

ННЦ «Інститут біології та медицини»
Кафедра загальної та медичної генетики



Мобільні генетичні елементи

Ярослав Ясінський,
Олександра Проценко,
Надія Морозова



Київ – 2026

УДК 577.21:575.21(072.8)

ББК 28.04я73

П78

*Рекомендовано Вченою радою ННЦ «Інститут біології та медицини»
Київського національного університету імені Тараса Шевченка (протокол №
7 від «24» березня 2026 року)*

Автори:

Ясінський Я.С., асистент кафедри загальної та медичної генетики

Проценко О.В., к.б.н. асистент кафедри загальної та медичної генетики

Морозова Н.М., к.б.н. асистент кафедри загальної та медичної генетики

Рецензенти:

Костюк О.С. к.б.н., доцент кафедри біохімії ННЦ «Інститут біології та медицини» Київського національного університету імені Тараса Шевченка.

Світін Р.С. к.б.н. асистент кафедри молекулярної біотехнології та біоінформатики Навчально-науковий інститут високих технологій Київського національного університету імені Тараса Шевченка.

“Мобільні генетичні елементи” методичні рекомендації до практичних занять/ Ясінський Я.С., Проценко О.В., Морозова Н.М. Київ, 2026. – 103 с.

Методичні рекомендації підготовлено для проведення лабораторних і практичних занять із курсу «Мобільні генетичні елементи». Матеріали спрямовані на формування у студентів практичних навичок застосування молекулярно-генетичних методів і біоінформатичних підходів, а також на закріплення теоретичних знань, отриманих під час вивчення лекційного курсу. У виданні наведено основні теоретичні відомості про мобільні генетичні елементи (МГЕ) як важливе джерело геномної мінливості. Особливу увагу приділено механізмам активності та функціонування МГЕ у геномах різних організмів, а також їхній ролі у регуляції геномної організації та взаємозв'язку з біологічними процесами в організмі хазяїна. Методичні рекомендації призначені для студентів ННЦ «Інститут біології та медицини» Київського національного університету імені Тараса Шевченка, які навчаються за освітньою програмою «Біологія», освітнього рівня «Магістр».

Зміст

Лабораторна робота 1: Визначення Р-М статусу природних популяцій <i>Drosophila melanogaster</i>	4
Лабораторна робота 2: Молекулярно-генетичний аналіз розповсюдження Р-елемента в природних популяціях <i>Drosophila melanogaster</i>	11
Практична робота 3. Ідентифікація транспозонів у геномі <i>Drosophila melanogaster</i>	20
Практична робота 4: Виявлення профагів у бактеріальному геномі (інтегровані віруси)	31
Практична робота 5: Визначення бактеріальних плазмід за допомогою біоінформатичних інструментів	40
Практична робота 6: Структурний аналіз LTR-ретротранспозона ONSEN (ATCOPIA78) у геномі <i>Arabidopsis thaliana</i>	53
Практична робота 7: Аналіз поліморфних ретротранспозонних вставок (Alu, LINE-1, SVA) у геномі <i>Homo sapiens</i> за допомогою dbRIP та UCSC Genome Browser	69
Практична робота 8: Біоінформатичний підбір праймерів для детекції Р-елементу у <i>Drosophila melanogaster</i>	84

Лабораторна робота 1: Визначення Р-М статусу природних популяцій ***Drosophila melanogaster***

Мета роботи

Визначити приналежність досліджуваної природної популяції до Р, М або Q типу шляхом проведення реципрокних схрещувань та аналізу фертильності гібридів F1.

Теоретична основа

Р-елемент – це ДНК-транспозон (мобільний генетичний елемент), який кодує фермент транспозазу. Цей фермент дозволяє елементу вирізатися з одного місця геному і вбудовуватися в інше (переміщуватися в геномі).

Структура: Повний Р-елемент має довжину близько 2,9 т.п.н. і містить чотири екзони.

Транскрипція: У клітинах тіла (соматичних) зчитується лише частина генів, утворюючи репресор, який блокує переміщення.

Сплайсинг у зародковій лінії: Тільки в зародкових клітинах (яйцеклітинах та сперматозоїдах) відбувається специфічний сплайсинг останнього інтрона, що дозволяє синтезувати активну транспозазу.

Характеристика Р-елемента

Р-елемент потрапив у геном *Drosophila melanogaster* приблизно в середині XX століття шляхом горизонтального перенесення від іншого виду (*Drosophila willistoni*).

У літературі показано, що цей мобільний елемент може викликати явище гібридного дизгенезу. Гібридний дизгенез виникає, коли самці з Р-елементами (Р-лінія) схрещуються з самицями, що їх не мають (М-лінія). У таких гібридів у зародкових клітинах активується транспозиція Р-елементів, що призводить до **атрофії гонад (стерильності)**, мутацій та хромосомних перебудов.

Поняття цитотипу (Р, М або Q)

Статус мухи визначається не лише наявністю Р-елемента в хромосомах, а й станом цитоплазми яйцеклітини.

Р-цитотип (Захищений): Самиці Р-ліній мають у цитоплазмі яєць велику кількість молекул-репресорів (PIWI-білки та piRNA). Вони миттєво блокують будь-яку активність Р-елементів, що потрапляють від батька. Мухи розвиваються нормально.

М-цитотип (Сприйнятливий): Самиці М-ліній не мають у цитоплазмі специфічних репресорів. Якщо така самиця зустрінеється з Р-самцем, його Р-елементи в зародкових клітинах нащадків починають неконтрольовано транспонуватися.

Окремо слід відзначити мух з Q-цитотипом.

Q-цитотип (Neutral) – це особливий стан популяції дрозофіл, який за своєю сутністю є нейтральним або нейтральним варіантом Р-системи. Якщо Р-тип – має повний набір Р-елементів та захисні механізми, а М-тип – не має Р-елементів та захисних механізмів, то Q-тип – це популяція, яка вміє захищатися, але не вміє нападати.

Особливість Q-мух полягає в тому, що вони мають у своєму геномі Р-елементи, але ці елементи зазвичай дефектні або мають низьку активність транспозиції.

Захисний цитотип (Р-цитотип): Самиці Q-типу мають у цитоплазмі яєць репресори (piRNA), які блокують будь-яку активність Р-елементів.

Відсутність дизгеного потенціалу: Самці Q-типу не викликають масової стерильності (дизгенезу) при схрещуванні з незахищеними М-самицями.

Механізм дизгенезу в схрещуваннях.

Дизгенез – це синдром генетичних порушень, який виникає тільки при певній комбінації батьків.

А. Дизгенне схрещування: ♀ М х ♂ Р

Сперматозоїд приносить Р-елементи в незахищену цитоплазму М-яйцеклітини.

У зародкових клітинах гібрида F1 активується транспозаза.

P-елементи починають масово вбудовуватися в нові ділянки ДНК, розриваючи хромосоми.

Наслідки: Масові мутації, стерильність (через загибель клітин-попередників гонад), розриви хромосом.

Б. Реципрокне (нормальне) схрещування: ♀ P x ♂ M

Самиця передає нащадкам P-цитотип (захисні фактори в цитоплазмі). Навіть якщо в геномі є P-елементи, вони перебувають у неактивному стані. Нащадки фертильні та здорові.

Матеріали та обладнання

- Досліджувана лінія (W): Мухи, зібрані в природі (Wild type).
- Тестерна P-лінія: Наприклад, Harwich (має сильний P-фактор).
- Тестерна M-лінія: Наприклад, Canton-S або Oregon-R (не мають P-елементів).
- Ефіризатори, мікроскопи, термостат на 25°C та 29°C.

Хід роботи

Етап I: Схрещування (Схема перевірки)

Для визначення статусу природної популяції (W) необхідно закласти два типи схрещувань у двох повторностях (25°C та 29°C, де дизгенез проявляється найсильніше) (табл. 1).

Таблиця 1

Типи дизгенних схрещувань

Тест	Схрещування (Самиця x Самець)	Мета
Тест А	Самиця (W) x Самець (P-лінія)	Перевірка цитотипу самиць (чи здатна їх цитоплазма пригнічувати P-елемент).

Тест	Схрещування (Самиця x Самець)	Мета
Тест Б	Самиця (М-лінія) x Самець (W)	Перевірка наявності активних Р-елементів у самців природної популяції.

Етап II: Отримання гібридів F1

1. Відібрати віргінних самиць з відповідних ліній.
2. Закласти культури та витримувати їх при 29°C (для максимізації прояву дизгенезу).
3. Після вильоту імаго F1 відібрати молодих мух для аналізу гонад. Аналіз проводиться не раніше ніж на 3-тю добу після вильоту, щоб бути певними, що гонади встигли сформуватися, але не пізніше 7-ї доби, щоб уникнути ситуації, коли оварії порожні через вік мухи..

Етап III: Аналіз атрофії гонад

Методика розтину самиць F1

1. Підготовка: Помістіть самицю на предметне скло у краплю фізіологічного розчину (0,7% NaCl).
2. Фіксація: Однією голкою притисніть муху в ділянці грудей (thorax).
3. Екстракція: Іншою голкою (або тонким пінцетом) обережно потягніть за кінчик черевця (анальні пластинки).
4. Оцінка: Витягнуті нутроці мають містити два яєчники (рис. 1).
 - Норма: Великі, овальні, з чіткими ооцитами (у вигляді структур з численними ооцитами).
 - Дизгенез: Маленькі, прозорі редуковані структури або майже повна їх відсутність. Гонади вважаються атрофованими, якщо в них немає жодного зрілого яйця.

