2012. Вважається, що консервативні залишки амінокислот, що знаходяться в центральному цитоплазматичному домені (CCD), беруть участь у каталітичній активності, – Roberts and Roberts 2009. Насправді, амінокислотна послідовність CCD, складена у конформацію, утворену 13  $\alpha$ -спіралями та 6  $\beta$ -листками, дозволяє утворювати каталітичну кишеню, характерну для ферментів класу GT-2, – Sethaphong et al. 2013. Клас-специфічний регіон (CSR) та рослинно-консервативний регіон (P-CRs), також знаходяться в рослинному центральному цитоплазматичному домені (CCD) CesaA. CSR є змінними серед паралогів CesaA, але зберігаються серед ортологів CesaA, – Vergara and Carpita 2001. Укладка як CSR, так і P-CRs, утворена різною кількістю  $\alpha$ -спіралей і просторово відокремлені від каталітичної кишені, і їх роль полягає у тому, що вони відіграють ключову роль в олігомеризації білків CesaA у целюлозосинтазний комплекс, – Sethaphong et al. 2013.

Як повідомлялося в недавньому дослідженні, посттрансляційні модифікації, такі як фосфорилування конкретних залишків ізотипів CesaA, можуть впливати на анізотропне розширення клітин, синтез і відкладання целюлози, а також на залежну від мікротрубочок двонаправлену рухливість комплексів CSC, – Chen et al. 2016. Ці висновки корисні з точки зору функціональних аналізів, що мають бути проведені в майбутньому за допомогою технологій редагування геномів, і можуть надати кращу інформацію, ніж дані, отримані за допомогою аналізу декількох відомих мутантів CesaA, всі з яких, як відомо, страждають від важких дефектних фенотипів у процесі росту та розвитку рослин, за рахунок, як правило, зменшують стабільність комплексів CSC до такої міри, що призводять до швидкого розкладання CesaA протеасоми, що унеможливорює біосинтез целюлози при з низькому рівні UDP, – Li et al. 2017.

Гени CesaA, що беруть участь у синтезі PCW, найбільш сильно транскрибуються у проростках, але рівень експресії поступово зменшується в стеблах на етапі швидкого росту, де він може значно знизитись і бути до у сотню

разів нижче, ніж у генів паралогів SCW *LusCesA4* та *LusCesA7*, – Mokshina et al. 2014; Chantreau et al. 2015; Galinousky et al. 2017. Однак, даний рівень експресії може у значній мірі залежати від індивідуальних особливостей гена, а також конкретного досліджуваного сорту льону. Таким чином, у підвидах *elongatum* та *mediterraneum* експресія одного з двох паралогів CESA1 зменшилась у 5,2–10,7 разів, при порівнянні гіпокотил з стеблом на стадії швидкого росту, але у льоні підвиду *dehiscent* змін не вдалося зафіксувати, – Galinousky et al. 2017. Експресія генів PCW CesaA не обмежується ранніми стадіями розвитку стебла, але може бути реактивована на будь-якій пізнішій стадії розвитку тканини, коли це потрібно, як показано для відкладання товстих третинних клітинних стінок і у відповідь на вплив сили тяжіння, – Mokshina et al. 2017. Вагомі докази важливості CesaA1 та CesaA6 у відкладанні PCW у процесі розвитку стебла були елегантно продемонстровані сайленсингом гену за допомогою вірусу, коли зменшення кількості транскриптів цих генів на 60% призвело до дезорганізації як ксилеми, так і флоєми, – Chantreau et al. 2015.

Гени CesaA вторинної клітинної стінки показали високий рівень експресії як в прикореневій частині стебла льону порівняно з верхівковою меристемою, – Zhang and Deyholos 2016, так і в стеблі порівняно з гіпокотиллями, – Mokshina et al. 2014; Galinousky et al. 2017. *LusCesA4*, *LusCesA7-A*, *B* та *LusCesA8-A*, однаково сильно експресуються як у клітинах ксилеми з лігніфікованими вторинними клітинними стінками, тобто типу ксилану, так і у волокнах флоєми з товстою драглистою клітинною стінкою – третинна клітинна стінка. Це свідчить про те, що волокна флоєми та ксилеми можуть використовувати подібні розетки CSC. Отже, різні властивості клітинних стінок драглистого та ксиланового типів навряд чи визначаються композицією білків CesaA, а, швидше, іншими білками, пов'язаними з целюлозосинтазним комплексом, – Bringmann et al. 2012a; Worden et al. 2015; Gu et al. 2010. Різні дослідники відзначали переважання або PCW CesaA, – Chantreau et al. 2015; Galinousky and Padvitski 2017, або SCW CesaA, – Mokshina et al. 2014; Guo et al. 2017, під час потовщення клітин луб'яних волокон. Недавні транскриптомні дослідження клітин волокон показують, що гени CesaA,

пов'язані як з первинною, так і вторинною клітинною стінкою, дійсно необхідні для біогенезу третинної клітинної стінки. На основі даних про коекспресію генів *CesA* було припущено, що два типи целюлозосинтазних комплексів приймали участь у відкладенні третинної клітинної стінки, – Mokshina et al. 2017.

Клітини волокон флоєми, які утворюють збагачену целюлозою третинну клітинну стінку, мають важливе значення для культури льону. Аналізи світлової та електронної мікроскопії трьох підвидів льону, що відрізняються за якістю волокон, показують відмінності в структурі тканини первинної флоєми та в діаметрі елементарних волокон – Titok et al. 2010. Високоякісні волокна, вироблені підвидами *elongatum*, щільно упаковані, переважно невеликі, діаметром 10–14 мкм, тоді як в стеблах великого насінневого льону (*subsp. Mediterraneum Vav. Et Ell.*) можна виділити два типи волокон, діаметром 17–20 та 27–30 мкм, відповідно. В свою чергу, льон стрибунець (*dehiscent, subsp. crepitans Boenn.*), Містить волокна діаметром 23–25 мкм, – Galinousky et al. 2017. Крім того, нещодавно були виявлені деякі відмінності в анатомії тканин флоєми стебла у чотирьох різновидів волокнистого льону, що також характеризувалося збільшенням кількості волокон і заповненням проміжків між пучками, – Goudenhooft et al. 2017.

Як показано на філогенетичному дереві з Рис. 1.2 і вже згадувалося вище, номенклатура генів *CesA* у льоні та відносне віднесення окремих генів до утворення PCW чи SCW, по суті, базується на даних, встановлених для генів сімейства *CesA Arabidopsis thaliana* – Desprez et al. 2007; Kumar et al. 2009; Atanassov et al. 2009; Kumar and Turner 2015.

### **1.5 Синтез целюлози. Тубуліни і *CesA*: дві сімейства генів, що сприяють відкладенню клітинної стінки**

Целюлозосинтазні комплекси, які локалізовані на мембрані, також, асоційовані з цитоскелетом і можуть рухатися вздовж кортикальних мікротрубочок, – Bringmann et al., 2012. За останні роки накопичено докази

участі мікротрубочок у доставці будівельного матеріалу для целюлозних волокон, правильної орієнтації та відкладання міцних целюлозних фібрил у рослинах, – Baskin et al., 2004; Baskin 2005; Boutté et al., 2007. Відповідно до існуючих концепцій, відмінності у функціонуванні генів, що кодують комплекс целюлозосинтазної розетки – гени целюлозосинтаз, та білки цитоскелета – гени тубуліну та актину можуть суттєво впливати на якість та орієнтацію фібрил целюлози. Відповідно, вивчення експресії цих генів у різних формах льону може допомогти виявити фактори, що визначають якість лляного волокна.

Мікрофібрили целюлози синтезуються мультиферментними целюлозосинтезуючими комплексами (CSC), які складаються з 18-36 субодиниць *CesA*, відповідно до різних джерел, – Doblin et al. 2002; Fernandes et al. 2011; Newman et al. 2013, що після доставки до плазматичної мембрани за допомогою везикул, що рухаються по тяжам F-актину, під час синтезу целюлози рухаються вже вздовж кортикальних мікротрубочок, – Wightman and Turner 2008; Watanabe et al. 2015. Функція рослинних генів *CesA* спочатку була передбачена на основі їх гомології з генами целюлозосинтаз бактерій, – Pear et al. 1996, а потім шляхом вивчення мутантних рослин *Arabidopsis thaliana*, – Arioli et al. 1998 та шляхом маркування імунним золотом (immune-gold labeling) *CesA*-пептидів, – Kimura et al. 1999. Синтез целюлози відбувається на плазматичній мембрані, а комплекси CSC приєднуються і переміщуються вздовж кортикальних мікротрубочок за допомогою проміжних білків, що взаємодіють з тубулінами та актином, – Bringmann et al. 2012a, що підтверджує важливість їх взаємодії у процесі біосинтезу клітинної стінки.

У цьому дослідженні ми, між іншим, кількісно оцінили експресію генів целюлозосинтаз (*CesA*) та генів, що кодують основні білки цитоскелета у різних підвидів льону. З усіх генів *CesA* ми кількісно визначали два – *LusCesA1*, *LusCesA6*, які пов'язані з синтезом первинної клітинної стінки, і два – *LusCesA4*, *LusCesA7*, які пов'язані з вторинною клітинною стінкою. Вибір базувався як на попередньому аналізі бібліотеки кДНК – Galinousky et al., 2014, так і на порівняльному аналізі – Rydiura et al., 2015. Ми показали диференціальну

експресію генів *CesA* між онтогенетичними стадіями, а також диференційну експресію гена *LusCesA1* між підвидами.

**Трансгенний льон.** Льон має довгу історію генетичної трансформації, починаючи з кінця вісімдесятих років минулого століття. Вміст жирних кислот і склад насіння, вміст лігнану, якість волокон, толерантність до гербіцидів та стійкість до хвороб належать, серед інших, до властивостей льону, які вже були модифіковані генною інженерією, – оглянуто у Ludvíková and Griga 2015. Незважаючи на це, до цього часу в Канаді зареєстровано лише один трансгенний сорт льону, – McHughen 1989; McHughen et al. 1997, що характеризується парадигматичною історією, що призвела до його ранньої зняття з реєстрації в 2001 р. та тривалого зменшення загальної кількості льону експортованого Канадою, – Ryan and Smyth 2012.

Цей негативний та такий, що викликає стурбованість, досвід, швидше за все, буде усунутий завдяки впровадженню та використанню технологій редагування геномів, які дозволяють модифікувати ендогенні гени, спрямовані на сайт, без введення чужорідних генів. Модифіковані рослини, вироблені цими способами, швидше за все, будуть прийняті тими країнами, які негативно ставляться на трансгенних культур, і це має відкрити нові перспективи для поліпшення агрономічних ознак у культурних рослин. Окрім свого потенціалу для підвищення врожайності сільськогосподарських культур, редагування геному в даний час є одним з найкращих інструментів для проведення зворотної генетики, і воно виступає як особливо універсальний інструмент для вивчення базової біології рослин.

Ряд сімейств генів, що беруть участь у подовженні клітин, а також у різних фазах відкладення та потовщення клітинної стінки, були охарактеризовані для льону за допомогою філогенетичних та експресійних досліджень, особливо в клітинах клітковини:  $\beta$ -галактозидаза (*GH35*), – Roach et al. 2011; Hobson and Deyholos 2013; UDP-глікозилтрансфераза, – Barvkar et al. 2012; рамногалактуронан ліази, – Hotte and Deyholos 2008; пектинестераза (PME та PME1), – Pinzón-Latorre and Deyholos 2013; целюлозосинтази (*CesA*), – Mokshina

et al. 2014; Galinousky et al. 2017; целюлозосинтазо-подібні білки (*CSL*), – Guo et al. 2017; хітиназоподібні білки, – Mokshina et al. 2014; тубулін та актин для цитоскелета, – Gavazzi et al. 2017; Pydiura et al. 2018.

У нашій роботі ми зосереджували свою увагу на двох із цих сімей білків, обидва яких, як відомо, відіграють фундаментальну роль у морфогенезі рослин, впливаючи на форму та подовження клітин: тубулін – будівельний матеріал кортикальних мікротрубочок (с-МТ) та білки CesA – субодиниці мультимерного целюлозо-синтезуючого комплексу (CSC), як забезпечує направлене осадження целюлозних мікрофібрил. Було показано, що дві надмолекулярні структури, МТ і CSC, співпрацюють у формуванні клітинної стінки в ряді елегантних і фундаментальних експериментів, де їх взаємна асоціація, гіпотеза, висунута ще Green 1962, або безпосередньо спостерігалася за допомогою конфокальної відеомікроскопії – Paredez et al. 2006; Li et al. 2012, або чітко зроблені висновки в експериментах, проведених з мутантами, в яких змінена орієнтація МТ в кінцевому підсумку порушує біосинтез целюлози, – Paredez et al. 2008 або її відкладення, – Bringmann et al. 2012а, що, в решті решт, гальмує подовження клітин. Однак жоден з цих експериментів не міг надати інформацію про конкретні білок-білкові взаємодії, питання, вирішення якого є особливо цікавим, оскільки обидва сімейства генів складаються з декількох членів, що групуються у різні класи, де паралоги можуть демонструвати досить різний рівень експресії в різних тканинах та на різних стадіях росту, – Breviario et al. 2013; Ermawar et al. 2015; Dhugga 2001.

## **1.6 Менеджмент біологічних даних**

**Біологічні типи даних.** Величезний виклик – один із найбільш важливих для біології XXI століття – це управління різноманітністю та складністю типів даних, ієрархією біології та неминучою потребою отримувати дані за допомогою різноманітних засобів. Біологічні дані бувають різних типів. Наприклад, біологічні дані можуть складатися з наступного:



*Дані про первинні послідовності* – історично першими типами даних, що стали загальнодоступними, були дані послідовності нуклеотидів. Бази даних цього типу містять дані первинної послідовності з базовою інформацією про послідовності ДНК, РНК та білків. Ці дані становлять загальну базу, на якій будується більшість інших біомолекулярних БД.

*Дані мотивів та доменів* – мотиви та білкові домени представляють біомолекулярні сутності, які зазвичай виявляються за допомогою методів розпізнавання образів, що застосовуються до основних даних первинних послідовностей, які широко використовуються в біоінформатиці та дослідженнях молекулярної біології для характеристики функцій та сімейств білків.

*Структурні дані* – структурні дані білків – просторові атомні координати атомів та амінокислот, що складають сам білок. Реконструкція тривимірної структури білка має першорядним значенням для розуміння його функції. Дані отримують за допомогою рентгенівської кристалографії або ЯМР-спектроскопії.

*Дані генного рівня* – хоча бази даних генів починалися з анотації баз даних первинних послідовностей, нещодавні досягнення міжнародних проектів щодо секвенування цілих геномів сприяли розробці конкретних генноцентричних даних. Надається вся інформація про контекст конкретного гена: інформація про транскрипти, продукти, геномні області, генотип, фенотип, пов'язані регуляторні та каталітичні шляхи (pathways) та терміни генної онтології пов'язані з досліджуванним геном.

*Геномні дані* – характеристики та властивості біомолекул можна досліджувати на рівні «оміки»: на основі вивчення та аналізу окремих генів або білків – нові біотехнологічні підходи, запроваджені наприкінці 90-х, дозволяють аналізувати весь набір генів (геном) або білки (протеоми) даного виду. Ці дані були отримані на основі послідовності та картографування геному цілих організмів і доступні як специфічні для певного виду ресурси.

*Транскриптомічні дані* – дані мікрочипів ДНК збирають рівні експресії генів (тобто рівні мРНК, експресовані в даній клітині в певний час) у масштабі

всього геному. Ці дані дозволяють проаналізувати мінливість експресії генів між різними тканинами, особинами або між різними функціональними або патологічними станами.

*Дані про поліморфізми та мутації* – дані поліморфізмів та мутацій тепер доступні у відкритих базах даних і дозволяють аналізувати на геномному рівні асоціації між мутаціями та фенотипами клініки, а також проводити дослідження в галузі популяційної генетики.

*Реляційні дані на системному рівні* – взаємозв'язки та взаємодії між різними сутностями та підсистемами в клітинах на різних рівнях (наприклад, генні мережі або метаболізм цілої клітини) представляють клас реляційних даних, за допомогою яких ми можемо моделювати поведінку складних біологічних систем. Ці дані, отримані в основному за допомогою високопродуктивних біотехнологій, можуть бути використані для висновку про складні взаємозв'язки між біомолекулами на "системному рівні", розглядаючи біологічні явища як результат інтеграції різних процесів та різних взаємодій, що включають весь геном і протеом.

***Підходи до управління даними у BioGrid.*** Загальновідомим фактом є те, що більша частина загальнодоступних біологічних даних надходить із різних джерел інформації, оскільки вони є гетерогенними та отримуються, зберігаються та робляться доступні у різні способи в різних місцях по всьому світу, таким чином, вони є розподіленими. Неоднорідність даних може бути синтактичною, оскільки вони відрізняються форматами файлів, мовах запису та протоколах доступу та ін., неоднорідність даних, також, може бути семантичною, тобто геномні, або та протеомні дані, а також схематичною, тобто такі, де різниця полягає у назвах таблиць бази даних та полях таблиць. Для цих гетерогенних та розподілених даних, щоб отримати доступ до єдиного та федеративного середовища, потрібно скористатися відповідними веб-сайтами Grid-технологій, щоб сформувати проміжний міст. Грідифікація біологічних баз даних та додаткових даних також мотивовано тим, що їх кількість, розмір та різноманітність швидко та постійно зростають. Це робить неможливим для

індивідуального, окремо взятого біолога збереження локальної копії будь-якої з основних баз даних та використання даних або комп'ютерних програм у локальному середовищі.

***Grid-інфраструктура для біоінформаційних досліджень.*** Великі дослідницькі проекти у галузі біології та біомедицини включають величезні гетерогенні та географічно розподілені ресурси. Такі ресурси генеруються та спільно використовуються різними партнерами. Більше того, їх дослідницька робота спирається на засоби, які вимагають обчислень та методів, що вимагають великих обсягів даних. Окрім традиційних питань інтеграції даних та оперативної сумісності, у цих сценаріях потрібно вирішити нові проблеми. Більшість з цих питань вирішуються за допомогою технологій Grid, які забезпечують базову технологію, завдяки якій можна будувати спільні та безпечні програми. Однією з основних особливостей, що надаються підходами на основі Grid, є віртуальна організація (VO). За допомогою віртуальної організації можна визначити тимчасову співпрацю між дослідниками. VO визначає партнерів та їх можливі групи, встановлює правила співпраці між ними та визначає правила конфіденційності та безпеки для всіх ресурсів Grid. Слід звернути увагу на те, що конфіденційність та безпека особливо актуальні в галузі біомедицини, де використовуються біологічні дані реальних хворих, а тому їх конфіденційність зберігається.

Ще одна головна особливість інструментів, які базуються на Grid, – це розповсюдження як складних даних, так і процесів. Однією з основних завдань Grid-інфраструктури є обробка дуже часозатратних процесів за допомогою їх розподілення по набору комп'ютерних ресурсів. Для цієї задачі, технологія Grid надає сервісний фреймворк (наприклад, веб-служби) для визначення та виконання складних завдань. Такий фреймворк дозволяє визначити складні робочі процеси, що включають набір зареєстрованих у Grid сервісів.

На даний момент є кілька спроб застосувати технологію Grid для обробки біологічних даних. Ймовірно, eScience Lab (<http://www.esciencelab.org.uk/>) – це найвидатніший проект, який включає як біологічні сервіси на основі XML і

технології Grid. Більше того, eScience Lab вирішує проблему поєднання біологічних сервісів за допомогою робочих процесів Taverna (<https://esciencelab.org.uk/>). eScience Lab також вирішує питання семантичної неоднорідності за рахунок включення специфічних для домену онтологій. Вони використовуються для анотації біосервісів, тобто даних та інтерфейсів на їх входах та виходах, для подальшого виявлення та перевірки сумісності для створення комбінованих біоресурсів. Даний тип інфраструктури, що поєднує семантичні анотації та Grid, отримав назву Semantic Grid.

### **1.6.1 Менеджмент структурно-біологічних та генетичних даних генів білків пов'язаних з цитоскелетом та генів целюлозосинтаз**

Накопичення величезних обсягів біологічних даних призвело до створення різноманітних баз даних. Вони варіюються від більш спеціалізованої бази даних послідовностей Genebank (<https://www.ncbi.nlm.nih.gov/genbank/>) із 221 467 827 анотованими послідовностями станом на грудень 2020 р. або Protein Data Bank (<https://www.rcsb.org>, Berman et al., 2000) для тривимірних структурних даних із 174 014 моделями станом на січень 2011 р. До більш інтегрованих баз даних, таких як UniProt (<https://www.uniprot.org>, Magrane and the UniProt consortium), що поєднують дані білка з декількох різних баз даних: Swiss-Prot (<https://www.expasy.org/resources/uniprotkb-swiss-prot>), TrEMBL (<https://www.ebi.ac.uk/training/online/glossary/uniprotkbtrembl>) та PIR-PSD (<https://www.ebi.ac.uk/training/online/glossary/pir-psd>), а також дані, отримані з літератури. Але відповідь навіть на певне просте запитання вимагає не тільки запитів до вказаних баз даних, а й подальшого аналізу результатів та інтеграції з похідними, об'єктно-специфічними та експериментальними даними. Це призводить до створення вузьких конкретних баз даних, що містять інформацію, що стосується конкретних видів, тканин або станів. На сьогодні існує понад тисяча таких баз даних – наукових, медичних, сільськогосподарських та промислових, і їх кількість продовжує зростати.

Білки цитоскелета – тубуліни, МАРи, тощо, є важливими мішенями для широкого спектра протипухлинних, фунгіцидних, гербіцидних, протипротозойних, протигельмінтних та інших засобів комерційного значення. Незважаючи на довгу історію та інтенсивне вивчення, розуміння більшості цитоскелетних процесів тісно пов'язане з обчислювальними методами *in silico*. Проект Grid «Створення віртуальної організації для вирішення обчислювальних проблем дослідження цитоскелету та високоефективного скринінгу антимітотичних сполук із застосуванням сучасних Grid-технологій», виконаний Інститутом харчових біотехнологій та геноміки та спонсорований Національною академією наук України (<http://grid.nas.gov.ua/>) спрямований на залучення до співпраці кількох цитоскелетних груп з України, Росії, Болгарії, Чехії, Німеччини, Канади. Віртуальна організація CytoLabGrid передбачає інтеграцію та спільне використання обчислювальних засобів, баз даних молекулярних структур та експериментальних результатів дослідницькими групами, пов'язаними спільністю наукових цілей та цілей у дослідженні цитоскелета. Віртуальне середовище співпраці повинно стати багатим сховищем даних, інформації та знань, що стосуються цитоскелету, включаючи геноміку, біоінформатику, протеоміку і дані про каскади реакцій, дані про захворювання та різні стани – здатні забезпечити можливість передбачення фармакофорних властивостей, розробки та дослідження лікарських засобів.

Така віртуальна Grid-організація в даний час активно розробляється, і вона, як планується, складатиметься з чотирьох основних частин: 1) інструмент запитів та звітування; 2) портал дослідницьких даних; 3) служба спільного використання процесів аналізу та робочих процесів; 4) база даних. Інструмент запитів та звітування дозволяє отримати доступ до вихідних даних із загальних онлайн-баз даних, а також до унікальних похідних даних, обчислених користувачами системи та збережених у локальній базі даних, що забезпечує легкий пошук та багаторазове повторне використання. Розробка та впровадження нових схем даних та поєднання відповідних елементів існуючих схем дає можливість створити інтегровані представлення, відповідні до цілей

досліджень. Портал дослідницьких даних забезпечує інтерфейс для зв'язку різних біоінформаційних та експериментальних даних у процесі дослідження. Сервіси спільного використання процесів аналізу та робочих процесів забезпечують типові аналітичні інструменти, такі як, наприклад, BLAST, результати роботи якого можуть бути включені як дані і стати частиною інтегрованого профілю об'єкту дослідження. Також дані сервіси зберігають інформацію про параметри обробки, яка служить для оцінки надійності даних, автоматизації, накопичення біоінформаційних знань та співпраці вчених.

Система призначена для обробки загальних біоінформаційних типів даних, таких як послідовності ДНК та білків, геноміка, доменна організація білків, мутації та поліморфізм, протеоміка, тривимірна структура, метаболічні та регуляторні мережі, подібні до ліків та фармакофорних малих органічних молекул, хвороби та дані про певні стани організму, та літературні дані. Анотація та управління даними інформаційних ресурсів базується на семантичних технологіях і передбачає широке використання наявних таксономій, а також розробку нових.

Розроблена система забезпечить легкий доступ до даних, що стосуються елементів цитоскелету, накопичення нової корисної інформації та її обмін в колі науковою спільноти, що бере участь в проекті.

## РОЗДІЛ 2

### МАТЕРІАЛИ ТА МЕТОДИ ДОСЛІДЖЕННЯ

#### 2.1 Біоінформатичні методи досліджень

Для ідентифікації генів тубуліну та актину у геномі *L. usitatissimum*, було виконано TBLASTN, – Gertz et al. 2006; Altschul et al. 1990, пошук у базах даних трансльованих нуклеотидів у базі даних Phytozome, версій 10–13 (<http://www.phytozome.net>), – Goodstein et al. 2012. У якості запитів було використано білкові послідовності актинів і тубулінів *Arabidopsis*, – Swarbreck et al. 2008: *TUA1–3* і *TUB1, 2, 4–9* та *ACT1, ACT2*. Використовувались наступні параметри: граничне значення E –  $1e-5$ , матриця порівняння BLOSUM62, – Henikoff and Henikoff 1992; Altschul 1993, довжина слова – 3. Ми проаналізували результати та відкинули короткі та незначущі співпадіння. Отримані послідовності були перевірені за допомогою алгоритму Hidden Markov Model у програмі HMMER (<http://hmmer.org>), – Finn et al. 2011, з використанням референтної бази даних родин білків Pfam-A, версія 25.0 (<http://pfam.xfam.org>), – El-Gebali et al. 2019. Генетичні локуси, транскрибовані послідовності яких були віднесені за даними бази даних Pfam до родини тубулінів, було визначено як гени тубулінів льону. Надалі, з бази даних Phytozome ми отримали інформацію про знайдені гени *L. usitatissimum*, включаючи розташування, ідентифікатор послідовності, геномну послідовність, послідовність білка та кодуєчу послідовність (CDS).

Множинне вирівнювання послідовностей передбачених генів *L. usitatissimum* та тубулінів та актинів *A. thaliana* було проведено за допомогою ClustalW, – Larkin et al. 2015, та ClustalO (<https://www.ebi.ac.uk/Tools/msa/clustalo/>), – Madeira et al. 2019. Філогенетичні дерева були побудовані з використанням алгоритму приєднання сусідів (Neighbor-Joining) у програмі MEGA7 v7.0.26, – Kumar et al. 2016. Стабільність внутрішніх вузлів вимірювали за допомогою аналізу методом самонастройки

(bootstrap analysis) 1000 повторень. Іншими параметрами були: модель заміщення амінокислотних залишків: модель Пуассона; коефіцієнти між сайтами: єдині значення; обробка пропусків: повне видалення. Парами генів з найвищим рівнем ідентичності, що знаходяться в різних скаффолдах, вважалися паралогами.

Філогенетичний аналіз послідовностей генів було проведено за допомогою сервера Phylogeny.fr, – Dereeper et al. 2008. Використаний алгоритм включав багаторазове вирівнювання за допомогою інструменту MUSCLE – Edgar et al 2004, обробку отриманого вирівнювання за допомогою Gblocks – Castresana et al. 2000, а також реконструкцію філогенетичного дерева та подальшу візуалізацію за допомогою програм PhyML – Guindon et al. 2010, та TreeDyn – Chevenet et al. 2006, відповідно. Алгоритм ClustalW2 був використаний для порівняння певних послідовностей між собою, – Larkin et al. 2007.

Візуалізація екзон-інтронної структури генів тубуліну та актину була отримана за допомогою сервера відображення структури генів Gene Structure Display Server (<http://gsds.cbi.pku.edu.cn/>), – Hu et al. 2015.

**Моделювання за гомологією.** Пошук шаблонів для моделювання за гомологією проводили за допомогою пошуку послідовностей у базі даних PDB – PDB-BLAST зі стандартними параметрами. Множинне вирівнювання послідовностей було зроблено у програмі ClustalW зі стандартними параметрами. Моделювання за гомологією та *ab initio* проводили за допомогою програмного забезпечення Modeller 9v10, – Eswar et al., 2006, зі стандартними параметрами. Пакети VMD, – Humphrey et al. 1996 та PyMOL, – Schrödinger, LLC, використовувались для молекулярного моделювання, аналізу та візуалізації моделей.

Моделювання тривимірних структур білків тубулінів разом із неупорядкованими С-кінцевими хвостами було здійснено за допомогою онлайнового серверу I-TASSER, – Yang et al. 2015; Roy et al. 2010; Zhang 2008, що призначений для передбачення структури і функції білків. Для всіх даних



моделей було проведено оптимізацію їхньої геометрії в силовому полі amber03, – Ponder and Case 2003.

Сервер ClusPro 2.0 (<https://cluspro.bu.edu/>), – Kozakov et al. 2017, зі стандартними параметрами моделювання був використаний для білок-білкового докінгу. Процедура аналіз якості моделей за допомогою MolProbity була аналогічною для моделювання за гомологією.

## **2.2 COMODORE – система менеджменту структурно-біологічних даних**

Передбачення просторової структури білка та білок-лігандних взаємодій є однією з найважливіших цілей обчислювальної структурної біології та біоінформатики. Основними обчислювальними підходами до структурного та функціонального аналізу біомолекул *in silico* є порівняльне моделювання структури, моделювання молекулярної динаміки та докінг, які часто послідовно використовуються для аналізу одних і тих самих об'єктів, або процесів. Однак, на кожній з цих стадій генерується велика кількість гетерогенних наборів даних, які певною мірою залежать від початкових координат та параметрів моделювання, що використовуються. Керування великими обсягами таких даних досить складна задача. Таким чином, ми розробили веб-систему COMODORE (COmparative MOdeling, DOcking and Rafinement Environment) для автоматизації рутинних етапів обчислювальних завдань та їх анотації, таких як: підготовка, аналіз результатів та управління даними.

Система розроблена як розширюваний фреймворк, який млідує парадигмі розробки програмного забезпечення модель-представлення-контролер (Model View Controller, MVC), написаний на PHP, а для зберігання даних використовується реляційна база даних MySQL. COMODORE встановлюється на веб-сервері Linux з набором програм Apache / PHP / MySQL (LAMP). Короткі аналітичні завдання виконуються локально за допомогою Perl, Java або bash-скриптів або зовнішніх веб-серверів. Для більш ресурсоемного аналізу та

виконання обчислювальних завдань система підключена через мережу до високопродуктивних обчислювальних систем (кластер та Grid).

***Система включає наступні компоненти:***

- *Компонент ModDB* – це база структурних даних, метаданих та аналітичних інструментів, які направлені на білкові моделі, ліганди, їх комплекси та дані про взаємодію. Це основний компонент COMODORE, який дає можливість виконувати теоретичні дослідження надаючи систему управління даними для основних типів даних - експериментальних та теоретичних структурних моделей білків та/або лігандів, а також –амінокислотних послідовностей, анотацій та метаданих молекулярних комплексів. COMODORE інтегрований з такими базами даних, як UniProt (<https://www.uniprot.org>), Protein Data Bank (<https://www.rcsb.org>) та PubChem (<https://pubchem.ncbi.nlm.nih.gov>, Bolton et al., 2008). Дані із зовнішніх баз даних завантажуються у форматі xml через протокол http, аналізуються та зберігаються на локальному сховищі. Інтеграція дозволяє початковий автоматичний пошук даних про білкові структури, ліганди та послідовності, а пізніше синхронізацію, наприклад, для отримання даних про домени та структурні мотиви та сайти.

Також, цей компонент інтегрований з набором аналітичних інструментів, орієнтованих на оцінку структурної моделі, тобто MolProbity, – Chen et al., 2010, та DSSP, – Kabsch and Sander 1983, встановлених локально, та веб-сервера POLYVIEW-2D, – Porollo et al., 2004. Аналіз за допомогою цих інструментів проводиться автоматично; результати аналізуються та зберігаються як метадані, що відносяться до даної структурної моделі. Представлення даних для кожної структурної моделі організовано у вкладки та включає вкладку «Оцінка», що містить дані аналізу MolProbity – Рис. 2.1, вкладку «Послідовність» – послідовність, мотиви, аналіз вторинної структури DSSP та POLYVIEW-2D – Рис. 2.2, та вкладку «Структура» з інтегрованим аплетом Jmol, – (<http://www.jmol.org/>), Jmol, що містить скрипти для візуалізації сайтів зв'язування лігандів – Рис. 2.3.

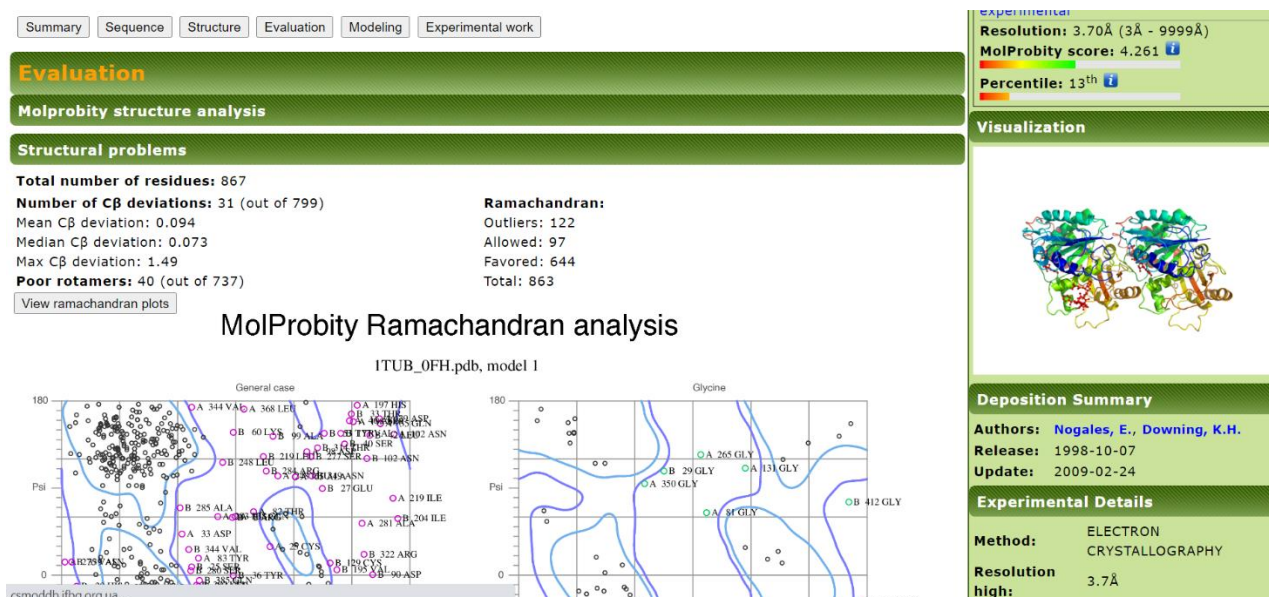


Рис. 2.1. Вкладка «Оцінка», що містить дані аналізу MolProbity.

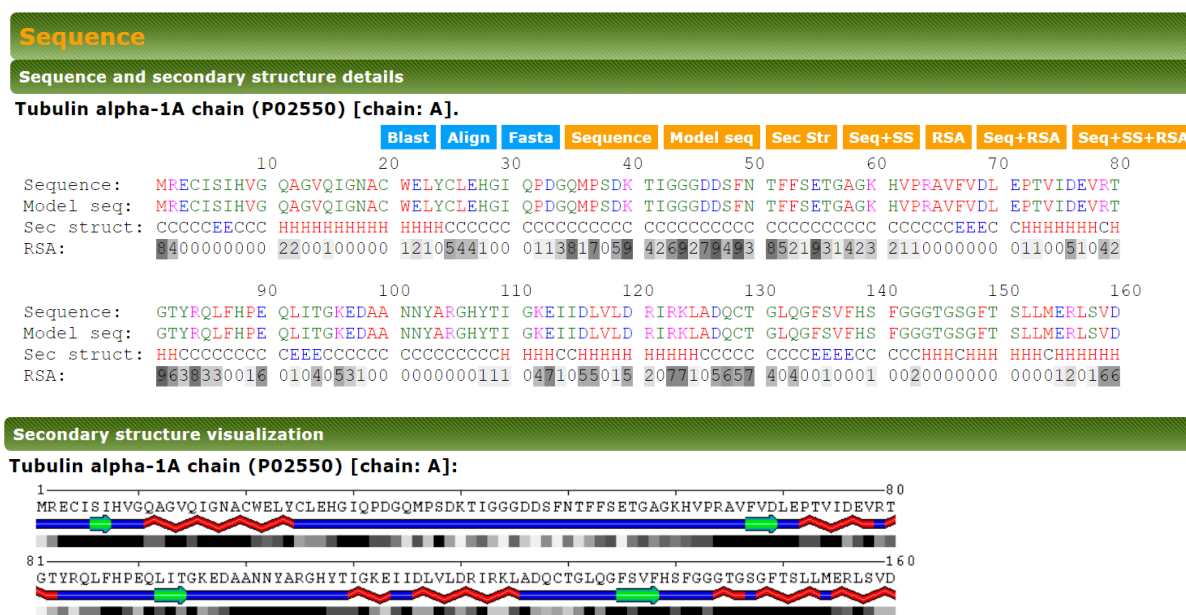
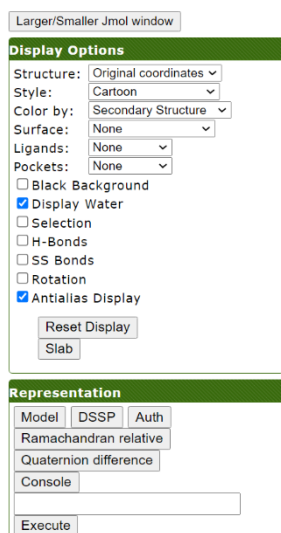


Рис. 2.2. Вкладка «Послідовність» – послідовність, мотиви, аналіз вторинної структури за допомогою програм DSSP та POLYVIEW-2D.



**Рис. 2.3.** Вкладка «Структура» з інтегрованим аплетом Jmol, – (<http://www.jmol.org/>), Jmol, що містить скрипти для візуалізації сайтів зв’язування лігандів.

- *Компонент TrajDB* – це база даних траєкторій молекулярної динаміки та похідні метадані. Компонент інтегрований з аналітичними інструментами програмного пакету GROMACS, – Hess et al., 2008. Конкретний набір проведених аналізів залежить від властивостей молекулярної системи.

- *Компонент ParamsAPI* служить для підготовки, збереження та управління наборами параметрів теоретичних експериментів. В даний час він підтримує моделювання за гомологією у Modeller, – Eswar et al., 2006, та програмні пакети молекулярної динаміки GROMACS.

Важливою проблемою при проведенні великої кількості обчислювальних експериментів є відстеження та настройка наборів параметрів для різних прогонів. Для полегшення підготовки експериментів та аналізу результатів цей компонент надає редактор параметрів, з опцією порівняння параметрів.

Набори параметрів зберігаються в базі даних та організовуються у “проекти”, які можуть містити дані параметрів для запуску різних послідовних завдань, забезпечуючи, таким чином, систему вбудованим механізмом планування робочого процесу (workflow system).

COMODORE інтегрований з високопродуктивною обчислювальною системою Ukrainian National Grid (<http://ung.in.ua>), завдяки чому обчислювальні завдання автоматично подаються із заздалегідь визначеними наборами параметрів з ParamsAPI, виконуються в Grid або на кластері і зберігаються в базі даних як записи у ModDB або TrajDB. Крім того, після завершення, набір визначених аналітичних процедур може бути автоматично виконаний на результатах обчислень і збережений як метадані, наприклад, – оцінка моделей MolProbity для завдання моделювання.

Таким чином, COMODORE дозволяє полегшити підготовку, аналіз та етапи управління даними високоефективних обчислювальних завдань порівняльного моделювання, моделювання молекулярної динаміки та стикування та уніфікувати ці процеси.

### **2.3 Менеджмент структурно-біологічних та генетичних даних генів білків пов'язаних з цитоскелетом та генів целюлозосинтаз**

Цитоскелет – високодинамічна структура клітин, функціонування якої забезпечує процеси мітозу, мейозу, цитокінезу, підтримки й адаптації форми клітин, екзо- та ендоцитоз, рух клітини, активний транспорт тощо, – Vignaud et al. 2012; Eren et al 2012. Завдяки здатності тубуліну специфічним чином зв'язувати низькомолекулярні сполуки різної природи мікротрубочки є відповідальними за реакцію еукаріотичних клітин на дію різноманітних хімічних чинників, – Sui et al. 2012; Zhao et al. 2012. Саме тому  $\alpha$ - і  $\beta$ -субодиниці тубуліну є важливими мішенями для широкого спектра комерційно значущих препаратів, що використовуються як протипухлинні, фунгіцидні, гербіцидні, антипротозойні, протигельмінтні засоби.

Для вирішення обчислювальних задач із дослідження цитоскелету у 2011 р. на основі Грід-вузлу Державної установи «Інституту харчової біотехнології і геноміки НАН України» була створена віртуальна організація (ВО)/лабораторія CSLabGrid (<http://ifbg.org.ua/uk/cslabgrid>), – Pydiura et al. 2011; Карпов та ін. 2015.