5. Підрахувати відсоток дизгенних самиць (GD – gonadal dysgenesis) (рис. 1)

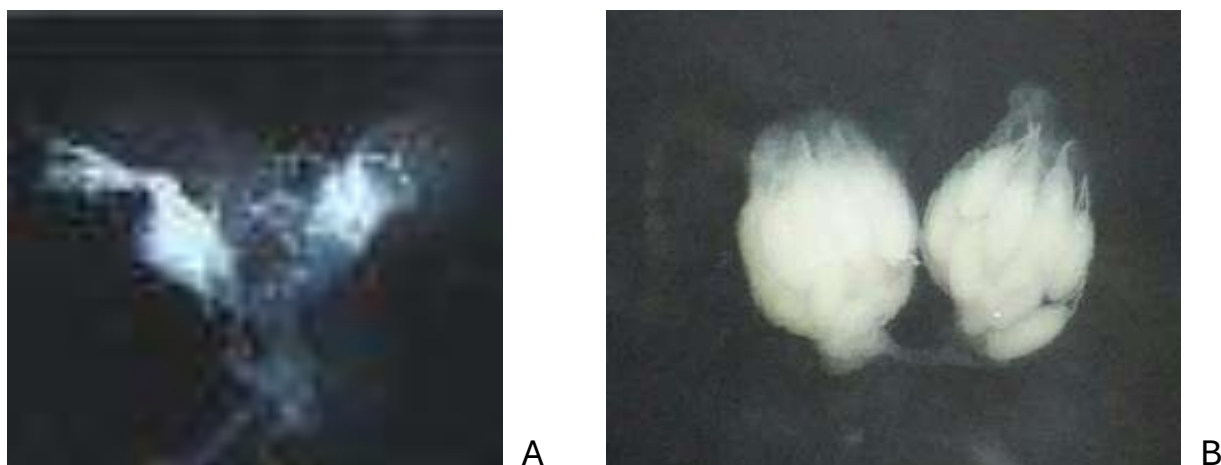


Рисунок 1. Атрофовані оварії (А), нормальні оварії (В).

Інтерпретація результатів

На основі відсотка стерильності в тестах А та Б робиться висновок про тип популяції:

- **Р-тип:** Низька стерильність у Тесті А (мають захисний цитотип) та висока стерильність у Тесті Б (мають активні Р-елементи).
- **М-тип:** Висока стерильність у Тесті А (немає захисту) та нульова у Тесті Б (немає Р-елементів).
- **Q-тип:** Низька стерильність в обох тестах (мають Р-елементи, які не викликають стерильності, але мають захисний цитотип).

Оформлення звіту

1. Заповніть таблицю з результатами підрахунків гонад.
2. Розрахуйте коефіцієнт дизгенезу для кожного варіанту.
3. Зробіть висновок: до якого типу належить досліджувана вами "дика" популяція та як це корелює з сучасним поширенням Р-елемента у світі.

Статистична обробка даних

Для кожної повторності (схрещування А та Б) необхідно проаналізувати не менше **50 самиць**.

Результати аналізу гонадного дизгенезу

Схрещування (♀x♂)	Самиці з обома нормальними яєчниками	Самиці з 1 атрофованим яєчником	Самиці з 2 атрофованими яєчниками	% Гібридного дизгенезу (GD)
Тест А: W x P				
Тест Б: M x W				

Формула розрахунку відсотка дизгенезу:

$$\%GD = (N1 \times 0.5 + N2) / N \times 100\%$$

де N1 – кількість самиць з одним атрофованим яєчником, N2 – з двома.

Якщо ваші результати вказують на проміжний тип (наприклад, 15% стерильності), варто перевірити, чи це не випадкове відхилення від очікуваного М-типу (100% дизгенезу в тесті А) або Р-типу (0% дизгенезу в тесті А).

Використовуємо критерій Пірсона:

$$\chi^2 = \sum((O - E)^2 / E)$$

де О – отримане значення, Е – очікуване значення.

Важливо: Оскільки Р-елемент зараз присутній майже у всіх природних популяціях світу (крім дуже старих лабораторних ліній), найімовірніше, ви отримаєте результат, характерний для **Р-типу** або **Q-типу** (слабкі Р-елементи, що не викликають дизгенезу, але захищають від нього).

Контрольні запитання

1. Чому дизгенез проявляється лише у зародкових клітинах (germline)?
2. Чим відрізняються реципрокні схрещування (♀M x ♂P та ♀P x ♂M)?
3. Чому експеримент проводять саме при 29°C?
4. Як ви відрізните Q-популяцію від Р-популяції за результатами тестів?

5. Що таке «горизонтальне перенесення» у контексті Р-елемента?

Рекомендована література

1. Kidwell, M. G., Kidwell, J. F., & Sved, J. A. (1977). Hybrid dysgenesis in *Drosophila melanogaster*: A syndrome of aberrant traits including mutation, sterility and male recombination. *Genetics*, 86(4), 813–833.
2. Engels, W. R. (1983). The P family of transposable elements in *Drosophila*. *Annual Review of Genetics*, 17(1), 315–344.
3. Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., & Hannon, G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*, 128(6), 1089–1103.
4. Kofler, R., Hill, T., Nolte, V., Betancourt, A. J., & Schlötterer, C. (2015). The invasion of *Drosophila melanogaster* by the P-елемент: A whole-genome sequence study of geographic populations. *Molecular Biology and Evolution*, 32(2), 458–469.
5. Mérel, V., Boulesteix, M., Fablet, M., Vieira, C., & Biémont, C. (2020). Transposable elements in *Drosophila*. *Mobile DNA*, 11, 23. <https://doi.org/10.1186/s13100-020-00213-z>
6. Kozeretska, I. A., Shulha, V. I., Serga, S. V., Rozhok, A. I., Protsenko, O. V., & Lau, N. C. (2018). A rapid change in P-елемент-induced hybrid dysgenesis status in Ukrainian populations of *Drosophila melanogaster*. *Biology Letters*, 14(8), 20180184. <https://doi.org/10.1098/rsbl.2018.0184>

Лабораторна робота 2: Молекулярно-генетичний аналіз розповсюдження Р-елемента в природних популяціях *Drosophila melanogaster*

Мета роботи

Ознайомитися з методами детекції мобільних генетичних елементів (МГЕ), опанувати техніку виділення ДНК, проведення ПЛР та принципи цитогенетичного картування за допомогою FISH.

Теоретична основа

Гібридний дизгенез

Р-елемент тісно пов'язаний із явищем гібридного дизгенезу. Воно виникає під час схрещування самиці М-лінії з самцем Р-лінії. У такому випадку батьківські Р-елементи потрапляють у цитоплазму яйцеклітини, яка не містить материнських рiRNA, здатних їх пригнічувати. У нащадків активується транспозиція Р-елементів у зародкових клітинах, що призводить до атрофії гонад, стерильності, мутацій і хромосомних перебудов. За реципрокного схрещування (самка Р х самець М) дизгенез, як правило, не виникає, бо самка передає нащадкам фактори захисту.

ПЛР є швидким і чутливим методом молекулярного скринінгу, який дозволяє встановити, чи присутня в геномі досліджуваної популяції послідовність Р-елемента. Для ампліфікації зазвичай обирають консервативні ділянки в районі ехон 0 / ехон 1 або в межах гена транспозази, оскільки такі ділянки краще підходять для специфічного виявлення як повних, так і багатьох дефектних копій Р-елемента.

FISH (флуоресцентна *in situ* гібридизація) – це цитогенетичний метод, за якого флуоресцентно мічений зонд зв'язується з комплементарною послідовністю ДНК безпосередньо на хромосомах. У дрозофіли для цього часто використовують політенні хромосоми слинних залоз личинок III віку, оскільки вони дуже великі, мають чіткий смугастий малюнок і зручні для картування. На відміну від ПЛР, FISH дозволяє побачити де саме локалізований Р-елемент

у хромосомах і скільки приблизно копій присутні в геномі за кількістю сигналів. Тому FISH є інформативнішим методом для дослідження розповсюдження мобільних елементів у популяціях.

Інвазія Р-елемента стала прикладом гострого конфлікту між геномом господаря та мобільним елементом. На початкових етапах така інвазія підвищує рівень мутацій і знижує репродуктивну успішність, але з часом у популяціях формується захист, зокрема piRNA-опосередковане пригнічення транспозиції. Отже, Р-елемент не лише викликає генетичну нестабільність, а й виступає важливим фактором еволюції геномів, їх адаптації та перебудови систем геномного захисту.

Частина 1. Молекулярно-генетична детекція Р-елемента

Матеріали та обладнання

- Біологічний матеріал:
 - мухи *D. melanogaster*, відловлені в різних географічних точках.
Контрольні лінії: Canton-S (М-лінія, без Р-елемента) та Harwich (Р-лінія, у геномі якої наявні повнорозмірні копії Р-елемента).
- Для виділення ДНК
 - *Обладнання*: Мікроцентрифуга. Вортекс. Термостат (для інкубації з протеїназою К). Мікропіпетки
 - *Реактиви та витратні матеріали*: Буфери для лізису, Протеїназа К (фермент для розщеплення білків), Ізопропанол або етанол (для осадження ДНК), Стерильні мікропробірки типу «Епендорф» (1.5 мл), Пластикові пестики (гомогенізатори) для розтирання мух у пробірці.
- Для проведення ПЛР-аналізу
 - *Обладнання*: ампліфікатор, бокс для ПЛР з УФ-лампкою для стерильної роботи.

- Реактиви (ПЛР-суміш): ДНК-матриця (виділена на 1-му етапі). Специфічні праймери, Таq-полімераза, Суміш дезоксинуклеозидтрифосфатів (dNTPs), ПЛР-буфер з іонами Mg^{2+} .
- Для проведення електрофорезу
 - *Обладнання:* Камера для горизонтального електрофорезу, джерело постійного струму, транслюмінатор (УФ-випромінювач) або система гель-документації для перегляду результатів.
 - *Реактиви:* Агароза, електродний буфер (наприклад, TAE або TBE), флуоресцентний барвник (наприклад, бромистий етидій або SafeView) для фарбування ДНК, маркер молекулярної маси (DNA ladder), буфер для нанесення проб (фарба-маркер, щоб бачити рух фронту).

Хід роботи

Етап I: Виділення геномної ДНК

1. Відібрати по 5-10 особин імаго з кожної дослідної популяції.
2. Провести гомогенізацію тканин у лізуючому буфері.
3. Очистити ДНК методом фенол-хлороформної екстракції або за допомогою сорбційних колонок.

Етап II: ПЛР-діагностика

Для ампліфікації використовуємо специфічні праймери.