Ця ВО займається розв'язанням наукових та прикладних завдань у галузі дослідження цитоскелету, які потребують спільних зусиль і значних обчислювальних ресурсів як для методів біології *in silico* (структурна біологія, біоінформатика та біофізика), так і для завдань, пов'язаних із використанням сучасних інструментальних методів дослідження, що також потребують ресурсоємної обробки первинних даних та/або створення доступних (у межах віртуальної організації) баз даних для зберігання оброблених результатів.

У межах віртуальної організації було, також, створено базу даних CSMoDB (CytoSceleton protein MODelling DataBase; <http://csmodb.ifbg.org.ua/comodore/index.php>), – Pydiura et al. 2013, яка містить репозиторії 3D-моделей ряду цитоскелетних білків (тубулінів, білків, асоційованих з мікротрубочками (БАМів), філаментних температуро-чутливих Z-білків (*FtsZ*) тощо), а також просторових структур канонічних антимікротрубочкових речовин, придатну для подальшого використання в експериментах, що виконуються в рамках спільних досліджень учасників ВО CSLabGrid. Наповнення цієї бази даних інформацією щодо структур молекул тубулінів різного походження сприятиме прискоренню пошуку та відбору розробок нових сполук з антимікротрубочковою активністю.

У результаті аналізу онлайнної бази даних UniProtKB, – The UniProt Consortium 2015 було відібрано наявні в ній на момент пошуку повні первинні послідовності  $\alpha$ -,  $\beta$ -, та  $\gamma$ -тубулінів як модельних організмів, так і паразитичних простіших та круглих і плоских червів.

Так, було вибрано 128 послідовностей  $\alpha$ -тубулінів, які належать окремим представникам таких родів, як *Ancylostoma*, *Ascaris*, *Candida*, *Chlamydomonas*, *Clonorchis*, *Cooperia*, *Dictyocaulus*, *Echinococcus*, *Emericella*, *Encephalitozoon*, *Entamoeba*, *Fasciola*, *Haemonchus*, *Homo*, *Hymenolepis*, *Lepidoglyphus*, *Loa*, *Naegleria*, *Necator*, *Neospora*, *Oesophagostomum*, *Onchocerca*, *Ostertagia*, *Plasmodium*, *Pneumocystis*, *Pristionchus*, *Schistosoma*, *Spirometra*, *Strongyloides*, *Toxocara*, *Toxoplasma*, *Trichinella*, *Trichuris*, *Trypanosoma*, *Wuchereria*.

Також, було відібрано 197 послідовностей  $\beta$ -тубулінів, що належать представникам родів *Ajellomyces*, *Ancylostoma*, *Ascaris*, *Babesia*, *Brugia*, *Candida*, *Chlamydomonas*, *Clonorchis*, *Cooperia*, *Coronocyclus*, *Cyathostomum*, *Cylicocyclus*, *Cylicostephanus*, *Dictyocaulus*, *Dirofilaria*, *Echinococcus*, *Emericella*, *Encephalitozoon*, *Fasciola*, *Giardia*, *Haemonchus*, *Homo*, *Hymenolepis*, *Leishmania*, *Loa*, *Moniezia*, *Necator*, *Neosartorya*, *Oesophagostomum*, *Onchocerca*, *Ostertagia*, *Parascaris*, *Plasmodium*, *Pneumocystis*, *Pristionchus*, *Schistosoma*, *Spirometra*, *Strongyloides*, *Teladorsagia*, *Toxocara*, *Toxoplasma*, *Trichinella*, *Trichostrongylus*, *Trichuris*, *Trypanosoma*.

Було відібрано також лише 36  $\gamma$ -тубулінів із представників родів *Ancylostoma*, *Ascaris*, *Brugia*, *Candida*, *Chlamydomonas*, *Dictyocaulus*, *Echinococcus*, *Emericella*, *Encephalitozoon*, *Entamoeba*, *Haemonchus*, *Homo*, *Loa*, *Necator*, *Oesophagostomum*, *Onchocerca*, *Opisthorchis*, *Plasmodium*, *Pristionchus*, *Schistosoma*, *Strongyloides*, *Toxocara*, *Trichinella* *Trichuris* та *Wuchereria*.

Моделювання тривимірних (3D) структур цих білків тубулінів було здійснено за допомогою онлайн-серверу I-TASSER, – Roy et al. 2010, що призначений для передбачення структури і функції білків. Для всіх даних моделей було проведено оптимізацію їхньої геометрії в силовому полі amber03.

Для заповнення репозиторія тубулінів бази даних CSMoDB ВО CSLabGrid моделями цитоскелетних білків у неї було внесено інформацію щодо  $\alpha$ -тубулінів, з яких 73 належать паразитичним круглим червам, зокрема, таким видам, як *Ascaris suum*, *Toxocara canis*, *Pristionchus pacificus*, *Strongyloides ratti*, *Ancylostoma ceylanicum*, *A. duodenale*, *Necator americanus*, *Oesophagostomum dentatum*, *Cooperia oncophora*, *Dictyocaulus viviparous*, *Haemonchus contortus*, *Ostertagia ostertagi*, *Loa loa*, *Onchocerca volvulus*, *Wuchereria bancrofti*, *Trichuris suis*, *T. trichiura*, *Trichinella spiralis*; 32 – паразитичним плоским червам *Echinococcus granulosus*, *E. multilocularis*, *Hymenolepis diminuta*, *H. microstoma*, *Spirometra erinaceieuropaei*, *Fasciola hepatica*, *Clonorchis sinensis*, *Schistosoma haematobium*, *S. japonicum*, *S. mansoni*; 10 тваринам, зокрема, 8 тублінів людини та ще два належать *Lepidoglyphus destructor*; 8 тубулінів належать простішим

*Trypanosoma brucei rhodesiense*, *T. cruzi*, *Toxoplasma gondii*, *Neospora caninum*, *Plasmodium falciparum*, *P. yoelii*, *Naegleria gruberi*, *Entamoeba histolytica*; 4 тубуліни – грибам *Candida albicans*, *Emericella nidulans*, *Encephalitozoon cuniculi*, *Pneumocystis carinii* та 1 тубулін – водорості *Chlamydomonas reinhardtii*.

Зі 197  $\beta$ -тубулінів, внесених у репозиторій, 127 3D-моделей тубулінів належать паразитичним круглим червам *Ascaris suum*, *Parascaris equorum*, *Toxocara canis*, *Pristionchus pacificus*, *Strongyloides papillosus*, *Strongyloides ratti*, *Strongyloides stercoralis*, *Ancylostoma caninum*, *A. ceylanicum*, *A. duodenale*, *Necator americanus*, *Oesophagostomum dentatum*, *Coronocylus coronatus*, *Cyathostomum catinatum*, *C. pateratum*, *Cylicocyclus elongates*, *C. insigne*, *C. nassatus*, *Cylicocyclus radiates*, *Cylicostephanus goldi*, *C. longibursatus*, *Cooperia oncophora*, *C. pectinata*, *Dictyocaulus viviparous*, *Haemonchus contortus*, *Ostertagia ostertagi*, *Teladorsagia circumcincta*, *Trichostrongylus colubriformis*, *Brugia malayi*, *B. pahangi*, *Dirofilaria immitis*, *Loa loa*, *Onchocerca volvulus*, *O. gibsoni*, *Trichuris trichiura*, *Trichinella spiralis*; 43 тубуліни – паразитичним плоским червам *Echinococcus granulosus*, *E. multilocularis*, *Moniezia expansa*, *Hymenolepis microstoma*, *Spirometra erinaceieuropaei*, *Fasciola hepatica*, *Clonorchis sinensis*, *Schistosoma haematobium*, *S. japonicum*, *S. mansoni*; 9 тубулінів є людськими, а ще 9 належать простішим *Giardia intestinalis*, *Leishmania mexicana*, *Trypanosoma brucei rhodesiense*, *T. cruzi*, *Toxoplasma gondii*, *Babesia bovis*, *Plasmodium falciparum*; 8 тубулінів належать грибам *Neosartorya fumigata*, *Ajellomyces capsulatus*, *Candida albicans*, *Emericella nidulans*, *Encephalitozoon cuniculi*, *E. hellem*, *E. intestinalis* та *Pneumocystis carinii*, а 1 тубулін – *Chlamydomonas reinhardtii*.

Серед 36  $\gamma$ -тубулінів, задепонованих у CSModDB, 20 білків належать паразитичним круглим червам, зокрема, таким видам, як *Ascaris suum*, *Toxocara canis*, *Pristionchus pacificus*, *Strongyloides ratti*, *Ancylostoma ceylanicum*, *Necator americanus*, *Oesophagostomum dentatum*, *Dictyocaulus viviparus*, *Haemonchus contortus*, *Brugia malayi*, *Loa loa*, *Onchocerca volvulus*, *Wuchereria bancrofti*, *Trichuris suis*, *Trichinella spiralis*; 7 – паразитичним плоским червам *Echinococcus*



*granulosus*, *E. multilocularis*, *Opisthorchis viverrini*, *Schistosoma haematobium*, *S. japonicum* та *S. mansoni*; 2 тубуліни є людськими; 4 тубуліни з грибів *Candida albicans*, *C. glabrata*, *Emericella nidulans*, *Encephalitozoon cuniculi*; 2 білкові структури належать найпростішим *Plasmodium falciparum* та *Entamoeba histolytica*, а 1 тубулін – водорості *Chlamydomonas reinhardtii*.

У результаті внесення інформації щодо відповідних послідовностей тубулінів та депонування в форматі \*.pdb безпосередньо файлів тривимірних структур вищевказаних білків у базу даних CModDB ВО CSLabGrid було сформовано відповідний інтерфейс доступу, фрагмент якого показаний на Рис. 2.4. Цей інтерфейс значно спрощує зареєстрованим членам ВО CSLabGrid доступ до конкретних об'єктів серед значної кількості змодельованих цитоскелетних білків, оскільки обладнаний пошуком як за видовою назвою організму, так і за Uniprot-кодом первинної послідовності відповідного білка.

**CSModDB**  
CytoSkeleton protein  
MODelling DataBase

Models DB Parameters API Login  
Models

List sorting

Type:	Object:	Organism:	Classification:	Authors:	Uniprot:	Ligands:
exp mod	EB EB1 tubulin	Ajiellomyces capsulatus Ancylostoma caninum Ancylostoma ceylanicum Ancylostoma duodenale Aquifex aeolicus (strain VF5)	Cytoskeleton CipX EB EB1 EzrA	Accelerated Technologies Center for Gene to 3D Structure (ATCG3D) Aiwale, S.T. Alfonso, C. Allegro, D. Almo, S.C.	AGA016SYM3 AGA016SZV7 AGA016UP54 AGA016WZQ2 AGA024N126	01G ACP ACT ADP AGS

Submit Discard Search: Submit Discard

Sort by: Update date Resolution Mol Probtity Score Percentile Discard sorting

<b>1TUB_0</b> experimental Resolution: 3.70Å (3Å - 9999Å) MolProbtity score: 4.261 Percentile: 13 <sup>th</sup>	<b>TUBULIN ALPHA-BETA DIMER, ELECTRON DIFFRACTION</b> Authors: Nogales, E., Downing, K.H. Classification: Tubulin, Microtubules, Cytoskeleton Released/Updated: 1998-10-07 / 2009-02-24 Experiment: ELECTRON CRYSTALLOGRAPHY Compound: 2 Polymers 3 Ligands Organisms: Sus scrofa Citation: Structure of the alpha beta tubulin dimer by electron crystallography. (PubMed) Nogales, E., Wolf, S.G., Downing, K.H. – Nature, 1998. – DOI: 10.1038/34465
<b>3DU7_0</b> experimental Resolution: 4.10Å (3Å - 9999Å) MolProbtity score: 4.243 Percentile: 14 <sup>th</sup>	<b>Tubulin-colchicine-phomopsis A: Stathmin-like domain complex</b> Authors: Cormier, A., Marchand, M., Ravelli, R.B., Knossow, M., Gigant, B. Classification: Tubulin, Microtubules, Cytoskeleton Released/Updated: 2008-10-21 / 2011-07-13 Experiment: X-RAY DIFFRACTION Compound: 3 Polymers 5 Ligands Organisms: Bos taurus, Rattus norvegicus Citation: Structural insight into the inhibition of tubulin by vinca domain peptide ligands (PubMed) Cormier, A., Marchand, M., Ravelli, R.B., Knossow, M., Gigant, B. – Embo Rep., 2008. – DOI: 10.1038/embor.2008.171

Рис. 2.4. Фрагмент інтерфейсу бази даних ВО CSLabGrid із внесеними 3D-моделями тубулінів.

**Висновки.** Наповнення бази даних CSModDB інформацією щодо тривимірних структур таких цитоскелетних білків, як тубуліни, з різних таксономічних груп як паразитичних організмів, так і модельних об'єктів полегшує доступ до такої інформації усіх членів ВО CSLabGrid. Це в свою чергу пришвидшуватиме вирішення наукових та прикладних завдань у галузі дослідження цитоскелету, зокрема, пошук і відбір нових сполук з антимікротрубочковою активністю для подальшого їх застосування як протипухлинних, фунгіцидних, гербіцидних, антипротозойних та протигельмінтних засобів.

## 2.4 Аналіз експресії генів

**Рослинний матеріал.** Для дослідження було обрано чотири сорти льону: Блакіт, Білорусь та Аріан, Франція, що належать до підвиду *elongatum* Vav. et Ell. і використовується для виробництва високоякісного волокна; льону великого насіння *Endress Olajlen, subsp. mediterraneum* Vav. et Ell.; та льону-стрибунця, *subsp. crepitans* Boenn. Два останні істотно відрізняються від *subsp. elongatum* за морфологічними та біологічними ознаками, і вони не використовуються в текстильній промисловості. Рослини льону вирощували на експериментальних ділянках Центрального ботанічного саду Національної академії наук Білорусі у 2015 році, використовуючи звичайні методи. Насіння висівали в оптимальний час, у першій декаді травня – 05.08.2015, щоб забезпечити формування нормальних стебел. Саджанці льону отримували пророщуванням насіння *in vitro* на вологому фільтрувальному папері. Поперечні зрізи стебел льону фарбували флороглюцинол-НСІ (Day *et al.*, 2005).

**Екстракція РНК та синтез кДНК.** РНК виділяли з гіпокотилів десятиденних саджанців льону, з яких видаляли сім'ядолі та коріння, та зі стебел рослин, зібраних на етапі швидкого росту (через 4 тижні після сходів), за допомогою TRI REAGENT – Sigma, США. Контроль якості вилученої РНК

проводили електрофорезом зразків РНК у 1,5% агарозному гелі та спектрофотометричними вимірами.

До зворотної транскрипції зразки РНК обробляли ДНКазою. Реакційний об'єм становив 20 мкл, включаючи 10 мкл розчину РНК, 1 мкл ДНКазі1 – Thermo Scientific, 2 мкл 10× ферментного буфера і 7 мкл води, обробленої DEPC. Суміш інкубували протягом 1 години при 37 ° С, тоді як фермент інактивували додаванням 2 мкл 0,5 М EDTA з подальшою інкубацією протягом 15 хв при 65°C.

Синтез кДНК проводили за допомогою RevertAid First Strand cDNA Synthesis Kit – Thermo Scientific, відповідно до інструкцій виробника, використовуючи праймери oligo dT18. Для кожного зразка проводили контрольну реакцію NRT, суміш складалася з усіх компонентів, крім зворотної транскриптази. Позитивний контроль проводили за допомогою матриці РНК гена GAPDH, який входить до комплекта. Реакційну суміш без матриці використовували як негативний контроль.

**qPCR.** Полімеразну ланцюгову реакцію в режимі реального часу проводили в ПЛР-пробірках або на 96-лунковому планшеті – Thermo Scientific, Великобританія. Реакційний об'єм на зразок становив 20 мкл і містив 10 мкл 2× буфера-Т (S) для ДНК-полімерази – Primetech, Білорусь, нуклеотиди в кінцевій концентрації 0,2 мМ для кожного нуклеотида, прямий і зворотний праймер у кінцевій концентрації 0,2 мкМ, 50-100 нг кДНК, 0,1 одиниці ДНК-полімерази Tornado – Primetech, Білорусь та 0,16 мкл флуоресцентного барвника Zubr Green 100× – Primetech, Білорусь. В якості альтернативи був використаний набір SYBR® Green JumpStart™ Taq ReadyMix™ – Sigma-Aldrich, США. Кожна експериментальна реакція супроводжувалась реакціями контролю NRT та NTC. Праймери генів, використані в цій роботі, наведені у Таблиці 2.1. Гени GAPDH та ETIF1 використовувались як еталонні гени – Huis et al., 2010.

Таблиця 2.1

## Опис праймерів, використаних у дослідженні

Ген	ID мішені	Продукт гену	Назва праймеру	Послідовність, 5'→3'
<i>LusCesA1</i>	<i>Lus10018902/</i> <i>Lus10028597</i>	cellulose synthase 1	BA12F5mod4	GGAGCAAAAAGTACTCTGACAA AAG
			BA12R4	CGCTTCTCGAGACTTCTCTG
<i>LusCesA4</i>	<i>Lus10008226</i>	cellulose synthase 4	BS58F14	TACGGTTATGATCCACCAGTGTC
			BS58R14	CGCTTCCCTCCCATTTTCTTCTT
<i>LusCesA7</i>	scaffold464	cellulose synthase 7	BS19F3	CGGCCAAAGATGGTAA
			BS19R5	GCAGGGCTAGAGGATGGA
<i>LusCesA6</i>	<i>Lus10002939/</i> <i>Lus10002940</i>	cellulose synthase 9	BS13F3	ACGAGCACTTTATGAGCTAT
			BS13R2	GGTTTTGATTCTTCTCTCC
<i>ACT</i>	<i>GR508908</i>	actin	ACT-F	TCCAGGCCGTTCTTTCTCTA
			ACT-R	CTGTAAGGTCACGACCAGCA
<i>GAPDH</i>	<i>CV478202</i>	glyceraldehyde 3-phosphate dehydrogenase	GAPDH-F	AGGTTCTTCCCGCTCTCAAT
			GAPDH-R	CCTCCTTGATAGCAGCCTTG
<i>TUA</i>	<i>CA482867</i>	$\alpha$ -tubulin	TUA-F	CCTGTTGGGAGCTTTACTGC
			TUA-R	AAGGTGTTGAAGGCATCGTC
<i>ETIF1</i>	<i>GR508906</i>	eEF1a1	ETF1-F	CCTTGTAGGGCTGAGGGATT
			ETF1-R	CTCATCAAGACCACCAGCAA

Реакції ампліфікації проводили за допомогою термоциклів CFX96 та ПЛР у реальному часі IQCycler – BioRad, США. Програма ПЛР в режимі реального часу з ферментом Tornado була такою: (1) 95,0° – 15:00 хв, (2) 99,0° – 1 с, (3) 58,0° – 10 с, (4) 72,0° – 10 с, (5) зчитування пластини, повторить дії ( 2) - (5) 44 рази, (6) 95,0° – 10с, (7) крива плавлення від 65°C до 95°C з кроком 0,1°C. Для набору SYBR® Green JumpStart™ Taq ReadyMix™ – Sigma-Aldrich, США, програма була змінена відповідно до інструкції з набору.

Для кожної реакції висока специфічність ампліфікації підтверджувалася наявністю одного піка на кривій плавлення. Аналіз експресії проводили за допомогою методу  $\Delta\Delta C_t$ , – Livak and Schmittgen, 2001.

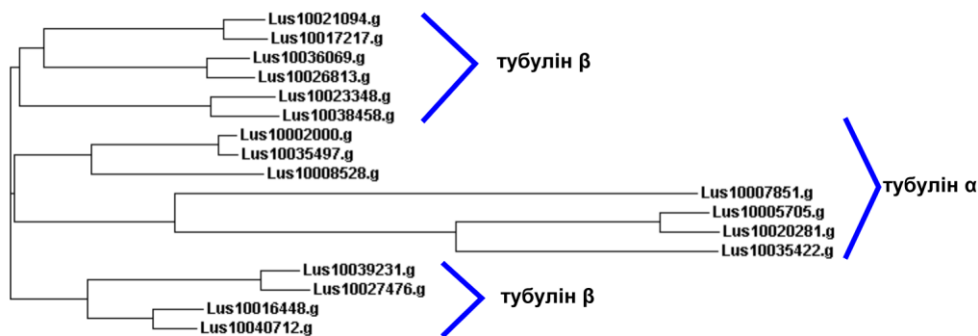
## РОЗДІЛ 3

### ПОШУК ТА АНАЛІЗ ГЕНІВ ЦИТОСКЕЛЕТНИХ БІЛКІВ У ГЕНОМІ ЛЬОНУ

#### 3.1 Пошук генів тубулінів у геномі льону

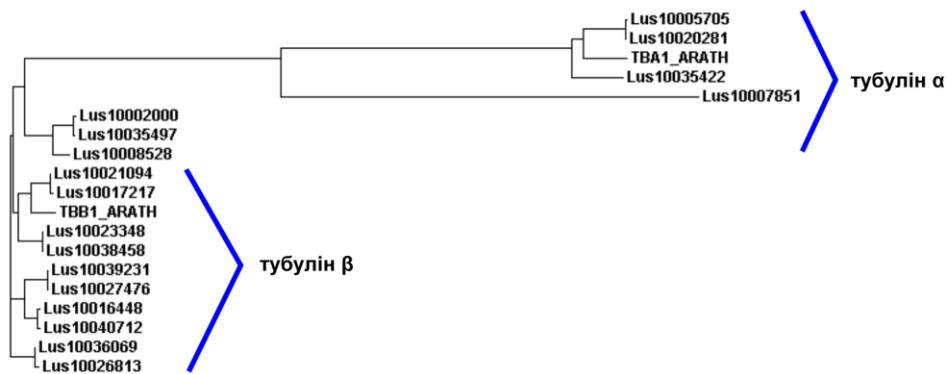
У базі даних UniProt (<http://www.uniprot.org/help/sequence-alignments>) на даний момент не міститься жодної достовірно анотованої послідовності  $\alpha$ - чи  $\beta$ -тубулінів, закодованих у геномі льону (*Linum usitatissimum*). Як правило, у геномах вищих рослин міститься від 3 до 11 ізоформ  $\alpha$ - і  $\beta$ - субодиниць тубулінів. Так, наприклад, у геномі різушки Таля – *A. thaliana*, який є досить добре дослідженим, анотовано 6 ізоформ  $\alpha$ - тубуліну і 9 ізоформ  $\beta$ - тубуліну, – Korczak et al. 1992; Snustad et al. 1992. Середня кількість амінокислот у послідовностях  $\alpha$ - і  $\beta$ -тубуліну вищих рослин складає близько 445–450 амінокислот, – Vlume et al. 2008. Відомо, що  $\alpha$ - і  $\beta$ -тубуліни філогенетично мають суттєві відмінності, тому для пошуку генів тубуліну у геномі льону нами було взято по одній послідовності  $\alpha$ - і  $\beta$ -тубулінію *A. thaliana*, а саме – ізоформи alpha-1 (*TBA1\_ARATH*) і beta-1 (*TBB1\_ARATH*). У результаті проведеного дослідження, у геномі льону було виявлено 17 генів, які кодують послідовності тубулінів. До них віднесено гени, які закодовані у локусах *Lus10021094*, *Lus10017217*, *Lus10036069*, *Lus10026813*, *Lus10002000*, *Lus10035497*, *Lus10008528*, *Lus10023348*, *Lus10038458*, *Lus10039231*, *Lus10027476*, *Lus10016448*, *Lus10040712*, *Lus10007851*, *Lus10005705*, *Lus10020281* і *Lus10035422*. Середня кількість амінокислот у послідовностях, які транслюються на основі даних генів тубулінів льону, складає близько 450 амінокислот, вказані гени мають невеликий розмір і подібну структуру. На основі філогенетичної подібності генів, що кодують  $\alpha$ - і  $\beta$ -тубуліни льону, виділено групи найбільш подібних послідовностей, – Рис. 3.1. На основі цих даних зроблено припущення, що у

геномі льону може бути закодовано до 7 ізоформ тубуліну  $\alpha$ - і до 10 ізоформ  $\beta$ -тубуліну.



**Рис. 3.1.** Філогенетичне дерево подібності визначених нами генів, що кодують  $\alpha$ - і  $\beta$ -тубуліни льону.

Далі, до множинного вирівнювання було включено по одній ізоформі  $\alpha$ - і  $\beta$ -тубулінів *A. thaliana* – ті ж послідовності, що брали для пошуку генів льону. У результаті вирівнювання виявлених генів  $\alpha$ - і  $\beta$ -тубуліну льону, трансльованих у амінокислотні послідовності та десяти типів  $\alpha$ - і  $\beta$ -тубуліну *A. thaliana* виділено дві групи генів тубуліну льону, які виявляють значну подібність до анотованих фрагментів  $\alpha$ - і  $\beta$ -тубулінів *A. thaliana* з бази даних UniProt, – Рис. 3.2, тобто, можна виділити локуси, що кодують тубуліни різних типів. Так, генетичні локуси *Lus10007851*, *Lus10005705*, *Lus10020281* і *Lus10035422* з високою долею ймовірності кодують ізоформи тубуліну  $\alpha$ , а локуси *Lus10021094*, *Lus10017217*, *Lus10036069*, *Lus10026813*, *Lus10023348*, *Lus10038458*, *Lus10039231*, *Lus10027476*, *Lus10016448* і *Lus10040712* кодують ізоформи  $\beta$ -тубуліну. Локуси *Lus10002000*, *Lus10035479* і *Lus10008528* найбільш ймовірно кодують послідовності ізоформ  $\alpha$ -тубуліну. Нами було висунуте припущення, що геном льону може містити 7 ізоформ  $\alpha$ -тубуліну і 10 ізоформ  $\beta$ -тубуліну.



**Рис. 3.2.** Філогенетичне дерево подібності виявлених генів, що кодують  $\alpha$ - і  $\beta$ -тубуліни льону, по одній ізоформі  $\alpha$ - і  $\beta$ -тубулінів *A. thaliana* – ті ж послідовності, які брали для пошуку генів льону – *TBA1\_ARATH* і *TBB1\_ARATH*.

Для всіх даних генів отримані генетичні та трансльовані амінокислотні послідовності. Виявлено, що дані гени мають у своїй послідовності 3 екзони і 2 інтрони. Використовуючи отримані амінокислотні послідовності та онлайн програму [Primer3: WWW primer tool](http://biotools.umassmed.edu/bioapps/primer3_www.cgi) ([http://biotools.umassmed.edu/bioapps/primer3\\_www.cgi](http://biotools.umassmed.edu/bioapps/primer3_www.cgi)) було проведено дизайн праймерів, які в подальшому були використані для оцінки експресії генів, що кодують  $\alpha$ - і  $\beta$ -тубуліни льону.

Таким чином, нами передбачено 17 генів тубуліну льону. За гомологією амінокислотних послідовностей до анотованих послідовностей *A. thaliana* гени льону вдалося віднести до 7 типів ізоформ  $\alpha$ -тубуліну і 10 ізоформ  $\beta$ -тубуліну. Пізніше, ці дані було скориговано наступним чином – повногеномний аналіз *in silico* сімейства генів тубуліну льону виявив наявність шести  $\alpha$ -, 13  $\beta$ - та двох локусів  $\gamma$ -тубуліну, – Rudiura *et al.*, 2018. На основі аналізу послідовності всі ці гени, крім одного, можна згрупувати в пари паралогів зі значним ступенем ідентичності, що, ймовірно, є результатом останньої повногеномної дуплікації, яка відбулась у геномі льону. Такі гени називають онологами.

Ми ідентифікували 6 генів (3 паралогічні пари), що, найімовірніше, кодують  $\alpha$ -тубуліни, у геномі *L. usitatissimum* (льон), локуси: *Lus10035422* – *Lus10031032*, *Lus10013765* – *Lus10039169* та *Lus10020281* – *Lus10005705*, – Таблиця 3.1, стовпці 1–4. Кількість генів  $\beta$ -тубуліну необхідно додатково скоригувати до 14,



як результат експериментів з клонування «вологих» кДНК, – Gavazzi *et al.* 2017, – Таблиця 3.2. У геномі льону було виявлено два локуси, пов’язані з  $\gamma$ -тубуліном.

Таблиця 3.1

Гени  $\alpha$ -тубуліну, ідентифіковані у геномі *L. usitatissimum* з бази даних

**Phytozome**

#	Ген. локус	Пар нукл.	К-сть екзонів	К-ість а.з.	Пеп./Ген. ідентичн., %	Ізотип А. <i>thaliana</i>	Ідентичність до А. <i>thaliana</i> , %
	-1-	-2-	-3-	-4-	-5-	-6-	-7-
<b>1</b>	<i>Lus10035422</i>	1995	4	450	100 /	$\alpha 2, \alpha 4$	96.44
					93.83	$\alpha 6$	96.22
<b>2</b>	<i>Lus10031032</i>	2010	4	450	100 /	$\alpha 2, \alpha 4$	96.44
					90.90	$\alpha 6$	96.22
<b>3</b>	<i>Lus10013765</i>	2107	5	450	100 /	$\alpha 3, \alpha 5$	97.11
<b>4</b>	<i>Lus10039169</i>	2110	5	450	90.90		
<b>5</b>	<i>Lus10020281</i>	2027	5	451	99.78 /	$\alpha 1$	92.00
<b>6</b>	<i>Lus10005705</i>	2049	5	451	93.06		

Примітка: ідентичність між паралогічними генами – стовпець 5; відповідний гомологічний ізотип  $\alpha$ -тубуліну *A. thaliana* – стовпець 6; відсоток ідентичності передбачуваного пептидного продукту гена льону з відповідним ізотипом  $\alpha$ -тубуліну *A. thaliana* – стовпець 7.

### 3.2 Пошук генів актинів у геномі льону

У депонованій версії геному льону виявлено 15 можливих генів актину: *Lus10006783*, *Lus10005820*, *Lus10006784*, *Lus10005819*, *Lus10005163*, *Lus10001693*, *Lus10001693*, *Lus10001694*, *Lus10004956*, *Lus10005457*, *Lus10016558*, *Lus10040826*, *Lus10029286*, *Lus10016259*, *Lus10004169*, *Lus10021057* – Таблиця 4.

У геномній послідовності локуса *Lus10001693* у базі даних Phytozome відсутній екзон 1 – 60 п.н., інтрон 1 – 321 п.н. та частина екзону 2 – 75 п.н. Але дані фрагменти присутні в послідовності генома серед не віднесених ділянок і були ідентифіковані за допомогою пошуку BLAST. Вони на 100% ідентичні *Lus10001694*. Послідовність цього гена із знайденими фрагментами буде вказана як *Lus10001693\**.

Геномні дані свідчать про те, що *Lus10004956* може мати два альтернативні варіанти залишків фрагмента з 20 а.з. – 198-217 в середині екзону 3. Ми будемо позначати ці два можливі варіанти цього гена як *Lus10004956\** – <sup>198</sup>RGYSFTTTAERGNVRDMKEK<sup>217</sup> і *Lus10004956\*\** – <sup>198</sup>RGYSFTPTAEREIVRDMKEK<sup>217</sup>.

Передбачувана білкова послідовність *Lus10006784* містить фрагмент з 11 а.з. <sup>152</sup>GKGLVMATWHK<sup>162</sup> на 3'-кінці екзону 2, якого немає в інших актинах льону. Ми припускаємо, що це може бути трансльована послідовність інтрону 2 через невпізнану пограничну екзон / інтронну послідовність <sup>542</sup>CGGTA<sup>546</sup>, – Hanley and Schuler, 1988, і ми розглядаємо нашу інтерпретацію даних цього гена як *Lus10006784\**.

Дані бази даних Phytozome вказують на те, що інтрон 3 локусів *Lus10004169* та *Lus10021057*, можливо, зазнав екзонізації, і, отже, екзон 4, що кодує фрагмент С-кінцевої ділянки білка з 21 а.з., може бути трансльован зі зміщенням рамки зчитування. Подібно до тубулінів, С-кінець актинів є неструктурованою областю, що бере участь у формуванні інтерфейсу між актиновими мономерами в актинових філаментах (Holmes 1990, Kim 2000) і утворює, також, сайти взаємодії для ряду актинозв'язуючих білків, – оглянуто у (dos Remedios, 2003). Пошук за гомологією за допомогою BLASTP у базі даних UniProt проти всіх наявних даних, виявив подібні С-кінцеві послідовності у *Populus tomentosa* – L0ASD9\_POPTO, *Dichantheium oligosanthes* – A0A1E5W3K8\_9POAL та *Gossypium raimondii* – A0A0D2TK24\_GOSRA. Чи дійсно інтрон 3 зазнає екзонізації у локусах *Lus10004169* та *Lus10021057* льону, а також інших видів, і

чи відповідають ці локуси функціонально активним генам і експресуються – це питання подальшого дослідження.

**Висновки.** Наші дані показують, що гени актину широко представлені у геномі льону. За допомогою досліджень *in silico*, проведеним на першому драфті зборки геному льону, ми виявили 15 генів актину. П'ятнадцять генів актину згруповані в 7 паралогічних пар – 7 ізотипів актину та послідовно дуплікована копія одного з генів одного з ізотипів.

### **3.3 Біоінформаційне дослідження білків, асоційованих з мікротрубочками вищих рослин та аналіз їх подібності**

Пошук послідовностей кандидатів у геномах відібраних рослин було проведено за допомогою інструменту TBLASTN проти бази даних “Nucleotide collection”, використовуючи анотовані послідовності *A. thaliana* та *O. sativa* як вхідні дані. Множинне вирівнювання, курування вирівнювання, побудова, візуалізація та аналіз філогенетичних дерев виконано за допомогою веб-сервісу Phylogeny.fr ([www.phylogeny.fr](http://www.phylogeny.fr)).

За допомогою інструменту TBLASTN нами було виконано пошук специфічних рослинних БАМ у геномах 6 родів вищих рослин та виконано аналіз їх подібності. Досліджені БАМ належать до трьох родів класу дводольних: ріпак (*Brassica*), соя (*Glycine*) та люцерна (*Medicago*) і трьох родів класу однодольних: ячмінь (*Hordeum*), пшениця (*Triticum*) та кукурудза (*Zea*), які мають сільськогосподарське значення. Для пошуку гомологів були використані анотовані послідовності БАМ родин GCP (Gamma-tubulin Complex Proteins), MAP215/MOR1, MAP65, MAP70, EB1, SPR1 та TOR1/SPR2 з *Arabidopsis thaliana* для дводольних і *Oryza sativa* для однодольних.

Було з'ясовано що білки GCP2, GCP3, GCP4, GCP6 із представників родів *Medicago*, *Hordeum*, *Triticum*, та *Zea* мають ідентичність амінокислотних послідовностей >70%. Білки GCP5 із представників *Medicago* і *Zea* мають виражену гомологію, але у *Hordeum* і *Triticum* було знайдено лише невеликі

консенсусні ділянки. Це може пояснюватись тим, що даний білок є найбільш варіабельним представником родини GCP – у геномах *A. thaliana* і *O. sativa* він предсталений двома генами, а не одним, як інші представники даної родини. Цікаво, що у геномах *Brassica* і *Glycine* відсутні гомологи БАМ даної родини окрім наявності невеликого консенсусного фрагменту до білку GCP6 що є найкоротшим із родини.

У *O. sativa* білки MOR1 представлені двома зв'язаними генами – 1998 і 326 а.з. Гомологи MOR1 виявлено у родах *Medicago* (85% подібність, 96% ідентичність і 61% подібність, 68% ідентичність) і *Zea* (33% подібність, 62% ідентичність і 87% подібність, 63% ідентичність).

Група гомологічних білків родини MAP65 можна розділити за довжинами амінокислотних послідовностей 2 групи: 1) *MAP65-1*, *MAP65-2* та *MAP65-5–9* (550-610 а.з.) і 2) *MAP65-3*, *MAP65-4* (670-710 а.з.). Гомологи більш коротких білків групи 1 були знайдені у всіх представників досліджених родів (ідентичність 40-90%, подібність 90-100%). При цьому виявилось що консенсусні ділянки білків *MAP65-5–9* із представників роду *Brassica* на 20-30% коротші за гомологічні ділянки інших 5 родів (подібність 70-80%). Гомологи аналогічної довжини до білків *MAP65-3* і *MAP65-4* були знайдені лише у геномах роду *Zea* (98% подібність, 51% ідентичності і 99% подібності, 44% ідентичності відповідно).

Гомологи до білків родини *MAP70* були знайдені (окрім *Triticum*) у геномах 5 з досліджених родів (подібність 60-100%, ідентичність 60-100%). Гомологи +TIPs білків родин EB1 та SPR1 знайдено також (окрім *Triticum*) у геномах 5 з досліджених родів. Ідентичність складала 50-90% і 35-80% відповідно при подібності 90-100% у обох випадках.

БАМ родини TOR1/SPR2 були знайдені (окрім *Brassica* та *Triticum*) у 4-х досліджених геномах. Подібність послідовності складає від 48% (*Hordeum*) до 94% (*Medicago*), ідентичність складає від 30% (*Glycine*) до 68% (*Medicago*).

**Висновки.** Проведено оцінку подібності специфічних рослинних БАМ у геномах 6 родів вищих рослин. Відсутність гомологів родини GCP і MOR1 у

геномах *Brassica* та *Glycine*, MOR1 у геномах *Medicago* і *Hordeum*, MAP70 та +TIPs родин білків EB1 та SPR1 у геномі роду *Triticum* і TOR1/SPR2 у геномах *Brassica* та *Triticum*, що потребує додаткового дослідження.

### 3.4 Аналіз генів цитоскелетних білків у геномі льону

#### 3.4.1 $\alpha$ -тубуліни

Ми виявили 6 генів (3 паралогічні пари) у геномі *L. usitatissimum* (льон), що, найімовірніше, кодують  $\alpha$ -тубуліни, локуси: *Lus10035422* – *Lus10031032*, *Lus10013765* – *Lus10039169* та *Lus10020281* – *Lus10005705*, – Таблиця 3.1, стовпці 1–4. Всі кодовані цими генами льону білки TUA містять С-кінцевий залишок Тур (Y-тип), як і більшість ідентифікованих рослинних білків TUA.

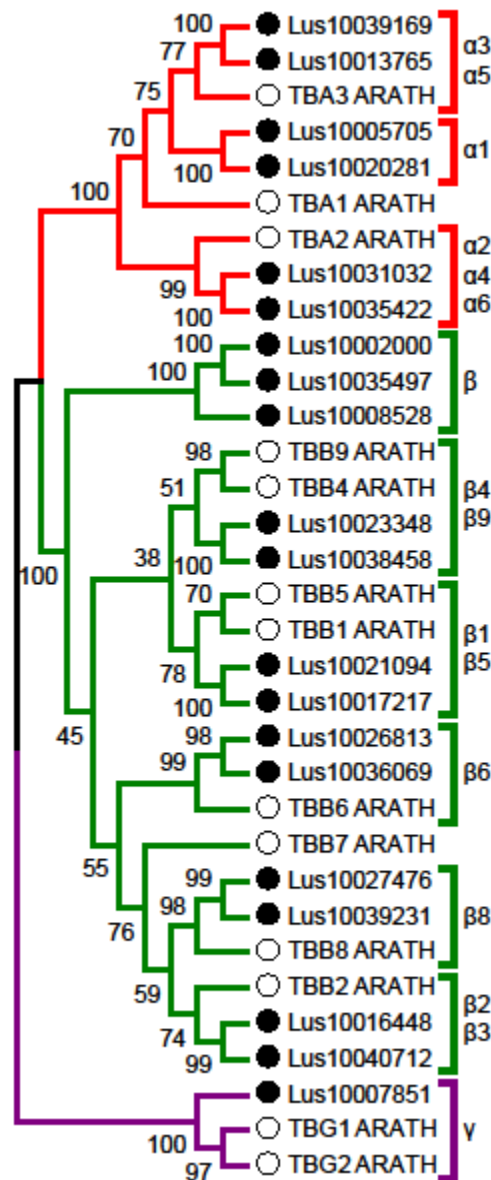
Аналіз ідентичності послідовностей пептидних продуктів цих шести генів  $\alpha$ -тубуліну, – Таблиця 3.1, колонки 1, 5, – показує, що пептидні продукти, кодовані локусами *Lus10035422*, *Lus10031032*, ідентичні таким локусам *Lus10013765*, *Lus10039169*. Послідовності ДНК локусів *Lus10035422* та *Lus10031032* ідентичні на 93,83% у зв'язку з наявністю п'яти пропусків (або вставки, або делеції) у вирівнюванні у послідовності в *Lus10031032* – один довжиною 9 п.н., два 2 п.н. і два 1 п.н. Крім того, послідовності ДНК *Lus10013765* та *Lus10039169* ідентичні на 90,90%, але *Lus10013765* має п'ять проміжків у вирівнюванні у порівнянні з *Lus10039169* – 21, 14, 9, 5 та 2 п.н., а *Lus10039169* – три проміжки – 27, 18 та 10 п.н.

Пептидні послідовності двох останніх локусів *Lus10020281* та *Lus10005705* ідентичні, за винятком одного амінокислотного поліморфізму (SAP) 444E/D, присутнього на мінливому змінному С-кінці білка. Ця заміна спричинена поліморфізмом одного нуклеотиду (SNP), що відбувається в третьому нуклеотиді триплета. Таким чином, пептидні продукти локусів *Lus10020281* та *Lus10005705* слід вважати майже однаковими. Послідовності ДНК *Lus10020281* та *Lus10005705* на 93,06% ідентичні з шістьма пробілами у вирівнюванні

послідовності *Lus10005705* – 14, 9, 3, 2 та два 1 п.н.; і п'ять розривів у вирівнюванні у *Lus10020281* – 5, і чотири 1 п.н.

У свою чергу, ідентичність амінокислотних послідовностей серед цих трьох різних ізотипів, кожен з яких представлений паралогічною парою генів, така: *Lus10035422* на 92% ідентичний *Lus10039169* та на 90,44% ідентичний *Lus10020281*, тоді як *Lus10039169* та *Lus10020281* ідентичні один одного на 94,22%.

Порівняння передбачуваної амінокислотної послідовності трьох ідентифікованих ізотипів  $\alpha$ -тубуліну льону з послідовністю анотованих ізотипів  $\alpha$ -тубуліну *A. thaliana* – Рис. 3.3; Таблиця 3.1, стовпці 1, 6 та 7 свідчить про наступне. Пара генів-онологів *Lus10020281* і *Lus10005705* є гомологічною ізотипу  $\alpha 1$  тубуліну альфа *A. thaliana* (клас альфа тубулінів IIb); тоді як пара генів-онологів *Lus10039169* та *Lus10013765* є гомологічною до ізотипів  $\alpha 3$ , так і  $\alpha 5$ , які ідентичні у *A. thaliana* (клас IIa). Подібним чином, онологи *Lus10035422* і *Lus10031032* льону є гомологічними ізотипам  $\alpha 2$  і  $\alpha 4$  тубуліну альфа, які ідентичні у *A. thaliana* і на 99,56% ідентичні ізотипу  $\alpha 6$ -тубуліну *A. thaliana* (клас Ia).



**Рис. 3.3.** Філогенетичне порівняння послідовностей пептидного продукту 6 генів  $\alpha$ -тубуліну, 13 генів  $\beta$ -тубуліну та 1 гена  $\gamma$ -тубуліну, ідентифіковані у геномі *L. usitatissimum* (●), із анотованими генами тубуліну *A. thaliana* (○).

Паралогічні пари генів  $\alpha$ -тубулінів льону *Lus10013765* – *Lus10039169* (клас IIa) та *Lus10020281* – *Lus10005705* (клас IIb) містять 4 інтрони в положеннях 37-2, 109-1, 176-0, 345-2, відносно кодону ініціації ATG – Рис. 3.4, що є типовим для генів  $\alpha$ -тубуліну класу II, наприклад у *A. thaliana* (Perumal, 2005), *Zea mays* – *GRMZM2G083243* та *Oryza sativa* – *LOC\_Os07g38730*. У свою чергу, паралогічна пара генів *Lus10035422* – *Lus10031032* (клас Ia) має 3 інтрони в положеннях 31-0, 109-1 та 233-0, відносно кодону ініціації ATG, що є типовим для генів  $\alpha$ -

тубуліну класу I, наприклад у генів *A. thaliana*, *Pisum sativum*, *O. sativa* та *Z. mays* (Perumal, 2005). Всі інтрони генів  $\alpha$ -тубуліну льону мають ідентичні канонічні послідовності GT-AG у донор-акцепторних сайтах сплайсингу.

**Висновки.** Сімейство генів  $\alpha$ -тубуліну складається з шести генів (Pydiura et al., 2018), організованих на геномному рівні, як це зазвичай зустрічається у багатьох інших видів рослин, у вигляді екзонів, розділених у консервативних положеннях трьома або чотирма інтронами, залежно від відповідного класу  $\alpha$ -тубуліну (Breviario et al. 2013). Відповідно, дві паралогічні пари генів  $\alpha$ -тубуліну *Lus10013765* – *Lus10039169* та *Lus10020281* – *Lus10005705* містять чотири інтрони у положеннях 37-2, 109-1, 176-0, 345-2, відносно кодону ініціації ATG, як це характерно для представників класу II тоді як третя пара *Lus10035422* – *Lus10031032* містить три інтрони в положеннях 31-0, 109-1 та 233-0, відносно кодону ініціації ATG, як це характерно для переважної більшості генів  $\alpha$ -тубуліну, що належать до класу I. Всі інтрони генів  $\alpha$ -тубуліну льону мають ідентичні канонічні послідовності GT-AG у донор-акцепторних сайтах сплайсингу.

### 3.4.2 $\beta$ -тубуліни

**Ідентифікація генів  $\beta$ -тубуліну.** У базі даних Phytozome v13 ми ідентифікували 13 генів  $\beta$ -тубуліну у геномі *L. usitatissimum* – Таблиця 3.2. Аналіз пептидних послідовностей, кодованих цими генами  $\beta$ -тубуліну, – Таблиця 3.2, стовпець 5, показує, що ті послідовності, які відповідають парам локусів *Lus10038458* – *Lus10023348* і *Lus10039231* – *Lus10027476* були попарно однаковими. Однак послідовності ДНК локусів *Lus10039231* та *Lus10027476* ідентичні на рівні 93,62%. Відмінності зумовлені численними нуклеотидними замінами та наявністю чотирьох пропусків у вирівнюванні *Lus10027476* – 16, 9, 5 та 1 п.н. та чотирьох пропусків у *Lus10039231* – 17, 4, 2 та 1 п.н. Подібним чином, хоча пептидні амінокислотні послідовності локусів *Lus10038458* та *Lus10023348* і ідентичні, їх послідовності ДНК мають лише 89,74% ідентичності.



Спостережувані відмінності включають численні нуклеотидні заміни та наявність трьох пропусків у вирівнюванні *Lus10038458* – 18, 12 та 1 п.н. та десять пропусків у *Lus10023348* – 131, 69, 18, 10, 7, 3 та чотирьох пропусків в 1 п.н.

Амінокислотні послідовності пептидів, кодовані локусами *Lus10002000* та *Lus10035497*, ідентичні, за винятком одного одиничного амінокислотного поліморфізму (SAP) <sup>440</sup>E/D на варіабельному С-кінці білка. Примітно, що <sup>440</sup>D також присутній у *Lus10008528*, тоді як <sup>440</sup>E зберігається у всіх інших тубулінах льону, але загалом даний SAP є незначним. Послідовності ДНК локусів *Lus10002000* та *Lus10035497* мають значну ідентичність у 98,11% з єдиним пропуском у 5 п.н. у *Lus10002000*.

Пептидні продукти, кодовані локусами *Lus10021094* та *Lus10017217*, ідентичні один одному, за винятком одного SAP <sup>263</sup>P/T у *Lus10017217*. <sup>263</sup>P є консервативним у всіх інших аналізованих β-тубулінах льону. Повні послідовності ДНК даних генів на 93,30% ідентичні, з чотирма пропусками у *Lus10017217* – 38, 7, 3 та 1 п.н. та п'ятьма проміжками у *Lus10021094* – 21, та чотирма пропусками 1 п.н. Примітно, що другі інтрони *Lus10017217* та *Lus10021094* із 102 п.н. ідентичні на 100%. Генетичні варіації між цими двома локусами накопичувались переважно у довшому інтроні 1, – 983 та 959 п.н., відповідно.

Таким же чином пептидні продукти *Lus10036069* та *Lus10026813* ідентичні, за винятком однієї консервативної заміни SAP <sup>437</sup>E/D та інделу двох залишків глютамінової кислоти, виявлених на С-кінці білка, багатому глютаміновою кислотою. Їх послідовності ДНК на 93,88% однакові, з одним пропуском у 7 п.н. у *Lus10036069* та трьома пропусками в *Lus10026813* – 6, 5 та 1 п.н.

У свою чергу, пептидні продукти локусів *Lus10016448* та *Lus10040712* ідентичні, за винятком двох одиничних амінокислотних поліморфізмів, SAP, <sup>443</sup>F/Y та <sup>450</sup>E/Q, що зустрічаються на змінному С-кінці білка. Їх відповідні послідовності ДНК на 95,09% ідентичні, з одним пропуском у 5 п.н. у *Lus10016448* та одним розривом 1 п.н. у *Lus10040712*.

Таким чином, серед 13 генів  $\beta$ -тубуліну, знайдених у даних про геном льону бази даних Phytozome v13, можна виділити 7 різних ізотипів, 6 з яких кодуються парою паралогічних генів та один ген-одиначку. У той же час в UniProtKB (<http://www.uniprot.org/>) 14 послідовностей  $\beta$ -тубуліну льону вже були анотовані групою D. Breviaro – Gavazzi, 2017, які належать до 7 різних ізотипів відповідно до Wang, 2012. Отже, ми проаналізували відповідність між білковими амінокислотними послідовностями цих двох джерел даних - Таблиця 3.2, стовпці 1, 6 та 7. Вищезазначена унікальна заміна  $^{263}\text{P/T}$  у *Lus10017217* відсутня ні в одній з послідовностей UniProtKB. Амінокислотні послідовності *Lus10035497*, *Lus10002000* та *TUB\_2b* ідентичні, тоді як *TUB\_2a* має дві консервативні заміни на варіабельному С-кінці  $\beta$ -тубуліна:  $^{438}\text{D/E}$  та  $^{448}\text{H/Q}$ . Консервативна заміна  $^{7}\text{I/V}$  присутня у *Lus10008528*, *TUB\_3a/b*, *TUB\_5*, *TUB7c* та *TUB\_7b*. У *A. thaliana*  $^{7}\text{V}$  присутні у  $\beta 1$ , в інших ізотипах  $^{7}\text{I}$  зберігається.

Таблиця 3.2

Гени  $\beta$ -тубуліну, виявлені у геномі *L. usitatissimum*.

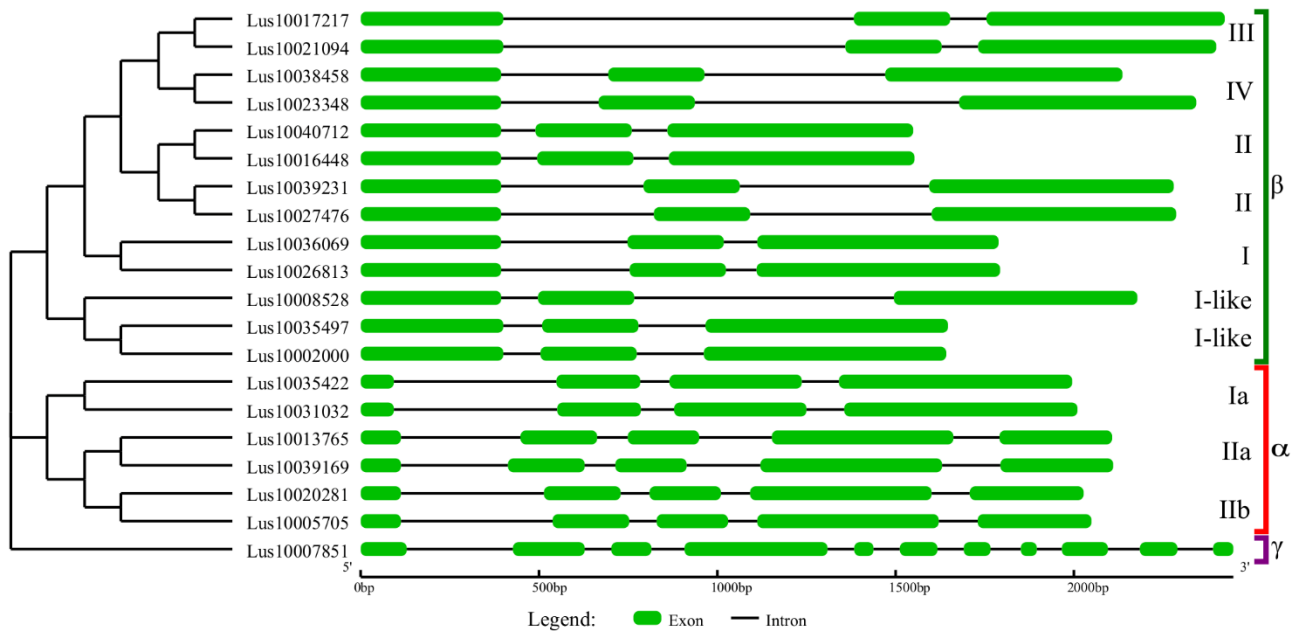
#	Ген. локус	Пар нукл.	К-сть екзонів	К-сть а.з.	Пеп./Ген. ідент., %	Gavazzi, 2017	Пеп. ідент. до -6-, %	Ізотип <i>A. thaliana</i>	Ідент. до <i>A. th.</i> , %
	-1-	-2-	-3-	-4-	-5-	-6-	-7-	-8-	-9-
1	<i>Lus10039231</i>	2280	3	449	100.00 /	<i>TUB_1a</i> ,	100.00	$\beta 7$	96.96
2	<i>Lus10027476</i>	2287	3	449	93.62	<i>TUB_1b</i>		$\beta 8$	96.65
3	<i>Lus10035497</i>	1647	3	449	99.78 /	<i>TUB_2a</i>	99.78	$\beta 6$	91.44
4	<i>Lus10002000</i>	1642	3	449	98.11	<i>TUB_2b</i>	100.00	$\beta 6$	91.87
5	<i>Lus10008528</i>	2178	3	448	-	<i>TUB_5</i>	100.00	$\beta 6$	93.02
6	<i>Lus10017217</i>	2423	3	445	99.78 /	<i>TUB_3a</i> ,	99.55	$\beta 1$	95.49
								$\beta 5$	94.36
7	<i>Lus10021094</i>	2399	3	445	93.30	<i>TUB_3b</i>	99.78	$\beta 1$	95.71
								$\beta 5$	94.58
8	<i>Lus10038458</i>	2136	3	442	100.00 /	<i>TUB_4a</i> ,	100.00	$\beta 4$	95.02
9	<i>Lus10023348</i>	2343	3	442	89.74	<i>TUB_4b</i>		$\beta 9$	94.34
10	<i>Lus10036069</i>	1789	3	446	99.55 /	<i>TUB_6a</i>	100.00	$\beta 6$	96.64
11	<i>Lus10026813</i>	1793	3	448	93.88	<i>TUB_6b</i>	100.00	$\beta 6$	96.21
12	<i>Lus10040712</i>	1549	3	450	99.56 /	<i>TUB_7a</i>	100.00	$\beta 2, \beta 3$	97.77
						<i>TUB7c</i>	99.33	$\beta 7$	97.20
13	<i>Lus10016448</i>	1553	3	450	95.09	<i>TUB_7b</i>	99.78	$\beta 2, \beta 3$	97.55
						<i>TUB7c</i>	99.33	$\beta 7$	97.20

\* ідентичність між паралогічними генами – стовпець 5; відповідний ізотип  $\beta$ -тубуліну за (Gavazzi et al., 2017) – стовпець 6, а відсоток ідентичності пептидних

послідовностей – стовпець 7; відповідний гомологічний ізотип  $\beta$ -тубуліну *A. thaliana* – стовпець 8; відсоток ідентичності передбачуваного пептидного продукту гена льону з відповідним ізотипом  $\beta$ -тубуліну *A. thaliana* – стовпець 9.

Далі ми провели філогенетичне порівняння ізотипів *L. usitatissimum* з ізотипами *A. thaliana* – Таблиця 3.2, стовпці 8, 9; Рис. 3.3. Це показує, що пептиди *Lus10017217* та *Lus10021094* гомологічні ізотипам  $\beta$ 1- і  $\beta$ 5-тубулінів *A. thaliana*, тоді як *Lus10040712* та *Lus10016448* гомологічні ізотипам  $\beta$ 2- і  $\beta$ 3-тубуліну *A. thaliana*, які ідентичні  $\beta$ 7 і майже ідентичні ізотипу тубуліну  $\beta$ 7. Подібним чином *Lus10038458* та *Lus10023348* гомологічні ізотипам  $\beta$ 4- та  $\beta$ 9-тубулінів *A. thaliana* та *Lus10036069*, а *Lus10026813* – ізотипу  $\beta$ 6-тубуліну *A. thaliana*. Більше того, *Lus10039231* та *Lus10027476* є гомологічними ізотипу  $\beta$ 8-тубуліну *A. thaliana*, тоді як послідовності льону *Lus10035497*, *Lus10002000*, *Lus10008528* утворюють окрему кладу. Ідентичність амінокислотної послідовності гомологічних ізотипів *L. usitatissimum* та *A. thaliana* представлена в Таблиці 3.2, стовпці 8, 9.

Кожен із ідентифікованих генів  $\beta$ -тубуліну льону містить три екзони – Таблиця 3.2, стовпець 3; Рис. 3.4. Всі інтрони генів  $\beta$ -тубуліну льону демонструють консенсусні конвенційні послідовності GT-AG у донор-акцепторних сайтах сплайсингу. Послідовності амінокислот, кодовані екзоном 2 довжиною 270 п.н., усіх аналізованих генів  $\beta$ -тубуліну льону повністю консервативні. Є два варіанти довжини екзона 1 (394 і 400 п.н.). Але відносно екзона 2 позиції обох цих інтронів  $\beta$ -тубуліну виявилися консервативними – 131-1 і 221-1, відносно кодону ініціації ATG – такими ж, як повідомлено для *Daucus carota*, *A. thaliana*, *Lupinus albus*, *Glycine max*, *Physcomitrella patens*, *Pisum sativum* та *Gossypium hirsutum* – Pegumal, 2005. Незважаючи на високий рівень ідентичності послідовностей, що кодують  $\beta$ -тубулін *L. usitatissimum* та *A. thaliana* – Таблиця 3.2, стовпці 8, 9, немає значної подібності між інтронами гомологічних генів (дані не наведені).



**Рис. 3.4.** Екзон-інтронна структура 6 генів  $\alpha$ -тубуліну, 13 генів  $\beta$ -тубуліну та 1 гена  $\gamma$ -тубуліну, виявлених у геномі *L. usitatissimum*. Класи тубуліну та підкласи  $\alpha$ - та  $\beta$ -тубуліну позначені праворуч. Підкласи  $\beta$ -тубуліну взяті з Gavazzi, 2017.

**Варіативність інтронів  $\beta$ -тубуліну.** Виходячи з довжини їх генетичних послідовностей, гени  $\beta$ -тубуліну льону можна розділити на дві групи: (a) *Lus10035497*, *Lus10002000*, *Lus10036069*, *Lus10026813*, *Lus10040712*, та *Lus10016448* з довжинами між 1549–1793 п.н.; та (b) *Lus10039231*, *Lus10027476*, *Lus10017217*, *Lus10021094*, *Lus10038458*, *Lus10023348*, та *Lus10008528* з довжинами 2136–2423 п.н. – Рис. 3.4.

Хоча всі гени  $\beta$ -тубуліну льону містять три інтрони, саме їх різниця в довжині пояснює зміни у довжині повної послідовності генів. Дійсно, обидва інтрони групи (b) довші, ніж групи (a). Наприклад, довжина інтронів 1 і 2 у *Lus10039231* становить 399 і 531 п.н. проти 109 і 188 п.н. у *Lus10035497*. Більше того, інтрони різних ізотипів  $\beta$ -тубуліну льону мають низьку подібність послідовностей. Їх множинне вирівнювання показує багато пропусків, а ідентичність нуклеотидів між ними становить близько 50%, в значній мірі сприяє високий вміст тиміну, який часто зустрічається в інтронах. Отже подібність є переважно випадковою. Така різниця в довжині та складі інтронів між двома

групами генів  $\beta$ -тубуліну льону може зумовлювати диференціацію даних ізотипів за профілем експресії (Chorev and Carmel, 2012; Jo, 2015).

**Характеристика родина  $\beta$ -тубулінів льону.** Нещодавно сімейство генів  $\beta$ -тубуліну льона було більш детально охарактеризовано за допомогою досліджень як методом клонування кДНК, так і аналізу експресії за допомогою qPCR, проведеними на різних стадіях розвитку квітки та росту стебла та гіпокотилу – Gavazzi *et al.* 2017. 14 генів  $\beta$ -тубуліну льону організовані в класичну екзон-інтронну структуру, як і у багатьох інших видів рослин. Всі ідентифіковані гени  $\beta$ -тубуліну льону мають три екзони, розділені двома інтронами у консервативних позиціях, які розмежовані консенсусними консервативними послідовностями GT-AG у донор-акцепторних сайтах сплайсингу. У той час як первинна амінокислотна послідовність, кодована екзоном 2 усіх генів  $\beta$ -тубуліну льону, ідентична, дослідження показало наявність двох варіантів екзону 1 які відрізняються наявністю вставки двох кодонів як у парах *LusTub2a/b*, так і *LusTub3a/b*. Взагалі, так само, як і це було показано для  $\alpha$ -тубуліну, 14 локусів  $\beta$ -тубуліну є попарно дуже гомологічними, і їх можна поєднати у п'ять пар генів-паралогів, а також *LusTub5*, який є синглетом, і триплет *LusTub 7a, b і c*. У даних послідовностях генів  $\beta$ -тубуліну є по два інтрони у консервативних положеннях 131-1 та 221-1, відносно кодону ініціації ATG. Загалом, 14  $\beta$ -тубулінових генів продукують лише 10 різних білків, довжиною від 442 до 450 амінокислот, із загальною ідентичністю послідовності 85,8%.

Більша частина варіабельності, виявленої в амінокислотних послідовностях  $\beta$ -тубулінів, була зосереджена у невпорядкованому варіабельному С-кінцевому хвості (C-terminal tail, СТТ) – а.з. 431-450. Поза цим доменом, у 45 з 430 позицій амінокислотних залишків були знайдені додаткові варіабельні позиції, більшість з яких згруповано в N-кінцевому домені, зокрема в межах N-петлі, між а.з. 33 і 44. У цій області *LusTub2* і *LusTub3* мають два додаткові а.з. так само, як це виявлено у більшості представників класу III у інших видів рослин. Роль такої мінливості довжини N-петлі була досліджена дуже слабо, і її внесок у особливості формування гетеродимера та його властивості, заслуговує

детальних досліджень. Дійсно, згідно з моделлю структури димеру тубуліну – Nogales *et al.* 1998, N-петля взаємодіє з M-петлями сусідніх латерально розташованих субодиниць. Нещодавні дослідження за допомогою моделювання за гомологією, з використанням моделі 1JFF  $\beta$ -тубуліну *Bos taurus*, показали, що додаткові амінокислотні залишки в цій області  $\beta$ -тубуліну сприяють вищій стабільності мікротрубочок – Selva Kumar *et al.*, 2015.

### 3.4.3 $\gamma$ -тубуліни

**Ідентифікація генів  $\gamma$ -тубуліну.** У геномі льону було виявлено три локуси, які гомологічні до  $\gamma$ -тубуліну: *Lus10007851* – 470 амінокислотних залишків, *Lus10010985* – 58 а.з. та *Lus10010986* – 429 а.з. *Lus10010985* насправді відповідає першому екзону *Lus10010986*. Дані бази даних Phytozome свідчать про те, що інтрон 1 з *Lus10010986* міг набути стартового кодону на 3'-кінці, а його 5'-частина зазнала екзонізації генетичної послідовності, таким чином кодуючи фрагмент 16 а.з. QIAKLAVVGNVILLTLM – нехарактерний для тубулінів. Нуклеотидний склад відповідної генетичної послідовності багатий тиміном, що є характерним для інтронів. Отже, отриманий фрагмент амінокислоти багатий на Leu, який кодується триплетом CTT, і Val – триплетом нуклеотидів GTT.

Амінокислотна послідовність пептиду, кодованого локусом *Lus10010986*, на 100% ідентична відповідній частині локусу *Lus10007851*. Однак їх послідовності ДНК ідентичні лише на 94,83% через наявність делеції у 14 п.н., вставки у 2 п.н. та множинних SNP. Амінокислотна послідовність пептиду, кодованого локусом *Lus10010985*, містить 3 SAP в порівнянні з *Lus10007851*. Таким чином, ми припускаємо, що *Lus10010985-6* та *Lus10007851* представляють пару паралогічних генів  $\gamma$ -тубуліну – Таблиця 3.3. Чи відповідають локуси *Lus10010985* та *Lus10010986* експресованим генам, це питання подальшого дослідження. Наразі, лише локус *Lus10007851* слід розглядати як дійсний функціональний ген  $\gamma$ -тубуліну *Linum*. Ідентичність послідовності пептидного

продукту *Lus10007851* та ізотипів *A. thaliana*  $\gamma 1$  та  $\gamma 2$  становить 94,26% та 92,98%, відповідно, – Таблиця 3.3, стовпці 5, 6.

У *Lus10007851* інтрони знаходяться в позиціях 43-0, 110-0, 147-1, 281-0, 299-0, 334-0, 359-0, 374-0, 417-0 та 452-0, відносно кодону ініціації ATG. Це збігається з положенням інтронів генів  $\gamma$ -тубуліну у *A. thaliana*, *Physcomitrella patens*, *Haemodorum mnioides* та *Lupinus albus* – Perumal, 2005. Всі інтрони генів  $\gamma$ -тубуліну льону мають ідентичні канонічні послідовності GT-AG у донор-акцепторних сайтах сплайсингу.

Таблиця 3.3

### Гени $\gamma$ -тубуліну, ідентифіковані у геномі *L. usitatissimum* з бази даних

#### Phytozome

#	Ген. локус	Пар нукл.	К-сть екзонів	К-ість а.з.	Ізотип <i>A. thaliana</i>	Ідентичн. <i>A. thaliana</i> , %
	-1-	-2-	-3-	-4-	-5-	-6-
1	<i>Lus10007851</i>	2447	11	470	$\gamma 1$ $\gamma 2$	94.26 92.98
2	<i>Lus10010985-6</i>	2482	11	487?	-	-

\* відсоток ідентичності передбачуваного пептидного продукту гена льону з відповідним ізотипом  $\gamma$ -тубуліну *A. thaliana* – стовпці 5, 6.

#### 3.4.4 Актини

**Ідентифікація генів актину.** Ми проаналізували п'ятнадцять передбачених білкових послідовностей, які кодуються 15 ідентифікованими генами актину льону та один додатковий варіант *Lus10004956* (вирівнювання не показано). Амінокислотні послідовності, закодовані *Lus10006783/84*\* ідентичні послідовностям *Lus10005819/20*, тоді як *Lus10001693*\* і *Lus10001694* ідентичні *Lus10005163*. Нарешті, амінокислотні послідовності, закодовані *Lus10005457*, *Lus10040826* та *Lus10016558* мають однакову амінокислотну послідовність.

Перші 356 N-кінцевих амінокислотних залишків білкових послідовностей всіх відібраних генів актину льону ідентичні один одному, за винятком 17

положень. Вони відповідають восьми унікальним замінам, виявленим у *Lus10006783/84\** та *Lus10005819/20*: дві радикальні заміни –  $^{131}\text{T/V}$ ,  $^{201}\text{S/M}$  та 6 консервативних –  $^3\text{E/D}$ ,  $^6\text{D/E}$ ,  $^{131}\text{T/V}$ ,  $^{201}\text{S/M}$ ,  $^{221}\text{I/V}$ ,  $^{234}\text{TS/SS}$ ,  $^{241}\text{S/N}$  та  $^{268}\text{M/L}$ ; чотири унікальні консервативні заміни, знайдені у *Lus10001693\**, *Lus10001694* та *Lus10005163*:  $^{79}\text{SN/NN}$ ,  $^{225}\text{Y/F}$ ,  $^{229}\text{L/M}$  та  $^{320}\text{T/S}$ . Є також дві консервативні заміни, однакові у всіх семи згаданих вище генах:  $^{190}\text{A/S}$ ,  $^{232}\text{S/A}$  – порівняно з іншими 8 генами актину льону. Крім того, у *Lus10016259* є дві унікальні консервативні заміни  $^{203}\text{TTT/TTS}$  та  $^{309}\text{PG/GG}$ , а в *Lus10004956\*\** радикальна заміна  $^{203}\text{TTT/TPT}$ . За винятком *Lus10004169* та *Lus10021057*, які обговорювались вище, амінокислотні послідовності області 357-377 мають високу консервативність у всіх послідовностях білків актину льону, за винятком двох консервативних замін A/S:  $^{360}\text{AKA/SKS}$ , присутніх у *Lus10006783/84\** та *Lus10005819/20*.

Білкова послідовність, кодована локусом *Lus10004956\**, містить радикальну заміну 209EI/GN, яка на основі аналізу моделей тубулінів, депонованих у Protein Data Bank (PDB, <http://www.rcsb.org/>), може мати вплив на структуру і функція білка, оскільки вона знаходиться в  $\alpha$ -спіралі. Дійсно, залишок гліцину може діяти як "руйнівник спіралей", що спричиняє структурні та функціональні наслідки і позначається на функціональних властивостях даного гена. Тому, для цього дослідження ми вибрали для аналізу можливий сплайсинг-варіант *Lus10004956\*\** цього гена, який не має заміни  $^{209}\text{EI/GN}$ .

Аналіз екзон-інтронної структури даних генів показує, що *Lus10004169* та *Lus10021057* містять три екзони, тоді як усі інші ідентифіковані гени актину льону містять чотири екзони – Таблиця 3.4. Це пояснюється тим, що ці два локуси не містять інтрон 3 на відміну від того, що спостерігається в інших генах актину – Рис. 3.5. У гені *Arabidopsis ACT2*, що характеризується подібним зменшенням кількості інтронів (McDowell, 1996b), де відсутній інтрон 1, таким чином це являється іншим аналогічним випадком втрати інтрону. Крім того, у *Lus10004169* та *Lus10021057*, *Lus10005457*, *Lus10029286* та *Lus10016259* інтрон 2 в три-чотири рази довший, ніж у інших семи генах актину. Подібним чином, у



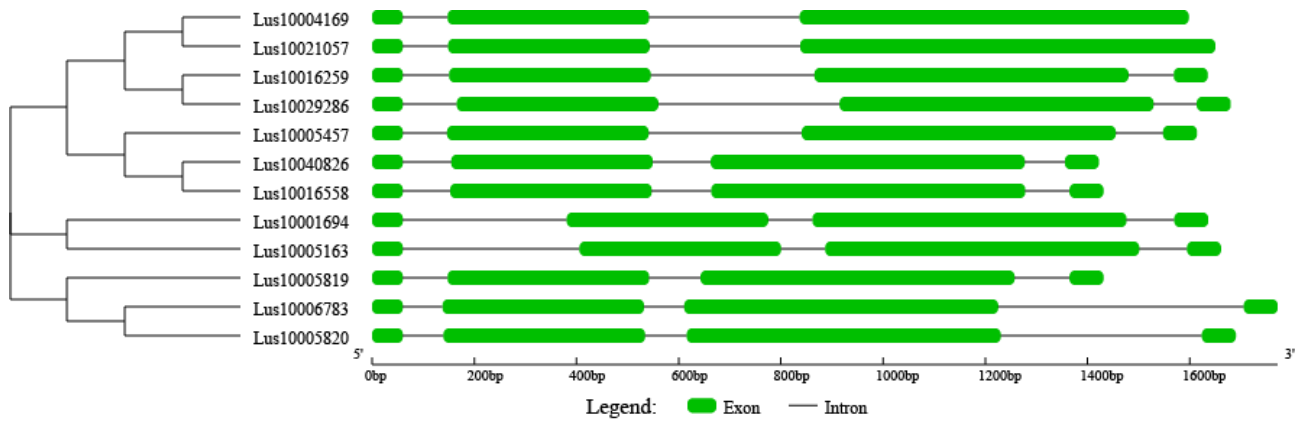
*Lus10005163* та *Lus10001694* інтрон 1 у три-чотири рази довший, ніж в інших проаналізованих 10 генах актину, що характеризуються ідентичною екзон-інтронною структурою, – Рис. 3.5.

Таблиця 3.4

Гени актину, виявлені у геномі *L. usitatissimum* з бази даних Phytozome

#	Локус	н. п.	Екзонів	а.з.
	-1-	-2-	-3-	-4-
<b>1</b>	<i>Lus10006783</i>	1772	4	377
<b>2</b>	<i>Lus10005820</i>	1690	4	377
	<i>Lus10006784</i>	1427?	4	388
<b>3</b>	<i>Lus10006784*</i>	-	4	377
<b>4</b>	<i>Lus10005819</i>	1431	4	377
<b>5</b>	<i>Lus10005163</i>	1661	4	377
	<i>Lus10001693</i>	1184?	3?	332
<b>6</b>	<i>Lus10001693*</i>	-	4	377
<b>7</b>	<i>Lus10001694</i>	1636	4	377
	<i>Lus10004956</i>	1607?	4?	414
	<i>Lus10004956*</i>	-	4	377
<b>8</b>	<i>Lus10004956**</i>	-	4	377
<b>9</b>	<i>Lus10005457</i>	1614	4	377
<b>10</b>	<i>Lus10016558</i>	1431	4	377
<b>11</b>	<i>Lus10040826</i>	1422	4	377
<b>12</b>	<i>Lus10029286</i>	1680	4	377
<b>13</b>	<i>Lus10016259</i>	1635	4	377
<b>14</b>	<i>Lus10004169</i>	1598	3	404
<b>15</b>	<i>Lus10021057</i>	1650	3	421

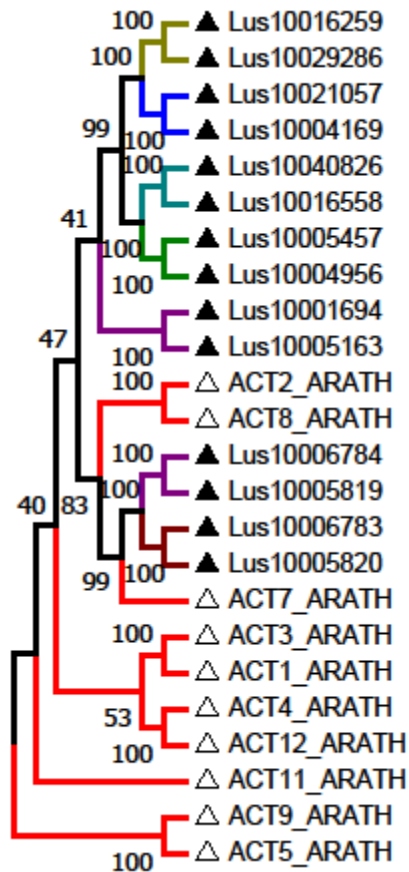
Примітка: Символи (?), (\*) та (\*\*\*) позначають альтернативну інтерпретацію геномних даних – див. опис у тексті.



**Рис. 3.5.** Екзон-інтронна структура 12 з 15 генів актину, ідентифікованих у геномі *L. usitatissimum* – дані для запропонованих генних послідовностей *Lus10006784\**, *Lus10001693\** та *Lus10004956\*\** не показані.

Вирівнювання послідовностей ДНК повних нуклеотидних послідовностей 12 актинових генів, де зберігається типова екзон-інтронна структура (дані не представлені), показує, що всі інтрони всіх цих генів мають консенсусні конвенційні послідовності GT-AG у донор-акцепторних сайтах сплайсингу, крім інтрону 2 локусів *Lus10005163* та *Lus10001694*, де спостерігається звичайний варіант сплайсинг послідовностей – GC-AG.

**Ізотипи генів актину та паралогія.** Через високий рівень ідентичності білкових послідовностей генів актину, для того, щоб диференціювати відібрані гени, були взяті повнорозмірні генетичні послідовності 14 з 15 ідентифікованих локусів актину льону. Локус *Lus10001693* було відкинуто. Отже, 14 послідовностей були вирівняні разом з 10 анотованими генами актину *A. thaliana* для побудови філогенетичного дерева, – Рис. 3.6.



**Рис. 3.6.** Філогенетичне порівняння генетичних послідовностей 14 з 15 генів актину, ідентифікованих у геномі *L. usitatissimum* (▲), з 10 генами актину *A. thaliana* (Δ).

**Філогенетичний аналіз** об'єднує 15 ідентифікованих генів актину льону у 7 пар генів-паралогів: *Lus10006783* – *Lus10005820*, *Lus10006784*\* – *Lus10005819*, *Lus10005163* – *Lus10001694*, *Lus10004956*\*\* – *Lus10005457*, *Lus10016558* – *Lus10040826*, *Lus10029286* – *Lus10016259*, та *Lus10004169* – *Lus10021057*, та одиночка *Lus10001693*\*, що є послідовною копією *Lus10001694*, – Рис. 3.6.

Філогенетичний аналіз послідовностей ДНК вказує на те, що дві паралогічні пари генів *Lus10006783* – *Lus10005820*, *Lus10006784*\* – *Lus10005819* більш гомологічні до вегетативних актинів *A. thaliana* АСТ7, АСТ2/8, ніж з рештою сімейства актинів льону – Рис. 3.6. Більш точна класифікація генів актину льону вимагає експериментальних досліджень.

**Висновки** Аналіз екзон-інтронної структури генів актину льону показав як мінливість кількості інтронів, яка може складати 2 або 3, так і поліморфізм

довжини інтронів. Інтронний аналіз важливий для філогенетичного та еволюційного аналізу генів актину. Інтронний поліморфізм генів актину *Linum* дає можливість висловити припущення про те, що пов'язані з інтронами регуляторні елементи можуть регулювати експресію даних генів. Варіації довжини інтрону 2 генів актинів можуть зробити їх життєздатними мішенями для дослідження поліморфізмів довжини інтронів (intron length polymorphisms, ILP) різних видів *Linum* методом молекулярних маркерних ILP, – Bierne et al. 2000, Breviario et al. 2008.

### 3.4.5 Висновки

Наші дані показують, що гени  $\alpha$ -,  $\beta$ -тубуліну та актину широко представлені у геномі льону. За допомогою *in silico* дослідженням, проведеним на першому драфті генома льону, ми виявили 6 генів  $\alpha$ -тубуліну, 13 генів  $\beta$ -тубуліну, 2 гени  $\gamma$ -тубуліну та 15 генів актину. Шість генів  $\alpha$ -тубуліну згруповані у три паралогічні пари. Серед 13 генів  $\beta$ -тубуліну можна виділити сім різних ізотипів, шість з яких кодуються двома паралогічними генами.  $\gamma$ -тубулін представлений паралогічною парою генів, один з яких, ймовірно, втратив свою функцію. П'ятнадцять генів актину представляють сім паралогічних пар – сім ізотипів актину та послідовну дуплікацію одного з генів одного із ізотипів.

Аналіз екзон-інтронної структури показав поліморфізм довжини інтронів в генах  $\beta$ -тубуліну та варіацію кількості інтронів у генах  $\alpha$ -тубуліну: три або чотири інтрони знаходяться у двох і чотирьох генах відповідно. Позиції інтронів у генах тубуліну льону є такими ж, як і у інших видах рослин. У генах актину льону спостерігається як мінливість кількості інтронів, яка може складати 2 або 3, так і поліморфізм довжини інтронів.

Інтронний аналіз важливий для філогенетичного та еволюційного аналізу генів тубуліну та актину. Інтронний поліморфізм генів тубуліну *Linum* свідчить про те, що пов'язані з інтронами регуляторні елементи можуть впливати на експресію цих генів, – Giani, 2009. Варіації довжини інтрону 3  $\alpha$ -тубуліну,

інтронів 1 і 2  $\beta$ -тубулінів та інтрону 2 актину можуть зробити їх життєздатними мішенями для дослідження поліморфізмів довжини інтронів (intron length polymorphisms, ILPs) різних видів *Linum* методом молекулярного маркера ILP, – Bierre, 2000, Breviario, 2008.

### 3.5 Моделювання рослинних тубулінів за гомологією

**Пошук гомологів.** Для аналізу були вибрані тубуліни alpha-1 (NCBI Reference Sequence: NP\_176654.1) – 447 а.з. і beta-1 (GenBank: AAA32893.1) – 450 а.з. *Arabidopsis thaliana*.

Пошук гомологів був здійснений в базі даних PDB за допомогою програми BLAST (PDB-BLAST). Результати представлені в Таблиці 3.5. Для подальшого аналізу були обрані найкращі одночасно для alpha і beta ланцюгів гомологи, якими є такі рентгеноструктурні моделі:

1Z2B – Tubulin-colchicine-vinblastine: stathmin-like domain complex. Ланцюги A, B, C, D – *Bos taurus*, E – *Rattus norvegicus*

1FFX – The 4 A X-ray structure of a tubulin: stathmin-like domain complex. Ланцюги A, B, C, D – *Bos taurus*, E – *Rattus norvegicus*

3DU7 – Tubulin-colchicine-phomopsin A: Stathmin-like domain complex. Ланцюги A, B, C, D – *Bos taurus*, E – *Rattus norvegicus*

Дані моделі представляють собою біологічні комплекси, що складаються з ланцюгів тубуліна alpha (A, C), beta (B, D) і об'єднує їх допоміжного ланцюга E. Дані біологічні ансамблі визначені авторами і побудовані програмою PDBePISA (<https://www.ebi.ac.uk/pdbe/pisa/>).

Таблиця 3.5

Гомологи тубуліна alpha-1 *A. thaliana* у базі даних PDB.