Найчастіше обирають консервативну ділянку в районі Exon 0 або Exon 1, що дозволяє виявити як повні, так і багато дефектних копій Р-елемента.

- P-Forward (257-276): 5'-CGG-TGA-AGT-CCC-AGT-CTC-GC-3'
- P-Reverse (698-717): 5'-GGT-GGT-CCC-GTT-CGA-TGT-TT-3'

Очікуваний розмір продукту: 461 п.н. (пари нуклеотидів).

Розрахунок реакційної суміші (на 1 зразок)

Загальний об'єм реакції – **25 мкл.**

Компонент	Концентрація (Stock)	Робочий об'єм	Кінцева конц.
ddH	-	17.3 мкл	
ПЛР-буфер	10	2.5 мкл	1
MgCl ₂	50 mM	0.75 мкл	1.5 mM
dNTPs	10 mM кожен	0.5 мкл	200 мкМ
Forward Primer	10 мкМ	0.5 мкл	0.2 мкМ
Reverse Primer	10 мкМ	0.5 мкл	0.2 мкМ
Taq-полімераза	5 од/мкл	0.2 мкл	1 од
ДНК (матриця)	50 нг/мкл	2.7 мкл	135 нг

Протокол ампліфікації (Термоциклювання)

Цей режим оптимізований для запобігання утворенню праймер-димерів:

1. Гарячий старт: 95°C - 5 хв.
2. Циклічна частина (30-35 циклів):
 - Денатурація: 95°C - 30 сек.
 - Відпал (Annealing): 58°C - 30 сек.
 - Елонгація: 72°C - 45 сек.
3. Фінальна елонгація: 72°C - 7 хв.
4. Зберігання: 4°C .

Етап III: Аналіз результатів ПЛР (Електрофорез)

Для перевірки результатів використовуємо 1.5% агарозний гель з додаванням бромистого етидію або SYBR Safe.

- Маркер (Ladder): 100 bp DNA Ladder.

- Очікувана картина: У Р-ліній (дикі популяції та лінія Harwich) - чітка смуга на рівні 461 п.н.
 - У М-ліній (лінія Canton-S) – смуги відсутні.

Частина 2. Цитогенетичне картування Р-елемента. Гібридизація in situ (FISH)

Цей етап дозволяє побачити, де саме в хромосомах знаходяться мобільні елементи.

Приготування препаратів політених хромосом із слинних залоз личинок III віку.

Матеріали та обладнання:

- Для отримання препаратів
 - Біологічний матеріал: Личинки дрозоділи 3-го віку.
 - Обладнання та реактиви: препарувальні голки, стереомікроскоп, предметні та покривні скельця, фізіологічний розчин, фіксатор (суміш етанолу та льодяної оцтової кислоти у співвідношенні 3:1), 45% оцтова кислота, рідкий азот, етиловий спирт 96%.
- Для проведення FISH гібридизації
 - ДНК-зонд мічений флуорохромом (наприклад, FITC - зелений або Cy3 - червоний), термошейкер, водяна баня, флуоресцентний мікроскоп.
 - 45% оцтова кислота; фіксатор (суміш 1/2/3 молочна кислота/вода/льодовикова оцтова кислота); етанол: 30%, 50%, 70%, 95%; 2х фізіологічний розчин-цитрат натрію (SSC); 0,1 N NaOH; розчин для гібридизації: 1х розчин Денхардта, 50 мМ фосфатний буфер натрію - рН 6,8, 0,6 М NaCl, 5 мМ MgCl₂; фосфатно-буферний фізіологічний розчин (PBS) 1х робочий розчин (8 г/л NaCl, 0,2 г/л KCl, 1,44 г/л Na₂HPO₄, 0,24 г/л KH₂PO₄); PBS/BSA (0,5% BSA); зонд, мічений дигоксигеніном; анти-дигоксигенін-

флуоресцеїн, фрагменти Fab' (Boehringer Mannheim); середовище для заливання (80 % гліцерин у PBS, 1 мг/мл парафенілендіаміну, 1 мкг/мл йодиду пропідію (PI).

Хід роботи

Приготування препарату

1. *Препарування:* Помістити личинку в краплю 0.7% розчину NaCl. За допомогою ентомологічних голок відокремити голову від тіла - слинні залози (прозорі парні органи) вийдуть разом із жировим тілом.
2. *Фіксація:* Залози перенести у свіжий фіксатор (суміш метанолу та крижаної оцтової кислоти 3:1) на 1-2 хвилини.
3. *Мацерація:* Перенести залози на предметне скло в краплю 45% оцтової кислоти. Це розм'якшує тканини для кращого розпластування хромосом.
4. *Виготовлення давленого препарату:* Накрити покривним склом і сильно натиснути великим пальцем через фільтрувальний папір.
5. *Заморожування:* Помістити скло в рідкий азот, після чого швидко зняти покривне скло лезом. Це залишає хромосоми прикріпленими до предметного скла.
6. Занурити препарати в ємність з етиловим спиртом 96%, для фіксації на 5 хв.
7. Відберіть якісні препарати для подальшої роботи.

Денатурація ДНК на склі.

1. Попередня обробка та денатурація політенових хромосом для гібридизації
2. Перед проведенням флуоресцентної гібридизації *in situ*, хромосомну ДНК слід денатурувати, щоб мічений зонд міг знайти цільову послідовність. Скельця переносимо в послідовні розчини етанолу з концентрацією: 95%, 70%, 50%, 30% (по 2 хвилини у кожному).
3. Для дегідратації політенових хромосом, промити в 2 х SSC протягом 2 хвилин.
4. Інкубувати в 2 х SSC протягом 45 хвилин при 70°C.

5. Денатурацію політенових хромосом проводимо шляхом інкубації скельць у 0,1 N NaOH (свіжоприготованому) протягом 2 хвилин.
6. Після промиваємо в 2 x SSC протягом 2 хвилин
7. Політенові хромосоми дегідратуємо шляхом занурення скельць у послідовні розчини зі зростаючою концентрацією етанолу: 30%, 50%, 70% та 95%, протягом 2 хвилин кожен.
8. Потім скельця сушимо на повітрі.

Гібридизація з флуоресцентно-міченим зондом, що комплементарний Р-елементу.

1. Зонд (0,8 мкл) змішуємо з 9,2 мкл гібридизаційного буфера, кип'ятимо протягом 5 хвилин на водяній бані та негайно охолоджуємо льодом.
2. Всі 10 мкл суміші зонд-гібридизація нанести на політенні хромосоми, зразки покрити несиліконізованими покривними скельцями розміром 18 мм x 18 мм, не утворюючи бульбашки повітря, та інкубувати у вологій камері протягом ночі при 65°C.
3. Потім покривні скельця змити 2 x SSC протягом 5 хвилин при кімнатній температурі.
4. Неспецифічно зв'язані молекули зонда видалити шляхом триразового промивання слайдів у 2 x SSC при 53°C, по 20 хвилин кожен раз.
5. Після промивання в 1 x PBS протягом 5 хвилин, слайди висушити на повітрі (але не повністю, оскільки це створює фон).
6. 5 мкл розчину антитіл проти диоксигеніну-флуоресцеїну, Fab'-фрагментів (35 мкг/мл у PBS, 0,5% BSA) наносимо на кожну ділянку політенових хромосом.
7. Препарати покрити несиліконізованими покривними скельцями розміром 20x20 мм та інкубували у вологій камері протягом 45 хвилин при 37°C.

Промивка та мікроскопія.

1. Покривні скельця зняти з поверхні в 1x PBS, потім предметні скельця промили 2 рази по 5 хвилин в 1x PBS при кімнатній температурі та висушили на повітрі, але не до кінця.

2. Препарати політенових хромосом фіксуємо в гліцериновому середовищі (80% гліцерин, 1 мг/мл п-фенілендіаміну, 1 мкг/мл йодиду пропідію)
3. Аналізуємо за допомогою флуоресцентного мікроскопа.

Оформлення результатів:

Заповніть таблицю за результатами дослідження:

Популяція (локація)	Результат ПЛР (+/-)	Кількість сигналів FISH (копійність)	Тип лінії (Р / М / М')
Контроль (Harwich)	+	Багато (30-50)	Р-лінія
Контроль (Canton-S)	-	0	М-лінія
Популяція №1	?	?	?

Контрольні запитання

1. Чим відрізняються Р-лінії від М-ліній на молекулярному рівні?
2. Чому FISH є більш інформативним методом для вивчення МГЕ, ніж звичайна ПЛР?
3. Які еволюційні наслідки має "вторгнення" Р-елемента в геном дрозофіли?
4. Методологія ПЛР: Чому для детекції Р-елемента ми використовуємо саме ген транспозази, а не термінальні інвертовані повтори (ITRs)?
5. Інтерпретація FISH: Якщо на препараті FISH ми бачимо багато сигналів, а ПЛР дає негативний результат – про що це може свідчити? *(Підказка: можлива мутація в місцях відпалу праймерів або технічна помилка).*

Рекомендована література

1. Pardue, M.-L. (2011). In Situ Hybridization to Polytene Chromosomes in *Drosophila* Using Digoxigenin-Labeled Probes. Cold Spring Harbor Protocols, 2011(8), pdb.prot5655. <https://doi.org/10.1101/pdb.prot5655>
2. Gavrila, L., Ecovoiu, Al. Al., & Georgescu, L. M. (2001). Localizing genes in *Drosophila melanogaster* polytene chromosomes by fluorescence in situ

hybridization. *Journal of Cellular and Molecular Medicine*, 5(1), 74–78. Portico.
<https://doi.org/10.1111/j.1582-4934.2001.tb00139.x>

3. Sambrook, J., & Russell, D. W. *Molecular Cloning: A Laboratory Manual*.
(Протоколи виділення ДНК та ПЛР).

Практична робота 3. Ідентифікація транспозонів у геномі *Drosophila melanogaster*

Мета роботи

Ознайомитися з транспозиційними елементами *Drosophila melanogaster* та навчитися виявляти копії відомого транспозону у геномі за допомогою біоінформатичних інструментів. Навчитися працювати з базами даних нуклеотидних послідовностей, використовувати BLASTn для пошуку мобільних елементів у геномі, інтерпретувати параметри вирівнювання та оцінювати кількість, повноту й розподіл копій транспозону.