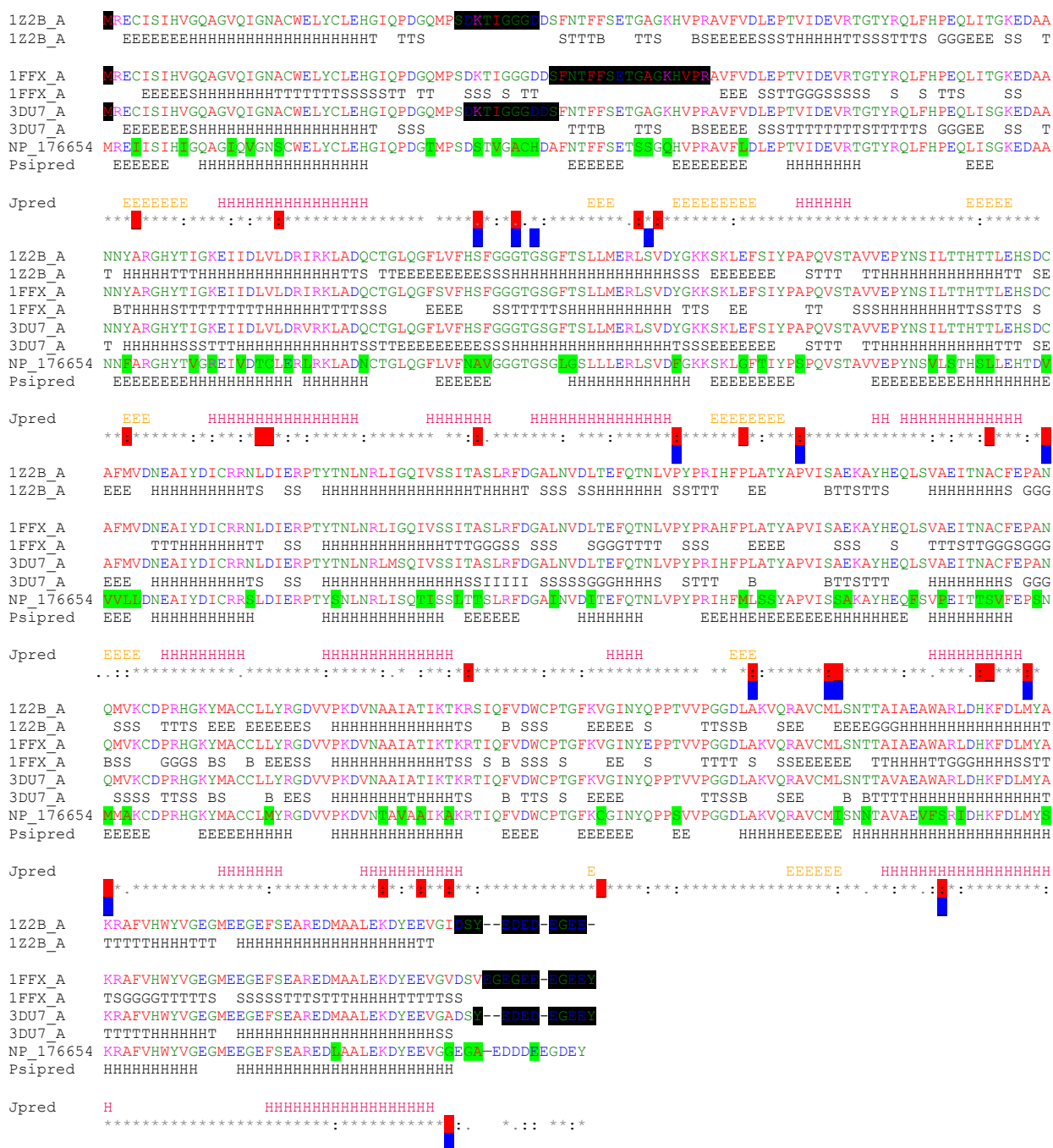
PDB ID	Оцінка ідентичності	Довжина вирівн.	PDB ID	Оцінка ідентичності	Довжина вирівн.
<i>alpha</i>			<i>beta</i>		
1Z2B_A	768	98%	1FFX_B	775	96%
3DU7_A	764	97%	3DU7_B	774	96%
3HKB_A	761	96%	1Z2B_B	774	96%
1SA0_A	757	96%	1TUB_B	768	95%
1FFX_A	756	96%			
1JFF_A	754	96%			
1TUB_A	754	96%			

**Множинне вирівнювання** Далі, для відібраних послідовностей було проведено множинне вирівнювання за допомогою сервера програми ClustalW. Результати оцінки множинного вирівнювання тубуліна alpha-1 *A.thaliana* з гомологами представлені в Таблиці 3.6. Відсоток ідентичності становить: 1Z2B\_A – 79%, 3DU7\_A – 80%, 1FFX\_A – 78%. Множинне вирівнювання послідовності тубуліна alpha-1 *A. thaliana* щодо шаблонів представлено на Рис. 3.7. Результати множинного вирівнювання тубуліна beta-1 *A.thaliana* з гомологами представлений у Таблиці 3.7. Відсоток ідентичності становить: 1Z2B\_A – 84%, 3DU7\_A – 84%, 1FFX\_A – 84%. Множинне вирівнювання послідовності тубуліна beta-1 *A.thaliana* щодо шаблонів представлено на Рис. 3.8.

Таблиця 3.6

Ідентичність тубуліну alpha-1 *A. thaliana* з обраними шаблонами.

№	Назва	Довжина, а.з.	№	Назва	Довжина, а.з.	Оцінка ідентичності
1	NP_176654	452	2	1Z2B_A	452	79
1	NP_176654	452	3	3DU7_A	452	80
1	NP_176654	452	4	1FFX_A	452	78
2	1Z2B_A	452	3	3DU7_A	452	98
2	1Z2B_A	452	4	1FFX_A	452	98
3	3DU7_A	452	4	1FFX_A	452	97



**Рис. 3.7.** Множинне вирівнювання послідовності тубуліна alpha-1 A. *thaliana* щодо шаблонів, PDB-коди: 1Z2B\_A, 1FFX\_A, 3DU7\_A. Позначення вторинної структури: H – альфа спіраль. E – бета-структура. S – бета-шар. T – поворот. G – 3/10 helix. Консервативність: «\*» – повністю консервативний залишок. «:» – напівконсервативний залишок, або того ж типу. «.» – слабо консервативний. «>» – не консервативний. «█» – не ідентичний залишок. «█» – неконсервативний залишок. «█» – неконсервативний залишок, який може впливати на локальну вторинну структуру. █ – координати залишку відсутні у PDB файлі.

Таблиця 3.7

Ідентичність тубуліну beta-1 *A. thaliana* з обраними шаблонами.

№	Назва	Довжина, а.з.	№	Назва	Довжина, а.з.	Оцінка ідентичності
1	AAA32893.1	447	2	1FFX_B	445	84
1	AAA32893.1	447	3	3DU7_B	445	84
1	AAA32893.1	447	4	1Z2B_B	445	84
2	1FFX_B	452	3	3DU7_B	445	98
2	1FFX_B	452	4	1Z2B_B	445	98
3	3DU7_B	452	4	1Z2B_B	445	99

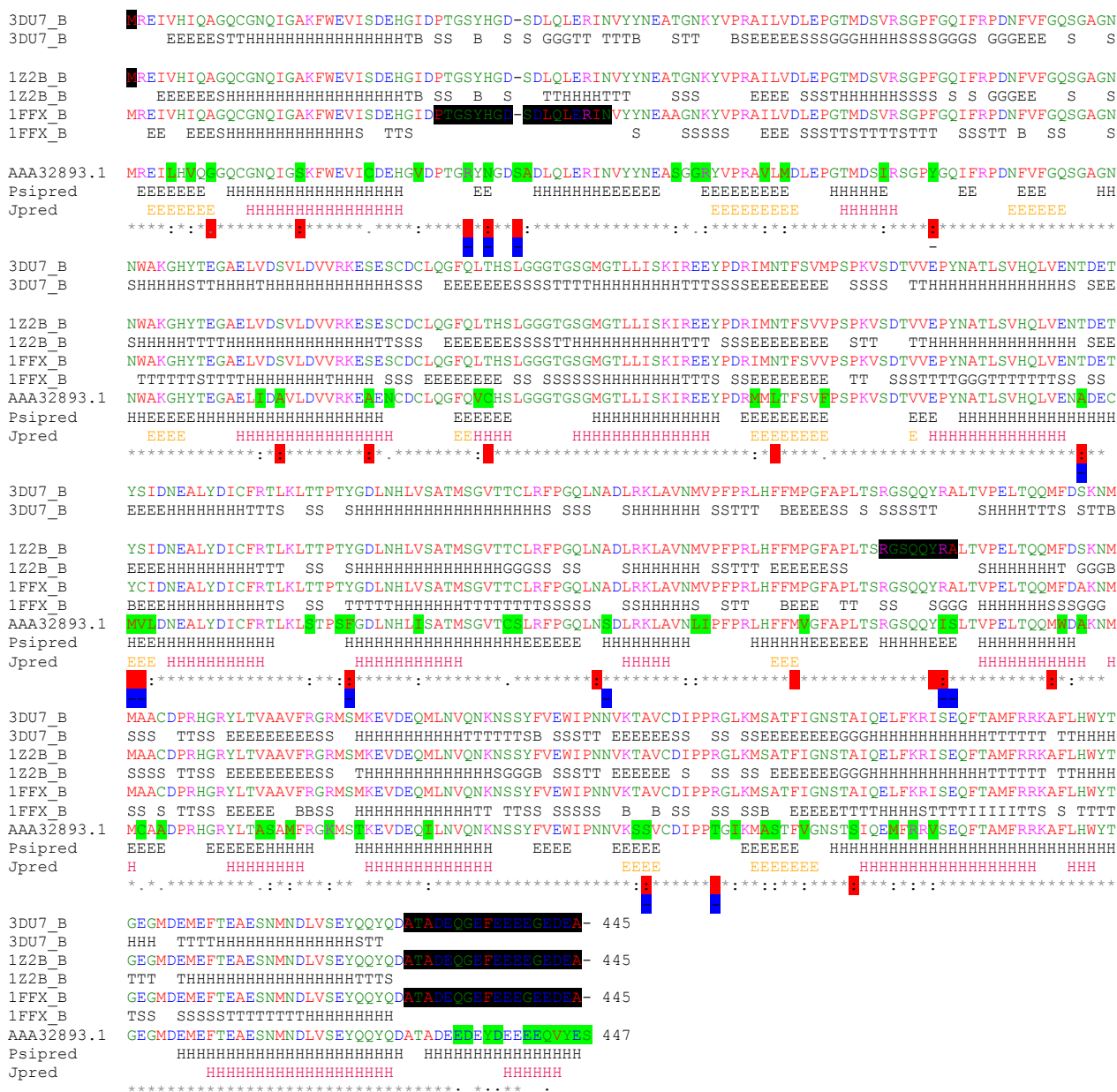


Рис. 3.8. Множинне вирівнювання послідовності тубуліна beta-1 *A. thaliana* щодо шаблонів, PDB-коди: 1Z2B\_A, 1FFX\_A, 3DU7\_A. Позначення вторинної



*структури*: Н – альфа спіраль. Е – бета-структура. S – бета-шар. Т – поворот. G – 3/10 helix. *Консервативність*: «\*» – повністю консервативний залишок. «:» – напівконсервативний залишок, або того ж типу. «.» – слабо консервативний. «>» – не консервативний. «█» – не ідентичний залишок. «█» – неконсервативний залишок. «█» – неконсервативний залишок, який може впливати на локальну вторинну структуру. **M** – координати залишку відсутні у PDB файлі.

Високий відсоток ідентичності послідовностей дозволяє говорити про високий ступінь гомології (80 і більше відсотків) в обох випадках.

Негомологічних а.о. залишки проаналізовані і найбільш цікаві з них, ті які можуть мати вплив на локальну структуру, відзначені кольором. Для оцінки їх можливого впливу на структуру був проведений аналіз вторинної структури.

Вторинна структура приведена паралельно з множинним вирівнюванням первинних послідовностей на Рис. 3.7 і Рис. 3.8. Для рентгенівських структур – це дані з бази даних PDB. Для послідовностей *A. thaliana* – результати аналізу за допомогою серверів Psipred, – Buchan and Jones 2019, и Jpred, – Drozdetskiy et al. 2015.

Аналіз вторинної структури послідовностей *A. thaliana* показує, що, не дивлячись на мутації, її основні елементи в більшості випадків досить добре відповідають вторинній структурі рентгенів. Окремі відмінності стосуються менш упорядкованих ділянок – поворотів і невеликих, можливо тимчасових флуктуючих елементів. Для побудови моделі це не суттєво і при необхідності певні ключові місця можуть бути додатково досліджені.

У рентгенівських структурах-шаблонах є кілька незакрісталізованих ділянок. Для гомологів alpha це ділянки: 1Z2B\_A Ser38-Asp46, 1FFX\_A Ser48-Arg63, 3DU7\_A Ser39-Ser48. Дані ділянки не збігаються у всіх трьох шаблонах і таким чином відсутність залишків в 1Z2B\_A і 3DU7\_A компенсується їх наявністю в 1FFX\_A, а відсутність залишків в 1FFX\_A компенсується їх наявністю в 1Z2B\_A і 3DU7\_A. Для гомологів beta відсутні окремі ділянки в різних шаблонах: 1FFX\_B Pro32-Asn50 і 1Z2B\_B Arg278-Ala285, що не суттєво.

Також, в обох субодиниць відсутні близько 10 залишків на рухомому С-кінці, що типово для невпорядкованих ділянок.

В цілому, для всіх амінокислотних залишків послідовностей *A. thaliana* є необхідні шаблони і набір шаблонів цілком задовільний для проведення моделювання.

**Моделювання** Процедура моделювання була проведена в програмі Modeller 9v7, – Eswar et al., 2006. Було побудовано 10 моделей. Моделі включали одночасно 2 ланцюга –  $\alpha$  і  $\beta$ . При необхідності можна змоделювати і весь комплекс з 5 ланцюгів, аналогічно шаблонами, але таке завдання поки не ставилось.

Моделі були спочатку оптимізовані за допомогою VTФМ оптимізації за вбудованим алгоритмом «slow», який полягає в множинних циклах мінімізації енергії за алгоритмом CG (conjugate gradients). Далі моделі були оптимізовані за допомогою молекулярної динаміки методом примусового заморожування (simulated annealing) за вбудованим алгоритмом «slow». Даний алгоритм передбачає поетапний розігрів і охолодження системи при температурах 150К, 250К, 400К, 700К, 1000К, 1000К, 800К, 600К, 500К, 400К, 300К. Цикл повторюється двічі, за чим слідує кінцева CG-мінімізація енергії.

**Аналіз моделей** Отримані моделі були проаналізовані вбудованими алгоритмами оцінки якості – функцією molpdf (геометрія) і DOPE score (енергетика) і його похідними. Значення оціночних параметрів побудованих моделей наведені у Таблиці 3.8.

Таблиця 3.8

## Значення оціночних параметрів побудованих моделей.

Модель	molpdf	DOPE score	DOPE-HR score	Normalized DOPE score
1	47611.36328	-104793.07813	-86900.88281	-1.10200
2	47702.75781	-104482.67188	-86345.39063	-1.07749
<b>3</b>	<b>47379.51172</b>	<b>-105001.51563</b>	<b>-87260.44531</b>	<b>-1.11845</b>
4	47689.40625	-104792.32813	-86941.00781	-1.10194
5	47652.23828	-103861.58594	-84385.46875	-1.02845
6	47881.33203	-104811.56250	-86871.46875	-1.10345
<b>7</b>	<b>47479.65625</b>	<b>-105004.59375</b>	<b>-87764.07813</b>	<b>-1.11870</b>
8	48876.04688	-102636.93750	-81153.30469	-0.93176
9	48085.92578	-104494.77344	-86700.71875	-1.07844
10	47854.98828	-104407.17188	-85990.59375	-1.07153

За даними параметрами найкращими є моделі №7 та №3. Якість даних моделей було проаналізовано за допомогою сервера MolProbity, – Chen et al., 2010. Даний інструмент використовується для аналізу рентгеноструктурних і ЯМР моделей. Він порівнює геометричні параметри моделі, – перетин радіусів атомів, неправильні ротамери, положення у графіку Рамачандран, відхилення C $\beta$  атомів, неправильні зв'язку і кути, з ідеальними. Таким чином, за його визначенням будь-яка модель має з точки зору алгоритму «недоліки». Тобто в такому разі більш застосовним є порівняльний аналіз одразу декількох моделей.

Тому для порівняння були також проаналізовані і всі шаблони. Дані результати представлені у Таблиці 3.9. Як видно, оцінка моделей за більшістю параметрів є не найгіршою, а за деякими параметрами і кращою. Що стосується сумарного оцінного параметра, який включає інші параметри з ваговими коефіцієнтами – «MolProbity score», то він для моделей приблизно на 20% нижче, ніж для рентгенів і знаходиться за оцінкою сервера в прийнятних межах, виділено жовтим кольором.

Таблиця 3.9

## Результати аналізу моделей і шаблонів сервером MolProbity.

Область	Параметр	Модель №3	Модель №7	1Z2B	3DU7	1FFX	опис
All-Atom Contacts	Clashscore, all atoms:	105.01	98.17	<b>52.22</b>	57.86	172.1	
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.	0 <sup>th</sup> percentile* (N=1784, all resolutions)	0 <sup>th</sup> percentile* (N=1784, all resolutions)	<b>51<sup>st</sup> percentile*</b> (N=37, 3Å - 9999Å)	32 <sup>nd</sup> percentile* (N=37, 3Å - 9999Å)	0 <sup>th</sup> percentile* (N=37, 3Å - 9999Å)	
Protein Geometry	Poor rotamers	2.33%	<b>2.20%</b>	34.50%	33.48%	20.83%	Goal: <1%
	Ramachandran outliers	3.36%	<b>3.36%</b>	7.55%	11.44%	16.85%	Goal: <0.2%
	Ramachandran favored	<b>90.26%</b>	89.81%	77.92%	71.46%	53.09%	Goal: >98%
	Cβ deviations >0.25Å	20	23	32	18	<b>0</b>	Goal: 0
	MolProbity score <sup>^</sup>	3.31 13 <sup>th</sup> percentile* (N=27675, 0Å - 99Å)	<b>3.27</b> <b>14<sup>th</sup> percentile*</b> (N=27675, 0Å - 99Å)	4.12 24th percentile* (N=342, 3.25Å - 4.35Å)	4.22 16 <sup>th</sup> percentile* (N=342, 3.25Å - 4.35Å)	4.65 3rd percentile* (N=342, 3.25Å - 4.2Å)	
	Residues with bad bonds:	0.89%	0.67%	0.08%	<b>0.00%</b>	<b>0.00%</b>	Goal: 0%
	Residues with bad angles:	2.68%	2.45%	1.02%	<b>0.82%</b>	3.48%	Goal: <0.1%

\* 100<sup>th</sup> percentile is the best among structures of comparable resolution; 0<sup>th</sup> percentile is the worst.

<sup>^</sup> MolProbity score is defined as the following:  $0.42574 \cdot \log(1 + \text{clashscore}) + 0.32996 \cdot \log(1 + \max(0, \text{pctRotOut} - 1)) + 0.24979 \cdot \log(1 + \max(0, 100 - \text{pctRamaFavored} - 2)) + 0.5$

*Примітка:* Найкращі значення параметрів виділені жирним шрифтом.

Таблиця 3.10

## RMSD, Å між моделями і шаблонами.

	Модель3	Модель7	1Z2B	13DU	1FFX
Модель3	-	0.28	0.42	0.49	1.05
Модель7	0.28	-	0.47	0.54	1.03
1Z2B	0.42	0.47	-	0.49	1.08
13DU	0.49	0.54	0.49	-	1.07
1FFX	1.05	1.03	1.08	1.07	-

Подібність геометрії між моделями і шаблонами проаналізовано за допомогою накладання та оцінки RMSD. Результати представлені у Таблиці 3.10. З таблиці видно, що RMSD між моделями 3 і 7 і шаблонами близька до RMSD між шаблонами 1Z2B і 3DU7 – близько 0,5Å. Шаблон 1FFX має гіршу якість структури, де неправильно визначаються багато елементів вторинної структури. RMSD між 1FFX і моделями і іншими шаблонами однаково високі – близько

1,05Å. Між собою моделі більш схожі. RMSD між моделями 3 і 7 нижче, ніж між моделями і шаблонами і моделями – 0,28Å. Невисокий RMSD між моделями і шаблонами цілком очікуваний і відповідає високому рівню гомології.

### **3.6 Структурні особливості цитоскелетних білків льону як основа їх функцій**

Рослинні мікротрубочки (МТ), утворені гетеродимерами  $\alpha$ - та  $\beta$ -тубулінів, контролюють ряд основних клітинних процесів, таких як поділ клітин, ріст, транспорт тощо. Для побудови функціональних мікротрубочкових структур в середині клітини довжина окремих МТ повинна бути динамічною і регульованою. Цей механізм забезпечується явищем «динамічної нестабільності» – постійним подовженням та вкороченням мікротрубочок, який забезпечується білками, щр приймають участь у формуванні та рості мікротрубочок. Рослинні є об'єктом біотехнології, тому процеси організації рослинних МТ детально вивчаються. У рослинних клітинах відсутні центріолі – органели, які організовують МТ у цитоплазмі клітин тварин. На відміну від тварин, ділянки зародження кортикальних МТ у рослинних клітинах мають дифузну локалізацію. Хоча структури МТ тварин відрізняються від рослин, процеси динамічної нестабільності регулюються подібним чином. Ці процеси регулюються білками, асоційованими з МТ: стабілізація відбувається за участю MAP65, EB1 та WVD2, а дестабілізація – за допомогою MAP215, катаніну та кінезину-13.

У цьому дослідженні, спираючись на кристалографічні структури тваринних тубулінів, переважно тубулінів ссавців, ми побудували моделі за гомологією рослинних  $\alpha$ -,  $\beta$ - та  $\gamma$ -тубулінів та таких MAPів, як: EB1, MAP65-1, катанін, кінезін-13A та кінезін-1. Також, був проведений порівняльний аналіз структури отриманих рослинних тубулінів та MAPів з гомологами тварин.

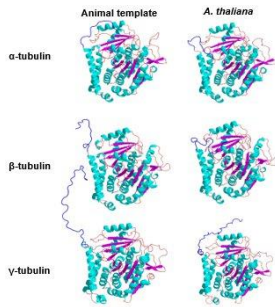
### 3.6.1 Тубуліни $\alpha$ , $\beta$ та $\gamma$

Послідовності тубулінів  $\alpha$ ,  $\beta$  та  $\gamma$  організмів з різних таксонів дуже гомологічні. За допомогою програмного пакету Modeller9v10, ми отримали моделі за гомологією гетеродимера тубулінів  $\alpha$  та  $\beta$  та тубулін  $\gamma$  *A. thaliana*. У якості шаблонів для моделювання за гомологією  $\alpha$ - та  $\beta$ -тубулінів було використано рентгеноструктурні PDB-моделі тубулінів великої рогатої худоби – 1Z2B, 1FFX та 3DU7, та вівці – 3RYC та 3HKV. Ці послідовності мають приблизно 80% ідентичності з тубулінами  $\alpha$  та  $\beta$  *A. thaliana*. Кристалографічна модель  $\gamma$ -тубуліну людини 3CB2 була використана як шаблон для моделювання  $\gamma$ -тубуліну *A. thaliana*. Ідентичність між цими послідовностями становить 71%.

Не зважаючи на високу ідентичність тубулінів, поодинокі заміни амінокислот у критичних регіонах визначають відмінності між різними таксонами та навіть видами. Такі відмінності можуть впливати на взаємодію тубулінів з MAP-ами та зв'язування фармакофор. Таким чином, визначення структурних відмінностей тубулінів різних видів та того, яким чином такі відмінності у структурі впливають на регуляторні механізми постало важливим питанням для біотехнології.

Раніше, активність близько 30 нових синтезованих фармакофорних сполук, що відносяться до класів динітроанілінів та фосфороамідів, були оцінені *in vivo* в нашій лабораторії. Деякі з них було визнано ефективними агентами для впливу на (де)полімеризацію мікротрубочок. Наступним кроком є дослідження моделі взаємодії найбільш перспективних з них шляхом молекулярного моделювання *in silico* та прогнозування їх можливого використання. Для цього необхідні посторові моделі тубулінів, білків пов'язаних з мікротрубочками, та цілих окремих фрагментів мікротрубочок.

Отримані моделі тубулінів  $\alpha$ ,  $\beta$  та  $\gamma$  *A. thaliana* використовувались для стикування білок-лігандного докінгу з гербіцидами динітроанілінового та фосфороамідного ряду, і частина отриманих результатів була визнана відповідною нашим експериментальним даним.



**Рис. 3.9.** Моделі тубулінів  $\alpha$ ,  $\beta$  та  $\gamma$  *A. thaliana* у порівнянні з тваринними моделями-шаблонами.

Структура глобулярної частини тубулінів є дуже консервативною, як і її послідовність. RMSD відхилення моделей від шаблону становить 0,34Å, 0,42Å та 0,28Å для  $\alpha$ -,  $\beta$ - та  $\gamma$ -тубулінів, відповідно.

У побудові моделях тубулінів особливу складність представляє С-кінцевий хвіст білка, який неструктурований – Рис. 3.9, показаний зеленим кольором. С-кінцевий фрагмент тубулінів розпізнається низкою МАРів і відіграє важливу роль у регулюванні росту та вкорочення МТ. Властивості С-кінцевих хвостів тубулінів будуть аналізуватися в майбутньому за допомогою молекулярної динаміки та гнучкого стикування. Властивості С-кінцевих хвостів тубулінів є складною для вирішення задачею і будуть продовжувати досліджуватись в майбутньому за допомогою молекулярної динаміки та гнучкого докінгу.

**Сайти зв'язування колхіцину, таксолу та вінбластину.** Похідні рослинних алкалоїдів, такі як колхіцин, вінбластин і таксол, зв'язуються з тубулінами та мають широке застосування в медицині та біотехнологіях, завдяки їх впливу на (де)полімеризацію мікротрубочок. Загальновідомо, що заміна ключових амінокислотних залишків у сайтах зв'язування визначає чутливість або стійкість до того самого препарату. Структура сайтів зв'язування колхіцину, вінбластину та таксолу у тваринних тубулінах відомі із наступних структур банку даних Protein Data Bank: 3DU7, 3Z2B – *Bos taurus*, та 1TUB – *Sus scrofa*, відповідно. Колхіцин зв'язується з тубуліном в одному сайті, утвореному Ser178 з  $\alpha$ -тубуліну та Leu255 і Ala316 з  $\beta$ -тубуліну. Сайт зв'язування вінбластину

утворена Pro325, Asn329, Val353 та Ile355 з  $\alpha$ -тубуліну та Val177, Tyr210, Tyr224 з  $\beta$ -тубуліну. Сайт таксолу утворений Val23, His229, Phe272, Thr276 та Pro360 з  $\beta$ -тубуліну.

Було проаналізовано 3 ізоформи  $\alpha$ -тубулінів і 8 ізоформ  $\beta$ -тубулінів *A. thaliana*, 3 і 8 ізоформ у *O. sativa*, 6 і 8 у *Z. mays*, 1 і 5 у *T. aestivum*, 0 і 3 у *G. max*, 5 і 10 *M. sativa* ізоформи  $\alpha$ - і  $\beta$ -тубулінів, відповідно. У всіх зазначених видах рослин у порівнянні з тубулінами  $\alpha$  і  $\beta$  *Bos taurus*  $\alpha$ Val353 заміщений на Cys,  $\beta$ Tyr224 заміщений на Phe, і  $\beta$ Ala316 заміщений на Ser – Рис. 3.10.

Ми порівняли амінокислотні послідовності тваринних тубулінів, для яких відомі сайти зв'язування колхіцину, вінбластину та таксолу, з тубулінами з дводольних – *Arabidopsis*, *Glycine*, *Medicago* та однодольних – *Oryza*, *Triticum*, *Zea* квіткових рослин.

Геноми рослин, зазвичай, містять більше – до 10 генів тубуліну, ніж геноми тварин. Але охоплення послідовності геномів вибраних видів рослин сиквенуванням різне. Як для *Arabidopsis*, так і для *Oryza* ми проаналізували 3 анатовані послідовності  $\alpha$ - та 8 послідовностей  $\beta$ -тубулінів з баз даних UniProtKB/Swiss-Prot. Для *Zea*, *Triticum* та *Glycine* було проаналізовано 6  $\alpha$  / 8  $\beta$ , 1  $\alpha$  / 5  $\beta$  та 3  $\beta$  послідовностей тубулінів, відповідно. Для роду *Medicago* ми проаналізували 5 неаннотованих послідовностей  $\alpha$ - та 10  $\beta$ -тубуліну з баз даних UniProtKB/TrEMBL. Множинне вирівнювання названих послідовностей ми проводили за допомогою онлайн програми ClustalW зі стандартними параметрами.

Загалом, було проаналізовано 3 ізоформи  $\alpha$ -тубулінів і 8 ізоформ  $\beta$ -тубулінів *A. thaliana*, 3 і 8 ізоформ *O. sativa*, 6 і 8 *Z. mays*, 1 і 5 *T. aestivum*, 0 і 3 у *G. max*, 5 і 10 *M. sativa* ізоформ  $\alpha$ - і  $\beta$ -тубулінів, відповідно. У всіх зазначених видах рослин у порівнянні з тубулінами  $\alpha$  і  $\beta$  *Bos taurus*  $\alpha$ Val353 заміщений на Cys,  $\beta$ Tyr224 заміщений на Phe та  $\beta$ Ala316 заміщений на Ser – Рис. 3.10 С.



A.				B.					
	$\alpha$ 178	$\beta$ 255	$\beta$ 316		$\beta$ 23	$\beta$ 229	$\beta$ 272	$\beta$ 276	$\beta$ 360
<i>B. taurus</i>	S	L	A	<i>B. taurus</i>	V	H	F	T	P
<i>Arabidopsis</i>	S	L	S	<i>Arabidopsis</i>	V	H	F	T	P
<i>Glycine</i>	S	L	S	<i>Glycine</i>	V	H	F	T/A	P
<i>Medicago</i>	S	L	S	<i>Medicago</i>	V/A/-	H	F	T	P
<i>Oryza</i>	S	L	S	<i>Oryza</i>	V	H	F	T	P
<i>Triticum</i>	S	L	S	<i>Triticum</i>	V	H	F	T	P
<i>Zea</i>	S	L	S	<i>Zea</i>	V	H	F	T	P

C.							
	$\alpha$ 325	$\alpha$ 329	$\alpha$ 353	$\alpha$ 355	$\beta$ 177	$\beta$ 210	$\beta$ 224
<i>B. taurus</i>	P	N	V	I	V	Y	Y
<i>Arabidopsis</i>	P	N	C	I	V	Y	F
<i>Glycine</i>	P	N	?	I	V	Y	F
<i>Medicago</i>	P	N	C	I	V	Y	F
<i>Oryza</i>	P	N	C	I	V	Y	F
<i>Triticum</i>	P	N	C	I	V	Y	F
<i>Zea</i>	P	N	C	I	V	Y	F

**Рис. 3.10.** Амінокислотні залишки тубулінів  $\alpha$  та  $\beta$ , які утворюють сайти зв'язування **A.** колхіцину; **B.** таксолу; **C.** вінбластину.

Було виявлено, що у рослинному сайті зв'язування колхіцину амінокислотний залишок  $\alpha$ -тубулін Val353 замінюється на Cys, а залишок  $\beta$ -тубуліну Ala316 замінюється на Ser. Інші а.з., які утворюють сайт зв'язування колхіцину є консервативними. У рослинному сайті зв'язування вінбластину ми виявили наступні заміни: Val353Cys в  $\alpha$ -тубуліні та Ala316Ser у  $\beta$ -тубуліні. Решта залишків ідентичні відповідним залишкам тубулінів тварин. У сайті таксолу Thr276 рослинного  $\alpha$ -тубуліну заміщений Ala лише в ізоформі бета-2 *Glycine*, але в інших послідовностях він зберігається. Також у 2 з 10 послідовностей  $\beta$ -тубулінів *Medicago*, – ідентифікатори UniProt Q949G6 та G7LDD, Val23 сайту зв'язування таксолу, відповідає делеції одного нуклеотиду та заміні на Ala, відповідно.

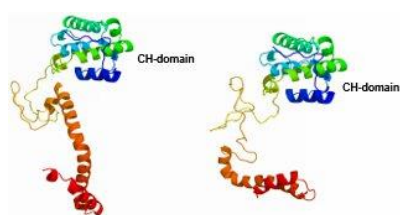
**Висновок** Наші результати демонструють, що сайти зв'язування колхіцину, вінбластину та таксолу практично повністю консервативні у всіх рослинних  $\alpha$ - та  $\beta$ -тубулінах. Заміни Val353Cys в  $\alpha$ -тубуліні та Ala316Ser у  $\beta$ -тубуліні у сайті зв'язування колхіцину у рослин можуть призвести до критичних відмінностей у природі цього сайту зв'язування у рослин порівняно з тваринами. У той же час, заміна у рослинному сайті зв'язування вінбластину Tyr224 у  $\beta$ -тубуліні подібним

ароматичним а.з. Phe може бути набагато менш значущою. У свою чергу, позиції амінокислот, які беруть участь у зв'язуванні таксолу в рослинних та тваринних тубулінах, виявились надзвичайно консервативними.

### 3.6.2 МАРи які стабілізують мікротрубочки

**EB1.** Білок EB1 складається з N-кінцевого домену гомологічного до кальпоніну (calponin homology domain CH-домен), що відповідає за зв'язування з МТ, головним чином, за рахунок електростатичних та гідрофобних взаємодій та консервативного С-кінцевого суперспірального домену димеризації, – домен EB гомології (EB homology domain). Кінцевий фрагмент С-кінцевої частини у приблизно 15-20 а.з. залишається неструктурованим. Домен гомології EB та неструктурований хвіст відповідають за розпізнавання мотивів SxIP та CAP-Gly, які забезпечують направлення білків +TIP, що містять ці мотиви, до наростаючих кінців мікротрубочок.

У геномі *A. thaliana* є дві основні ізоформи білка EB1: EB1A – 276а.з. та EB1C – 329а.з., що мають лише 52% ідентичності. Це пояснюється тим, що EB1C має 4 істотні за розміром вставки в С-кінцевому альфа-спіральному домені димеризації – Рис. 3.11.



**Рис. 3.11.** Моделі білків EB1A, зліва, і EB1C, справа, *A. thaliana*.

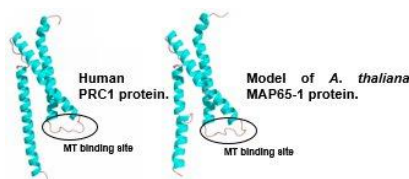
Ми порівняли гомологію білків EB1 людини та *A. thaliana*. Слід зазначити, що фрагменти EB1 *A. thaliana*, що відповідають залишкам 250–257 людського EB1C, які відповідають за розпізнавання мотивів SxIP, дуже дивергентні – Рис. 3.12.

Human EB1c	250	EGFVIPDE	257
At EB1A	238	ANESALED	245
At EB1C	258	GEDVGAAE	265

**Рис. 3.12.** Порівняння послідовностей білків EB1 людини та *A. thaliana*.

**MAP65** Білки сімейства MAP65 приймають участь в організації мікротрубочок в цитоплазмі клітини, зшиваючи нитки з антипаралельною орієнтацією. Вони містять три структурно різні домени: два домени, що зв'язують МТ, з характерною спектрин-укладкою та неструктурований сполучний домен, багатий на Lys і Arg.

Структуру домену зі спектрин-укладкою людського білка Prc1, віддаленого гомолога сімейства MAP65, було визначено методом кристалографії – модель PDBID: 3NRX. Його послідовність відповідає послідовності фрагменту 366-496 білку MAP65-1 *A. thaliana*, а ідентичність між даними послідовностями складає 24%. Але для доменів, які побудовані за подібним принципом, і складаються з трьох альфа-спіралей, які утворюють суперспіраль, важливим є не конкретний амінокислотний склад, а строге чередування гідрофільних та гідрофобних а.з. При збереженні даної умови допустимий велика ступінь варіації амінокислотного складу – Рис. 3.13.



**Рис. 3.13. А.** Білок *PRC1* людини. **В.** Модель білку *MAP65-1 A. thaliana*.

Тим не менш, низький рівень гомології амінокислотних послідовностей відображається в структурі. RMSD моделі за гомологією MAP65 *A. thaliana* від структури людського білку PRC1 складає 1.84Å. Родина білків MAP65 рослин, загалом, набагато різноманітніша порівняно з тваринами. Структурний аналіз різних представників рослинного сімейства білків MAP65 може дати більше розуміння в питанні про те, як за відсутності центріолей рослинні МТ

організуються з дифузних центрів організації мікротрубочок (microtubule organizing centers, МТОС).

### 3.6.3 МАРи, які дестабілізують мікротрубочки

**Кінезін-13А** На відміну від інших кінезінів, члени тваринного сімейства білків кінезін-13 активно деполімеризують МТ, а не активно рухаються вздовж них. Було показано, що ці білки утворюють кільцеві або спіральні олігомери навколо мікротрубочок. Такі комплекси можуть бути необхідними для утримання білків кінезін-13 на +кінці МТ і для продовження АТФ-залежної деполімеризації МТ. Ми побудували модель кінезину-13А *A. thaliana*, RMSD якої від шаблону становить 0,64А – Рис. 3.14.



Рис. 3.14. Модель кінезину-13А *A. thaliana*.

**Катанін** – це білок, що складається з двох субодиниць: катаніну р60 – АТФазавмісна субодиниця А1, яка відповідає за АТФ-залежну деполімеризацію МТ; і катаніну р80 – субодиниця В1, що містить домен WD40 і відповідає за +кінцеву локалізацію цього комплексу – Рис. 3.15.

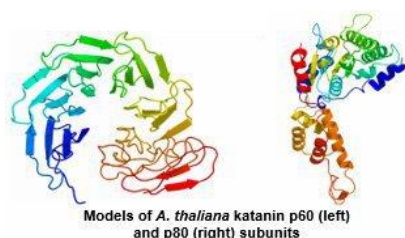
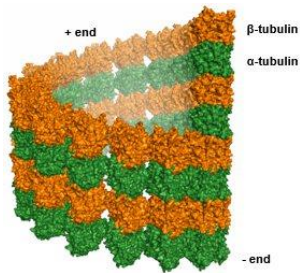


Рис. 3.15. Моделі субодиниць р60 (зліва) і р80 (зправа) катаніну *A. thaliana*.

**Модель мікротрубочки** За допомогою високомолекулярного моделювання *in silico* ми отримали можливу модель рослинної мікротрубочки. Використовуючи у якості шаблону модель кільцевого комплексу кінезину 13 та мікротрубочки – PDBID: 3EDL ми замінили координати тубулінів  $\alpha$  та  $\beta$  великої рогатої худоби на координати моделей за гомологією тубулінів *A. thaliana*. Отримана модель складається з 33  $\alpha$  /  $\beta$  димерів тубуліну.



**Рис. 3.16.** Модель мікротрубочки *A. thaliana* за гомологією, що складається з 33  $\alpha$  /  $\beta$  димерів тубуліну.

Отримана модель є підходящим інструментом для моделювання взаємодії з МАРами рослин та механізму (де)полімеризації рослинних МТ і може бути використана в подальшому для аналізу зв'язування МАРів рослин з МТ – Рис. 3.16.

### 3.7 Аналіз експресії цитоскелетних білків льону

#### 3.7.1 $\alpha$ -тубуліни

Ми кількісно визначили відносну експресію генів *CesA* та генів актину та альфа-тубуліну у стеблах льону на двох різних стадіях росту, а саме «проростання» та «швидкого росту», використовуючи RT-qPCR.

Вимірjana експресія пулів генів тубуліну та актину була приблизно на одному рівні у всіх досліджуваних зразках. Це може пояснюватись фактичною відсутністю коливань або через недостатню роздільну здатність експерименту, оскільки експресію окремих генів не було виміряно. Перша можливість корелює з тим фактом, що експресія генів актину та тубуліну, зазвичай, стала і часто

використовується для нормалізації результатів експресії (Gonzalez-Verdejo *et al.*, 2008). Однак, низка фактів підтверджує другу можливість. Раніше було показано, що експресія окремих генів актину нестабільна на різних стадіях розвитку рослин (Guenin *et al.*, 2009).

Подібні властивості було зареєстровано і для генів тубуліну: результати аналізу за допомогою geNorm довели, що експресія генів тубуліну є нестабільною при абіотичному стресі та паростку *Lolium perenne* (Martin *et al.*, 2008). Подібно до вищезазначених спостережень, дослідження, в якому шукали кандидатів-референтних генів для qPCR у льону, показує нестабільність експресії генів актину та тубуліну (Huis *et al.*, 2010). Як ми вже передбачали раніше, у геномі льону є 7 генів тубуліну та 10 актинів (Bayer *et al.*, 2014). Тому в майбутньому було б цікаво вивчити експресію окремих генів ізотипів актину та тубуліну у паростках льону.

Порівняльний аналіз передбачуваних амінокислотних послідовностей шести генів  $\alpha$ -тубуліну льону показує, що пара генів *Lus10020281* – *Lus10005705* (*LusTua1a*; *LusTua1b*) дуже гомологічна до  $\alpha 1$ -тубуліну *A. thaliana* – TUA1, який, хоча і значною мірою експресується в пилку (Carpenter *et al.*, 1992), також був візуалізований за допомогою експериментів з міченням синім флуоресцентним білком (CFP) у тій частині МТ, яка була асоційованій з CesA під час відкладення клітинної стінки (Paredes *et al.*, 2006).