Теоретичні відомості

Мобільні генетичні елементи – це ділянки ДНК, здатні переміщуватися в межах геному або між геномами. До них належать транспозони, ретротранспозони, деякі вірусні та плазмідні послідовності. Значна частка повторюваної ДНК у геномах еукаріот представлена саме транспозонами. У геномі *Drosophila melanogaster* транспозони становлять істотну частину повторюваного матеріалу, у людини повтори загалом займають майже половину геному. Такі елементи часто називають «егоїстичними», оскільки вони здатні до самопоширення навіть без очевидної користі для організму-хазяїна.

За механізмом переміщення транспозони поділяють на два основні класи. **DNA-транспозони** переміщуються на рівні ДНК і реалізують механізм «вирізати й вставити» за участі транспозази. **Ретротранспозони** спочатку транскрибуються в РНК, після чого копія знову перетворюється на ДНК і вставляється в новий сайт геному. Такий механізм називають «копіювати й вставити», тому саме ретротранспозони часто швидко збільшують число своїх копій. Серед ретротранспозонів у дрозофіли найпоширенішими є LTR- та LINE-подібні елементи; серед DNA-транспозонів важливе місце посідають елементи з кінцевими інвертованими повторами.

Drosophila melanogaster є класичним об'єктом генетики, у геномі якого описано багато родин транспозонів. Кількість копій окремих родин може змінюватися від поодиноких вставок до десятків і сотень. Розподіл транспозонів у геномі нерівномірний: частина елементів частіше накопичується в гетерохроматинових і прицентромерних ділянках, тоді як інші трапляються також у межах генів або поруч із ними. Через це транспозони можуть бути як зручними маркерами еволюційних подій, так і джерелом мутацій, що впливають на функцію генів.

Одним із найвідоміших транспозонів дрозофіли є *P*-елемент. Це DNA-транспозон довжиною близько 2,9 тис. пар основ із кінцевими інвертованими повторами, який кодує транспозазу.

Відкриття «стрибаючих генів» пов'язане з роботами Барбари Маклінток на кукурудзі. Її спостереження показали, що геном не є статичною структурою: окремі ділянки ДНК можуть змінювати своє положення, впливаючи на експресію генів і фенотип. Надалі транспозони були виявлені в усіх великих групах організмів, а їх вивчення стало важливою частиною молекулярної генетики, геноміки та еволюційної біології.

Для виявлення копій транспозону в геномі застосовують алгоритми пошуку послідовностей на основі локального вирівнювання. Найпоширенішим інструментом є BLAST – *Basic Local Alignment Search Tool*. У цій роботі використовується режим BLASTn, який порівнює нуклеотидну послідовність-запит із нуклеотидною базою даних і знаходить ділянки локальної подібності. У результатах особливе значення мають показники *Query Cover*, *Per. Ident*, довжина вирівнювання, наявність розривів і координати збігу на послідовності-мішені.

Окрім BLAST, для підготовки до аналізу та перевірки назви родини транспозону корисно використовувати спеціалізовані ресурси. Набір канонічних послідовностей транспозонів *D. melanogaster* доступний у репозиторії Bergman Lab, а база Dfam містить консенсусні послідовності повторів і профілі для їх розпізнавання. FlyBase та браузері геному дають

змогу інтерпретувати знайдені вставки в геномному контексті, тобто визначати, чи лежить копія в міжгенній ділянці, інтроні, екзоні або поблизу регуляторних елементів.

Матеріали та методи

- комп'ютер із доступом до Інтернету;
- веб-сервіс NCBI BLAST: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>;
- база нуклеотидних послідовностей NCBI Nucleotide: <https://www.ncbi.nlm.nih.gov/nucleotide/>;
- набір канонічних послідовностей транспозонів *D. melanogaster*: https://github.com/bergmanlab/drosophila-transposons/blob/master/current/D_mel_transposon_sequence_set.fa;
- база повторів Dfam: <https://dfam.org/>;
- ресурс FlyBase: <https://flybase.org/>;
- браузер FlyBase JBrowse: <https://flybase.org/jbrowse/>;
- альтернативний браузер UCSC Genome Browser для збірки dm6: <https://genome.ucsc.edu/cgi-bin/hgGateway?db=dm6>;
- нуклеотидна послідовність одного транспозиційного елемента з додатка 1;
- референтний геном *Drosophila melanogaster*.

Завдання

1. обрати транспозиційний елемент дрозофіли з додатка 1;
2. отримати нуклеотидну послідовність вибраного елемента у форматі FASTA;
3. виконати пошук цієї послідовності по геному *D. melanogaster* за допомогою BLASTn;
4. визначити кількість копій транспозону за результатами пошуку;

5. проаналізувати відсоток тотожності, довжину вирівнювання та повноту знайдених копій;
6. визначити хромосомну локалізацію знайдених копій;
7. оцінити частку геному, що припадає на вибраний мобільний елемент, і сформулювати висновки.

Хід роботи

Фіксація об'єкта аналізу

1. Оберіть один мобільний елемент із Додатка 1.
2. Занесіть до робочого протоколу такі атрибути обраного елемента: робоча назва, клас транспозону, очікувана канонічна довжина, джерело еталонної послідовності, дата виконання аналізу.
3. Створіть робочу таблицю результатів із такими полями: № локусу; назва елемента; клас; довжина запиту; ідентифікатор послідовності-мішені; хромосома або плече хромосоми; початок; кінець; орієнтація; сумарна довжина вирівнювання; сумарне покриття запиту; середньозважена тотожність; E-value; категорія повноти; геномний контекст.

Отримання еталонної нуклеотидної послідовності

1. Використайте канонічний набір транспозонів *Drosophila melanogaster* з репозиторію Bergman Lab як основне джерело еталонної послідовності.
2. Виберіть у FASTA-файлі запис, що відповідає обраній родині транспозону.
3. Скопіюйте повний FASTA-запис без змін.
4. Збережіть окремий робочий файл із одним FASTA-записом.
5. Зафіксуйте довжину послідовності запиту в парах основ.
6. Присвойте цій послідовності статус еталонний запит для BLASTn.

Верифікація родини транспозону

1. Зіставте назву родини обраного елемента з записом у Dfam.
2. Зафіксуйте у протоколі тип елемента: DNA-транспозон, LTR-ретротранспозон або non-LTR ретротранспозон.
3. Зафіксуйте структурні ознаки родини:

- наявність TIR для DNA-транспозонів;
 - наявність LTR для LTR-ретротранспозонів;
 - відсутність LTR і наявність LINE-архітектури для non-LTR ретротранспозонів.
4. Занесіть до протоколу канонічну назву родини без історичних синонімів і без назв окремих клонів.

Постановка пошуку BLASTn

1. Використайте режим BLASTn для порівняння нуклеотидного запиту з нуклеотидною базою даних.
2. Як запит задайте повний FASTA-запис обраного елемента.
3. Як організм-мішень задайте *Drosophila melanogaster* (taxid: 7227).
4. Як простір пошуку задайте референтний геном *Drosophila melanogaster*.
5. Як режим пошуку задайте megablast, оскільки завдання полягає у виявленні відомих копій елемента в межах одного виду.
6. Встановіть таку кількість цільових послідовностей, щоб у результатах відобразився повний набір вирівнювань для всіх знайдених локусів.
7. Зафіксуйте у протоколі дату запуску, назву запиту, довжину запиту та параметри пошуку.

Первинне вилучення релевантних збігів

1. Після завершення пошуку сформууйте список усіх вирівнювань, що відповідають обраному елементу.
2. Відкиньте службові, векторні та нерелевантні мішені.
3. Для подальшого аналізу залиште лише ті збіги, що мають одночасно такі ознаки:
 - походять з геному *Drosophila melanogaster*;
 - мають чітко визначені координати на хромосомі або на зібраному геномному фрагменті;
 - мають довжину вирівнювання не менше 200 п.о.;
 - мають статистичну значущість не гіршу за E-value 1e-20.

4. Перенесіть до робочої таблиці всі збережені HSP із такими параметрами:

- Subject title,
- Query start,
- Query end,
- Sbjct start,
- Sbjct end,
- alignment length,
- identities, gaps,
- strand, E-value.

Об'єднання HSP у геномні локуси

1. Впорядкуйте всі HSP за трьома ключами: ідентифікатор мішені → орієнтація → координата початку на мішені.
2. Об'єднайте HSP в один локус за сукупністю таких правил:
 - HSP належать одній і тій самій послідовності-мішені;
 - HSP мають однакову орієнтацію;
 - геномні інтервали HSP перекриваються або розділені проміжком не більше 100 п.о.;
 - порядок HSP уздовж запиту зберігає колінеарність.
3. Вважайте результат такого об'єднання однією копією транспозону в одному локусі.
4. Не рахуйте окремі HSP як окремі копії.
5. Присвойте кожному локусу власний порядковий номер.

Розрахунок параметрів для кожного локусу

1. Для кожного локусу визначте:
 - хромосому або плече хромосоми;
 - мінімальну координату початку;
 - максимальну координату кінця;
 - довжину локусу на мішені;
 - орієнтацію.

2. Розрахуйте сумарне покриття запиту так:
 - об'єднайте всі інтервали на запиті, покриті HSP цього локусу;
 - підсумуйте довжини неперекривних інтервалів;
 - поділіть отриману суму на повну довжину запиту, помножте результат на 100.
3. Розрахуйте середньозважену тотожність локусу за формулою:

$$\Sigma (\text{довжина HSP} \times \% \text{ тотожності HSP}) / \Sigma (\text{довжина HSP})$$
4. Зафіксуйте найкраще E-value серед HSP, що входять до локусу.
5. Занесіть усі обчислені параметри до робочої таблиці.

Класифікація копій за повнотою

1. Присвойте кожному локусу одну з трьох категорій:
 - повнорозмірна копія – сумарне покриття запиту $\geq 90\%$;
 - частково збережена копія – сумарне покриття запиту від 60% до 89,99%;
 - фрагментована копія – сумарне покриття запиту $< 60\%$.
2. Окремо позначте локуси з великими внутрішніми делеціями, коли сумарне покриття зберігається на рівні повнорозмірної копії, але структура складається з кількох розділених блоків вирівнювання.
3. Підрахуйте кількість локусів у кожній категорії.