Пара генів льону *Lus10039169* та *Lus10013765* (*LusTua3a*, *LusTua3b*) дуже гомологічна до ізотипів  $\alpha 3$ -/ $\alpha 5$ -тубуліну *A. thaliana* (*TUA3/TUA5*), тоді як амінокислотна послідовність пари генів льону *Lus10035422* та *Lus10031032* (*LusTua2a*; *LusTua2b*) гомологічна до ізотипів тубуліну *A. thaliana* *TUA2/TUA4/TUA6*. Кожна з даних двох пар локусів кодує ідентичні білки. Попередні дані транскрипції qRT-PCR, отримані шляхом порівняння різних сортів льону, показують, що онологи *LusTua2* переважно експресуються в апексі та флоемі, тоді як пара *LusTua3* переважно експресується у листі (Galinousky *et al.*, 2017). Згідно з тими ж даними, спостерігався певний рівень дисбалансу в експресії  $\alpha$ - та  $\beta$ -тубуліну, що призводить до загальної більшої кількості ізотипів

*LusTua2* та *LusTua3* у апексі, листках та флоемі. Це нагадує те, що вже спостерігалось у *Populus* (Oakley *et al.*, 2007) та наводить на думку про регуляторні механізми, які можуть по-різному впливати на експресію генів  $\alpha$ - та  $\beta$ -тубуліну (Breviario 2008).

### 3.7.2 $\beta$ -тубуліни

14 генів  $\beta$ -тубуліну льону експресуються, хоча і в різній мірі, в різних тканинах або на різних стадіях росту. Аналіз експресії генів у стеблі, гіпокотилі та квітці дав можливість припустити, що пули різних ізотипів тубуліну можуть утворювати різні МТ з можливими спеціалізованими функціями. Встановлено, що відносний рівень транскриптів  $\beta$ -тубуліну модулюється під час подовження стовбура та гіпокотилів. Однак, обидва онологи кожної пари виявляють подібний рівень експресії у всіх аналізованих тканинах, за винятком пари генів ізотипу *LusTub4*, де є помітні відмінності між двома членами пари, – Gavazzi *et al.*, 2017. У верхівці стебла, де більшість клітин активно ділиться, був виявлений високий рівень експресії для пар генів ізотипів *LusTub3* та *LusTub7*. І навпаки, рівень експресії пар генів ізотипів *LusTub2* і *LusTub5* був дуже низьким. Подібна закономірність диференціальної експресії спостерігалась також у корневих кінчиках, де гени ізотипу *LusTub7* мали найвищий рівень експресії, тоді як транскрипти ізотипів *LusTub2* були ледь помітними – Morello, неопубліковано. У зв'язку з цим можна зробити висновок, що мікротрубочки, що утворюють пре-профазне веретина або мітотичне веретено, можуть формуватися переважно ізотипами білка TUB7. Подібні дані були отримані, також, і з досліджень, проведених на ячмені, яка показала різницю в картині експресії  $\alpha$ -тубулінів, яка спостерігалась у різних ділянках листя: експресія транскриптів HVATUB2 та HVATUB4 була вище в базальній меристематичній тканині, ніж у ділянках швидкого подовження листя (Schröder *et al.*, 2001).

Використовуючи дані по апексам в якості контролю, також, спостерігалось підвищення регуляції експресії більшості генів у верхній частині стебла –

Gavazzi *et al.*, 2017, яке характеризується значним подовженням клітин клітковини та інших типів клітин. Цей процес пов'язаний з поперечною орієнтацією кортикальних МТ, які допомагають правильному розташуванню комплексів CesA в мембрані. Далі вниз по стеблу, експресія деяких генів  $\beta$ -тубуліну, *LusTub3a* та *b*, *LusTub5a* та *b* та *LusTub6*, була чітко регуляторно зниженою у наступних секціях, що характеризуються переорієнтацією кортикальних МТ, одночасно з зупинкою подовження та поступовим потовщення клітинної стінки у клітинах флоєми. І навпаки, регуляторне підвищення експресії *LusTub1*, *2* і *4* у верхній частині стебла зберігалось по всій довжині стебла, навіть у самих базальних його областях, де потовщення клітинної стінки було вже практично завершено. Паралоги *LusTub7* продемонстрували протилежну картину експресії – вони експресувались переважно у апексі, а у всіх інших частинах стебла експресія була регуляційно зниженою.

У квіток льону було помічено, що повне дозрівання пилку під час цвітіння викликає зміни рівня експресії різних ізотипів  $\beta$ -тубуліну при сталій загальній кількості транскриптів, зі значним регуляторним підвищенням експресії пар генів *LusTub3* та *4*, разом з паралельним зниженням кількості мРНК *LusTub2* та *5* – Gavazzi *et al.*, 2017. *LusTub3* належать до  $\beta$ -тубулінів класу III, члени якого переважно експресуються в пилку багатьох інших видів рослин. Це може свідчити про те, що специфічний для пилку пул транскриптів тубуліну збирається у спеціалізовані МТ, більш придатні для тонкої доставки везикул, що несуть білки і ліпіди до мембрани кінчика пилкової трубки, коли ріст поляризованої пилкової трубки повинен безперервно підтримуватися.



## РОЗДІЛ 4

### ПОШУК ТА АНАЛІЗ ГЕНІВ ЦЕЛЮЛОЗОСИНТАЗ У ГЕНОМІ ЛЬОНУ

#### 4.1 Пошук генів целюлозосинтаз у геномі льону

Щоб здійснити пошук *in silico* генів целюлозосинтаз льону, ми використали базу даних UniProt, яка на той момент містить шість анотованих послідовностей, закодованих у геномі льону, розміром приблизно 157 амінокислотних залишків кожна. Ці послідовності вважаються фрагментами генів целюлозосинтаз. Вищезазначені послідовності не є повнорозмірними, оскільки, наприклад, у *A. thaliana* є десять типів целюлозосинтаз – Hamann et al. 2004, кожна з яких містить в середньому 1000 амінокислотних залишків. Використовуючи ці шість послідовностей, ми провели пошук серед 43384 генів, що кодують білки, передбачених у геномі льону Wang et al. 2012, доступних у базі даних Phytozome v10. Застосований у цій роботі інструмент TBLASTN дозволяє порівнювати конкретні амінокислотні послідовності з пулом даних амінокислотних послідовностей, які отримані внаслідок розрахунку трансляції нуклеотидної послідовності геному у всіх шести можливих рамках зчитування. Пошук подібності у геномі льону привів до виявлення 29 локусів кандидатів на роль генів целюлозосинтаз, – Таблиця 4.1. На додаток до них, ми виявили послідовність-кандидат, віднесену до різних локусів – *Lus10008225* і *Lus10008226* і надалі іменуємо її *Lus100082256*, і два фрагменти, для яких поки що немає даних про можливі транскрипти (scaffold57 1381011–1385072 і scaffold464 1450325–1454110), доводячи кількість ідентифікованих послідовностей до загальної кількості 32.

Таблиця 4.1

**Дані про послідовностях генів-кандидатів на роль целюлозосинтаз і білків, подібних до целюлозосинтаз льону і *A. thaliana* (AT)**

Ген. локус	Сімейство	К-сть екзонів	Пар нукл.	К-сть а.з.	Ген. локус	Сімейство	К-сть екзонів	Пар нукл.	К-сть а.з.
<b>AT4G32410</b>	<i>CesA</i>	<b>14</b>	<b>6046</b>	<b>1081</b>	<b>AT3G03050</b>	<i>CsID3</i>	<b>4</b>	<b>4847</b>	<b>1145</b>
<i>Lus10018902</i>	<i>CesA</i>	13	5088	1079	<i>Lus10001619</i>	<i>CsID</i>	4	3431	1069
<i>Lus10028597</i>	<i>CesA</i>	13	5153	1079	<i>Lus10002134</i>	<i>CsID</i>	4	3572	1060
<i>Lus10007538</i>	<i>CesA</i>	14	5836	1092	<i>Lus10009248</i>	<i>CsID</i>	3	3809	1095
<i>Lus10012198</i>	<i>CesA</i>	14	5804	1094	<i>Lus10010024</i>	<i>CsID</i>	3	4339	1167
<i>Lus10039607</i>	<i>CesA</i>	13	4944	1069	<i>Lus10012119</i>	<i>CsID</i>	5	4667	1065
<i>Lus10008225-6</i>	<i>CesA</i>	10	3796	987	<i>Lus10013851</i>	<i>CsID</i>	3	4052	1128
<i>Lus10006161</i>	<i>CesA</i>	14	5118	1074	<i>Lus10022982</i>	<i>CsID</i>	3	3587	1127
<i>Lus10041063</i>	<i>CesA</i>	14	5329	1096	<i>Lus10026609</i>	<i>CsID</i>	4	3392	1031
<i>Lus10002939</i>	<i>CesA</i>	13	4844	1097	<i>Lus10026610</i>	<i>CsID</i>	4	3532	1056
<i>Lus10002940</i>	<i>CesA</i>	11	5513	906	<i>Lus10030453</i>	<i>CsID</i>	2	2717	869
<i>Lus10003526</i>	<i>CesA</i>	13	4855	1097	<i>Lus10026568</i>	<i>CsID</i>	3	3589	1135
<i>Lus10022449</i>	<i>CesA</i>	11	4616	1035	<i>Lus10000755</i>	<i>CsID</i>	4	3296	991
<i>scaffold57</i>	<i>CesA</i>	14	4015	978	<i>Lus10011736</i>	<i>CsID</i>	4	3164	949
<i>scaffold464</i>	<i>CesA</i>	13	3393	959	<i>Lus10038008</i>	<i>CsID</i>	3	3752	1064
<i>Lus10029245</i>	<i>CesA</i>	11	3867	988	<i>Lus10025046</i>	<i>CsID</i>	3	4137	1073
<i>Lus10007296</i>	<i>CesA</i>	11	3867	988	<i>Lus10030455</i>	<i>CsID</i>	3	3303	953

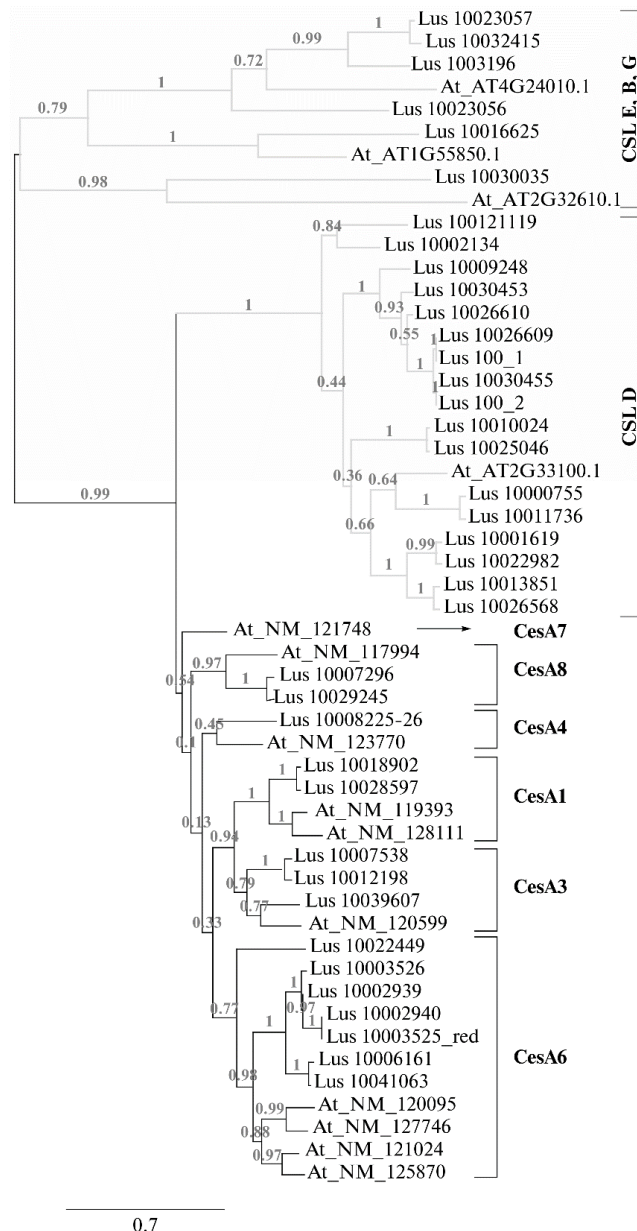
Загальновідомо, що геном рослин містять як гени целюлозосинтаз, так і гени, що кодують білки, подібні до целюлозосинтаз, які виконують синтез різних β-глюканів Richmond et al. 2000; Hazen et al. 2002; Fincher 2009;. Для розрізнення генів целюлозосинтаз та білків, подібних до целюлозосинтаз ми провели аналіз структури ідентифікованих генів льону з метою виділення генів справжніх целюлозосинтаз. Ми визначили, що ідентифіковані послідовності нуклеотидів при трансляції дають поліпептидні ланцюги, що складаються з не менше 869 амінокислотних залишків. Інформація про кількість пар основ у відповідних послідовностях генів та про кількість екзонів та амінокислотних залишків у трансльованих з них поліпептидах наведена в Таблиці 4.1. Для послідовностей у

scaffold57 та scaffold464 ми визначили кількість екзонів та залишків амінокислот на основі їх порівняння з генами *A. thaliana*.

Аналіз отриманих даних дозволив розділити ідентифіковані гени на дві групи за кількістю екзонів. Перша група охоплювала послідовності з 11–14 екзонами, а друга група – з 2–5 екзонами. Відмінності в генній структурі дозволили розрізнити гени целюлозосинтаз та гени целюлозосинтазо-подібних білків. У рослин гени целюлозосинтаз мають більш складну екзон-інтронну структуру, порівняно з генами целюлозосинтазо-подібних білків, що необхідно для диференціальної регуляції експресії генів целюлозосинтаз Richmond et al. 2000. Наприклад, ген *A. thaliana CesA1*, який відповідає локусу AT4G32410, містить 14 екзонів, тоді як ген *CslD3*, що відповідає локусу AT3G03050, що кодує целюлозосинтазо-подібний білок, містить лише чотири екзони. Виходячи з цього, можна припустити, що ідентифіковані гени льону, які містять 11–14 екзонів – *Lus10018902*, *Lus10028597*, *Lus10007538*, *Lus10012198*, *Lus10039607*, *Lu100082256*, *Lus10006161*, *Lus10041063*, *Lus10002939*, *Lus100029260*, *Lus100029354*, і *Lus10007296* – це гени целюлозосинтаз, а решта генів з 4–5 екзонами – гени целюлозосинтазо-подібних білків.

Для підтвердження цього припущення було проведено філогенетичний аналіз 32 передбачених послідовностей генів-кандидатів. Послідовності, наведені в Таблиці 4.1, використовувались для побудови філогенетичного дерева. Послідовність *AT3G03050*, що відповідає гену, подібному до целюлозосинтаз *CslD3*, у геномі *A. thaliana* була взята у якості зовнішньої групи. Багаторазове вирівнювання послідовностей з подальшою реконструкцією філогенетичного дерева показало, що послідовності згрупувались у дві різні клади, – Рис. 4.1. Одна клада включала послідовності генів льону, які містили 11–14 екзонів разом із послідовністю *AT4G32410 CesA1 A. thaliana*, тоді як інша клада містила послідовності *AT3G03050 CslD3 A. thaliana* та послідовності генів льону, які містять по 2–5 екзонів, – Рис. 4.1. На підставі результатів філогенетичного аналізу всі 16 генів льону, що потрапили в кладу *CslD*, ми вважаємо генами білків, подібних до целюлозосинтаз, – Таблиця 4.1, і з цієї

причини вони були виключені з подальшого аналізу. В результаті, серед 32 послідовностей, ідентифікованих у геномі льону, послідовності *Lus10018902*, *Lus10028597*, *Lus10007538*, *Lus10012198*, *Lus10039607*, *Lus100082256*, *Lus10006161*, *Lus0041063*, *Lus10002939*, *Lus10002940*, *Lus10003526*, *Lus10022449*, *scaffold57*, *scaffold464*, *Lus10029245*, і *Lus10007296* було обрано як найбільш вірогідні гени, що кодують целюлозосинтази льону.

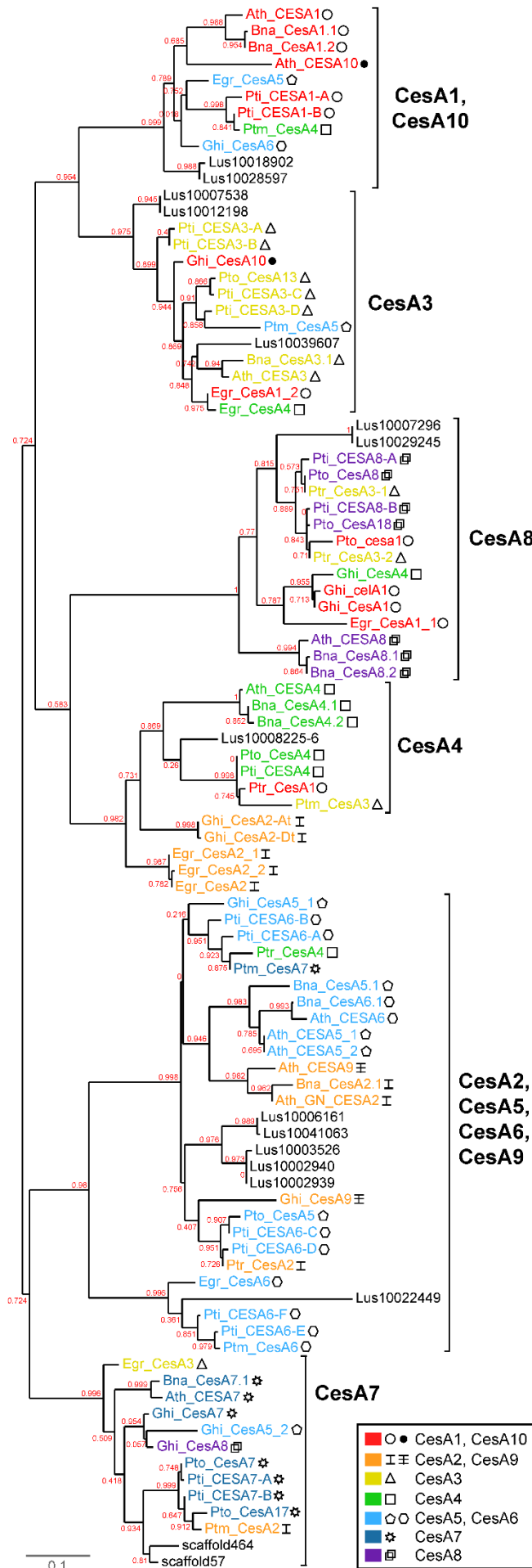


**Рис. 4.1.** Філогенетичні відношення між генами-кандидатами целюлозосинтаз льна, позначені *Lus* або *scaffold*, і *Arabidopsis thaliana* (*At*).

Проаналізовані послідовності розподіляються на три великі групи: перша включає сімейства генів *CslE,B,G*, друга – *CslD*, а третя – *CesA*.

#### 4.2 Аналіз генів целюлозосинтаз льону

**Філогенетичний аналіз генів-кандидатів целюлозосинтаз льону** При спробі розподілити гени целюлозосинтаз на класи, певні труднощі можуть виникнути навіть у межах одного виду. Ці труднощі спричинені тим, що гени целюлозосинтаз льону класів *CesA6*, *CesA9* та *CesA10* філогенетично наближені до відповідних генів класів *CesA5*, *CesA2* та *CesA1* – Рис. 4.2, що може бути пояснено двома останніми подіями дуплікації геному в рослинах метаспермів, зокрема льону Sveinsson et al. 2014. З цієї причини, також, неможливо розрізнити гени целюлозосинтаз типів *CesA2*, *CesA5*, *CesA6* та *CesA9* у роду *Populus*, філогенетично найбільш тісно спорідненого з родом *Linum*, серед видів, для яких нуклеотидна послідовність геному вже визначена. Наразі ці гени входять до однієї групи, *CesA6X* Kumar et al. 2009. Тому, коли ми відносимо гени целюлозосинтаз льону до того чи іншого класу за їх схожістю з даними генами *Populus*, гени льону, подібні до генів, що належать до розглянутої вище групи, також позначаються як *CesA6X* Mokshina et al. 2014, проте, насправді, ця група охоплює гени целюлозосинтаз кількох класів, а саме *CesA2*, *CesA5*, *CesA6* та *CesA9*.



**Рис. 4.2.** Філогенетичне дерево послідовностей генів целюлозосинтаз льна і генів целюлозосинтаз генетично близьких до *Linum usitatissimum* видів, для яких у базі даних UniProt міститься інформація про віднесення генів целюлозосинтаз до певних типів: *Populus tomentosa* (Pto – 8 генів), *Populus tremula/tremuloides* (Ptm – 11 генів), *Eucalyptus grandis* (Egr – 10 генів), *Gossypium hirsutum* (Ghi – 12 генів), *Gossypium hirsutum* (Ghi – 12 генів), *Brassica napus* (Bna – 11 генів) и *Arabidopsis thaliana* (Ath – 11 генів). 3 генів кандидатів целюлозосинтаз льна, позначені Lus або scaffold, два гени гомологічні послідовностям групи *Cesa1,10*, чотири – генам *Cesa3*, два гени гомологічні *Cesa8*, один – *Cesa4*, сім – послідовностям групи *Cesa2,5,6,9*, два – *Cesa7*.

Ми провели філогенетичний аналіз продуктів генів целюлозосинтаз видів, генетично споріднених з льном, для яких в базі даних UniProt є інформація про гени целюлозосинтаз, та їх приналежність до певних класів, а саме: *Populus tomentosa* (вісім генів), *Populus tremula*

/ *tremuloides* (11 генів), *Eucalyptus grandis* (десять генів), *Gossypium hirsutum* (12 генів), *Brassica napus* (11 генів) та *Arabidopsis thaliana* (11 генів) – Таблиця 4.2, Рис. 4.2. Для філогенетичного аналізу, загалом, було використано 63 гени целюлозосинтаз. За даними бази даних UniProt, для 11 генів *A. thaliana* було визначено наявність білкових продуктів, тоді як для решти 52 генів спостерігали наявність транскриптів відповідних генам *A. thaliana*. Здається очевидним, що чим більше різних видів рослин береться для філогенетичного аналізу генів целюлозосинтаз, тим точніше будуть класифіковані гени целюлозосинтаз льону.

Таблиця 4.2

**Взяті для філогенетичного аналізу гени целюлозосинтаз генетично близьких до *Linum usitatissimum* видів, для яких у базі даних UniProt є інформація про віднесення генів целюлозосинтаз до певних типів**

UniProt id	Назва гену	Клас Cesa	К-ість а.з.	UniProt id	Назва гену	Клас Cesa	К-ість а.з.
<i>P. tomentosa</i>				<i>G. hirsutum</i>			
Q4ZJ80_POPTO	<i>Pto_cesa1</i>	<i>Cesa1</i>	978	H2D036_GOSHI	<i>Ghi_Cesa2-At</i>	<i>Cesa2-At</i>	1039
F6KQG4_POPTO	<i>Pto_Cesa13</i>	<i>Cesa13</i>	1079	H2D037_GOSHI	<i>Ghi_Cesa2-Dt</i>	<i>Cesa2-Dt</i>	1040
U5L350_POPTO	<i>Pto_Cesa4</i>	<i>Cesa4</i>	1042	Q8W1W0_GOSHI	<i>Ghi_Cesa4</i>	<i>Cesa4</i>	974
F6KQG5_POPTO	<i>Pto_Cesa5</i>	<i>Cesa5</i>	1100	L7NUN4_GOSHI	<i>Ghi_Cesa5_1</i>	<i>Cesa5</i>	1095
LOATN1_POPTO	<i>Pto_Cesa7</i>	<i>Cesa7</i>	1036	E0WVS1_GOSHI	<i>Ghi_Cesa5_2</i>	<i>Cesa5</i>	1042
F6KQG3_POPTO	<i>Pto_Cesa17</i>	<i>Cesa17</i>	1032	L7NU96_GOSHI	<i>Ghi_Cesa6</i>	<i>Cesa6</i>	1083
F6KQF9_POPTO	<i>Pto_Cesa8</i>	<i>Cesa8</i>	978	L7NUA2_GOSHI	<i>Ghi_Cesa7</i>	<i>Cesa7</i>	1042
F6KQG0_POPTO	<i>Pto_Cesa18</i>	<i>Cesa18</i>	978	L7NUG2_GOSHI	<i>Ghi_Cesa8</i>	<i>Cesa8</i>	1039
<i>P. tremuloides</i>				L7NUN6_GOSHI	<i>Ghi_Cesa9</i>	<i>Cesa9</i>	1090
Q6J8X2_9ROSI	<i>Ptr_Cesa1</i>	<i>Cesa1</i>	1042	L7NUA3_GOSHI	<i>Ghi_Cesa10</i>	<i>Cesa10</i>	1068
Q6J8X1_9ROSI	<i>Ptr_Cesa2</i>	<i>Cesa2</i>	1095	<i>B. napus</i>			
Q8GSW2_POPTM	<i>Ptm_Cesa2</i>	<i>Cesa2</i>	1032	D8LIW1_BRANA	<i>Bna_Cesa1.1</i>	<i>Cesa1.1</i>	1083
Q7XAV3_POPTM	<i>Ptm_Cesa3</i>	<i>Cesa3</i>	1042	D8LIW2_BRANA	<i>Bna_Cesa1.2</i>	<i>Cesa1.2</i>	1083
Q6J8X0_9ROSI	<i>Ptr_Cesa3-1</i>	<i>Cesa3-1</i>	978	D8LIW3_BRANA	<i>Bna_Cesa2.1</i>	<i>Cesa2.1</i>	1074
Q6J8W9_9ROSI	<i>Ptr_Cesa3-2</i>	<i>Cesa3-2</i>	978	D8LIW4_BRANA	<i>Bna_Cesa3.1</i>	<i>Cesa3.1</i>	1066
Q6J8W8_9ROSI	<i>Ptr_Cesa4</i>	<i>Cesa4</i>	1096	D8LIW5_BRANA	<i>Bna_Cesa4.1</i>	<i>Cesa4.1</i>	1038
Q6YBV2_POPTM	<i>Ptm_Cesa4</i>	<i>Cesa4</i>	1083	D8LIW6_BRANA	<i>Bna_Cesa4.2</i>	<i>Cesa4.2</i>	1052
Q8H2C6_POPTM	<i>Ptm_Cesa5</i>	<i>Cesa5</i>	1078	D8LIW7_BRANA	<i>Bna_Cesa5.1</i>	<i>Cesa5.1</i>	1070
Q6XZC2_POPTM	<i>Ptm_Cesa6</i>	<i>Cesa6</i>	1087	D8LIW8_BRANA	<i>Bna_Cesa6.1</i>	<i>Cesa6.1</i>	1084
Q6YBV3_POPTM	<i>Ptm_Cesa7</i>	<i>Cesa7</i>	1096	D8LIW9_BRANA	<i>Bna_Cesa7.1</i>	<i>Cesa7.1</i>	1031
<i>E. grandis</i>				D8LIX0_BRANA	<i>Bna_Cesa8.1</i>	<i>Cesa8.1</i>	984
Q2IB43_EUCGR	<i>Egr_Cesa1_1</i>	<i>Cesa1</i>	978	D8LIX1_BRANA	<i>Bna_Cesa8.2</i>	<i>Cesa8.2</i>	984
B1NYI6_EUCGR	<i>Egr_Cesa1_2</i>	<i>Cesa1</i>	1080	<i>A. thaliana</i>			
B1NYI9_EUCGR	<i>(Egr_Cesa1_2)</i>	<i>Cesa1</i>	1080	CESA1_ARATH	<i>Ath_CESA1</i>	<i>Cesa1</i>	1081
B1NYI7_EUCGR	<i>Egr_Cesa2_1</i>	<i>Cesa2</i>	1045	CESA2_ARATH	<i>Ath_GN_CESA2</i>	<i>Cesa2</i>	1084
Q2IB42_EUCGR	<i>Egr_Cesa2_2</i>	<i>Cesa2</i>	1045	CESA3_ARATH	<i>Ath_CESA3</i>	<i>Cesa3</i>	1065
B1NYJ0_EUCGR	<i>Egr_Cesa2</i>	<i>Cesa2</i>	1045	CESA4_ARATH	<i>Ath_CESA4</i>	<i>Cesa4</i>	1049
Q2IB41_EUCGR	<i>Egr_Cesa3</i>	<i>Cesa3</i>	1040	CESA5_ARATH	<i>Ath_CESA5_1</i>	<i>Cesa5</i>	1069
Q2IB40_EUCGR	<i>Egr_Cesa4</i>	<i>Cesa4</i>	1080	G1FKV9_ARATH	<i>Ath_CESA5_2</i>	<i>Cesa5</i>	996
Q2IB39_EUCGR	<i>Egr_Cesa5</i>	<i>Cesa5</i>	1085	CESA6_ARATH	<i>Ath_CESA6</i>	<i>Cesa6</i>	1084
Q2IB38_EUCGR	<i>Egr_Cesa6</i>	<i>Cesa6</i>	1097	CESA7_ARATH	<i>Ath_CESA7</i>	<i>Cesa7</i>	1026
<i>G. hirsutum</i>				CESA8_ARATH	<i>Ath_CESA8</i>	<i>Cesa8</i>	985
P93155_GOSHI	<i>Ghi_celA1</i>	<i>Cesa1</i>	974	CESA9_ARATH	<i>Ath_CESA9</i>	<i>Cesa9</i>	1088
Q1EG93_GOSHI	<i>Ghi_Cesa1</i>	<i>Cesa1</i>	884	CESAA_ARATH	<i>Ath_CESA10</i>	<i>Cesa10</i>	1065

На філогенетичному дереві, отриманому в результаті проведеного аналізу, ми можемо спостерігати, що окремі гілки, як правило, охоплюють гени целюлозосинтаз, що належать до одного класу, хоча є кілька винятків, – Рис. 4.2. Клада *CesA1/10* утворена шістьма генами целюлозосинтаз, що належать до класів 1/10, але вона також містить гени целюлозосинтази *Ptm\_CesA4*, *Egr\_CesA5* та *Ghi\_CesA6*, що належать до класів *CesA4*, *CesA5* та *CesA6*, відповідно. Клада *CesA3* містить гени, що кодують целюлозосинтази типу 3, але вона включає і гени целюлозосинтаз *Ghi\_CesA10*, *Ptm\_CesA5*, *Egr\_CesA1\_2* та *Egr\_CesA4*, які належать до класів *CesA10*, *CesA5*, *CesA1* та *CesA4*, відповідно. На відміну від інших представників сімейства целюлозосинтаз, гени класів *CesA2*, *CesA5*, *CesA6* та *CesA9* разом утворюють єдину кладу. Можна припустити, що ці гени належать до певного класу або підкласу в єдиній групі целюлозосинтаз *CesA2/5/6/9*. Слід зазначити, що ця кладу відокремлюється від інших і містить гени чотирьох різних видів рослин. Клада *CesA7* охоплює сім генів целюлозосинтаз типу 7, а також містить гени целюлозосинтаз *Egr\_CesA3*, *Ghi\_CesA8*, *Ghi\_CesA5\_2* та *Ptm\_CesA2*, що належать до класів *CesA2*, *CesA8*, *CesA5* та *CesA2*, відповідно.

Спираючись на результати філогенетичного аналізу кандидатів генів целюлозосинтаз льону, разом із аннотованими генами целюлозосинтаз близькоспоріднених видів рослин, ми можемо зробити припущення, до якого класу віднести гени целюлозосинтаз льону, – Рис. 4.2. Гени льону *Lus10028597* та *Lus10018902* належать до близькоспорідненої групи генів *CesA1/10*. Гени *Lus10007538*, *Lus10012198* та *Lus10039607* належать до групи генів *CesA3*; гени *Lus10007538* та *Lus10012198* є продуктами дублювання генів, оскільки їх амінокислотні послідовності мають ідентичність 98,5%, в той час, як обидва вони мають лише 78,0% схожості з третім геном з цієї групи – *Lus10039607*; *Lus10007296* та *Lus10029245* належать до групи *CesA8*. Порівняння нуклеотидних послідовностей генів *Lus10007296* та *Lus10029245* показало 95,6% ідентичності між ними при порівнянні повних нуклеотидних послідовностей, та 93,8% ідентичності, коли для порівняння були взяті лише специфічні для класу



області. У той же, час амінокислотні послідовності білкових продуктів цих генів мають 98,3% подібності. Тому ці гени можна вважати синонімами, які являють між собою дві копії одного гена. Гени целюлозосинтаз, кодовані локусами *Lus10008225* та *Lus10008226*, належать до групи генів *CesA4*.

Гени *Lus10002939*, *Lus10002940*, *Lus10003526*, *Lus10006161*, *Lus10041063* та *Lus10022449* утворили єдину кладу з генами целюлозосинтаз класів *CesA2/5/6/9* інших видів. Гени льону показали вищу схожість між собою – 89,0% і вище, ніж гени целюлозосинтаз інших видів і вони утворили окрему підкладу в межах описаної кледи. Єдиним винятком тут був ген *Lus10022449*. Гени *Lus10006161* та *Lus10041063* продемонстрували 92,8% ідентичності та склали єдину підкладу, тоді як гени *Lus10002939*, *Lus10002940* та *Lus10003526* були ідентичними на 98% і утворили іншу підкладу. Слід також зазначити, що ген *Lus10002940* містить істотно великі делеції в регіонах 1170–1554 bp та 2165–2363 bp, що, можливо, може позначитись на целюлозосинтазній активності білкового продукту цього гена. Послідовності *Lus10003526* та *Lus10002939* ми вважаємо синонімами, оскільки вони ідентичні на 97,6% при порівнянні повномасштабних генів між собою та 97,9% ідентичних при порівнянні специфічних для класу областей, тоді як амінокислотні послідовності, кодовані цими генами, мають 98,9% подібності.

У той же час ген *Lus10022449* класу *CesA2/5/6/9* відокремлюється від інших генів льону цієї групи і виявляє лише 60,4–64,4% схожості з ними. Ген *Lus10022449* утворює єдину гілку з генами, позначеними як *Egr\_CesA6*, *Pti\_CESA6F*, *Pti\_CESA6E* та *Ptm\_CesA6*, причому подібність між цими чотирма генами становить 88,3–98,7%, а подібність між цими генами та *Lus10022449* становить у середньому 71% – від 70,7% до 71,1%. Гени, названі *scaffold57* і *scaffold464*, включені до кледи гена *CesA7*; амінокислотні послідовності білкових продуктів цих генів мають 75,7% схожості та становлять єдину пару на філогенетичному дереві, окремо від генів *CesA7* всіх інших видів.

**Віднесення генів целюлозосинтаз льону до певних типів целюлозосинтаз**  
Раніше вже були зроблені спроби класифікувати гени целюлозосинтаз льону та віднести їх до певних класів Sveinsson et al. 2014. Тому ми вважали важливим

доповнити отримані дані та більш точно визначити, до яких класів належать целюлозосинтази льону, на основі результатів нашого аналізу. При віднесенні целюлозосинтаз льону до того чи іншого класу та присвоєння їм офіційних назв слід враховувати наступні міркування. Перш за все, у випадку целюлозосинтаз льону, які, імовірно, належать до класів *CesA3*, *CesA8* та *CesA2/5/6/9*, існують дуже тісно спорідненого послідовності, які, швидше, слід вважати синонімічними копіями одного варіанту целюлозосинтаз певного класу, ніж різними варіантами целюлозосинтаз. Гени, що представляють різні варіанти целюлозосинтаз одного типу, позначаються цифрами, відокремленими від основної назви нижньою лінією, тобто «\_1», починаючи з найменш диференційованої послідовності; синонімічні послідовності позначаються додаванням номера індексу, наприклад “-1”. По-друге, важливо позначити ті гени, які належать до однієї групи, але представляють різні типи целюлозосинтаз. Це актуально для таких випадків, як комбіновані групи *CesA1/10* та *CesA5/6/2/9*, для яких, поки, досить складно визначити окремі класи. Такі гени позначаються латинськими літерами, такими як А, В та ін. Гени-кандидати целюлозосинтаз, відібрані на основі попереднього філогенетичного аналізу, ми класифікували, групуючи їх з відповідними ортологами з інших видів рослин. Конкретні типи кандидатів генів целюлозосинтаз льону та їх формальні назви наведені в Таблиці 4.3.

Таблиця 4.3

## Передбачені послідовності генів целюлозосинтаз льону

Ген, локус	К-ість а.з.	Клас Cesa	Клас Cesa, згідно Mokshina et al., 2014
<i>Lus10018902</i>	1079	<i>CesA1/10_A</i>	<i>LusCESA1-A</i>
<i>Lus10028597</i>	1079	<i>CesA1/10_B</i>	<i>LusCESA1-B</i>
<i>Lus10007538</i>	1092	<i>CesA3_1-1</i>	<i>LusCESA3-B</i>
<i>Lus10012198</i>	1094	<i>CesA3_1-2</i>	<i>LusCESA3-C</i>
<i>Lus10039607</i>	1069	<i>CesA3-2</i>	<i>LusCESA3-A</i>
<i>Lus10008225-6</i>	987	<i>CesA4</i>	<i>LusCESA4</i>
<i>Lus10006161</i>	1074	<i>CesA5/6/2/9_A</i>	<i>LusCESA6-A</i>
<i>Lus10041063</i>	1096	<i>CesA5/6/2/9_B</i>	<i>LusCESA6-B</i>
<i>Lus10002939</i>	1097	<i>CesA5/6/2/9_C-2</i>	<i>LusCESA6-D</i>
<i>Lus10002940</i>	906	<i>CesA5/6/2/9_C-3</i>	<i>LusCESA6-E</i>
<i>Lus10003526</i>	1097	<i>CesA5/6/2/9_C-1</i>	<i>LusCESA6-C</i>
<i>Lus10022449</i>	1035	<i>CesA5/6/2/9_D</i>	<i>LusCESA6-F</i>
<i>scaffold57</i>	978	<i>CesA7_1</i>	<i>LusCESA7-A</i>
<i>scaffold464</i>	959	<i>CesA7_2</i>	<i>LusCESA7-B</i>
<i>Lus10029245</i>	988	<i>CesA8_1-1</i>	<i>LusCESA8-B</i>
<i>Lus10007296</i>	988	<i>CesA8_1-2</i>	<i>LusCESA8-A</i>

Дві послідовності геному льону віднесені до групи *CesA1/10* і позначені як гени *CesA1* та *CesA10*. На цьому етапі представляється неможливим визначити, до якого типу целюлозосинтаз, *CesA1* або *CesA10*, належать тісно спорідненого гени *Lus10018902* та *Lus10028597*. Вони позначаються як *CesA1/10\_A* та *CesA1/10\_B*.

На наш погляд, у класі *CesA5/6/2/9* можуть існувати чотири окремі підгрупи генів льону: підгрупа А, що включає *Lus10006161*; підгрупа В, що включає *Lus10041063*; підгрупа С, що охоплює *Lus10002939*, *Lus10002940* та *Lus10003526*; підгрупа D з *Lus1002244*. Гени в кожній з цих підгруп можуть

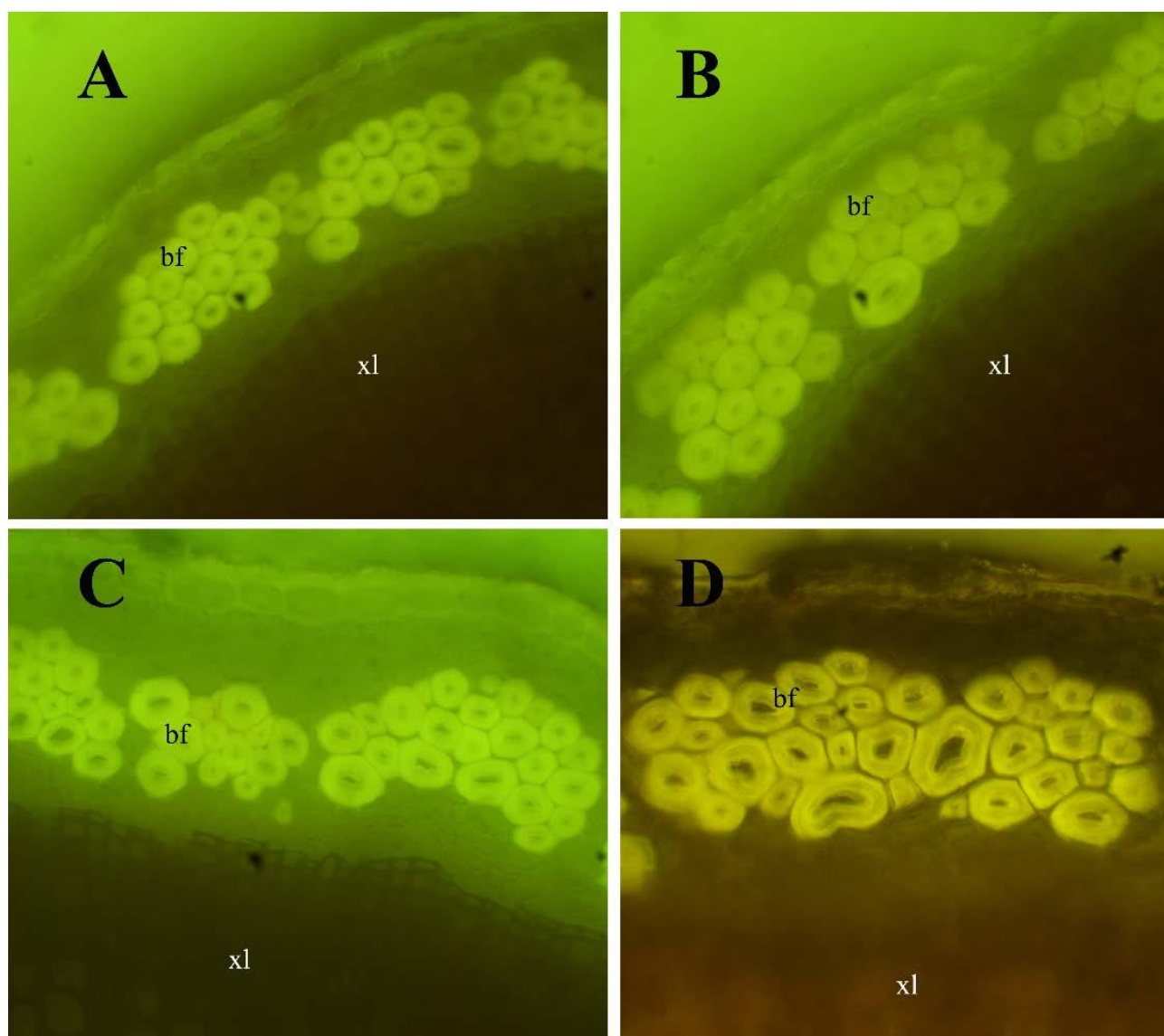
представляти целюлозосинтази певних окремих класів, причому підгрупа С представлена трьома тісно спорідненого копіями генів, 98% схожість між ними дає можливість вважати їх синонімічними. Таким чином, у групі *CesA5/6/2/9* ми визначили чотири класи генів *CesA*: *CesA2*, *CesA5*, *CesA6* та *CesA9*. Клас *CesA3* представлений у льоні двома варіантами генів: *CesA3\_1*, який існує як дві тісно спорідненого копії *CesA3\_11 – Lus10007538* та *CesA3\_12 – Lus10039607* та *CesA3\_2 – Lus10039607*. На даний момент, ми можемо зробити висновок, що клас *CesA4* представлений лише одним геном, *Lus100082256*. Кластер *CesA7* включає два фрагменти *CesA* геному льону, які представляють два варіанти гена целюлозосинтаз типу 7, варіант *CesA7\_1 – scaffold57* та варіант *CesA7\_2 – scaffold464*. У той же час, група *CesA8* містить один варіант гена *CesA* льону, представлений двома тісно спорідненого копіями *CesA8\_11 – Lus10029245* та *CesA8\_12 – Lus10007296*.