Визначення хромосомної локалізації

1. Для кожного локусу встановіть належність до однієї з геномних одиниць: X, 2L, 2R, 3L, 3R, 4, Y або до іншого зібраного референтного фрагмента.
2. Побудуйте зведення за хромосомами у вигляді переліку:
назва хромосоми / плеча – кількість локусів – частка від загальної кількості локусів.
3. Для кожного локусу запишіть координати у стандартизованому форматі:
хромосома: початок–кінець.
4. У підсумку окремо зазначте хромосому з максимальною кількістю копій.

Визначення геномного контексту вставок

1. Для кожного локусу використайте його точні координати у браузері геному FlyBase JBrowse.
2. Увімкніть анотації генів і визначте тип розташування локусу за жорстким правилом:
 - екзонна вставка – локус перекриває екзон;
 - інтронна вставка – локус розташований у межах гена, але не перекриває екзон;
 - пригенна вставка – локус розташований у межах 2000 п.о. від початку або кінця анотованого гена;
 - міжгенна вставка – локус не відповідає жодній з трьох попередніх категорій.
3. Занесіть категорію геномного контексту для кожного локусу до робочої таблиці.
4. Підрахуйте кількість локусів у кожній категорії контексту.

Оцінка частки геному, зайнятої вибраним елементом

1. Для всіх підтверджених локусів підсумуйте довжини їхніх неперекривних геномних інтервалів.
2. Отриману суму прийміть як сукупну довжину геному, зайняту копіями обраного елемента.
3. Обчисліть частку геному за формулою:

$$\text{Частка геному (\%)} = (\text{сукупна довжина локусів} / 140\,000\,000) \times 100$$

4. Запишіть окремо абсолютне значення в парах основ і відсоткове значення.

Формування висновку

1. вкажіть назву транспозону, його клас і довжину еталонного запиту.
2. загальну кількість локусів, виявлених у геномі.
3. подайте розподіл копій за повнотою:
 - кількість повнорозмірних,
 - частково збережених і
 - фрагментованих копій.

4. подайте узагальнену характеристику тотожності:
 - мінімальна,
 - максимальна та
 - середньозважена тотожність.
5. У п'ятому реченні подайте розподіл копій за хромосомами.
6. переважний геномний контекст вставок.
7. оцінку частки геному, зайнятої цим елементом.
8. Завершіть висновок короткою інтерпретацією стану родини в референтному геномі:
 - домінування повнорозмірних копій,
 - домінування внутрішньо делеційних копій або
 - домінування фрагментованих копій.

Контрольні запитання:

1. Що таке мобільні генетичні елементи і які їх основні типи? Наведіть приклади у еукаріот.
2. Яку частку геному *Drosophila melanogaster* складають транспозони? Які класи транспозонів переважають у дрозофіли та в людини?
3. У чому різниця між DNA-транспозонами і ретротранспозонами? Поясніть механізми “cut-and-paste” та “copy-and-paste” на прикладі.
4. Хто відкрив явище “стрибаючих генів” і яке біологічне значення цього відкриття?
5. Що собою являє Р-елемент дрозофіли і чому він відіграє важливу роль у генетиці (явище гібридної дисгенезії, застосування в генетичних експериментах)?
6. Який інструмент використовують для пошуку послідовностей транспозонів у геномі? Який алгоритм лежить в основі BLAST і як він працює?

7. Як, використовуючи результати BLAST, визначити кількість копій транспозону у геномі? На що звертати увагу при аналізі вирівнювань (параметри % ідентичності, довжина)?
8. Чому деякі копії транспозону можуть бути коротшими за довжину еталонної послідовності? Що означають фрагментовані копії і як вони могли утворитися?
9. Де у геномі зазвичай знаходяться ретротранспозони? Чому, наприклад, повнорозмірні соріа або gypsy часто зустрічаються поблизу центромер?
10. Практичне: Якщо BLAST показав 120 hits транспозону, а в графічному резюме видно, що на одній хромосомі згруповано багато влучань – як ви будете рахувати копії? На що звернете увагу, щоб не подвійно порахувати один і той же вставлений елемент?

Рекомендована література

1. O'Hare, K., & Rubin, G. M. (1983). Structures of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. *Cell*, 34 (1), 25-35.
<https://doi.org/10.1016/0092-8674%2883%2990133-2>
2. Marlor, R. L., Parkhurst, S. M., & Corces, V. G. (1986). The *Drosophila melanogaster* gypsy transposable element encodes putative gene products homologous to retroviral proteins. *Molecular and Cellular Biology*, 6(4), 1129-1134.
<https://doi.org/10.1128/MCB.6.4.1129-1134.1986>
3. Castro, J. P., & Carareto, C. M. A. (2004). *Drosophila melanogaster* P transposable elements: Mechanisms of transposition and regulation. *Genetica*, 121(2), 107-118.
<https://doi.org/10.1023/B:GENE.0000040382.48039.A2>
4. Rizzon, C., Marais, G., Gouy, M., & Biémont, C. (2002). Recombination rate and the distribution of transposable elements in the *Drosophila melanogaster* genome. *Genome Research*, 12(3), 400-407.

<https://doi.org/10.1101/gr.210802>

5. Hubley, R., Finn, R. D., Clements, J., Eddy, S. R., Jones, T. A., Bao, W., Smit, A. F. A., & Wheeler, T. J. (2016). The Dfam database of repetitive DNA families. *Nucleic Acids Research*, 44(D1), D81-D89. <https://doi.org/10.1093/nar/gkv1272>
6. McCullers, T. J., & Steiniger, M. (2017). Transposable elements in *Drosophila*. *Mobile Genetic Elements*, 7(3), 1-18.
<https://doi.org/10.1080/2159256X.2017.1318201>

Практична робота 4: Виявлення профагів у бактеріальному геномі (інтегровані віруси)

Мета роботи

Засвоїти методи пошуку мобільних елементів у прокаріотичних геномах на прикладі профагів – вірусів бактерій, інтегрованих у бактеріальну хромосому. Студенти навчатися використовувати онлайн-інструмент PHASTER для автоматичної ідентифікації фагових послідовностей у геномі бактерії, а також інтерпретувати отримані результати (кількість профагів, їхню повноту та можливі гени). Це дозволить зрозуміти роль вірусних елементів як чинника геномної варіабельності бактерій.

Теоретичні відомості

Профаги та стани лізогенії. Профагом називають геном помірного бактеріофага, який підтримується у клітині-хазяїні як стабільний спадковий елемент: найчастіше у вигляді інтегрованої ділянки бактеріальної хромосоми, але також можливе екстрахромосомне існування у плазмідному стані (phage-plasmids, plasmid prophage). Лізогенія є альтернативою негайному літичному циклу: фагова ДНК переходить у стан репресії (пригнічення експресії генів літичної програми), і клітина ділиться, передаючи профаг нащадкам. Індукція лізогенії (перехід у літичний цикл) може запускатися стресами, зокрема ушкодженням ДНК та активацією SOS-відповіді; для окремих патогенних систем це має прямий зв'язок із фенотипом вірулентності (наприклад, для Shiga-токсин-конвертуючих профагів).

Структурно-функціональна організація профагів.

Повнорозмірні (функціонально цілісні) профаги мають модульну архітектуру:

- інтеграція/екцизія (site-specific recombination; інтегрази/рекомбінази та att-сайти),

- реплікація та регуляція (репресор/антирепресор, транскрипційні регулятори),
- упаковка геному (терміназа, порталний білок),
- морфогенез віріона (головка/капсид, хвіст, базальна пластинка),
- вихід із клітини (holin-endolysin каскад; для грамнегативних бактерій також можливі додаткові компоненти руйнування оболонки).

Наявність цих модулів у повному наборі підвищує ймовірність того, що регіон є “інтактним” у сенсі збереження фагового геномного плану.

Інтактні, дефектні та “криптичні” профаги. Значна частина профагів у бактеріальних геномах є дефектними або деградованими: вони втратили критичні гени (наприклад, структурні або лізисні) і не формують інфекційних частинок, однак можуть зберігати окремі регуляторні або метаболічно важливі функції для хазяїна. Криптичні профаги інколи “одомашнюються” бактерією та впливають на стресостійкість, біоплівкоутворення або чутливість до антибіотиків; водночас носійство профагів має і потенційні витрати, передусім ризик загибелі клітини при індукції лізогенії та метаболічний тягар підтримання/експресії елементів.

Лізогенна конверсія та еволюційні наслідки. Профаги є важливим драйвером геномної мінливості бактерій через інтеграцію нових генетичних блоків і передачу генів між лініями. *Класичний механізм* – лізогенна конверсія, коли профаг приносить хазяїну фенотипно значущі гени (часто фактори вірулентності). Дифтерійний токсин визначається наявністю tox-гена, що переноситься помірними корінефагами; холерний токсин stxAB кодується лізогенним філаментозним фагом СТХФ у *Vibrio cholerae*; Shiga-токсини у STEC зазвичай пов’язані з lambdoid-профагами. Окрім конверсії, профаги беруть участь у горизонтальному переносі через спеціалізовану та генералізовану трансдукцію (помилкове захоплення бактеріальної ДНК під час екцизиї або пакування).

Принципи *in silico*-ідентифікації профагів. Комп'ютерне виявлення профагів у бактеріальному геномі спирається на поєднання ознак: збагачення регіону генами з фагоподібними гомологіями, наявність характерних модулів (інтеграза, структурні білки, лізис), кластеризація фагових CDS у безперервний інтервал та відповідність довжині/щільності генів, типовій для фагових геномів. Важливо трактувати прогноз як “кандидатний профаговий регіон”: у деградованих елементів межі можуть бути нечіткі, а наявність окремих фагових генів не завжди означає збереження повного фагового циклу.

PHASTER як інструмент пошуку профагів. PHASTER (PHAge Search Tool – Enhanced Release) – веб-сервіс для швидкої ідентифікації та анотації профагових регіонів у бактеріальних геномах і плазмідах; це розвиток PHAST із оптимізацією обчислювального конвеєра та інтерфейсу. Сервіс приймає повні геноми у форматах GenBank/FASTA або за accession-ідентифікатором з NCBI та формує зведення кандидатних регіонів із координатами, довжиною, кількістю CDS, оцінкою “повноти” та статусом. У PHASTER статус регіону визначається за сумарним балом: incomplete при score < 70, questionable при 70–90, intact при score > 90; ці пороги є операційними правилами сервісу та відображають ступінь насиченості регіону фаговими ознаками, а не експериментально підтверджену індукованість.

Вихідні компоненти результатів PHASTER. Типовий результат включає зведену таблицю профагових регіонів (координати, довжина, score, статус), візуалізацію розташування регіонів на геномі та детальні анотації генів усередині кожного регіону з функціональними позначеннями/гомологіями до фагових баз. Саме деталізація генів дозволяє перейти від констатації “є профаг” до інтерпретації його потенційної біологічної ролі: наявності інтегрази, структурних і лізисних модулів, а також “додаткових” генів, що можуть модифікувати фенотип хазяїна.

Матеріали та методи

- Комп'ютер з Інтернетом.

- Веб-сервіс PHASTER (PHAge Search Tool – Enhanced Release):
<https://phaster.ca>.
- База даних повних бактеріальних геномів (наприклад, NCBI Assembly або GenBank для вибору бактеріального геному).
- Набір бактеріальних геномів (або штамів) у форматі GenBank або FASTA для аналізу.

Завдання

1. **Вибрати бактеріальний геном** для аналізу – кожен студент бере окремий вид або штам (Додаток 2);
2. **Запустити аналіз профагів через PHASTER:** Завантажити обраний геном на сервер PHASTER;
3. **Визначити кількість профагів** у геномі: за результатами PHASTER з'ясувати, скільки ділянок ідентифіковано як фагові інтегровані елементи;
4. Для кожного знайденого профага зафіксуйте його характеристики: довжину (у парах основ, bp), положення у хромосомі (координати початку/кінця), статус (інтактний – містить більшість необхідних генів, або неповний/дефектний), а також кількість генів у межах регіону.
5. **Проаналізувати склад одного профага:** визначити наявність ключових генів – наприклад, *integrase* (інтеграза), *capsid* або *tail proteins*, регуляторні гени, або ж гени, що можуть надавати бактерії переваги (наприклад, токсини, фактори вірулентності, резистентності і т.п., якщо такі є).
6. **Зробити висновки** про роль профагів у вибраному штамі: скільки % геному займають профаги; чи містять вони відомі корисні або шкідливі для хазяїна гени; чи є серед них інтактні (можливо здатні до активної лізогенії/літичного циклу) або всі дефектні.

Хід роботи

Фіксація об'єкта аналізу

1. Обрати один бактеріальний геном із Додатка 2.
2. Зафіксувати у звіті:
 - повну назву організму (вид, серовар/підвид, штам);
 - код доступу RefSeq (NC_...);
 - довжину хромосоми (bp).

Підготовка вхідних даних для PHASTER

1. Використати RefSeq-код доступу (NC_...) як єдиний формат подання геному для аналізу.
2. Підготувати у звіті рядок ідентифікації аналізу:
 - Genome ID: NC_...
 - Organism: ...
 - Genome length (bp): ...

Запуск аналізу PHASTER

1. У PHASTER задати тип введення Enter accession.
2. Внести RefSeq-код доступу (NC_...) у поле accession.
3. Активувати режим використання попередньо обчислених результатів сервісу.
4. Ініціювати обробку запиту на сервері PHASTER.
5. Зафіксувати ідентифікатор завдання (Job ID), який присвоєно аналізу.
6. Зафіксувати унікальну сторінку результатів для цього Job ID.

Вилучення зведених результатів (Summary)

1. У блоці результатів відкрити таблицю **Prophage regions**.
2. Для кожного рядка таблиці зафіксувати у робочій таблиці (Таблиця 1) такі поля:
 - номер регіону (Region);
 - координати у хромосомі (Start–End);
 - довжина регіону (Length, bp);
 - сумарний бал (Score);
 - кількість генів у регіоні (# of CDS);

- статус (Status).
3. Виконати класифікацію профагів за правилом:
 - Intact: Score > 90
 - Questionable: $70 \leq \text{Score} \leq 90$
 - Incomplete: Score < 70.
 4. Зафіксувати загальну кількість профагових регіонів як число рядків таблиці Prophage regions.

Розрахунок частки геному, що припадає на профаги

1. Обчислити сумарну довжину профагових регіонів:
 - $L_total = \sum \text{Length}(\text{region}_i)$.
2. Обчислити частку геному, зайняту профагами:
 - $P(\%) = (L_total / \text{Genome_length}) \times 100$.
3. Зафіксувати у звіті:
 - L_total (bp);
 - Genome_length (bp);
 - P (%).

Деталізація одного профага (DETAILS)

1. Обрати для структурного аналізу профаг зі статусом **Intact**.
2. Відкрити для цього регіону вкладку **DETAILS**.
3. Сформувати перелік ключових функціональних категорій генів (Таблиця 2) за анотаціями PHASTER:
 - інтеграція/екцизія (integrase, recombinase);
 - структурні компоненти фага (capsid, portal, tail, baseplate);
 - упаковка ДНК (terminase);
 - лізис клітини (holin, endolysin, lysozyme);
 - регуляція (transcriptional regulator, repressor, antirepressor);
 - додаткові гени, пов'язані з фенотипом хазяїна (virulence, toxin, resistance, effector).
4. Для кожної знайденої категорії зафіксувати у Таблиці 2:

- назву гена/продукту (як у PHASTER);
- координати гена (Start–End);
- функціональну категорію зі списку вище;
- позначення подібності/ідентифікації (як наведено у PHASTER для цього гена).

Формулювання висновків

1. Подати числові підсумки:
 - загальна кількість профагових регіонів;
 - кількість Intact, Questionable, Incomplete;
 - L_total (bp) і P (%).
2. Подати структурну характеристику одного профага:
 - наявність генів інтеграції/екцизиї;
 - наявність структурних компонентів;
 - наявність лізисного модуля;
 - наявність додаткових генів, пов'язаних із фенотипом хазяїна.
3. Сформулювати роль профагів у геномі цього штаму через два твердження:
 - внесок у розмір геному (P, %);
 - потенційна біологічна значущість за складом генів у DETAILS.

Контрольні запитання

1. Дайте визначення поняттю профаг. Чим профаг відрізняється від вільного фага?
2. Які цикли розвитку бактеріофага вам відомі? Поясніть відмінність між літичним і лізогенним циклом.
3. Яку роль відіграють профаги у геномі бактерій? Наведіть приклади генів, що можуть приносити профаги (корисні чи шкідливі для бактерії).
4. Який онлайн-інструмент використовується для виявлення профагів у бактеріальних геномах? Яку інформацію він надає на виході?

5. В чому полягає різниця між статусами профагів Intact, Questionable та Incomplete, які визначає PHASTER? За якими критеріями PHASTER відносить знайдений фаг до інтактного чи дефектного?
6. Які основні гени повинні бути присутні в інтактному профазі? Назвіть не менше трьох (наприклад, інтеграза...).
7. Ви проаналізували геном *E. coli* і PHASTER знайшов 2 інтактних профаги та 3 неповних. Як визначити, яку частку геному становлять всі ці профаги разом?
8. Якщо профаг інтегровано в бактеріальну хромосому, як називається специфічне місце вбудови (на рівні ДНК-послідовності)? *Підказка: Більшість фагів інтегруються в специфічний сайт, наприклад tRNA-ген.*
9. Чому, на вашу думку, різні штами того самого виду бактерії можуть мати різну кількість профагів? Як це пов'язано з умовами існування бактерій або їх патогенністю?
10. Практичне: Отримавши результат PHASTER, ви бачите “Region 3: 45000–78000 bp, 52 CDS, Score 110, Intact”. Поясніть, що означають ці цифри і які подальші кроки вашого аналізу цього регіону.

Рекомендована література

1. Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., & Wishart, D. S. (2016). PHASTER: A better, faster version of the PHAST phage search tool. *Nucleic Acids Research*, 44(W1), W16–W21.
<https://doi.org/10.1093/nar/gkw387>
2. Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., & Wishart, D. S. (2011). PHAST: A fast phage search tool. *Nucleic Acids Research*, 39(suppl_2), W347–W352.
<https://doi.org/10.1093/nar/gkr485>
3. Wishart, D. S., Han, S., Saha, S., Oler, E., & others. (2023). PHASTEST: Faster than PHASTER, better than PHAST. *Nucleic Acids Research*, 51(W1), W443–W450.

<https://doi.org/10.1093/nar/gkad382>

4. Casjens, S. (2003). Prophages and bacterial genomics: What have we learned so far? *Molecular Microbiology*, 49(2), 277–300.
<https://doi.org/10.1046/j.1365-2958.2003.03580.x>
5. Brüssow, H., Canchaya, C., & Hardt, W.-D. (2004). Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion. *Microbiology and Molecular Biology Reviews*, 68(3), 560–602.
<https://doi.org/10.1128/MMBR.68.3.560-602.2004>
6. Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.-L., & Brüssow, H. (2003). Phage as agents of lateral gene transfer. *Current Opinion in Microbiology*, 6(4), 417–424.
[https://doi.org/10.1016/S1369-5274\(03\)00086-9](https://doi.org/10.1016/S1369-5274(03)00086-9)

Практична робота 5: Визначення бактеріальних плазмід за допомогою біоінформатичних інструментів

Мета роботи

Навчитися виявляти плазмід – автономні мобільні генетичні елементи бактерій – на основі геномних даних *in silico*. Студенти будуть використовувати веб-інструмент PlasmidFinder для ідентифікації репліконів плазмід у складі бактеріального геному або секвенованих контигів. Буде розглянуто поняття репліконів (генів, відповідальних за початок реплікації плазмід) і їх використання для типування плазмід. Результатом стане визначення наявності (чи відсутності) плазмід у різних бактеріальних ізолятах та їх типів, що дозволяє оцінити мобільний генофонд (наприклад, перенос генів антибіотикорезистентності).