**Висновок** Підводячи підсумок, проведений аналіз дозволив нам ідентифікувати 32 гени-кандидати целюлозосинтаз льону, 16 з яких з великою вірогідністю кодують саме целюлозосинтази, а решта 16 кодують білки, подібні до целюлозосинтаз. Кількість екзонів у гені може служити критерієм точного віднесення генів льону до генів целюлозосинтаз – 11–14 екзонів та гени білків, подібних до целюлозосинтаз – 2–5 екзонів). Філогенетичний аналіз білкових продуктів ідентифікованих генів целюлозосинтаз льону дозволив визначити шість груп, що охоплюють целюлозосинтази різних класів, а саме *CesA1/10*, *CesA3*, *CesA4*, *CesA5/6/2/9*, *CesA7* та *CesA8*. Однак, паралогічні послідовності, що належать до класів *CesA1/10* та *CesA5/6/2/9*, характеризувалися вищим рівнем подібності в цих класах, ніж ортологічні послідовності інших видів. З цієї причини ідентифікація окремих генів *in silico* була складною, навіть якщо в аналізі приймали участь гени целюлозосинтаз з геномів кількох видів рослин. Гени, що входять до кластерів *CesA1/10* і *CesA5/6/2/9*, пов'язані з утворенням первинної клітинної стінки, тоді як гени, що відповідають за біосинтез целюлози вторинної клітинної стінки, утворюють окремі від них клади – *CesA4*, *CesA7* і *CesA8*. Аналіз 16 відібраних генів-кандидатів целюлозосинтаз льону показав, що

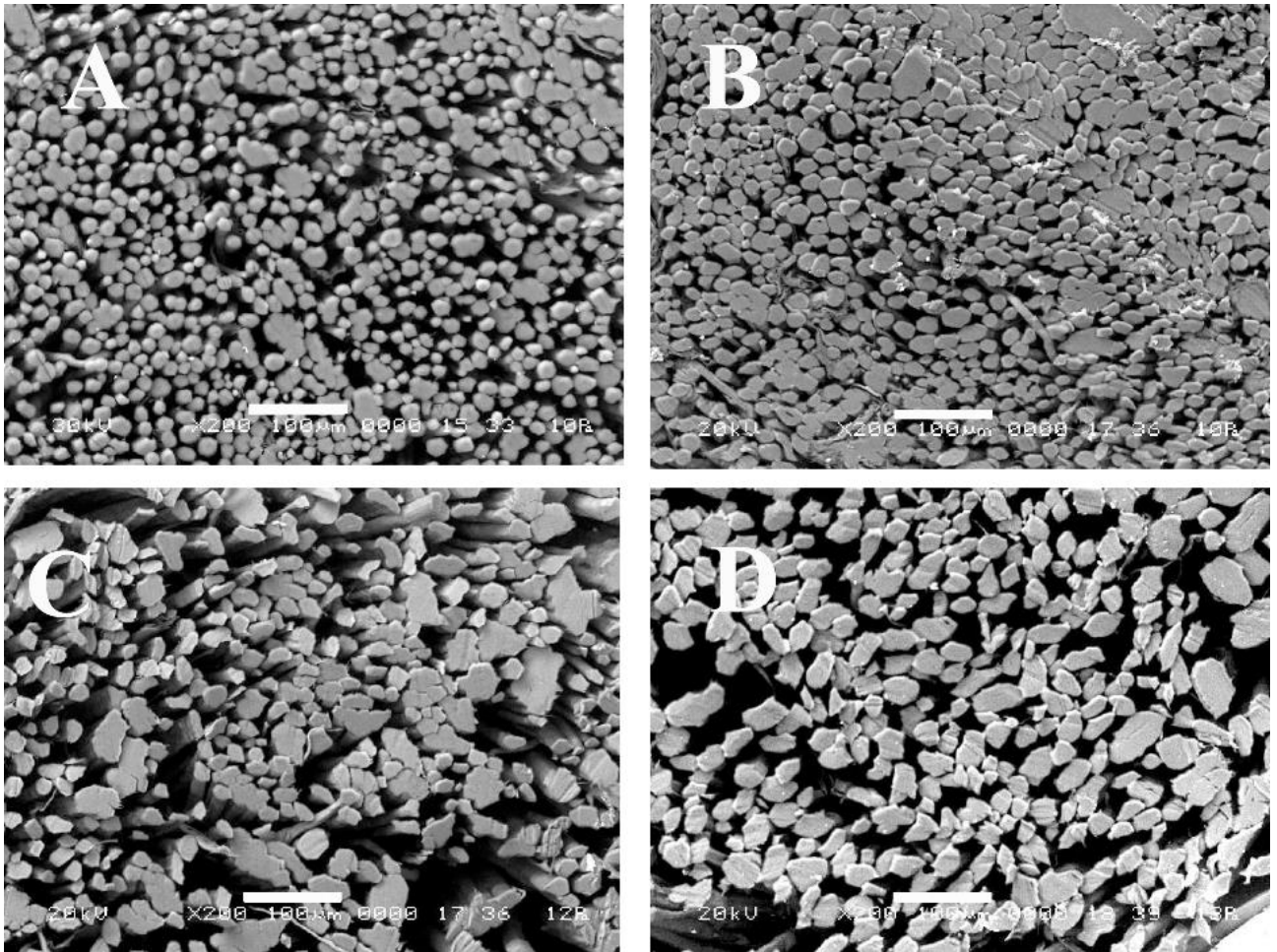
існує принаймні 12 різних варіантів генів целюлозосинтаз у геномі льону, що представляють усі шість класів. Виявлені нами варіанти генів целюлозосинтаз належать до всіх десяти відомих класів генів *CesA*.

### **4.3 Аналіз експресії генів целюлозосинтаз льону**

В якості предмету даного дослідження були використані чотири різновиди льону, які відрізняються за якістю волокна. Аналіз зображень, отриманих за допомогою світлової – Рис. 4.3 та електронної мікроскопії – Рис. 4.4, показує відмінності в структурі тканини первинної флоєми та в діаметрі елементарних волокон між цими зразками. У волокнистих льонових стеблах, – сорти *Blakit*, *Agiane*, більшість елементарних фібрил мають діаметр 10-14 мкм, у льону-стрибунці, *dehiscent* – 23–25 мкм, у стеблах великого насінневого льону, *Endress Olajlen*, домінують дві групи елементарних волокон: діаметром 17-20 мкм і діаметром 27-30 мкм, – Titok et al., 2010. Відмінності в розмірах клітин клітковини можуть свідчити про відмінності в їх біогенезі.

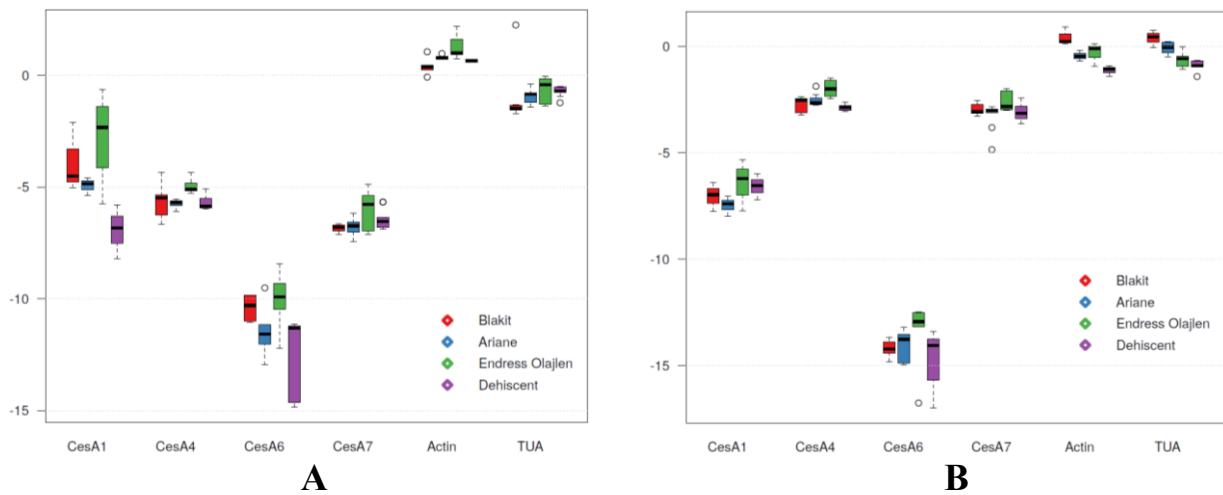


**Рис. 4.3.** Поперечні зрізи стебел льону: Blakit – **A**, Ariane – **B**, Endress Olajlen – **C**, льон-стрибунець, dehiscent – **D**; bf – луб'яне волокно (флоема), xl – ксилема. Збільшення –  $\times 300$ .



**Рис. 4.4.** Поперечні перерізи лляного волокна, отримані за допомогою низько вакуумної SEM з детектором зворотного розсіювання електронів, провідні покриття не наносились: Blakit – **A**, Ariane – **B**, Endress Olajlen – **C**, льон-стрибунець, dehiscent – **D**. Збільшення –  $\times 200$ , біла смужка – 100 мкм.

Ми кількісно оцінили відносну експресію генів *CesA* та генів актину та  $\alpha$ -тубуліну (TUA) у стеблах льону на двох різних стадіях росту, а саме «проростках» та «швидкому зростанні» за допомогою RT-qPCR. Рівні експресії генів *LusCesA1*, *LusCesA4*, *LusCesA7*, *LusCesA6*, актину та TUA були виявлені у всіх чотирьох різновидах на обох стадіях. Рівні експресії досліджуваних генів значно варіювали, – Рис. 4.5 А, В. Загалом, рівні експресії генів *CesA* були нижчими порівняно з еталонними генами, тоді як експресія генів актину та тубуліну була сильнішою – приблизно на рівні еталонних генів.



**Рис. 4.5.** Експресія *CesA* та цитоскелетних генів у саджанцях – **A**, та у стеблах – **B** 4 сортів льону. Низ і верх рамки – це нижній і верхній кuartилі, відповідно, смуга біля середини рамки є медіаною. Верхня риска =  $\min(\max(x), Q_3 + 1.5 * IQR)$ ; нижня риска =  $\max(\min(x), Q_1 - 1.5 * IQR)$ , де  $IQR = Q_3 - Q_1$ , довжина рамки.

Гени целюлозосинтаз *LusCesA4*, *LusCesA7* та *LusCesA1* були експресовані набагато сильніше, ніж ген *LusCesA6* у всіх досліджуваних зразках. Зокрема, експресія *LusCesA4* та *LusCesA7* у стеблах була на три порядки вищою – Рис. 4.5. А, а у проростках у 13,3–45,8 рази вищою – Рис. 4.5. А, ніж експресія *LusCesA6*. Експресія *LusCesA1* була на два порядки вищою у проростках та у 38,7–194,3 рази вищою у стеблах льону, порівняно з *LusCesA6*. Отже, *LusCesA6* виявляється найменш експресованим геном целюлозосинтаз в проростках і стеблах.

Загалом *LusCesA4* та *LusCesA7* були сильно експресовані на обох стадіях – Рис. 4.5. А, В. Ці результати узгоджуються з даними, отриманими раніше Mokshina et al., 2014. Вони спостерігали високу експресію цих генів у верхній і середній частинах пагону. Високий рівень експресії *CesA*-генів, що відповідають за відкладення вторинних клітинних стінок також був виявлений у базальній частині проростків льону Zhang and Deyholos 2016. Однак було досить несподівано виявити відносно невелику кількість транскриптів *LusCesA1* та *LusCesA6* у стеблах на стадії швидкого росту – Рис. 4.5. В, оскільки кілька авторів



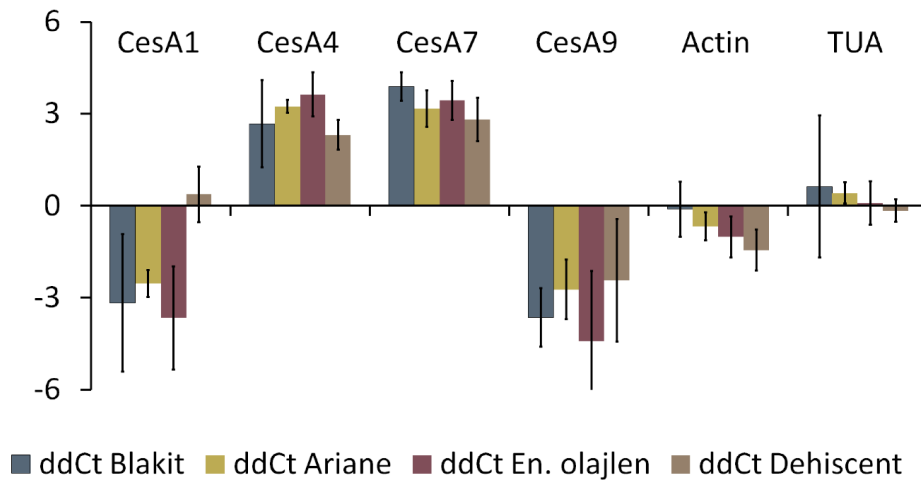
повідомляли про високу експресію цих генів на цій стадії – Mokshina et al., 2014; Chantreau et al., 2015.

Далі ми виявили, що напрямок зміни експресії генів *CesA* між досліджуваними стадіями залежить від функціонального підкласу продукованої целюлозосинтази. Гени, пов'язані із вторинною клітинною стінкою – *CesA4* та *CesA7*, були експресовані у 4,6-14,2 рази більше у стеблах, ніж у проростках, у всіх досліджуваних зразках – Рис. 4.5. Раніше було показано, що ці гени мають подібні профілі експресії у флоємних волокнах і в ксилемі – Mokshina et al., 2014. Отже, посилення їх експресії може відображати активацію біосинтезу целюлози в цих двох тканинах. Навпаки, рівні експресії целюлозосинтаз, пов'язаних з РСW *CesA1* та *CesA6*, за винятком *CesA1* у льоні-стрибунці, були у 4,3-10,7 разів нижчими у стеблах на стадії швидкого росту, ніж у гіпокотиліях проростка. Функціональний аналіз *CesA1* та *CesA6* вказує на їх вирішальне значення у формуванні стебла: відсутність їх функції призводить до глобальної дезорганізації як ксилеми, так і флоєми, – Chantreau et al., 2015. Наші результати дозволяють припустити, що функціональна важливість цих генів може знизитися на більш пізніх стадіях або, менш вірогідно, що вони можуть бути замінені паралогічними генами, експресію яких не були виміряно у нашій роботі.

Нарешті, ми порівняли експресію генів *CesA* на двох однакових стадіях розвитку між різними підвидами льону, – Рис. 4.6. Підвиди *elongatum* та *mediterraneum* продемонстрували подібні зміни в експресії досліджуваних генів між двома стадіями. Зміни в льоні-стрибунці, підвид *crepitans* були подібні до інших підвидів, за винятком гена *CesA1*, який мав менші зміни. Дійсно, експресія гена *CesA1* зменшилась у підвидах *elongatum* та *mediterraneum* у 5,2-10,7 рази протягом онтогенезу, а в *dehiscent* експресія цього гена не змінилася, – Рис. 4.6.

Підводячи підсумок, було показано, що гени *CesA* змінюють експресію між стадіями розвитку льону в залежності від групи – пов'язані з первинною і вторинною клітинною стінкою. Ми не виявили статистично значущих відмінностей між сортами, коли порівняли рівень експресії,  $\Delta$ St, див. методи, генів *CesA* окремо в проростках та стеблах, – Рис. 4.5. Однак, різниця в рівні

експресії гена *CesA1* була виявлена, коли порівнювали зміни експресії,  $\Delta\Delta Ct$ , між двома стадіями розвитку, – Рис. 4.6.



**Рис. 4.6.** Експресія генів *CesA*, актину та альфа-тубуліну у стеблі по відношенню з їх експресією у проростках (порівняння  $\Delta\Delta Ct$ ); риски показують стандартне відхилення середнього значення  $\Delta\Delta Ct$ .

Виміряна експресія пулів генів  $\alpha$ -тубуліну та актину була приблизно на одному рівні у всіх досліджуваних зразках. Це може спостерігатись через фактичну відсутність коливань або через недостатню роздільну здатність експерименту, оскільки експресію окремих генів не вимірювали. Перша можливість відповідає тому факту, що гени актину та тубуліну часто використовуються для нормалізації результатів експресії, – Gonzalex Verdejo et al., 2008. Однак, є і деякі факти, які підтверджують другу можливість. Раніше було показано, що експресія окремих генів актину нестабільна на різних стадіях розвитку рослин, – Guenin et al., 2009. Подібна картина була зареєстрована і для генів тубуліну: результати аналізу з geNorm довели, що експресія генів тубуліну нестабільна при абіотичному стресі та при розвитку у *Lolium perenne*, – Martin et al., 2008. Згідно з вищезазначеними спостереженнями, дослідження, де визначали гени кандидати на роль референтних для qPCR у льону показує нестабільність експресії генів актину та тубуліну, – Huis et al., 2010. Як ми вже передбачали, у геномі льону є 7 генів тубуліну та 10 актинів, – Bayer et al., 2014.

Тому в майбутньому було б цікаво вивчити експресію окремих генів ізотипів актину та тубуліну в паростках льону на різних стадіях розвитку.

**Висновки** Отже, у цьому дослідженні ми виміряли рівні транскриптів генів *CesA*, тубуліну та актину на двох стадіях росту у чотирьох сортів культивованого льону, що належать до різних підвидів. Було показано, що рівні транскриптів генів *CesA* залежать від 1) функціонального підкласу продукованого білка, 2) стадії розвитку рослини та 3) підвиду льону, що використовується як рослинний матеріал. Гени, відповідальні за синтез вторинної клітинної стінки – *LusCesA4* та *LusCesA7*, були досить сильно експресовані в проростках, і їх експресія зростала у стеблах на стадії швидкого росту. У той же час, експресія генів, пов'язаних з синтезом первинної клітинної стінки – *LusCesA1*, *LusCesA6*, зменшилась у стеблах на стадії швидкого росту. Отже, експресія *LusCesA4*, *LusCesA7* була у 200-250 разів вища, порівняно із середньою експресією *LusCesA1* і *LusCesA6*, у стеблах на стадії швидкого росту.

Суттєва різниця у експресії гена *LusCesA1* була виявлена між різними підвидами льону. У стеблах підвидів *elongatum* і *mediterraneum* експресія гена *LusCesA1* знизилася у 5,2-10,7 рази, і навпаки, експресія цього ж гена у підвиді *crepitans* не змінилася при порівнянні стадій швидкого росту і проростка.

Середня експресія генів цитоскелету була вищою, ніж середня експресія генів целюлозосинтаз, і була стабільною між онтогенетичними стадіями та сортами льону. Оцінка експресії окремих ізотипів тубуліну та актину буде предметом подальших досліджень.

#### **4.4 Гени суперсімейств тубулінів та целюлозосинтаз льону як перспективні мішені для покращення кількості та якості целюлозних волокон біотехнологічними методами**

*Стратегії редагування геному для розкриття ролей та взаємодії між членами надродини тубуліну та CesA, які можуть сприяти відкладенню клітинної стінки у льону. У той час як і сімейство тубуліну і CesA*

характеризуються чітко визначеною клітинною функцією, що полягає, відповідно, у підтримці росту і біосинтезі клітинної стінки шляхом утворення різних масивів кортикальних міротрубочок та целюлозосинтазних розеток, що синтезують мікрОВОлокна целюлози, роль різних членів кожної родини білків у наданні різних особливостей даним синтетичним комплексам з точки зору структури, динаміки, активності та взаємодії залишаються в значній мірі неясними. Це стосується кожної родини загалом, а також, що найголовніше, окремих білків всередині родин, тобто, що означає, що особливості взаємодії між MT та CSC, які, ймовірно, опосередковані додатковими білками та МАРами, в основному не визначені. Вирішити цю задачу – непросте завдання через наявність декількох генів та можливу функціональну надмірність їх закодованих продуктів, але нещодавно розроблені підходи до редагування геномів представляють надійний експериментальний інструмент, за допомогою якого можна встановити перший ступінь упорядкування в різних експериментальних напрямках, які можуть включати регуляторні послідовності, кодуючі послідовності, сайти пост трансляційних модифікацій (PTM) та потенційні сайти білок-білкових взаємодій. У цьому відношенні льон справді надає майже унікальну можливість завдяки своїй подвійній ідентичності, будучи модельною рослиною, геном якої в значній мірі розшифрований, та рослиною, що має важливе значення для багатьох промислових цілей. Тут ми не будемо вдаватися до деталей стратегій редагування геномів, пристосованих до конкретних генів. Це було б занадто спекулятивно і навіть нудно з огляду на відповідну кількість задіяних локусів, але ми скоріше окреслимо групу з декількох значущих стратегій, які можна було б застосувати.

Дотепер, нам відомий один єдиний приклад успішного редагування геному льону. Це стосується використання одноланцюгових олігодезоксинуклеотидів (ssODN) та системи CRISPR-Cas9 для розробки стійких до гліфосату рослин, отриманих шляхом репарації на основі гомології (HDR) двох дуже гомологічних локусів ENOLPYRUVYL SHIKIMATE-3-PHOSPHATE SYNTHASE (EPSPS), – Sauer et al. 2016. Регеновані рослини були родючими, і ново успадковані

відредаговані алелі EPSPS могли передаватися наступному поколінню за прогнозованим менделівським розподілом. Більше того, не виявлено нецільових мутацій. Частота точно відредагованих алелів EPSPS була досить низькою для обох локусів, але можуть бути розроблені шляхи її поліпшення, – Yang et al. 2013. Для підвищення ефективності HDR для точного редагування геному, нещодавно було досягнуто значного успіху шляхом впровадження гомологічної рекомбінації, опосередкованої використанням двох синтетичних однонаправляючих РНК (sgRNA), множинних дискретних точкових мутацій у гені рису що кодує ацетолактатсинтазу, – Sun et al. 2016. Повідомлений урожай рослин T<sub>0</sub>, які перенесли HDR, був вражаючим – 100%, але він, також, може залежати від поки невідомих особливостей цільових видів рослин, – Svitashv et al. 2015. Більш високі частоти відредагованих алелей також можна отримати альтернативним підходом редагування нуклеотидних залишків, коли мутовані форми Cas9 або Cas12 (спершу Cpf1) сплавляються з нуклеотидною дезаміназою, що призводить до перетворення нуклеотидів, що в кінцевому підсумку призводить до заміщення одиничних амінокислот, які можуть впливати на зв'язування білків, посттрансляційні модифікації та активність, – Lu and Zhu 2017; Wolter and Puchta 2018.

Усі ці результати є актуальними для того, щоб бути застосованими на нових різних сімействах генів льону, таких як тубуліни та Cesa, що характеризуються наявністю майже однакових копій дубльованих локусів (онологів) у тому контексті, що серед різних членів сімейства зберігається високий рівень гомології послідовностей. Мультиплексна здатність системи CRISPR-Cas9 редагувати високо гомологічні цільові послідовності була додатково продемонстрована у пшениці шляхом кнок-ауту трьох надлишкових генів MLO з метою отримання стійкої до борошнистої роси лінії, – Wang et al. 2014. Подібним чином, було отримано редагування геному трьох копій гена *FAD2* гексаплоїду *Camelina sativa* і відредаговані ознаки поширювались до покоління T<sub>3</sub>, – Morineau et al. 2016; Jiang et al. 2016. Поліплоїдія, здається, не є стримуючим фактором, як показано як для тетраплоїдної твердої пшениці, так і для

гексаплоїдної м'якої пшениці, обидві з яких було відредаговано методом CRISPR-Cas9, – Wang et al. 2014; Shan et al. 2013; Zhang et al. 2016. Навпаки, це може надати можливість вивчити ефекти дозування генів, порівнюючи одиночні та подвійні мутанти. З огляду на всі ці докази практичності, очікується, що націлювання методу одночасно на декілька генів може дати можливість краще зрозуміти ролі кожного гена в сімействі та в контексті більш складних мереж взаємодій, заснованих на конкретних взаємодіях, які є предметом інтересу. Крім того, ця сама стратегія може дати можливість виробити нові ознаки, що мають агрономічне значення.

Подальший революційний внесок у вивчення проблеми функціональної надмірності великих родин генів може бути внесений внаслідок ідентифікації абсолютно нового ефектора класу 2 – Cas13, який діє виключно на одноланцюгову РНК, відкриваючи, таким чином, величезну різноманітність застосувань у різних сферах регуляції експресії генів, нокауту РНК, зв'язування та візуалізація РНК та редагування та модифікація залишків РНК, – Abudayyeh et al. 2017; Wolter and Puchta 2018. Чудовою особливістю цієї системи, яка робить її особливо придатною для розв'язування нескінченної проблеми присутності множинних ізотипів тубулінів, є те, що вона також ефективна для некодуючих РНК, таких як ті, що можуть походити з інтронів, і за допомогою неї можна зробити нокдаун специфічної РНК у тканині, що представляє інтерес, якщо буде націлена на тканиноспецифічний промотор Wolter and Puchta 2018. Отже, подібний підхід є багатообіцяючим для розкриття як регуляторної ролі інтронів у експресії тубуліну, особливо тих, які присутні в 5' нетрансльовані області мРНК тубуліну, так і ключової ролі, які можуть мати деякі промотори в підтримці експресії тканиноспецифічних генів тубуліну, які, можливо, недостатньо замінні будь-якими іншими ізотипами тубуліну (див. також нижче).

У більш загальних рисах, вибір більш зручної стратегії редагування геному, яка може базуватися на мегануклеази, – Smith et al. 2006, ендонуклеазах цинкових пальців (ZFN), – Kim 1996, транскрипційно-активаторних ефекторних нуклеаз (TALEN) – Christian et al. 2010, або системи CRISPR-Cas, – Zhu et al. 2018, для

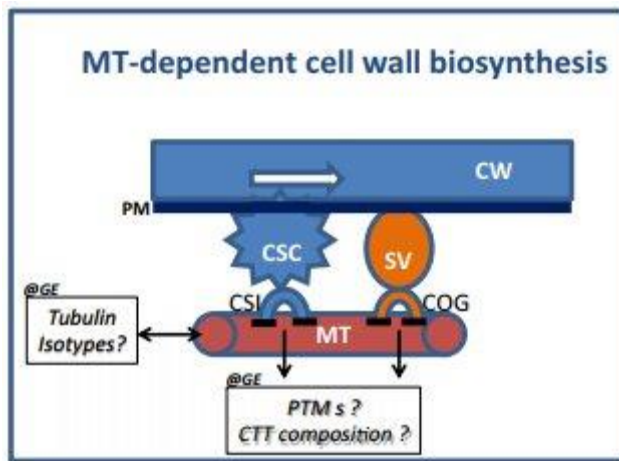
отримання мутантів різних представників сімейства генів, слід ретельно враховувати наступні різні фактори: наявність метильованих генів (для рослинних тубулінів відомо, що гени метильовані, – Lund et al. 1995); нецільові ефекти (частіше виникають у сімействі генів, якщо мутагенез, орієнтований на олігонуклеотиди, не був достатньо ретельно розроблений); наявність посусідству послідовності мотиву протоспейсеру (PAM) або потреба у більш високій специфічності цілі, як правило, відображається у довшій послідовності ДНК, – Malzahn et al. 2017. Системи CRISPR-Cas, які покладаються на синтетичні однонаправляючі РНК (sgRNA) для націлювання на конкретний геномний домен, як правило, є кращими в порівнянні з трьома іншими системами, що керуються білком, але різні інструментальні особливості можна знайти в численних типах і підтипах, що визначають клас 1 та клас 2, як вони фактично класифікуються. Наприклад: у той час як система CRISPR-Cas9, для розрізання кінців, в основному використовувалась для редагування геному, корисні альтернативи, засновані на розпізнаванні різних мотивів PAM та генерації дволанцюгових розривів, таких які забезпечує CRISPR-Cas12, – Schunder et al. 2013, колись відомий як Cpf1, може забезпечити кращу експериментальну стратегію.

В принципі, через тривалу відсутність знань про конкретну роль, яку мають окремі члени одного й того ж сімейства генів, зворотний генетичний підхід, заснований на принципі репарації ДНК з'єднанням негомологічних кінців (NHEJ), повинен бути дуже перспективним для підготовки першого раунду експериментів, виходячи з простої логіки: спочатку зруйнувати, щоб довідатись функцію, а потім редагувати, щоб змінити. Це може дати можливість виробництва нокаutowаних ліній рослин, що містять послідовно додані, множинні та руйнуючі мутації, щоб розкрити конкретну роль та можливе функціональне значення, таким чином забезпечуючи вперше відповідь на давні та досі невирішені питання про роль мультитубулінової гіпотези. В ідеалі, мутації у вигляді інсерцій або делецій, або заміщення, також, можуть бути введені в ті цільові ділянки ДНК, які кодують найбільш варіабельні

амінокислотні домени, такі як С-кінцевий хвіст тубуліну (СТТ) – Рис. 1.1, який, як вважається, є основною областю, яка бере участь у взаємодії з іншими білками, білками-моторами та МАРами, – Freedman et al. 2011. Нещодавно, такий підхід до редагування застосовували до дослідження людського тубуліну, виявляючи, що  $\beta$ -тубуліновий СТТ диференційовано координує динамічну поведінку мікротрубочок, – Parker et al., 2018. Таким чином, можуть бути виявлені можливі взаємодії з білками екзоцистного комплексу, які, слідуючи кортикальній мережі МТ, рухають секреторні пухирці локалізовано до вторинної клітинної стінки у розвиваючому трахеарному елементі, або ж можуть бути розкриті взаємодії з білками, асоційованими з целюлозосинтазними комплексами (CSC) первинної клітинної стінки, – Рис. 4.7.

Кандидатами на перший тип взаємодій можуть бути білки VETH1 і VETH2, які регулюють асоціації секреторних везикул з кортикальними МТ за допомогою консервативного олігомерного комплексу Гольджі (COG2), як це показано в культивованих клітинах ксилеми, – Vukašinić et al. 2017; Oda et al. 2015. З іншого боку, мембранно-зв'язана целюлаза KORRIGAN1 (KOR1), – Lei et al. 2012; Vain et al. 2014 та целюлозосинтаз взаємодіючі білки (CSI), є найбільш імовірними кандидатами на другий тип взаємодії, – Li et al. 2012. Ця складна взаємодія між синтезом клітинної стінки та масивами кортикальних мікротрубочок дійсно може залежати від специфічних ізотипів тубуліну, як показано для *Eucalyptus*, де, як повідомляється, надмірна експресія EgTub1 впливає на орієнтацію мікрофібрил целюлози в волокнах ксилеми, – Spokevicius et al. 2007. У *Populus* гени, що кодують тубуліни, які переважають в експресії у ксилемі, також мають регуляторно підвищений профіль експресії у деревині, що деформується (tension wood), у якій дуже активно відкладається целюлоза, – Oakley et al. 2007, і у хлопчатнику, *Gossypium*, де підгрупа генів тубуліну переважно експресується у хлопкових волокнах, що подовжуються, – Li et al. 2002; He et al. 2008.





**Рис. 4.7.** Спрощене представлення зв'язку між мікротрубочками (MT), целюлозосинтазами (CSC) та секреторними везикулами (SV) для формуванні клітинної стінки (CW), при чому медіатором або білком-посередником целюлозосинтази (CSI), або консервативним олігомерним комплексом Гольджі (COG). Показані три основні цілі для редагування генома (@GE), такі як ізотип тубуліну, С-кінцевий хвіст тубуліну (СТТ) та пост-трансляційні модифікації (PTM). Вміст різних ізотипів тубуліну у пулі генів тубуліну може бути змінений або порушенням кодуєчих послідовностей, або за рахунок механізмів надмірної експресії / репресії. Біла стрілка вказує напрямок відкладення мікрофібрил целюлози.

Відповідно до цих джерел та відповідно до транскрипційних досліджень експресії, проведених на льоні, з метою виявлення можливих субпопуляцій тубуліну, наявних у MT, що підтримують подовження клітин та потовщення клітинної стінки, NHEJ репарація ДНК може бути застосована до тих ізотипів, які, будучи експресовані диференційовано, можна запідозрити у впливі на подовження та диференціювання целюлозного волокна. Подібний підхід можна застосувати і для розшифрування їх ролі у дозріванні пилку. Результати, отримані за допомогою даної нокаут-стратегії, можуть бути додатково перевірені та підтверджені додатковим набором експериментів, де експресією генів тубуліну, кандидатів на певні ролі, можна маніпулювати, використовуючи деактивованій Cas9 (dCas9), злитий або з транскрипційним активатором, таким

як VP64, – Beerli et al. 1998, для надмірної експресії або з транскрипційним репресором, такий як SRDX, – Hiratsu et al. 2003 для регуляторного зниження експресії. Транскрипційні активації та репресії експресії на основі CRISPR-dCas9 були продемонстровані у тимчасово трансформованому (transiently transformed) тютюні, – Piatek et al. 2015 та у стабільно трансформованому *Arabidopsis*, – Lowder et al. 2015. Дані експерименти можуть також бути покращені за рахунок застосування системи перешкод експресії CRISPRi на основі вже описаної комбінації CRISPR-Cas13, – Wolter and Puchta 2018.

Нарешті, засновані на CRISPR-Cas9 зворотні генетичні дослідження можуть бути навмисно націлені на регуляторні елементи, такі як промотори або інтрони, які, як відомо, впливають на рівень і навіть на місце експресії специфічних генів тубуліну, – Gianì et al. 2009; Morello et al. 2011, як це було зроблено для кількох цис-регуляторних алелів трьох генів, які в помідорах регулюють розмір плодів, архітектуру суцвіть та характер росту рослин, – Rodriguez-Leal et al. 2017. Наведений підхід врешті-решт дасть можливість виявити, чи переважна експресія специфічних ізотипів тубуліну в органах і тканинах є результатом використання кращого промотору або функціональної потреби в конкретній ізоформі тубуліну. А це останнє, в свою чергу, проллє більше світла на можливий внесок інтронів та, можливо, споріднених мітронів – мікроРНК, що генеруються інтронами, у контроль експресії генів тубуліну. Питання регуляції є особливо ітнригуючим, з огляду на консервативну геномну структуру як  $\alpha$ -, так і  $\beta$ -тубулінових генів, яка зберігається, також, і у льоні на рівні екзон-інтронної організації з невеликою мінливістю довжин інтронів серед онологів.

Другий раунд стратегій редагування нуклеотидів, можливо, запроваджених як на рівнях ДНК, так і на РНК, як описано вище, врешті-решт скористається даними, отриманими за допомогою нокаут ліній та ліній з інсерціями-делеціями, і буде в основному присвячений модифікації сайтів, для яких було визначено, що вони є цільовими сайтами для одного з декількох типів пост-трансляційних модифікацій (PTM) – фосфорилування, нітрозилування, глутамілація, тирозинування, – Рис. 1.1, функціональна роль яких залишається в основному

невизначеною, хоча деякі підказки щодо взаємного розпізнавання, яке забезпечується специфічним фосфорилюванням між мікротрубочками та цлюлозосинтазним комплексом, до чого вже було повернуто широку увагу, – Chen et al. 2016.

## УЗАГАЛЬНЕННЯ

У дисертаційній роботі представлені результати досліджень родин генів цитоскелетних білків та генів целюлозосинтаз льону як перспективних мішеней для покращення кількості та якості целюлозних волокон біотехнологічними методами. У рамках даного дослідження було проведено біоінформатичний пошук генів сімейств тубулінів  $\alpha$ ,  $\beta$  та  $\gamma$ , актину, а також – целюлозосинтаз, проаналізовано особливості їх екзон-інтронної структури, проведено порівняльний філогенетичний аналіз і віднесено гени даних сімейств до різних ізотипів та функціональних класів. Надалі, було визначено рівні експресії окремих генів *CesA*,  $\alpha$ -тубуліну та актину у проростках та у стеблі на стадії швидкого росту у чотирьох сортів льону різних підвидів і зроблено висновки про функціональну роль різних ізотипів, кодованих цими генами білків. На основі отриманих результатів запропоновано стратегічні підходи для застосування методів редагування генів цитоскелетних білків, перш за все різних генів тубуліну, з метою посилення їх функціональної ролі для покращення целюлозних волокон льону.

На першому етапі дисертаційного дослідження здійснено біоінформатичний пошук генів цитоскелетних білків, – тубулінів  $\alpha$ ,  $\beta$ ,  $\gamma$  та актину, у геномі льону з використанням загальнодоступної бази даних Phytozome. Наші дані показують, що гени  $\alpha$ -,  $\beta$ -тубуліну та актину широко представлені у геномі льону. Було виявлено 6 генів  $\alpha$ -тубуліну, 14 генів  $\beta$ -тубуліну, 2 гени  $\gamma$ -тубуліну та 15 генів актину.

На наступному етапі було проведено аналіз структури та філогенії знайдених генів у порівнянні зі спорідненими видами і дані гени були віднесені до різних ізотипів і класів. Шість генів  $\alpha$ -тубуліну згруповані у три паралогічні пари. Серед 14 генів  $\beta$ -тубуліну можна виділити сім різних ізотипів, шість з яких кодуються двома паралогічними генами.  $\gamma$ -тубулін представлений паралогічною парою генів, один з яких, ймовірно, втратив свою функцію. П'ятнадцять генів актину представляють сім паралогічних пар – сім ізотипів актину та послідовну дуплікацію одного з генів одного із ізотипів. Віднесення генів до певних

функціональних класів важливе для проектування експериментів по аналізу експресії, оскільки гени різних класів по-різному експресуються у різних тканинах та на різних етапах розвитку рослин. І для того, щоб дослідити експресію генів різних класів необхідно підібрати відповідні праймери.

В свою чергу, аналіз екзон-інтронної структури показав поліморфізм довжини інтронів у генах  $\beta$ -тубуліну та варіацію кількості інтронів у генах  $\alpha$ -тубуліну: три або чотири інтрони знаходяться у двох (клас I) і чотирьох (клас II) генах відповідно. Позиції інтронів у генах тубуліну льону є такими ж, як і у інших видах рослин. У генах актину льону спостерігається як мінливість кількості інтронів, яка може складати 2 або 3, так і поліморфізм довжини інтронів. Проведений інтронний аналіз важливий для філогенетичного та еволюційного аналізу генів тубуліну та актину. Інтронний поліморфізм генів тубуліну *Linum* свідчить про те, що пов'язані з інтронами регуляторні елементи можуть впливати на експресію цих генів. Варіації довжини інтрону 3  $\alpha$ -тубуліну, інтронів 1 і 2  $\beta$ -тубулінів та інтрону 2 актину можуть зробити їх життєздатними мішенями для дослідження поліморфізмів довжини інтронів (intron length polymorphisms, ILPs) різних видів *Linum* методом молекулярного маркера ILP.

Надалі, було побудовано моделі за гомологією рослинних тубулінів  $\alpha$ ,  $\beta$  та  $\gamma$  та на їх основі проаналізовано амінокислотні залишки тубулінів  $\alpha$  та  $\beta$ , які утворюють сайти зв'язування колхіцину, таксолу та вінбластину у порівнянні з аналогічними тваринними сайтами. Було показано, що сайти зв'язування колхіцину, вінбластину та таксолу практично повністю консервативні у всіх рослинних  $\alpha$ - та  $\beta$ -тубулінах. Заміни Val353Cys в  $\alpha$ -тубуліні та Ala316Ser у  $\beta$ -тубуліні у сайті зв'язування колхіцину у рослин можуть призвести до критичних відмінностей у природі цього сайту зв'язування у рослин порівняно з тваринами. У той же час, заміна у рослинному сайті зв'язування вінбластину Tyr224 у  $\beta$ -тубуліні подібним ароматичним а.з. Phe може бути набагато менш значущою. У свою чергу, позиції амінокислот, які беруть участь у зв'язуванні таксолу в рослинних та тваринних тубулінах, виявились надзвичайно консервативними. Окрім того, побудовано модель рослинної мікротрубочки з 13 гетеродимерів  $\alpha$ -

$\beta$ -тубулінів. Побудовані тривимірні моделі дали нам можливість оцінити зв'язування фармакофорних сполук із рослинними тубулінами та оцінити їх можливе використання.