Теоретичні відомості

Плазмід – це позахромосомні реплікони, переважно кільцеві, хоча в окремих мікроорганізмів трапляються і лінійні форми, здатні до автономної та регульованої реплікації в клітині-хазяїні. На відміну від бактеріальної хромосоми, плазмід зазвичай не містять генів, критично необхідних для базового клітинного метаболізму за стандартних умов, проте суттєво розширюють адаптивний потенціал бактерій. Вони можуть переносити гени антибіотикорезистентності, факторів вірулентності, бактериоцинів, детоксикаційних систем, ферментів утилізації незвичних субстратів та інші допоміжні детермінанти, що підвищують екологічну й клінічну успішність штамів. Плазмід розглядають як один із центральних носіїв горизонтального перенесення генетичної інформації у прокаріотів.

Молекулярною основою автономного існування плазмід є її **реплікон** – функціональна одиниця, що включає origin of replication (*ori*) та пов'язані з ним гени і регуляторні елементи, які забезпечують ініціацію й контроль реплікації. У більшості випадків ключову роль відіграє плазмідно-кодований

ініціаторний білок Rep, який розпізнає специфічні ділянки *ori*, організовує початок синтезу ДНК і залучає клітинний реплікативний апарат. Для кільцевих бактеріальних плазмід описано кілька базових механізмів реплікації, зокрема θ -реплікацію, strand-displacement-реплікацію та rolling-circle-реплікацію. Не менш важливою є система контролю числа копій: реплікація плазмиди має бути узгоджена з поділом клітини, і тому її регуляція часто здійснюється за участю антисмислових РНК, ітеронів або комбінованих РНК-білкових механізмів.

Поширення плазмід у бактеріальних популяціях зумовлене не лише їхньою здатністю до самопідтримання, а й мобільністю. Кон'югативні плазмиди кодують апарат перенесення ДНК між клітинами, а мобілізовні плазмиди використовують системи переносу, заковані іншими елементами. Унаслідок цього плазмиди є ефективними векторами дисемінації генів стійкості, вірулентності та екологічної адаптації між штамми, видами й навіть вищими таксономічними групами бактерій. У контексті клінічної мікробіології саме плазмід-опосередкований горизонтальний перенос розглядають як одну з головних причин швидкого поширення мультирезистентності.

Класифікація плазмід традиційно пов'язана з поняттям груп несумісності (Inc-groups). Історично плазмиди об'єднували в одну групу тоді, коли вони не могли стабільно співіснувати в одній клітинній лінії; така несумісність зазвичай пов'язана зі спорідненими системами ініціації реплікації та її контролю. Отже, тип реплікона є не лише молекулярною ознакою конкретної плазмиди, а й проксі-показником її еволюційної та функціональної належності до певної групи. Саме ця властивість зробила реплікон-типсування фундаментом сучасної *in silico*-ідентифікації плазмід у геномних даних. Водночас Inc-класифікація не вичерпує всієї різноманітності плазмід і не завжди відображає їхню мобільність, оскільки остання залежить також від *relaxase*, систем формування спаровувального апарату та інших модулів переносу.

PlasmidFinder – це розроблений Center for Genomic Epidemiology інструмент для *in silico*-виявлення плазмідних репліконів у повних геномах, драфтових збірках або сирих секвенувальних даних. Принцип його роботи полягає не у реконструкції всієї плазмиди *de novo*, а у пошуку збігів між вхідними послідовностями та курованою базою відомих репліконних послідовностей. У веб-реалізації сервіс пропонує окремі бази для грампозитивних бактерій і для Enterobacterales/історично позначеної в інтерфейсі групи Enterobacteriales, а також дозволяє аналізувати як зібрані FASTA-файли з контигами, так і сирі FASTQ-зчитування. Для сирих даних у серверній реалізації використовується KMA; програмна реалізація сервісу передбачає використання KMA та BLAST+, залежно від типу вхідних даних і режиму запуску. Типовими стартовими параметрами є мінімум 95% нуклеотидної ідентичності та 60% покриття репліконної послідовності.

Методологічно PlasmidFinder є маркерним інструментом. Він не встановлює повну структуру плазмиди, не визначає автоматично її замкненість, топологію чи повний набір функціональних модулів, а лише фіксує наявність у зразку відомих репліконних детермінант. Тому коректне трактування негативного результату має бути обмеженим: відсутність хітів означає відсутність виявлених відомих репліконів у межах обраної бази й установлених порогів, але не є абсолютним доказом відсутності будь-якої плазмиди. Причинами такого результату можуть бути новий або недостатньо представлений у базі тип реплікона, сильна дивергенція послідовності, фрагментація збірки, низька якість даних або таксономічна невідповідність між досліджуваним організмом і обраною базою.

Інтерпретація виходу PlasmidFinder ґрунтується на кількох ключових метриках. Назва хіта відповідає референтній плазмідній/репліконній послідовності з бази. Показник % identity відображає нуклеотидну тотожність між знайденою ділянкою та еталоном, тоді як Query/HSP length характеризує, відповідно, довжину референтної послідовності та довжину вирівняного високооціненого сегмента (high-scoring segment pair, HSP). У таблиці також

наводяться назва контига та позиція збігу в ньому. Поєднання високої тотожності з майже повним покриттям свідчить про надійне виявлення реплікона; натомість неповне покриття за доброї тотожності частіше вказує на фрагментований або частково зібраний реплікон, а не на повністю охарактеризовану плазмиду.

Наявність кількох репліконів потребує обережної інтерпретації. Якщо два або більше реплікони локалізуються в одному контизі, це може відображати **багаторепліконну плазмиду** або коінтеграт, тобто молекулу, що виникла внаслідок злиття різних плазмідних репліконів в одну ДНК-структуру. Якщо ж реплікони виявлено на різних контигах, найбільш імовірним є існування кількох окремих плазмід, хоча фрагментація короткочитної збірки не завжди дозволяє ототожнити кожен контиг з окремою фізичною молекулою. Саме тому реплікон-типування слід трактувати як сильний індикатор плазмідного складу, але не як повну реконструкцію плазмідному ізоляту. Для остаточного відмежування плазмідних і хромосомних фрагментів бажані замкнені плазмідні контиги, гібридні збірки або додаткові спеціалізовані інструменти реконструкції.

pMLST є наступним рівнем типування, який застосовують не до всіх плазмід, а лише до тих сімейств, для яких існують валідовані схемі алельного профілювання. У сервісі CGE доступні окремі конфігурації для низки плазмідних груп, зокрема IncA/C, IncF, IncHI1, IncHI2, IncI1, IncN та pBSSB1-family. Алельні послідовності та профілі сервіс отримує з PubMLST, а результати інтерпретує консервативно: для веб-версії зазначено, що враховуються збіги з ідентичністю понад 85% і довжиною вирівнювання понад 66%. На відміну від реплікон-типування, pMLST дає змогу диференціювати споріднені плазмиди всередині однієї Inc-групи, що є особливо цінним у молекулярній епідеміології та трасуванні поширення конкретних плазмідних ліній.

Для поглибленого аналізу плазмідного вмісту драфтових збірок доцільно поєднувати PlasmidFinder із інструментами, що враховують не лише реплікон, а й модулі переносу та загальну подібність повних послідовностей.

MOB-suite є прикладом такого підходу: він орієнтований на реконструкцію плазмід із драфтових збірок, визначення родини реплікона, типу relaxase, типу системи формування спарувального апарату, прогнозованої мобільності та, в окремих випадках, імовірного діапазону хазяїв. Отже, якщо PlasmidFinder відповідає насамперед на запитання, чи є у зразку відомий реплікон, то pMLST і MOB-suite дозволяють точніше відповісти на запитання, якою саме є ця плазміда, до якої лінії вона належить і який її потенціал до горизонтального поширення.

Матеріали та методи

- Комп'ютер, Інтернет.
- Веб-сервіс PlasmidFinder (Center for Genomic Epidemiology, CGE):
<https://cge.food.dtu.dk/services/PlasmidFinder/> .

Примітка: частина сервісів CGE поступово переноситься на нову платформу, тому на момент виконання вами роботи сервіс може давати помилку доступу. Якщо сторінка недоступна, скористайтесь переліком сервісів на <https://www.genomicepidemiology.org/services/> або запускайте PlasmidFinder локально/через Galaxy.

- Набір секвованих бактеріальних геномів або контигів (у форматі FASTA) для різних ізолятів (Додаток 3)
- (Опціонально) Доступ до додаткових інструментів типування плазмід (pMLST) для поглибленого аналізу.

Завдання

1. **Обрати бактеріальний ізолят (геном) та отримати його зібрану послідовність.** Кожен студент обирає свій штам/ізолят (Додаток 3).

2. **Запустити PlasmidFinder** для свого файлу: завантажити FASTA з контигами до веб-сервісу, встановити пороги ідентичності та покриття (за замовчуванням 95% ідентичності і 60% покриття є оптимальними для початку) і вказати базу даних (Gram-positive чи Enterobacterales) відповідно до виду бактерії.
3. **Визначити наявність репліконів плазмід.** Після запуску аналізу PlasmidFinder видасть список знайдених **реплікон-генів** (якщо вони присутні) разом з відсотком ідентичності та покриття.
4. **Зафіксувати результати:** якщо жодного реплікона не знайдено, зазначити, що **плазмід не виявлено**. Якщо знайдено реплікон(и), вписати їх назви і до яких груп плазмід вони належать.
5. **Порівняти довжину контигів** з репліконами: визначити, чи контиг з репліконом є повною плазмідною чи фрагментом.
6. **(Опціонально)** Використати інструмент **pMLST** (плазмідний MLST), доступний також через CGE, для типування знайденої плазмід (для деяких типів, як IncF, IncN тощо).
7. **Зробити висновки** щодо плазмідного профілю ізоляту: чи містить він плазмід і які саме (назви репліконів = тип плазмід). За наявності кількох плазмід – які з них великі/малі (орієнтовно за типом або довжиною контига). Вказати, чи асоціюються ці плазмід з перенесенням певних генів.