На даному етапі досліджень для менеджменту та повторного використання біологічних даних, отриманих в результаті даної роботи, створено базу даних по цитоскелетним білкам CSMoDB. Обробка та збереження структурної-біологічної інформації у даній БД полегшує її первинну обробку та повторне використання. Це, в свою чергу, пришвидшує вирішення наукових та прикладних завдань у галузі дослідження цитоскелету, зокрема, пошук і відбір нових сполук з антимікротрубочковою активністю для подальшого їх застосування як протипухлинних, фунгіцидних, гербіцидних, антипротозойних та протигельмінтних засобів.

Показано, що пари генів онологів  $\alpha$ -тубуліну льону *Lus10039169* та *Lus10013765* (*LusTua3a*, *LusTua3b*) і *Lus10035422* та *Lus10031032* (*LusTua2a*; *LusTua2b*) кодують ідентичні білки і, як продемонстровано за допомогою qRT-PCR, для різних сортів льону, онологи *LusTua2* переважно експресуються в апексі та флоемі, тоді як пара *LusTua3* переважно експресується у листі. Спостерігається певний рівень дисбалансу в експресії  $\alpha$ - та  $\beta$ -тубуліну, що призводить до загальної більшої кількості ізотипів *LusTua2* та *LusTua3* у апексі, листках та флоемі. Це свідчить на користь припущення про наявність регуляторних механізмів, які можуть по-різному впливати на експресію генів  $\alpha$ - та  $\beta$ -тубуліну.

В свою чергу для  $\beta$ -тубулінів показано, що 14 генів  $\beta$ -тубуліну льону експресуються, хоча і у різній мірі, у різних тканинах або на різних стадіях росту. Аналіз експресії генів у стеблі, гіпокотилі та квітці дав можливість припустити, що пули різних ізотипів тубуліну можуть утворювати різні МТ з можливими спеціалізованими функціями. Встановлено, що відносний рівень транскриптів  $\beta$ -тубуліну модулюється під час подовження стовбура та гіпокотилів. Однак, обидва онологи кожної пари виявляють подібний рівень експресії у всіх аналізованих тканинах, за винятком пари генів ізотипу *LusTub4*, де є помітні

відмінності між двома членами пари. У верхівці стебла, де більшість клітин активно ділиться, був виявлений високий рівень експресії для пар генів ізотипів *LusTub3* та *LusTub7*. І навпаки, рівень експресії пар генів ізотипів *LusTub2* і *LusTub5* був дуже низьким. Подібна закономірність диференціальної експресії спостерігалась також у корневих кінчиках, де гени ізотипу *LusTub7* мали найвищий рівень експресії, тоді як транскрипти ізотипів *LusTub2* були ледь помітними. У зв'язку з цим можна зробити висновок, що мікротрубочки, що утворюють пре-профазне веретено або мітотичне веретено, можуть формуватися переважно ізотипами білка *TUB7*.

Використовуючи дані по апексам в якості контролю, також, спостерігалось підвищення регуляції експресії більшості генів у верхній частині стебла, яке характеризується значним подовженням клітин клітковини та інших типів клітин. Цей процес пов'язаний з поперечною орієнтацією кортикальних МТ, які допомагають правильному розташуванню комплексів Cesa в мембрані. Далі вниз по стеблу, експресія деяких генів  $\beta$ -тубуліну, *LusTub3a* та *b*, *LusTub5a* та *b* та *LusTub6*, була чітко регуляторно зниженою у наступних секціях, що характеризуються переорієнтацією кортикальних МТ, одночасно з зупинкою подовження та поступовим потовщенням клітинної стінки у клітинах флоєми. І навпаки, регуляторне підвищення експресії *LusTub1*, *2* і *4* у верхній частині стебла зберігалось по всій довжині стебла, навіть у самих базальних його областях, де потовщення клітинної стінки було вже практично завершено. Паралоги *LusTub7* продемонстрували протилежну картину експресії – вони експресувались переважно у апексі, а у всіх інших частинах стебла експресія була регуляційно зниженою.

У квіток льону було помічено, що повне дозрівання пилку під час цвітіння викликає зміни рівня експресії різних ізотипів  $\beta$ -тубуліну при сталій загальній кількості транскриптів, зі значним регуляторним підвищенням експресії пар генів *LusTub3* та *4*, разом з паралельним зниженням кількості мРНК *LusTub2* та *5*. *LusTub3* належать до  $\beta$ -тубулінів класу III, члени якого переважно експресуються в пилку багатьох інших видів рослин. Це може свідчити про те,

що специфічний для пилку пул транскриптів тубуліну збирається у спеціалізовані МТ, більш придатні для тонкої доставки везикул, що несуть білки і ліпіди до мембрани кінчика пилкової трубки, коли ріст поляризованої пилкової трубки повинен безперервно підтримуватися.

У роботі, також, проведено аналіз родини генів целюлозосинтаз льону і ідентифіковано 16 генів целюлозосинтаз і 16 генів білків, подібних до целюлозосинтаз. Філогенетичний аналіз білкових продуктів ідентифікованих генів дозволив визначити шість груп, що відповідають целюлозосинтазам різних класів – *CesA1/10*, *CesA3*, *CesA4*, *CesA5/6/2/9*, *CesA7* та *CesA8*.

Паралогічні послідовності генів льону, що належать до класів *CesA1/10* та *CesA5/6/2/9*, характеризувалися високим рівнем подібності. З цієї причини ідентифікація окремих генів *in silico* була складною, навіть якщо в аналізі приймали участь гени целюлозосинтаз з геномів кількох видів рослин. *CesA1/10* і *CesA5/6/2/9*, пов'язані з утворенням первинної клітинної стінки, тоді як гени, що відповідають за біосинтез целюлози вторинної клітинної стінки, утворюють окремі від них клади – *CesA4*, *CesA7* і *CesA8*. Аналіз 16 відібраних генів-кандидатів целюлозосинтаз льону показав, що існує принаймні 12 різних варіантів генів целюлозосинтаз у геномі льону, що представляють усі шість класів. Виявлені нами варіанти генів целюлозосинтаз належать до всіх десяти відомих класів генів *CesA*.

На наступному етапі досліджень було кількісно визначено відносну експресію генів *CesA* та генів актину та  $\alpha$ -тубуліну на двох різних стадіях росту, а саме «проростання» та «швидкого росту», використовуючи RT-qPCR у чотирьох сортів культивованого льону, що належать до різних підвидів. Було показано, що рівні транскриптів генів *CesA* залежать від: 1) функціонального підкласу продукованого білка, 2) стадії розвитку рослини та 3) підвиду льону, що використовується як рослинний матеріал. Гени, відповідальні за синтез вторинної клітинної стінки – *LusCesA4* та *LusCesA7*, були досить сильно експресовані в проростках, і їх експресія зростала у стеблах на стадії швидкого росту. У той же час, експресія генів, пов'язаних з синтезом первинної клітинної



стілки – *LusCesA1*, *LusCesA6*, зменшилась у стеблах на стадії швидкого росту. Отже, експресія *LusCesA4*, *LusCesA7* була у 200-250 разів вища, порівняно із середньою експресією *LusCesA1* і *LusCesA6*, у стеблах на стадії швидкого росту.

Суттєва різниця у експресії гена *LusCesA1* була виявлена між різними підвидами льону. У стеблах підвидів *elongatum* і *mediterraneum* експресія гена *LusCesA1* знизилася у 5,2-10,7 рази, і навпаки, експресія цього ж гена у підвиді *crepitans* не змінилася, коли порівнювали стадію швидкого росту і проростка. Середня експресія генів цитоскелету була вищою, ніж середня експресія генів целюлозосинтаз, і була стабільною між онтогенетичними стадіями та сортами льону.

На заключному етапі дисертаційного дослідження було запропоновано стратегічні підходи для застосування методів редагування генів цитоскелетних білків, перш за все різних генів тубуліну, з метою посилення їх функціональної ролі для покращення целюлозних волокон льону.

Отже, у дисертаційній роботі представлені результати біоінформатичного пошуку та характеристики генів родин цитоскелетних білків ( $\alpha$ -,  $\beta$ -,  $\gamma$ -тубулінів і актину) та целюлозосинтаз у геномі льону. На основі структурного аналізу та філогенетичної подібності різні гени даних сімейств білків віднесено до відповідних функціональних ізотипів. Показано, що поліморфізм довжини інтронів генів тубуліну *Linum* свідчить про те, що пов'язані з інтронами регуляторні елементи можуть впливати на експресію цих генів, а варіації довжини інтрону 3  $\alpha$ -тубуліну, інтронів 1 і 2  $\beta$ -тубулінів та інтрону 2 актину позиціонують їх як важливі мішені для дослідження поліморфізму довжини інтронів різних видів *Linum* (метод ILP). Надалі, було виміряно рівні експресії генів окремих генів *CesA*,  $\alpha$ -тубуліну та актину у проростках та у стеблі на стадії швидкого росту у чотирьох сортів льону різних підвидів і зроблено висновки про функціональну роль різних ізотипів, кодованих цими генами білків.

Загалом, отримані у роботі результати досліджень дали можливість запропонувати стратегії застосування методів редагування генів цитоскелетних

білків, перш за все різних генів тубуліну, з метою посилення їх функціональної ролі для покращення целюлозних волокон льону.

## ВИСНОВКИ

В роботі вперше визначено та охарактеризовано гени цитоскелетних білків – тубулінів  $\alpha$ ,  $\beta$ ,  $\gamma$  і актину,- та гени целюлозосинтаз у геномі льону, білкові продукти яких були віднесені до різних функціональних ізотипів на основі їх структури та філогенетичної подібності. Було проаналізовано структурні особливості генів цих цитоскелетних білків і проведено аналіз рівнів експресії генів тубулінів, актину та целюлозосинтаз на різних стадіях росту льону і зроблено висновок про функціональну роль їх різних генів, що дозволило узагальнити можливості покращення кількості та якості целюлозних волокон льону за допомогою біотехнологічних методів. В цілому, отримані результати дозволяють сформулювати наступні висновки:

1. Виявлено 6-ть генів  $\alpha$ -тубуліну, 13-ть генів  $\beta$ -тубуліну, 2-а гени  $\gamma$ -тубуліну та 15-ть генів актину у геномі льону. Шість генів  $\alpha$ -тубуліну згруповані у три паралогічні пари. Серед 13-ти генів  $\beta$ -тубуліну можна виділити сім різних ізотипів, шість з яких кодуються двома паралогічними генами, а  $\gamma$ -тубулін представлений паралогічною парою генів, один з яких, ймовірно, втратив свою функцію. П'ятнадцять генів актину представляють сім паралогічних пар – сім ізотипів актину та послідовну дуплікацію одного з генів одного з ізотипів.

2. За результатами аналізу екзон-інтронної структури генів основних білків цитоскелету виявлено поліморфізм довжини інтронів в генах  $\beta$ -тубуліну та варіацію кількості інтронів у генах  $\alpha$ -тубуліну: три або чотири інтрони знаходяться у двох і чотирьох генах відповідно. Позиції інтронів у генах тубуліну льону є такими ж, як і у інших видів рослин.

3. Вперше встановлено, що у генів актину льону спостерігається як мінливість кількості інтронів, яка може складати 2 або 3, так і поліморфізм довжини інтронів.

4. Встановлено, що інтронний аналіз важливий для філогенетичного аналізу генів тубуліну та актину. Інтронний поліморфізм генів тубуліну *Linum* свідчить про те, що пов'язані з інтронами регуляторні елементи можуть впливати на експресію цих генів. Варіації довжини інтронів 3  $\alpha$ -тубуліну, інтронів 1 і 2  $\beta$ -

тубуліну та інтрону 2 актину можуть зробити їх життєздатними мішенями для дослідження поліморфізму довжини інтронів (ILP) у різних видів *Linum*.

5. Проведено аналіз родини генів целюлозосинтаз льону і ідентифіковано 16-ть генів целюлозосинтаз і 16-ть генів білків, подібних до целюлозосинтаз. Філогенетичний аналіз білкових продуктів ідентифікованих генів целюлозосинтаз льону дозволив визначити шість груп, що охоплюють целюлозосинтази різних класів: *CesA1/10*, *CesA3*, *CesA4*, *CesA5/6/2/9*, *CesA7* та *CesA8*.

6. Встановлено, що гени целюлозосинтетаз льону, які належать до класів *CesA1/10* та *CesA5/6/2/9*, пов'язані з утворенням первинної клітинної стінки, тоді як гени, що відповідають за біосинтез целюлози вторинної клітинної стінки, утворюють окремі від них клади – *CesA4*, *CesA7* і *CesA8*. Аналіз 16-ти відібраних генів-кандидатів *CesA* льону показав, що існує принаймні 12-ть різних варіантів генів *CesA* у геномі льону, котрі представляють усі шість класів. Виявлені варіанти генів целюлозосинтаз належать до всіх десяти відомих класів генів *CesA*.

7. Виміряно рівні експресії генів *CesA*, тубуліну та актину на двох стадіях росту у чотирьох сортів культивованого льону, що належать до різних підвидів. Було показано, що рівні експресії генів *CesA* залежать від 1) функціонального підкласу продукованого білка, 2) стадії розвитку рослини та 3) підвиду льону, що використовується як рослинний матеріал. Середні рівні експресії генів білків цитоскелету були вищими, ніж середні значення експресії генів целюлозосинтаз, а також були стабільними на різних стадіях онтогенезу та у різних сортів льону.

8. Встановлено, що гени, відповідальні за синтез вторинної клітинної стінки – *LusCesA4* та *LusCesA7*, – були досить сильно експресовані в проростках, і рівні їх експресії зростали у стеблах на стадії швидкого росту. У той же час рівні експресії генів, пов'язаних з синтезом первинної клітинної стінки – *LusCesA1*, *LusCesA6*, - знижувались у стеблах на стадії швидкого росту.

9. Виявлена суттєва різниця у рівнях експресії гена *LusCesA1* між різними підвидами льону. У стеблах підвидів *elongatum* і *mediterraneum* рівні експресії

гена *LusCesA1* знижувалась у 5,2-10,7 рази, і навпаки, експресія цього ж гена у підвиду *crepitans* не змінювалась, якщо порівнювали стадію швидкого росту і проростка.

10. На основі отриманих результатів запропоновано стратегічні підходи для застосування методів редагування генів цитоскелетних білків, перш за все різних генів тубуліну, з метою посилення їх функціональної ролі для покращення кількості та якості целюлозних волокон льону.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Dissanayake NPJ, Summerscales J, Grove SM, Singh MM. (2009). Life cycle assessment of flax fiber for the reinforcement of composites. *J Biobased Mater Bioenergy*;3:245-48.
2. Kvavadze E, Bar-Yosef O, Belfer-Cohen A, Boaretto E, Jakeli N, Matskevich Z, Meshveliani T. (2009). 30,000-Year-Old Wild Flax Fibers. *Science*;325(5946):1359.
3. Wang Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, Yang L, Hawkins S, Neutelings G, Datla R, Lambert G, Galbraith DW, Grassa CJ, Geraldес A, Cronk QC, Cullis C, Dash PK, Kumar PA, Cloutier S, Sharpe AG, Wong GK, Wang J, Deyholos MK. (2012) The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads. *Plant J*;72(3):461-73.
4. Hotte NSC, Deyholos M (2008) A flax fiber proteome: identification of proteins enriched in bast fibers. *BMC Plant Biology*;8:52
5. Zhang N, Deyholos MK. (2016) RNASeq Analysis of the Shoot Apex of Flax (*Linum usitatissimum*) to Identify Phloem Fiber Specification Genes. *Frontiers in Plant Science*;7:950.
6. Bringmann M, Landrein B, Schudoma C, Hamant O, Hauser M-T, Persson S. (2012). Cracking the elusive alignment hypothesis: the microtubule-cellulose synthase nexus unraveled. *Trends Plant Sci*;17(11):666-74.
7. Galinousky DV, Anisimova NV, Raiski AP, Leontiev VN, Titok VV, Khotyleva LV. (2014) Cellulose synthase genes that control the fiber formation of flax (*Linum usitatissimum* L.) // *Russ J Genet*;50(1):20–7.
8. Mokshina N, Gorshkova T, Deyholos MK. (2014) Chitinase-Like (*CTL*) and Cellulose Synthase (*CESA*) Gene Expression in Gelatinous-Type Cellulosic Walls of Flax (*Linum usitatissimum* L.) Bast Fibers. *PLoS ONE*;9(6):e97949.
9. Burton RA, Gibeauta DM, Bacicb A, Findlayc K, Robertsc K, Hamiltond A, Baulcombed DS, Finchera S. (2000) Virus induced silencing of a plant cellulose synthase gene. *Plant Cell*;12(5):691-705.

10. Kumar M, Thammannagowda S, Bulone V, Chiang V, Han KH, Joshi CP, Mansfield SD, Mellerowicz E, Sundberg B, Teeri T, Ellis BE. (2009) An update on the nomenclature for the cellulose synthase genes in *Populus*. *Trends Plant Sci*;14(5):248-54.
11. Baskin TI, Beemster GT, Judy-March JE, Marda F. (2004) Disorganization of cortical microtubules stimulates tangential expansion and reduces the uniformity of cellulose microfibril alignment among cells in the root of *Arabidopsis*. *Plant Physiol*;135(4):2279-90.
12. Baskin TI. (2005) Anisotropic expansion of the plant cell wall. *Annu Rev Cell Dev Biol*;21:203-22.
13. Boutté Y, Vernhettes S, Satiat-Jeunemaitre B. (2007). Involvement of the cytoskeleton in the secretory pathway and plasma membrane organisation of higher plant cells. *Cell Biol Int*;31(7):649-54.
14. Pydiura NA, Bayer GYA, Galinousky DV, Yemets AI, Pirko YaV, Padvitski TA, Anisimova NV, Khotyleva LV, Kilchevsky AV, Blume YaB. (2015) Bioinformatic search for cellulose synthase genes in flax (*Linum usitatissimum*) and their phylogenetic analysis. *Tsitol Genet*;49(5):279-87.
15. van Zeist W, Bakker-Heeres JAH (1975) Evidence for linseed cultivation before 6000 BC. *J Archaeol Sci* 2:215–219. doi: 10.1016/0305-4403(75)90059-X
16. Singh KK, Mridula D, Rehal J, Barnwal P (2011) Flaxseed: a potential source of food, feed and fiber. *Crit Rev Food Sci Nutr* 51:210–222. doi: 10.1080/10408390903537241
17. Zhu J, Zhu H, Njuguna J, Abhyankar H (2013) Recent development of flax fibres and their reinforced composites based on different polymeric matrices. *Materials (Basel)* 6:5171–5198. doi: 10.3390/ma6115171
18. Oomah BD (2001) Flaxseed as a functional food source. *J Sci Food Agric* 81:889–894. doi: 10.1002/jsfa.898
19. Titok V, Leontiev V, Shostak L, Khotyleva L (2006) Thermogravimetric analysis of the flax bast fibre bundle. *J Nat Fibres* 3:35–41. doi: 10.1300/J395v03n01\_04

20. Bringmann M, Landrein B, Schudoma C, Hamant O (2012b) Cracking the elusive alignment hypothesis: the microtubule–cellulose synthase nexus unraveled. *Trends Plant Sci* 17:666–674. doi: 10.1016/j.tplants.2012.06.003
21. Goudenhoofdt C, Bourmaud A, Baley C (2017) Varietal selection of flax over time: evolution of plant architecture related to influence on the mechanical properties of fibres. *Industr Crops Products* 97:56–64. doi: 10.1016/j.indcrop.2016.11.062
22. Gorshkova T, Sal'nikova V, Chemikosova S, Ageeva M, Pavlencheva N, van Dam J (2003) The snap point: a transition point in *Linum usitatissimum* bast fibre development. *Industr Crops Products* 18: 213–221. doi: 10.1016/S0926-6690(03)00043-8
23. Roach MJ, Deyholos MK (2007) Microarray analysis of flax (*Linum usitatissimum* L.) stems identifies transcripts enriched in fibrebearing phloem tissues. *Mol Gen Genomics* 278:149–165. doi: 10.1007/s00438-007-0241-1
24. Mellerowicz EJ, Gorshkova TA (2012) Tensional stress generation in gelatinous fibres: a review and possible mechanism based on cell-wall structure and composition. *J Exp Bot* 63:551–565. doi: 10.1093/jxb/err339
25. Gorshkova T, Chernova T, Mokshina N, Ageeva M, Mikshina P (2018) Plant 'muscles': fibres with a tertiary cell wall. *New Phytol* 218:66–72. doi: 10.1111/nph.14997
26. Gorshkova TA, Mikshina PV, Gurjanov OP, Chemikosova SB (2010) Formation of plant cell wall supramolecular structure. *Biochemistry (Mosc)* 75:159–172
27. Wang Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, Yang L, Hawkins S, Neutelings G, Datla R, Lambert G, Galbraith DW, Grassa CJ, Geraldine A, Cronk QC, Cullis C, Dash PK, Kumar PA, Cloutier S, Sharpe AG, Wong GK-S, Wang J, Deyholos MK (2012) The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads. *Plant J* 72:461–473. doi: 10.1111/j.1365-313X.2012.05093.x
28. Gorshkov O, Mokshina N, Gorshkov V, Chemikosova S, Gogolev Y, Gorshkova T (2017) Transcriptome portrait of cellulose-enriched flax fibres at advanced



- stage of specialization. *Plant Mol Biol* 93:431–449. doi: 10.1007/s11103-016-0571-7
29. Venglat P, Xiang D, Qiu S, Stone SL, Tibiche C, Cram D, Alting-Mees M, Nowak J, Cloutier S, Deyholos M, Bekkaoui F, Sharpe A, Wang E, Rowland G, Selvaraj G, Datla R (2011) Gene expression analysis of flax seed development. *BMC Plant Biol* 11:74. doi: 10.1186/1471-2229-11-74
  30. Dash PK, Rai R, Mahato AK, Gaikwad K, Singh NK (2017) Transcriptome landscape at different developmental stages of a drought tolerant cultivar of flax (*Linum usitatissimum*). *Front Chem* 5:82. doi: 10.3389/fchem.2017.00082
  31. Ludvíková M, Griga M (2015) Transgenic flax/linseed (*Linum usitatissimum* L.) – expectations and reality. *Czech J Genet Plant Breed* 51:123–141. doi: 10.17221/104/2015-CJGPB
  32. McHughen A (1989) Agrobacterium mediated transfer of chlorsulfuron resistance to commercial flax cultivars. *Plant Cell Rep* 8:445–449. doi: 10.1007/BF00269045
  33. McHughen A, Rowland GG, Holm FA, Bhatti RS, Kenaschuk EO (1997) CDC Triffid transgenic flax. *Can J Plant Sci* 77:641–643. doi: 10.4141/P96-188
  34. Ryan CD, Smyth SJ (2012) Economic implications of low-level presence in a zero-tolerance European import market: the case of Canadian Triffid flax. *AgBioForum* 15:21–30
  35. Roach MJ, Mokshina NY, Badhan A, Snegireva AV, Hobson N, Deyholos MK, Gorshkova TA (2011) Development of cellulosic secondary walls in flax fibres requires beta-galactosidase. *Plant Physiol* 156:1351–1363. doi: 10.1104/pp.111.172676
  36. Hobson N, Deyholos MK (2013) Genomic and expression analysis of the flax (*Linum usitatissimum*) family of glycosyl hydrolase 35 genes. *BMC Genomics* 14:344. doi: 10.1186/1471-2164-14-344
  37. Barvkar VT, Pardeshi VC, Kale SM, Kadoo NY, Gupta VS (2012) Phylogenomic analysis of UDP glycosyltransferase 1 multigene family in *Linum usitatissimum*

- identified genes with varied expression patterns. *BMC Genomics* 13:175. doi: 10.1186/1471-2164-13-175
38. Pinzón-Latorre D, Deyholos MK (2013) Characterization and transcript profiling of the pectin methylesterase (PME) and pectin methylesterase inhibitor (PMEI) gene families in flax (*Linum usitatissimum*). *BMC Genomics* 14:742. doi: 10.1186/1471-2164-14-742
39. Mokshina N, Gorshkova T, Deyholos MK (2014) Chitinase-like (CTL) and cellulose synthase (CESA) gene expression in gelatinous-type cellulosic walls of flax (*Linum usitatissimum* L.) bast fibres. *PLoS One* 9:e97949. doi: 10.1371/journal.pone.0097949
40. Galinousky D, Padvitski T, Bayer G, Pirko Y, Pydiura N, Anisimova N, Nikitinskaya T, Khotyleva L, Yemets A, Kilchevsky A, Blume Y (2017a) Expression analysis of cellulose synthase and main cytoskeletal protein genes in flax (*Linum usitatissimum* L.). *Cell Biol Int*. doi: 10.1002/cbin.10837
41. Guo Y, Qiu C, Long S, Chen P, Hao D, Preisner M, Wang H, Wang Y (2017) Digital gene expression profiling of flax (*Linum usitatissimum* L.) stem peel identifies genes enriched in fibrebearing phloem tissue. *Gene* 626:32–40. doi: 10.1016/j.gene.2017.05.002
42. Gavazzi F, Pigna G, Braglia L, Gianì S, Breviario D, Morello L (2017) Evolutionary characterization and transcript profiling of  $\beta$ -tubulin genes in flax (*Linum usitatissimum* L.) during plant development. *Plant Biol* 17(237):237. doi: 10.1186/s12870-017-1186-0
43. Pydiura N, Pirko Y, Galinousky D, Postovoitova A, Yemets A, Kilchevsky A, Blume Y (2018) Genome-wide identification, phylogenetic classification, and exon-intron structure characterisation of the tubulin and actin genes in flax (*Linum usitatissimum*). *Cell Biol Intl*. doi: 10.1002/cbin.11001
44. Green PB (1962) Mechanism for plant cellular morphogenesis. *Science* 138:1404–1405. doi: 10.1126/science.138.3548.1404

45. Paredez A, Somerville C, Ehrhardt D (2006) Visualization of cellulose synthase demonstrates functional association with microtubules. *Science* 312:1491–1495. doi: 10.1126/science.1126551
46. Li S, Lei L, Somerville C, Gu Y (2012) Cellulose synthase interactive protein 1 (CSI1) links microtubules and cellulose synthase complexes. *Proc Natl Acad Sci U S A* 109:185–190. doi: 10.1073/pnas.1118560109
47. Paredez AR, Persson S, Ehrhardt DW, Somerville CR (2008) Genetic evidence that cellulose synthase activity influences microtubule cortical array organization. *Plant Physiol* 147:1723–1734. doi: 10.1104/pp.108.120196
48. Bringmann M, Li E, Sampathkumar A, Kocabek T, Hauser M-T, Persson S (2012a) POM-POM2/CELLULOSE SYNTHASE INTERACTING1 is essential for the functional association of cellulose synthase and microtubules in *Arabidopsis*. *Plant Cell* 24:163–177. doi: 10.1105/tpc.111.093575
49. Breviario D, Giani S, Morello L (2013) Multiple tubulins: evolutionary aspects and biological implications. *Plant J* 75:202–218. doi: 10.1111/tpj.12243
50. Ermawar RA, Collins HM, Byrt CS, Henderson M, O'Donovan LA, Shirley NJ, Schwerdt JG, Lahnstein J, Fincher GB, Burton RA (2015) Genetics and physiology of cell wall polysaccharides in the model C4 grass, *Setaria viridis* spp. *BMC Plant Biol* 15:236. doi: 10.1186/s12870-015-0624-0
51. Dhugga KS (2001) Building the wall: genes and enzyme complexes for polysaccharide synthases. *Curr Opin Plant Biol* 2:488–493
52. Kopczak, S. D., Haas, N. A., Hussey, P. J., Silflow, C. D., Snustad, D. P. (1992). The small genome of *Arabidopsis* contains at least six expressed alpha-tubulin genes. *Plant Cell*, 4, 539–47. doi: 10.1105/tpc.4.5.539.
53. Snustad, D. P., Haas, N. A., Kopczak, S. D., Silflow, C. D. (1992). The small genome of *Arabidopsis* contains at least nine expressed beta-tubulin genes. *Plant Cell*, 4, 549–56. doi: 10.1105/tpc.4.5.549
54. Liu, B., Joshi, H. C., Wilson, T. J., Silflow, C. D., Palevitz, B. A., Snustad, D. P. (1994). gammaTubulin in *Arabidopsis*: gene sequence, immunoblot, and immunofluorescence studies. *Plant Cell*, 62(2), 303–14. doi: 10.1105/tpc.6.2.303.

55. Meagher, R. B., McKinney, E. C., Vitale, A. V. (1999). The evolution of new structures: clues from plant cytoskeletal genes. *Trends in Genetics*, 15(7), 278–84.
56. Meagher, R. B., Fechheimer, M. (2003). The Arabidopsis cytoskeletal genome. *Arabidopsis Book*, 2, e0096. doi: 10.1199/tab.0096.
57. Zhang D, Waldsworth P, Hepler PK (1990) Microtubule dynamics in living dividing plant cells: confocal imaging of microinjected fluorescent brain tubulin. *Proc Natl Acad Sci U S A* 87:8820–8824
58. Yuan M, Shaw PJ, Warn RM, Lloyd CW (1994) Dynamic reorientation of cortical microtubules from transverse to longitudinal, in living plant cells. *Proc Natl Acad Sci U S A* 91:6050–6053
59. Oda Y (2015) Cortical microtubule rearrangements and cell wall patterning. *Front Plant Sci* 6:236. doi: 10.3389/fpls.2015.00236
60. Gardiner J (2013) The evolution and diversification of plant microtubule-associated proteins. *Plant J* 75:219–229. doi: 10.1111/tpj.12189
61. Serrano L, de la Torre J, Maccioni RB, Avila J (1984) Involvement of the carboxyl-terminal domain of tubulin in the regulation of its assembly. *Proc Natl Acad Sci U S A* 81:5989–5993
62. Hamada T (2007) Microtubule-associated proteins in higher plants. *J Plant Res* 120(1):79–98. doi: 10.1007/s10265-006-0057-9
63. Smertenko A, Blume Y, Viklicky V, Opatrny Z, Draber P (1997) Posttranslational modifications and multiple tubulin isoforms in *Nicotiana tabacum* L. cells. *Planta* 201:349–358. doi: 10.1007/s004250050077
64. Jovanović AM, Durst S, Nick P (2010) Plant cell division is specifically affected by nitrotyrosine. *J Exp Bot* 61:901–909. doi: 10.1093/jxb/erp369
65. Blume YB, Krasnylenko YA, Demchuk OM, Yemets AI (2013) Tubulin tyrosine nitration regulates microtubule organization in plant cells. *Front Plant Sci* 4:530. doi: 10.3389/fpls.2013.00530
66. Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, Qiu JL (2014) Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable

- resistance to powdery mildew. *Nat Biotechnol* 32:947–951. doi: 10.1038/nbt.2969
67. Dostál V, Libusová L (2014) Microtubule drugs: action, selectivity, and resistance across the kingdoms of life. *Protoplasma* 251:991–1005. doi: 10.1007/s00709-014-0633-0
  68. Wilson PG, Borisy GG (1997) Evolution of the multi-tubulin hypothesis. *Bioessays* 19:451–454. doi: 10.1002/bies.950190603
  69. Gadadhar S, Bodakuntla S, Natarajan K, Janke C (2017) The tubulin code at a glance. *J Cell Sci* 130:1347–1353. doi: 10.1242/jcs.199471
  70. Freedman H, Luchko T, Luduena RF, Tuszynski JA (2011) Molecular dynamics modeling of tubulin C-terminal tail interactions with the microtubule surface. *Proteins* 79:2968–2982. doi: 10.1002/prot.23155
  71. Bao Y, Kost B, Chua NH (2001) Reduced expression of alpha-tubulin genes in *Arabidopsis thaliana* specifically affects root growth and morphology, root hair development and root gravitropism. *Plant J* 28:145–157
  72. Thitamadee S, Tuchiara K, Hashimoto T (2002) Microtubule basis for left-handed helical growth in *Arabidopsis*. *Nature* 417:193–196. doi: 10.1038/417193a
  73. Mayer TU, Kapoor TM, Haggarty SJ, King RW, Schreiber SL, Mitchison TJ (1999) Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. *Science* 286:971–974
  74. Ishida T, Hashimoto TJ (2007) An *Arabidopsis thaliana* tubulin mutant with conditional root-skewing phenotype. *Plant Res* 120:635–640. doi: 10.1007/s10265-007-0105-0
  75. Spokevicius AV, Southerton SG, MacMillan CP, Qiu D, Gan S, Tibbits JFG, Moran GF, Bossinger G (2007)  $\beta$ -Tubulin affects cellulose microfibril orientation in plant secondary fibre cell walls. *Plant J* 51:717–726. doi: 10.1111/j.1365-313X.2007.03176.x

76. McDowell, J. M., Huang, S., McKinney, E. C., An, Y. Q., Meagher, R. B. (1996b). Structure and evolution of the actin gene family in *Arabidopsis thaliana*. *Genetics*, 142(2), 587–602.
77. Šljajčeroová, K., Fišerová, J., Fischer, L., Schwarzerová, K. (2012). Multiple actin isoforms in plants: diverse genes for diverse roles? *Frontiers in Plant Science*, 3, 226. doi: 10.3389/fpls.2012.00226
78. McDowell, J. M., An, Y. Q., Huang, S., McKinney, E. C., Meagher, R. B. (1996a). The *Arabidopsis* ACT7 actin gene is expressed in rapidly developing tissues and responds to several external stimuli. *Plant Physiology*, 111(3), 699–711.
79. Kandasamy, M. K., McKinney, E. C., Meagher, R. B. (2002). Functional nonequivalency of actin isoforms in *Arabidopsis*. *Molecular Biology of the Cell*, 13, 251–61. doi: 10.1091/mbc.01-07-0342
80. Kandasamy, M. K., McKinney, E. C., Roy, E., Meagher, R. B. (2012). Plant vegetative and animal cytoplasmic actins share functional competence for spatial development with protists. *Plant Cell*, 24, 2041–57. doi: 10.1105/tpc.111.095281
81. Kandasamy, M. K., Burgos-Rivera, B., McKinney, E. C., Ruzicka, D. R., Meagher, R. B. (2007). Class-specific interaction of profilin and ADF isoforms with actin in the regulation of plant development. *Plant Cell*, 19, 3111–26. doi: 10.1105/tpc.107.052621
82. Demchuk, O. N., Blume, Ya. B. (2005). Construction of phylogenetic tree of plant tubulins basing on the homology of their protein sequences. *Tsitologiya i Genetika*, 39(2), 3–9.
83. Dominguez, R., Holmes, K. C. (2011). Actin structure and function. *Annual Review of Biophysics*, 40, 169–86. doi: 10.1146/annurev-biophys-042910-155359
84. Gavazzi, F., Braglia, L., Mastromauro, F., Gianì, S., Morello, L., Breviaro, D. (2016). The tubulin-based-polymorphism method provides a simple and effective alternative to the genomic profiling of grape. *PLoS One*, 11(9), e0163335. doi: 10.1371/journal.pone.0163335

85. Hanley, B. A., Schuler, M. A. (1988). Plant intron sequences: evidence for distinct groups of introns. *Nucleic Acids Research*, 16(14B), 7159–76.
86. Chorev, M., Carmel, L. (2012). The function of introns. *Frontiers in Genetics*, 3, 55. doi: 10.3389/fgene.2012.00055.
87. Jo, B. S., Choi, S. S. (2015). Introns: The functional benefits of introns in genomes. *Genomics & Informatics*, 13(4), 112–118. doi: 10.5808/GI.2015.13.4.112
88. Mascarenhas, D., Mettler, I. J., Pierce, D. A., Lowe, H. W. (1990). Intron-mediated enhancement of heterologous gene expression in maize. *Plant Molecular Biology*, 15, 913–20.
89. Luehrsen, K. R., Walbot, V. (1991). Intron enhancement of gene expression and the splicing efficiency of introns in maize cells. *Molecular Genetics and Genomics*, 225, 81–93.
90. Akua, T., Berezin, I., Shaul, O. (2010). The leader intron of AtMHX can elicit, in the absence of splicing, low-level intron-mediated enhancement that depends on the internal intron sequence. *BMC Plant Biology*, 10, 93. doi: 10.1186/1471-2229-10-93
91. Rose, A. B. (2008). Intron-mediated regulation of gene expression. *Current Topics in Microbiology and Immunology*, 326, 277–90.
92. Morello, L., Bardini, M., Sala, F., Breviario, D. (2002). A long leader intron of the Ostub16 rice beta-tubulin gene is required for high-level gene expression and can autonomously promote transcription both in vivo and in vitro. *The Plant Journal*, 29(1), 33–44.
93. Parra, G., Bradnam, K., Rose, A. B., Korf, I. (2011). Comparative and functional analysis of intron-mediated enhancement signals reveals conserved features among plants. *Nucleic Acids Research*, 39, 5328–37. doi: 10.1093/nar/gkr043
94. Gallegos, J. E., Rose, A. B. (2015). The enduring mystery of intron-mediated enhancement. *Plant Science*, 237, 8–15. doi: 10.1016/j.plantsci.2015.04.017

95. Morello, L., Giani, S., Troina, F., Breviario, D. (2011). Testing the IMETER on rice introns and other aspects of intron-mediated enhancement of gene expression. *Journal of Experimental Botany*, 62, 533–44. doi: 10.1093/jxb/erq273
96. Fiume, E., Christou, P., Giani, S., Breviario, D. (2004). Introns are key regulatory elements of rice tubulin expression. *Planta*, 218, 693–703. doi: 10.1007/s00425-003-1150-0.
97. Giani, S., Altana, A., Campanoni, P., Morello, L., Breviario, D. (2009). In transgenic rice, alpha and beta-tubulin regulatory sequences control GUS amount and distribution through intron mediated enhancement and intron dependent spatial expression. *Transgenic Research*, 18(2), 151–62. doi: 10.1007/s11248-008-9202-7
98. P.A. Karpov, Y.B. Blume (2008) Bioinformatic search for plant homologues of animal structural MAPs in the *Arabidopsis thaliana* genome, In: *The Plant Cytoskeleton: a Key Tool for Agro-Biotechnology*. Blume YB *et al.*, 373-397 (Springer). DOI: 10.1007/978-1-4020-8843-8\_18
99. Doblin MS, Kurek I, Jacob-Wilk D, Delmer DP (2002) Cellulose biosynthesis in plants: from genes to rosettes. *Plant Cell Physiol* 43(12): 1407–1420
100. Fernandes AN, Thomas LH, Altaner CM, Callow P, Forsyth VT, Apperley DC, Kennedy CJ, Jarvis MC (2011) Nanostructure of cellulose microfibrils in spruce wood. *Proc Natl Acad Sci U S A* 108:E1195–E1203. doi: 10.1073/pnas.1108942108
101. Newman RH, Hill SJ, Harris PJ (2013) Wide-angle X-ray scattering and solid-state nuclear magnetic resonance data combined to test models for cellulose microfibrils in mung bean cell walls. *Plant Physiol* 163: 1558–1567. doi: 10.1104/pp.113.228262
102. Wightman R, Turner SR (2008) The roles of the cytoskeleton during cellulose deposition at the secondary cell wall. *Plant J* 54:794–805. doi: 10.1111/j.1365-313X.2008.03444.x
103. Watanabe Y, Meents MJ, McDonnell LM, Barkwill S, Sampathkumar A, Cartwright HN, Demura T, Ehrhardt DW, Samuels AL, Mansfield SD (2015)



- Visualization of cellulose synthases in Arabidopsis secondary cell walls. *Science* 350:198–203. doi: 10.1126/science.aac7446
104. Pear JR, Kawagoe Y, Schreckengost WE, Delmer DP, Stalker DM (1996) Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase. *Proc Natl Acad Sci U S A* 93:12637–12642
105. Arioli T, Peng LC, Betzner AS, Burn J, Wittke W, Herth W, Camilleri C, Hofte H, Plazinski J, Birch R, Cork A, Glover J, Redmond J, Williamson RE (1998) Molecular analysis of cellulose biosynthesis in Arabidopsis. *Science* 279:717–720
106. Kimura S, Laosinchai W, Itoh T, Cui X, Linder CR, Brown RM Jr (1999) Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant *Vigna angularis*. *Plant Cell* 11: 2075–2086. doi: 10.1105/tpc.11.11.2075
107. Pydiura NA, Bayer GY, Galinousky DV, Yemets AI, Pirko YV, Podvitski TA, Anisimova NV, Khotyleva LV, Kilchevski AV, Blume YB (2015) Bioinformatic search for cellulose synthase genes in flax (*Linum usitatissimum*) and their phylogenetic analysis. *Cytol Genetics* 49:279–287. doi: 10.3103/S0095452715050084
108. Chantreau M, Chabbert B, Billiard S, Hawkins S, Neutelings G (2015) Functional analyses of cellulose synthase genes in flax (*Linum usitatissimum*) by virus-induced gene silencing. *Plant Biotechnol J* 13(9):1312–1324. doi: 10.1111/pbi.12350
109. Holland N, Holland D, Helentjaris T, Dhugga KS, Xoconostle-Cazares B, Delmer DP (2000) A comparative analysis of the plant cellulose synthase (CesA) gene family. *Plant Physiol* 123:1313–1324. doi: 10.1104/pp.123.4.1313
110. Long SH, Deng X, Wang YF, Li X, Qiao RQ, Qiu CS, Guo Y, Hao DM, Jia WQ, Chen XB (2012) Analysis of 2,297 expressed sequence tags (ESTs) from a cDNA library of flax (*Linum usitatissimum* L.) bark tissue. *Mol Biol Rep* 39:6289–6296. doi: 10.1007/s11033-012-1450-1

111. Roberts E, Roberts AW (2009) A cellulose synthase gene from the red alga *Porphyra yezoensis* (Rhodophyta). *J Phycol* 45:203–212. doi: 10.1111/j.1529-8817.2008.00626.x
112. Sethaphong L, Haigler CH, Kubicki JD, Zimmer J, Bonetta D, DeBolt S, Yingling YG (2013) Plant cellulose synthase structure. *Proc Natl Acad Sci U S A* 110:7512–7517. doi: 10.1073/pnas.1301027110
113. Vergara CE, Carpita NC (2001)  $\beta$ -D-glycan synthases and the CesA gene family: lessons to be learned from the mixed-linkage (1 $\rightarrow$ 3), (1 $\rightarrow$ 4)  $\beta$ -D-glucan synthase. *Plant Mol Biol* 47:145–160. doi: 10.1023/A:1010631431620
114. Chen S, Jia H, Zhao H, Liu D, Liu Y, Liu B, Bauer S, Somerville CR (2016) Anisotropic cell expansion is affected through the bidirectional mobility of cellulose synthase complexes and phosphorylation at two critical residues on CESA3. *Plant Physiol* 171:242–250. doi: 10.1104/pp.15.01874
115. Li F, Xie G, Huang J, Zhang R, Li Y, Zhang M, Wang Y, Li A, Li X, Xia T, Qu C, Hu F, Ragauskas AJ, Peng L (2017) OsCESA9 conserved site mutation leads to largely enhanced plant lodging resistance and biomass enzymatic saccharification by reducing cellulose DP and crystallinity in rice. *Plant Biotechnol J* 15:1093–1104. doi: 10.1111/pbi.12700
116. Mokshina N, Gorshkov O, Ibragimova N, Chernova T, Gorshkova T (2017) Cellulosic fibres of flax recruit both primary and secondary cell wall cellulose synthases during deposition of thick tertiary cell walls and in the course of graviresponse. *Funct Plant Biol* 44:820–831. doi: 10.1071/FP17105
117. Zhang N, Deyholos MK (2016) RNASeq analysis of the shoot apex of flax (*Linum usitatissimum*) to identify phloem fibre specification genes. *Front Plant Sci* 7(950). doi: 10.3389/fpls.2016.00950
118. Worden N, Wilkop TE, Esteve VE, Jeannotte R, Lathe R, Vernhettes S, Weimer B, Hicks G, Alonso J, Labavitch J, Persson S, Ehrhardt D, Drakakaki G (2015) CESA TRAFFICKING INHIBITOR inhibits cellulose deposition and interferes with the trafficking of cellulose synthase complexes and their associated proteins

- KORRIGAN1 and POM2/CELLULOSE SYNTHASE INTERACTIVE PROTEIN1. *Plant Physiol* 167:381–393. doi: 10.1104/pp.114.249003
119. Gu Y, Kaplinsky N, Bringmann M, Cobb A, Carroll A, Sampathkumar A, Baskin TI, Persson S, Somerville CR (2010) Identification of a cellulose synthase-associated protein required for cellulose biosynthesis. *Proc Natl Acad Sci U S A* 107:12866–12871. doi: 10.1073/pnas.1007092107
120. Galinousky D, Padvitski T (2017b) Analysis of public RNA-seq data in studies of flax fibre biogenesis. *EuroBiotech J* 1:177–179. doi: 10.24190/ISSN2564-615X/2017/02.10
121. Titok VV, Lemesh VA, Jurenkova SI, Khotyleva LV (2010) Genetics, physiology and biochemistry of flax. *Belaruskaya navuka*, Minsk In Russian
122. Desprez T, Juraniec M, Crowell EF, Jouy H, Pochylova Z, Parcy F, Hofte H, Gonneau M, Vernhettes S (2007) Organization of cellulose synthase complexes involved in primary cell wall synthesis in *Arabidopsis thaliana*. *Proc Natl Acad Sci* 104:15572–15577. doi: 10.1073/pnas.0706569104
123. Kumar M, Thammannagowda S, Bulone V, Chiang V, Han KH, Joshi CP, Mansfield SD, Mellerowicz E, Sundberg B, Teeri T, Ellis BE (2009) An update on the nomenclature for the cellulose synthase genes in *Populus*. *Trends Plant Sci* 14:248–254. doi: 10.1016/j.tplants.2009.02.004
124. Atanassov II, Pittman JK, Turner SR (2009) Elucidating the mechanisms of assembly and subunit interaction of the cellulose synthase complex of *Arabidopsis* secondary cell walls. *J Biol Chem* 284:3833–3841. doi: 10.1074/jbc.M807456200
125. Kumar M, Turner S (2015) Plant cellulose synthesis: CESA proteins crossing kingdoms. *Phytochem* 112:91–99. doi: 10.1016/j.phytochem.2014.07.009
126. Brown, R.M., Cellulose structure and biosynthesis: what is in store for the 21st century?, *J. Polym. Sci. A. Polym. Chem.*, 2004, vol. 42, pp. 487–495.
127. Scheible, W.R., Eshed, R., Richmond, T., et al., Modifications of cellulose synthase confer resistance to isoxaben and thiazolidinone herbicides in

- Arabidopsis CFPxr1 mutants, Proc. Nat. Acad. Sci. USA, 2001, vol. 98, pp. 10079–10084.
128. Taylor, N.G., Howells, R.M., Huttly, A.K., et al., Interactions among three distinct cesa proteins essential for cellulose synthesis, Proc. Nat. Acad. Sci. USA, 2003, vol. 100, pp. 1450–1455.
129. Ross, P., Mayer, R., and Benziman, M., Cellulose biosynthesis and function in bacteria, Microbiol. Rev., 1991, vol. 55, no. 1, pp. 35–58.
130. Hamann, T., Osborne, E., Youngs, H.L., et al., Global expression analysis of CESA and CSL genes in Arabidopsis, Cellulose, 2004, vol. 11, pp. 279–286.
131. Ranik, M. and Myburg, A.A., Six new cellulose synthase genes from eucalyptus associated with primary and secondary cell wall biosynthesis, Tree Physiol., 2006, vol. 26, pp. 545–556.
132. Carroll, A. and Specht, C., Understanding plant cellulose synthases through a comprehensive investigation of the cellulose synthase family sequences, Front. Plant Sci., 2011, vol. 2, pp. 1–11.
133. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct 5;215(3):403-10. doi: 10.1016/S0022-2836(05)80360-2
134. Gertz EM, Yu YK, Agarwala R, Schäffer AA, Altschul SF. Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. BMC Biol. 2006;4:41. Published 2006 Dec 7. doi:10.1186/1741-7007-4-41
135. Goodstein DM, Shu S, Howson R, et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 2012;40(Database issue):D1178-D1186. doi:10.1093/nar/gkr944
136. Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T. Z., Garcia-Hernandez, M., Foerster, H., Li, D., Meyer, T., Muller, R., Ploetz, L., Radenbaugh, A., Singh, S., Swing, V., Tissier, C., Zhang, P., Huala, E.(2008). The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Research, 36(Database issue),D1009–14. doi:10.1093/nar/gkm965

137. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A*. 1992 Nov 15;89(22):10915-9. doi: 10.1073/pnas.89.22.10915
138. Altschul SF. A protein alignment scoring system sensitive at all evolutionary distances. *J Mol Evol*. 1993 Mar;36(3):290-300. doi: 10.1007/BF00160485
139. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23, 2947-2948.
140. Madeira F, Park YM, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*. 2019 Jul;47(W1):W636-W641. DOI: 10.1093/nar/gkz268
141. Kumar, S., Stecher, G., Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, 33(7), 1870-4. doi:10.1093/molbev/msw054
142. Dereeper, A., Guignon, V., Blanc, G., et al., Phylogeny. fr: robust phylogenetic analysis for the nonspecialist, *Nucleic Acids Res.*, 2008, vol. 36 (Web Server issue), pp. W465–W469.
143. Edgar, R.C., MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.*, 2004, vol. 32, no. 5, pp. 1792–1797.
144. Castresana, J., Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis, *Mol. Biol. Evol.*, 2000, vol. 17, pp. 540–552.
145. Guindon, S., Dufayard, J.F., Lefort, V., et al., New algorithms and methods to estimate maximumlikelihood phylogenies: assessing the performance of PhyML 3.0, *System. Biol.*, 2010, vol. 59, no. 3, pp. 307–321.
146. Chevenet, F., Brun, C., Banuls, A.L., et al., TreeDyn: towards dynamic graphics and annotations for analyses of trees, *MBC Bioinform.*, 2006, vol. 7, p. 439.
147. Bo Hu, Jinpu Jin, An-Yuan Guo, He Zhang, Jingchu Luo and Ge Gao. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*, 31(8):1296-1297.

148. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 2011 Jul;39(Web Server issue):W29-37. doi: 10.1093/nar/gkr367
149. S. El-Gebali, J. Mistry, A. Bateman, S.R. Eddy, A. Luciani, S.C. Potter, M. Qureshi, L.J. Richardson, G.A. Salazar, A. Smart, E.L.L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S.C.E. Tosatto, R.D. Finn / The Pfam protein families database in 2019 // *Nucleic Acids Research* (2019), Volume 47, Issue D1, 08 January 2019, Pages D427–D432. doi: 10.1093/nar/gky995
150. N. Eswar, M.A. Marti-Renom, B. Webb, M.S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, A. Sali. Comparative protein structure modeling with MODELLER. *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc., Supplement 15, 5.6.1-5.6.30, 2006. DOI: 10.1002/cpbi.3
151. Humphrey, W., Dalke, A. and Schulten, K., VMD - Visual Molecular Dynamics, *J. Molec. Graphics*, 1996, vol. 14, pp. 33-38. DOI: 10.1016/0263-7855(96)00018-5 (<http://www.ks.uiuc.edu/Research/vmd/>)
152. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. (<https://pymol.org/>)
153. J Yang, R Yan, A Roy, D Xu, J Poisson, Y Zhang. The I-TASSER Suite: Protein structure and function prediction. *Nature Methods*, 12: 7-8 (2015). (Download the PDF file).
154. A Roy, A Kucukural, Y Zhang. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, 5: 725-738 (2010) (download the PDF file)
155. Y Zhang. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, vol 9, 40 (2008).
156. J.W. Ponder and D.A. Case. (2003) Force fields for protein simulations. *Adv. Prot. Chem.* 66, 27-85.
157. Buchan DWA, Jones DT (2019). The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Research*. doi: 10.1093/nar/gkz297

158. Drozdetskiy A, Cole Ch, Procter J, Barton G, JPred4: a protein secondary structure prediction server, *Nucleic Acids Research*, Volume 43, Issue W1, 1 July 2015, Pages W389–W394, doi: 10.1093/nar/gkv332
159. V.B. Chen, W.B. Arendall 3rd, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica*, D66:12-21, 2010. doi: 10.1107/S0907444909042073 (<http://molprobity.biochem.duke.edu/>)
160. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, Beglov D, Vajda S. The ClusPro web server for protein-protein docking. *Nature Protocols*. 2017 Feb;12(2):255-278
161. A. Goble, S. Pettifer and R. Stevens, Knowledge Integration: In silico Experiments in Bioinformatics, in I. Foster and C. Kesselman, eds., *The Grid: Blueprint for a New Computing Infrastructure Second Edition*, Morgan Kaufman, 2003.
162. Katherine Wolstencroft, Robert Haines, Donal Fellows, Alan Williams, David Withers, Stuart Owen, Stian Soiland-Reyes, Ian Dunlop, Aleksandra Nenadic, Paul Fisher, Jiten Bhagat, Khalid Belhajjame, Finn Bacall, Alex Hardisty, Abraham Nieva de la Hidalga, Maria P. Balcazar Vargas, Shoaib Sufi, and Carole Goble (2013): The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud, *Nucleic Acids Research*, 41(W1): W557-W561. doi:10.1093/nar/gkt328
163. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. The Protein Data Bank. 2000 *Nucleic Acids Research*, 28: 235-242. DOI: 10.1093/nar/28.1.235
164. M. Magrane and the UniProt consortium. UniProt Knowledgebase: a hub of integrated protein data. // *Database*, 2011. DOI: 10.1093/database/bar009
165. E. Bolton, Y. Wang, P.A. Thiessen, S.H. Bryant. PubChem: Integrated Platform of Small Molecules and Biological Activities. *Annual Reports in Computational*

- Chemistry*, Volume 4, Chapter 12, American Chemical Society, Washington, DC, 2008. DOI: 10.1016/S1574-1400(08)00012-1
166. W. Kabsch, C. Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22 2577-2637, 1983. DOI: 10.1002/bip.360221211
167. A. Porollo, R. Adamczak, J. Meller. POLYVIEW: A Flexible Visualization Tool for Structural and Functional Annotations of Proteins. *Bioinformatics*, 20: 2460-2462, 2004. (<http://polyview.cchmc.org>). DOI: 10.1093/bioinformatics/bth248
168. Jmol: an open-source Java viewer for chemical structures in 3D. (<http://www.jmol.org/>)
169. B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.*, 4 (3), pp 435–447, 2008. DOI: 10.1021/ct700301q
170. Vignaud T., Blanchoin L., Théry M. Directed cytoskeleton self-organization // *Trends Cell Biol.* – 2012. – 22, № 12. – P. 671–682.
171. Eren E.C., Gautam N., Dixit R. Computer simulation and mathematical models of the noncentrosomal plant cortical microtubule cytoskeleton // *Cytoskeleton* (Hoboken). – 2012. – 69, № 3. – P. 144–154.
172. Sui M., Zhang H., Di X., Chang J., Shen Y., Fan W. G2 checkpoint abrogator abates the antagonistic interaction between antimicrotubule drugs and radiation therapy // *Radiother. Oncol.* – 2012. – 104, № 2. – P. 243–248.
173. Zhao Y., Wu F., Wang Y., Chen S., Han G., Liu M., Jin D. Inhibitory action of chamaejasmin A against human HEP-2 epithelial cells: effect on tubulin protein // *Mol. Biol. Rep.* – 2012. – 39, № 12. – P. 11105–11112.
174. Pydiura N., Karpov P., Blume Ya. Design of specific cytoskeleton related biological database and data management environment for bioinformatical cytoskeleton investigation and collaboration within virtual Grid-organisation // *Proc. of the Int. Moscow Conference on Computational Molecular Biology (MCCMB'11)*, July 21 – 24 2011. – Moscow, Russia, 2011. – P. 297–298.



175. Карпов П.А., Брицун В.М., Раєвський О.В., Демчук О.М., Пидюра М.О., Ожерєдов С.П., Самофалова Д.О., Співак С.І., Ємець А.І., Кальченко В.І., Блюм Я.Б. Високопропускний скринінг речовин з анти мітотичною активністю на базі віртуальної організації CSLabGrid // Наука та інновації. – 2015. – 11, № 1. – С. 92–100.
176. Pydiura N., Karpov P., Blume Y. COMODORE: a web system for high-throughput computational investigations in Grid environment // On-line Proc. of the 9th Int. Symposium on Integrative Bioinformatics 2013. Ed. by F. Schreiber F., Hofestädt R., Lange M., Scholz U., Czauderna T., Schnee R. and Weise S, 18–20 March 2013. – Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany, 2013. – P. 190–191.
177. The UniProt Consortium. UniProt: a hub for protein information // Nucl. Acids Res. – 2015. – 43 (D1). – P. D204–D212.
178. Roy A., Kucukural A., Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction // Nat. Protoc. – 2010. – 5, № 4. – P. 725–738.
179. James C. Global Status of Commercialized Biotech/GM Crops// ISAAA Brief. – ISAAA: Ithaca, NY, 2014. – № 49.
180. Kozub N.O., Pilipenko L.A., Sozinov I.O., Blume Ya.B., Sozinov O.O. Genetically modified plants and plant protection problems: progress and estimation of potential risks // Cytol. Genet. – 2012. – 46, № 4. – P. 251–262. DOI: 10.3103/S0095452712040081
181. Novozhylov O.V., Blume Ya.B. Preliminary risk assessment of the transgenic plant use in Ukraine // AgBioForum. – 2014. – 17, № 3. – P. 191–196.
182. Корховий В.І., Ємець А.І., Яцишин В.Ю. Методи відбору проб і підготовки зразків для детекції та визначення ГМО у сільськогосподарських рослинах і продуктах харчування, вироблених із них [Електронний ресурс] // Наукові доп. Національного університету біоресурсів і природокористування України. – 2015. – № 8 (57). – Режим доступу: [http://nd.nubip.edu.ua/2015\\_8/15.pdf](http://nd.nubip.edu.ua/2015_8/15.pdf).

183. Miraglia M., Berdal K.G., Brera C. et al. Detection and traceability of genetically modified organisms in the food production chain // *Food Chem. Toxicol.* – 2004. – 42. – P. 1157–1180. DOI: 10.1016/j.fct.2004.02.018
184. Blume Ya.B., Baird W.V., Yemets A.I., Breviario D. *The Plant Cytoskeleton: Key Tool for Agro-Biotechnology.* – Springer-Verlag, Berlin, Heidelberg, NY, 2008. – 457 p.
185. Holmes, K. C., Popp, D., Gebhard, W., Kabsch, W. (1990). Atomic model of the actin filament. *Nature*, 347, 44–9. doi: 10.1038/347044a0.
186. Kim, E., Wriggers, W., Phillips, M., Kokabi, K., Rubenstein, P. A., Reisler, E. (2000). Crosslinking constraints on F-actin structure. *Journal of Molecular Biology*, 299, 421–29. doi: 10.1006/jmbi.2000.3727.
187. dos Remedios, C. G., Chhabra, D., Kekic, M., Dedova, I. V., Tsubakihara, M., Berry, D. A., Nosworthy, N. J. (2003). Actin binding proteins: regulation of cytoskeletal microfilaments. *Physiological Reviews*, 83(2), 433–73. doi: 10.1152/physrev.00026.2002.
188. Perumal, B. S., Sakharkar, K. R., Chow, V. T., Pandjassarame, K., Sakharkar, M. K. (2005). Intron position conservation across eukaryotic lineages in tubulin genes. *Frontiers in Bioscience*, 10, 2412–9.
189. Nogales E, Wolf SG, Downing KH (1998) Structure of the alpha beta tubulin dimer by electron crystallography. *Nature* 391:199–203. doi: 10.1038/34465
190. Selvaa Kumar C, Gadewal N, Mohammed SM (2015) Seminal role of deletion of amino acid residues in H1-S2 and S-loop regions in eukaryotic  $\beta$ -tubulin investigated from docking and dynamics perspective. *J Theor Biol* 378:79–88. doi: 10.1016/j.jtbi.2015.04.035
191. Bierne, N., Lehnert, S. A., Bedier, E., Bonhomme, F., Moore, S. S. (2000). Screening for intronlength polymorphisms in penaeid shrimps using exon-primed intron-crossing (EPIC)-PCR. *Molecular Ecology*, 9, 233–35.
192. Breviario, D., Giani, S., Ponzoni, E., Mastromauro, F., Morello, L. (2008). Plant tubulin intronics. *Cell Biology International*, 32(5), 571–3. doi: 10.1016/j.cellbi.2007.11.013

193. Gonzalez-Verdejo C, Die J, Nadal S, Jimenez-Marin A, Moreno M, Romain B (2008). Selection of housekeeping genes for normalization by real-time RTPCR: Analysis of Or-MYB1 gene expression in *Orobanche ramosa* development. *Anal Biochem*;379(2):38-43.
194. Guenin S, Mauriat M, Pelloux J, Van Wuytswinkel O, Bellini C, Gutierrez L (2009). Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. *J Exp Bot*;60(2):487-93.
195. Martin R, Hollenbeck V, Dombrowski J. (2008). Evaluation of Reference Genes for Quantitative RT-PCR in *Lolium perenne*. *Crop Sci*;48:1881-87.
196. Huis R, Hawkins S, Neutelings G. (2010). Selection of reference genes for quantitative gene expression normalization in flax (*Linum usitatissimum* L.). *BMC Plant Biol*;10:71(1).
197. Bayer GYa, Pydiura N., Yemets AI, Pirko YaV, Blume YaB. (2014). Bioinformatics search of the tubulin genes in *Linum* genome. *Factors in experimental evolution of organisms*;14:146-50.
198. Carpenter JL, Ploense SE, Snustad DP, Silflow CD (1992) Preferential expression of an  $\alpha$ -tubulin gene of *Arabidopsis* in pollen. *Plant Cell* 4:557–571. doi: 10.1105/tpc.4.5.557
199. Oakley RV, Wang YS, Ramakrishna W, Harding SA, Tsai CJ (2007) Differential expansion and expression of alpha- and beta-tubulin gene families in *Populus*. *Plant Physiol* 145:961–973. doi: 10.1104/pp.107.107086
200. Breviario D (2008) Plant tubulin genes: regulatory and evolutionary aspects. In: Nick P (ed) *Plant microtubules*. Plant cell monographs, vol 11. Springer, Berlin, pp 207–232.
201. Schröder J, Stenger H, Wernicke W (2001)  $\alpha$ -Tubulin genes are differentially expressed during leaf cell development in barley (*Hordeum vulgare* L.). *Plant Mol Biol* 45:723–730. doi: 10.1023/A:1010648519206
202. Richmond, T.A. and Somerville, C.R., The cellulose synthase superfamily, *Plant Physiol.*, 2000, vol. 124, pp. 495–498.

203. Hazen, S., ScottCraig, J.S., and Walton, J.D., Cellulose synthaselike (CSL) genes of rice, *Plant Physiol.*, 2002, vol. 128, pp. 336–340.
204. Fincher, G.B., Revolutionary times on our understanding of cell wall biosynthesis and remodeling in the grasses, *Plant Physiol.*, 2009, vol. 149, pp. 27–37.
205. Sveinsson, S., McDill, J., Wong, G.K., et al., Phylogenetic pinpointing of a paleopolyploidy event within the flax genus (*Linum*) using transcriptomics, *Ann. Bot.*, 2014, vol. 113, no. 5, pp. 753–761.
206. Mokshina, N., Gorshkova, T., and Deyholos, M.K., Chitinase-like (CTL) and cellulose synthase (CESA) gene expression in gelatinous type cellulosic walls of flax (*Linum usitatissimum* L.) bast fibers, *PLoS One*, 2014, vol. 9, no. 2, p. e94949.
207. Day A, Ruel K, Neutelings G, Cronier D, David H, Hawkins S, Chabbert B. (2005) Lignification in the flax stem: evidence for an unusual lignin in bast fibers. *Planta*;222(2):234-245.
208. Livak KJ, Schmittgen TD. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C_T}$  method. *Methods*;25(4):402-8.
209. Sauer NJ, Narváez-Vásquez J, Mozoruk J, Miller RB, Warburg ZJ, Woodward MJ, Mihiret YA, Lincoln TA, Segami RE, Sanders SL, Walker KA, Beetham PR, Schöpke CR, Gocal GFW (2016) Oligonucleotide-mediated genome editing provides precision and function to engineered nucleases and antibiotics in plants. *Plant Physiol* 170:1917–1928. doi: 10.1104/pp.15.01696
210. Yang L, Guell M, Byrne S, Yang JL, De Los Angeles A, Mali P, Aach J, Kim-Kiselak C, Briggs AW, Rios X, Huang PY, Daley G, Church G (2013) Optimization of scarless human stem cell genome editing. *Nucleic Acids Res* 41:9049–9061. doi: 10.1093/nar/gkt555
211. Sun Y, Zhang X, Wu C, He Y, Ma Y, Hou H, Guo X, Du W, Zhao Y, Xia L (2016) Engineering herbicide-resistant rice plants through CRISPR-Cas9-mediated homologous recombination of acetolactate synthase. *Mol Plant* 9:628–631. doi: 10.1016/j.molp.2016.01.001

212. Svitashv S, Young JK, Schwartz C, Gao H, Falco SC, Cigan AM (2015) Targeted mutagenesis, precise gene editing, and site specific gene insertion in maize using Cas9 and guide RNA. *Plant Physiol* 169:931–945. doi: 10.1104/pp.15.00793
213. Lu Y, Zhu JK (2017) Precise editing of a target base in the rice genome using a modified CRISPR-Cas9 system. *Mol Plant* 10:523–525. doi: 10.1016/j.molp.2016.11.013
214. Wolter F, Puchta H (2018) The CRISPR-Cas revolution reaches the RNA world: Cas13, a new Swiss Army knife for plant biologists. *Plant J* 94:767–775. doi: 10.1111/tpj.13899
215. Morineau C, Bellec Y, Tellier F, Gissot L, Kelemen Z, Nogue F, Faure JD (2016) Selective gene dosage by CRISPR-Cas9 genome editing in hexaploid *Camelina sativa*. *Plant Biotechnol J* 15:729–739. doi: 10.1111/pbi.12671
216. Jiang WZ, Henry IM, Lynagh PG, Comai L, Cahoon EB, Weeks DP (2016) Significant enhancement of fatty acid composition in seeds of the allohexaploid, *Camelina sativa*, using CRISPR-Cas9 gene editing. *Plant Biotechnol J* 15:648–657. doi: 10.1111/pbi.12663
217. Shan Q, Wang Y, Li J, Zhang Y, Chen K, Liang Z, Zhang K, Liu J, Xi JJ, Qiu JL (2013) Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol* 31:686–688. doi: 10.1038/nbt.2650
218. Zhang Y, Liang Z, Zong Y, Wang Y, Liu J, Chen K, Qiu JL, Gao C (2016) Efficient and transgene-free genome editing in wheat through transient expression of CRISPR-Cas9 DNA or RNA. *Nat Commun* 7: 12617. doi: 10.1038/ncomms12617
219. Abudayyeh OO, Gootenberg JS, Essletzbichler P, Han S, Joung J, Belanto JJ, Verdine V, Cox DBT, Kellner MJ, Regev A, Lander ES, Voytas DF, Ting AY, Zhang F (2017) RNA targeting with CRISPR-Cas13. *Nat Lett* 550:280–284. doi: 10.1038/nature24049
220. Smith J, Grizot S, Arnold S, Duclert A, Epinat JC, Chames P, Prieto J, Redondo P, Blanco FJ, Bravo J, Montoya G, Pâques F, Duchateau P (2006) A combinatorial

- approach to create artificial homing endonuclease cleaving chosen sequences. *Nucleic Acids Res* 34:e149. doi: 10.1093/nar/gkl720
221. Kim YG (1996) Hybrid restriction enzymes: zinc finger fusions to FokI cleavage domain. *Proc Natl Acad Sci U S A* 93:1156–1180. doi: 10.1073/pnas.93.3.1156
222. Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF (2010) Targeting DNA double strand breaks with TAL effector nucleases. *Genetics* 186:757–761. doi: 10.1534/genetics.110.120717
223. Zhu Y, Klompe SE, Vlot M, van der Oost J, Staals RHJ (2018) Shooting the messenger: RNA-targeting CRISPR-Cas systems. *Biosci Rep* 38:BSR20170788. doi: 10.1042/BSR20170788
224. Lund G, Messing J, Viotti A (1995) Endosperm-specific demethylation and activation of specific alleles of alpha-tubulin genes of *Zea mays* L. *Mol Gen Genet* 246:716–722. doi: 10.1007/BF00290717
225. Malzahn A, Lowder L, Qi Y (2017) Plant genome editing with TALEN and CRISPR. *Cell Biosci* 7:21. doi: 10.1186/s13578-017-0148-4
226. Schunder E, Rydsewski GR, Heuner K (2013) First indication for a functional CRISPR-Cas system in *Francisella tularensis*. *Int J Med Microbiol* 303:51–60. doi: 10.1016/j.ijmm.2012.11.004
227. Parker AL, Teo WS, Pandzi E, Vicente J, McCarroll JA, Wordeman L, Kavallaris M (2018)  $\beta$ -Tubulin carboxy-terminal tails exhibit isotype-specific effects on microtubule dynamics in human geneedited cells. *Life Sci Alliance* 1:e201800059. doi: 10.26508/lsa.201800059
228. Vukašinović N, Oda Y, Pejchar P, Synek L, Pečenková T, Rawat A, Sekereš J, Potocký M, Žárský V (2017) Microtubule-dependent targeting of the exocyst complex is necessary for xylem development in *Arabidopsis*. *New Phytol* 213:1052–1067. doi: 10.1111/nph.14267
229. Oda Y, Iida Y, Nagashima Y, Sugiyama Y, Fukuda H (2015) Novel coiled-coil proteins regulate exocyst association with cortical microtubules in xylem cells via the conserved oligomeric Golgi-complex 2 protein. *Plant Cell Physiol* 56:277–286. doi: 10.1093/pcp/pcu197

230. Lei L, Li S, Gu Y (2012) Cellulose synthase complexes: composition and regulation. *Front Plant Sci* 3:75. doi: 10.3389/fpls.2012.00075
231. Vain T, Crowell EF, Timpano H, Biot E, Desprez T, Mansoori N, Trindade LM, Pagant S, Robert S, Höfte H, Gonneau M, Vernhettes S (2014) The cellulase KORRIGAN is part of the cellulose synthase complex. *Plant Physiol* 165:1521–1532. doi: 10.1104/pp.114.241216
232. Li S, Lei L, Somerville C, Gu Y (2012) Cellulose synthase interactive protein 1 (CSII) links microtubules and cellulose synthase complexes. *Proc Natl Acad Sci U S A* 109:185–190. doi: 10.1073/pnas.1118560109
233. Spokevicius AV, Southerton SG, MacMillan CP, Qiu D, Gan S, Tibbits JFG, Moran GF, Bossinger G (2007)  $\beta$ -Tubulin affects cellulose microfibril orientation in plant secondary fibre cell walls. *Plant J* 51:717–726. doi: 10.1111/j.1365-313X.2007.03176.x
234. Li X, Cai L, Cheng N, Liu J (2002) Molecular characterization of the cotton GhTUB1 gene that is preferentially expressed in fiber. *Plant Physiol* 130:666–674. doi: 10.1104/pp.005538
235. He X, Qin Y, Xu Y, Hu C, Zhu Y (2008) Molecular cloning, expression profiling, and yeast complementation of 19 beta-tubulin cDNAs from developing cotton ovules. *J Exp Bot* 59:2687–2695. doi: 10.1093/jxb/ern127
236. Beerli RR, Segal DJ, Dreier B, Barbas CF III (1998) Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc Natl Acad Sci U S A* 95:14628–14633
237. Hiratsu K, Matsui K, Koyama T, Ohme-Takagi M (2003) Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in *Arabidopsis*. *Plant J* 34:733–739. doi: 10.1046/j.1365-313X.2003.01759.x
238. Piatek A, Ali Z, Baazim H, Li L, Abulfaraj A, Al-Shareef S, Aouida M, Mahfouz MM (2015) RNA-guided transcriptional regulation in planta via synthetic dCas9-

- based transcription factors. *Plant Biotechnol J* 13:578–589. doi: 10.1111/pbi.12284
239. Lowder LG, Zhang D, Baltes NJ, Paul JW, Tang X, Zheng X, Voytas DF, Hsieh TF, Zhang Y, Qi Y (2015) A CRISPR-Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. *Plant Physiol* 169:971–985. doi: 10.1104/pp.15.00636
240. Rodriguez-Leal D, Lemmon ZH, Man J, Bartlett ME, Lippman ZB (2017) Engineering quantitative trait variation for crop improvement by genome editing. *Cell* 171:470–480. doi: 10.1016/j.cell.2017.08.030



### Список публікацій здобувача

1. Morello L., **Pydiura N.**, Galinousky D., Blume Ya., Breviario D. Flax tubulin and CesA superfamilies represent attractive and challenging targets for a variety of genome- and base-editing applications. *Funct. Integrat. Genomics*. 2020, 20(1), p. 163–176. DOI: <https://doi.org/10.1007/s10142-019-00667-2> (Здобувачем разом зі співавторами проведено дослідження, опрацьовано дані, написано статтю). **Q2**
2. **Pydiura N.**, Pirko Ya., Galinousky D., Postovoitova A., Yemets A., Kilchevsky A., Blume Ya. Genome-wide identification, phylogenetic classification, and exon-intron structure characterisation of the tubulin and actin genes in flax (*Linum usitatissimum*). *Cell Biol. Int.* 2019, 43(9), p. 1010–1019. DOI: <https://doi.org/10.1002/cbin.11001> (Здобувачем особисто проведено дослідження, опрацьовано отримані дані та написано статтю). **Q2**
3. Galinousky D., Padvitski Ts., Bayer G., Pirko Ya., **Pydiura N.**, Anisimova N., Nikitinskaya T., Khotyleva L., Yemets A., Kilchevsky A. Expression analysis of cellulose synthase and main cytoskeletal protein genes in flax (*Linum usitatissimum L.*). *Cell Biol. Int.* 2019, 43(9), p. 1065–1071. DOI: <https://doi.org/10.1002/cbin.10837> (Здобувачем разом зі співавторами проведено дослідження, опрацьовано отримані дані та написано статтю). **Q2**
4. Блюм, Я.Б., Ємець А.І., Пірко Я.В., Баєр Г.Я., Рабоконь А.М., Буй Д.Д., **Пидюра М.О.**, Галіновський Д.В., Анісімова Н.В., Хотильова Л.В., Кільчевський А.В. Оцінка експресії різних генів β-тубуліну і хітиназо-подібних генів при біогенезі волокон льону. *Анотований збірник проектів спільного конкурсу ДФФД-БРФФД*, 2017, Академперіодика: К., с. 66–69. (Здобувачем разом зі співавторами проведено дослідження, опрацьовано отримані дані та написано статтю).
5. Карпов, Р.; Demchuk, O.; Britsun, V.; Lytvyn, D.; **Pydiura, N.**; Rayevsky, A.; Samofalova, D.; Spivak, D.; Volochnyuk, D.; Yemets, A.; Blume, Ya. New

- Imidazole Inhibitors of Mycobacterial FtsZ: the Way from High-Throughput Molecular Screening in Grid up to in vitro Verification. *Nauka innov.* 2016, 12(3), p. 44–59. DOI: <https://doi.org/10.15407/scin12.03.044> (Здобувачем разом зі співавторами проведено дослідження, опрацьовано отримані дані та написано статтю).
6. Демчук О.М., Карпов П.А., Ожерєдов С.П., Співак С.І., **Пидюра М.О.**, Самофалова Д.О., Блюм Я.Б. Створення репозиторія моделей просторової структури тубулінів як реалізація одного із стратегічних завдань віртуальної організації CSLabGrid. *Фактори експериментальної еволюції організмів.* 2016, 19, с. 208–211. (Здобувачем разом зі співавторами проведено дослідження, опрацьовано отримані дані та написано статтю).
  7. Співак, С.І., Ємець, А.І., Пірко, Я.В., **Пидюра, М.О.**, Карпов, П.А., Ісаєнков, С.В., Блюм, Я.Б. Розроблення українського порталу даних щодо генетично модифікованих рослин “Genetically modified plants gateway”. *Фактори експериментальної еволюції організмів.* 2016, 19, с. 238–242. (Здобувачем особисто проведено дослідження, опрацьовано отримані дані та написано статтю).
  8. Karpov P.A., Brytsun V.M., Rayevsky A.V., Demchuk O.M., **Pydiura N.O.**, Ozheredov S.P., Samofalova D.A., Spivak S.I., Yemets A.I., Kalchenko V.I., Blume Ya.B. High-throughput screening of new antimetabolic compounds based on CSLabGrid virtual organization. *Sci. Innov.* 2015, 11(1), p. 85–93. DOI: <https://doi.org/10.15407/scine11.01.085>) (Здобувачем разом зі співавторами проведено дослідження, опрацьовано отримані дані та написано статтю).
  9. **Pydiura N.A.**, Bayer G.Ya., Galinousky D.V., Yemets A.I., Pirko Ya.V., Padvitski T.A., Anisimova N.V., Khotyleva L.V., Kilchevsky A.V., Blume Ya.B. Bioinformatic search for cellulose synthase genes in flax (*Linum usitatissimum*) and their phylogenetic analysis. *Cytol. Genetics.* 2015, 49(5), p. 279–287. DOI: <https://doi.org/10.3103/S0095452715050084> (Здобувачем разом зі співавторами проведено дослідження, опрацьовано отримані дані та написано статтю).

10. Novosylina, O., Jurewicz, E., **Pydiura, N.**, Goral, A., Filipek, A., Negrutskii, B., El'skaya, A. Translation elongation factor eEF1A1 is a novel partner of a multifunctional protein Sgt1. *Biochimie*. 2015, 119, p. 137–145. DOI: <https://doi.org/10.1016/j.biochi.2015.10.026> (Здобувачем разом зі співавторами проведено дослідження, опрацьовано отримані дані та написано статтю).
11. Баєр Г.Я., **Пидюра М.О.**, Пірко Я.В., Ємець А.І., Блюм Я. Б. Біоінформаційний пошук послідовностей генів, що кодують тубуліни у геномі льону *Фактори експериментальної еволюції організмів*. 2014, 14, с. 14–17. (Здобувачем разом зі співавторами проведено дослідження, опрацьовано дані та написано статтю).
12. **Pydiura N.**, Karpov P., Blume Ya. On the efficiency of CPU and hybrid CPU-GPU systems in computational biology tasks. *Comput. Sci. Appl.* 2014, 1(1), p. 48–59. (Здобувачем разом зі співавторами проведено дослідження, опрацьовано отримані дані та написано статтю).
13. **Pydiura N.**, Karpov P., Blume Ya. Hybrid CPU-GPU calculations – a promising future for computational biology. Third Int. Conf. "High Performance Computing" (Kyiv, Ukraine, October 7–11, 2013), Nat. Technical Univ. of Ukraine «Kyiv Polytechnic Institute», p. 330–335. DOI: <https://doi.org/10.13140/2.1.1709.5366> (Здобувачем разом зі співавторами проведено дослідження, опрацьовано отримано дані та написано статтю).
14. **Pydiura N.**, Karpov P., Blume Ya. Hardware environment for CSLabGrid: Reaching maximum efficacy of computations in structural biology and bioinformatics. In: Cluster Computing: Proc. II Int. Conf. (Lviv, Ukraine, June 3–5, 2013), Lviv Polytechnic Nat. Univ., p. 191–194. (Здобувачем разом зі співавторами проведено дослідження, опрацьовано отримано дані та написано статтю).
15. Галиновский Д.В., Подвицкий Т.А., Анисимова Н.В., **Пыдюра Н.А.**, Баєр Г.Я., Ємець А.И., Пірко Я.В., Блюм Я.Б., Хотылева Л.В., Кильчевский А.В. Экспрессия генов цитоскелета и целлюлозосинтезирующего комплекса клетки при формировании волокон льна. *Технология органических веществ*.

- Тезисы докл. 79-й научно-техн. конф-ции профессорско-преподавательского состава, научных сотрудников и аспирантов (Минск, Беларусь, 2–6 февраля 2015), Белорусский гос. технологический университет: Минск, 2015; с. 6.
16. **Pydiura N.**, Karpov P., Blume Y. COMODORE: a web system for high-throughput computational investigations in Grid environment. On-line Proc. of the 9th Int. Symposium on Integrative Bioinformatics, Gatersleben, Germany (March 18–20, 2013), Leibniz Inst. Plant Genetics and Crop Plant Research (IPK), 2013; pp 190–191.
  17. **Pydiura N.A.**, Karpov P.A., Blume Y.B. The comparison of colchicine, vinblastine and taxol binding sites on animal and plant tubulins. III Int. Symposium «Intracellular Signaling and Bioactive Molecules Design» (Lviv, Ukraine, September 17–23, 2012), Ivan Franko Nat. University of Lviv, 2012, p. 129.
  18. **Pydiura N.A.**, Karpov P.A., Blume Y.B. Bioinformatic search and phylogenetic analysis of the plant-specific MAPs in genomes of monocots and dicots. 8th Int. Conference on the Bioinformatics of Genome Regulation and Structure/Systems Biology (Novosibirsk, Russia, June 25–29, 2012), Institute of Cytology and Genetics, Novosibirsk, p. 260.
  19. **Pydiura N.**, Karpov P., Blume Ya. Design of specific cytoskeleton related biological database and data management environment for bioinformatical cytoskeleton investigation and collaboration within virtual grid-organisation. Proc. Int. Moscow Conf. on Comput. Mol. Biol. (21–24 July, 2011, Moscow, Russia), НИИ Физико-химической биологии им. А. Н. Белозерского МГУ; p. 297–298.
  20. **Pydiura N.A.**, Karpov P.A., Blume Ya.B. Modeling of the regulation of microtubule dynamic instability by row of associated proteins. 2nd IRB Barcelona PhD Student Symposium: Life in Motion: Dynamics of Molecules and Systems (Barcelona, Spain, November 17–18, 2011) Inst. Res. Biomedicine, Barcelona, p. 82–83.
  21. **Пидюра М.О.**, Карпов П.А. Біоінформаційне дослідження білків, асоційованих з мікротрубочками вищих рослин та аналіз їх подібності. Збірник тез: Перша конф-ція мол. учених «Біологія рослин та біотехнологія»

(5–7 жовтня 2011, Біла Церква), 2011, Київ: Ін-т харчової біотехнології та геноміки, с. 69.