Хід роботи

Підготовка вихідних даних

1. Для аналізу обирають **один бактеріальний ізолят** і формують для нього єдиний набір вхідних даних у форматі **FASTA**: або повну зібрану послідовність, або набір контигів draft-геному.
2. Якщо для вибраного контрольного прикладу в базі NCBI доступні окремі записи хромосоми і плазмід, їх доцільно використовувати як референтний матеріал для перевірки логіки інтерпретації: такий формат дозволяє

співвіднести хіти PlasmidFinder не лише з типом реплікону, а й з реальною структурою геному.

3. Якщо аналізують draft-геном, назви контигів у FASTA не змінюють і не скорочують, оскільки саме вони використовуються для подальшого зіставлення реплікону з конкретним фрагментом збірки.

Таксономічна верифікація ізоляту та вибір бази репліконів

1. Перед запуском аналізу визначають, до якої групи належить досліджуваний організм, оскільки в актуальній реалізації сервісу використовуються **окремі бази Enterobacteriales та Gram Positive**.
2. Для ізолятів Enterobacterales застосовують ентеробактеріальну базу репліконів; для Staphylococcus, Enterococcus та інших грампозитивних бактерій – базу Gram Positive. Помилка на цьому етапі змінює простір пошуку і може призвести до хибного негативного результату.
3. Обрану базу даних зазначають у протоколі окремим рядком, оскільки це є частиною відтворюваності аналізу так само, як accession-номер файлу і використані пороги ідентичності та покриття.

Параметризація запуску PlasmidFinder

1. Для вже зібраних геномів використовують режим аналізу assembled or draft genome/contigs (FASTA). Для цієї лабораторної саме такий формат є методично коректним, тому що завдання полягає не в аналізі сирих рідів, а в пошуку плазмідних репліконів у зібраній послідовності.
2. Перший запуск виконують зі стандартними порогамі 95% identity і 60% coverage. Саме ці значення сервіс рекомендує як базовий режим, і вони відповідають класичному підходу до первинного виявлення відомих репліконів.
3. Для кожного ізоляту проводять окремий запуск. Якщо після основного аналізу залишаються прикордонні або біологічно сумнівні хіти, їх можна перевіряти додатковим прогоном, але формальний висновок для звіту доцільно будувати саме на стандартному запуску.

4. У протоколі обов'язково фіксують три параметри: тип вхідних даних, вибрану базу репліконів і пороги ідентичності/покриття. Без цього відтворити аналіз у наступному семестрі або порівняти результати між студентами буде неможливо.

Первинне читання результатів PlasmidFinder

1. Після завершення аналізу результати зчитують не як загальний список «наявних плазмід», а як список знайдених репліконів, тобто маркерів систем реплікації плазмід. Це принципово, оскільки PlasmidFinder не реконструює плазмиду повністю, а виявляє її реплікативний маркер.
2. У поточному виході сервісу ключовими для інтерпретації є поля Plasmid, % Identity, Query/HSP length, Contig, Position in contig та Accession number. Саме ці поля мають бути перенесені до робочої таблиці або чернетки звіту.
3. Для кожного знайденого хіта окремо виписують назву реплікону, відсоток ідентичності, покриття/довжину вирівнювання, назву контига та координати вирівнювання на контизі. Без прив'язки до контига зробити коректний висновок про кількість плазмід і ступінь повноти плазмідного елемента неможливо.
4. Якщо сервіс не виявляє жодного хіта, у протоколі фіксують не «плазмід немає», а формулювання «відомі плазмідні реплікони не виявлено». Таке формулювання є коректним, бо негативний результат PlasmidFinder не виключає наявності нових, рідкісних або не представлених у базі плазмід.

Інтерпретація знайдених репліконів

1. Реплікон розглядають як підтверджений, якщо він проходить прийняті пороги ідентичності та покриття. У межах цієї роботи базовим критерієм присутності є $\geq 95\%$ identity та $\geq 60\%$ coverage.
2. Один знайдений реплікон означає, що в геномних даних присутній принаймні один відомий плазмідний маркер, але це ще не доводить, що плазмідна реконструйована повністю. Коректний висновок має відрізняти «наявність реплікону» від «повної плазмід».

3. Якщо на одному й тому самому контизі виявлено кілька репліконів, їх насамперед інтерпретують як ознаку багаторепліконної плазмиди, а не автоматично як кілька окремих плазмід. Для ентеробактерій така ситуація добре описана в літературі для F-подібних плазмідних фонів.
4. Якщо різні реплікони знайдено на різних контигах, це є підставою вважати, що в ізоляті присутні щонайменше кілька плазмідних елементів або плазмідних фрагментів. Остаточний висновок у такому випадку роблять лише після зіставлення з довжиною відповідних контигів.
5. Для грампозитивних бактерій у результатах частіше трапляються не класичні Inc-позначення ентеробактерій, а гер-типи або родинні їм позначення реплікативних систем; це треба враховувати під час написання висновку.

Зіставлення реплікону з контигом і оцінка повноти плазмідного елемента

1. Для кожного підтвердженого реплікону визначають, у якому саме контизі він локалізований, і звіряють координати вирівнювання з довжиною цього контига. Це дозволяє відрізнити короткий фрагмент плазмідної послідовності від великого окремого плазмідного контига.
2. Якщо реплікон присутній майже повністю, а сам контиг має довжину, співмірну з відомими плазмідами відповідного типу, та не виглядає як короткий фрагмент збірки, його можна описувати як кандидат на повну або майже повну плазмиду. Якщо ж континг короткий або містить лише частину реплікону, результат описують обережніше – як плазмідний фрагмент або континг плазмідного походження.
3. Для complete-genome прикладів із окремими плазмідними accession-записами довжину контига доцільно порівнювати з довжиною референтної плазмиди. Для draft-геномів такий висновок роблять обережніше і не підміняють його твердженням про повністю встановлену структуру плазмиди.

Розмежування кількох плазмід у межах одного ізоляту

1. Після первинного читання результатів усі хіти групують за контигами, а не лише за назвами репліконів. Саме контиг є мінімальною структурною одиницею, з якою можна пов'язувати конкретний плазмідний елемент у збірці.
2. Якщо на одному контизі виявлено споріднені або сумісні реплікони, у висновку їх описують як одну плазмиду з кількома реплікативними маркерами. Якщо хіти рознесені по різних контигах, у висновку вказують мінімальну оцінку кількості плазмідних елементів.
3. Асоціювати знайдену плазмиду з перенесенням генів резистентності або вірулентності дозволено лише тоді, коли така асоціація підтверджується або анотацією конкретного референтного прикладу, або публікаціями саме для цього типу плазмиди. Загальні фрази без опори на джерело у висновок не вносять.

Додаткове типування pMLST

1. pMLST використовують не для всіх плазмід, а лише для тих, для яких у сервісі підтримується відповідна схема типування. У поточній реалізації це, зокрема, IncA/C, IncF RST, IncHI1, IncHI2, IncI1, IncN та pBSSB1-family.
2. Якщо знайдений реплікон належить до однієї з підтримуваних груп, для відповідного контига або плазмідного запису виконують pMLST і фіксують алельний профіль та присвоєний тип. Якщо повний профіль не визначається, це також записують окремо, а не залишають поза протоколом.
3. Якщо для знайденого плазмідного типу pMLST-схеми немає, у звіті прямо зазначають, що додаткове плазмідне типування для цього реплікону не застосовується. Це є коректним негативним результатом опційного етапу.

Формування висновків

1. для кожного ізоляту формують у стандартизованому вигляді:
 - чи виявлено плазмідні реплікони;
 - які саме реплікони знайдено;

- їхні показники identity і coverage;
 - у яких контигах вони локалізовані;
 - чи є підстава вважати відповідний континг повною плазмідною або лише її фрагментом;
 - чи застосовувався pMLST і що він показав.
2. У висновку не ототожнюють назву реплікону з уже доведеною повною структурою плазмід. Коректне формулювання має вигляд: «виявлено реплікон плазмідного типу...», «ймовірно присутня плазміда цього типу», «континг відповідає плазмідному елементу».
 3. Для позитивного результату висновок доцільно оформлювати як короткий аналітичний абзац

Контрольні запитання:

1. Що таке плазміда? Чим плазмідні відрізняються від хромосоми за будовою та функціями?
2. Які корисні гени можуть переносити бактеріальні плазмідні? Наведіть кілька прикладів, чому плазмідні важливі у медицині та біотехнологіях.
3. Поясніть поняття реплікон плазмідного. Яку роль відіграє реплікаційний ген (ген *rep*)?
4. Що таке групи несумісності плазмід (*Inc-groups*) і чому плазмідні однієї групи не можуть співіснувати в одній клітині?
5. Який веб-інструмент використовується для виявлення репліконів плазмід у геномних послідовностях? Які параметри пошуку в ньому можна налаштовувати?
6. Ви запустили PlasmidFinder і отримали результат: “*IncFII*: Identity 99.8%, Coverage 100%, Length 846 bp”. Поясніть, що означають ці показники.
7. Якщо в одному контізі знайдено два різні реплікони (наприклад, *IncFII* та *IncFIB*), що це може означати щодо будови плазмідного?

8. PlasmidFinder не знайшов жодного реплікон-гену у геномі. Які можливі причини цього і як ви перевірите, чи дійсно у штаму немає плазмід?
9. Практичне: Ви виявили в штаму *Klebsiella* реплікони IncFII, IncR та ColE1. Скільки окремих плазмід, ймовірно, присутні? Які з них, імовірно, великі, а які – маленькі? (Відповідь обґрунтуйте).
10. Як наявність плазмід у бактеріального ізолята може вплинути на його стійкість до антибіотиків або інші фенотипові ознаки? Чому важливо вміти ідентифікувати плазмід в клінічних ізолятах?

Рекомендована література

1. Carattoli, A., Zankari, E., García-Fernández, A., Voldby Larsen, M., Lund, O., Villa, L., Aarestrup, F. M., & Hasman, H. (2014). In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrobial Agents and Chemotherapy*, 58(7), 3895–3903.
<https://doi.org/10.1128/AAC.02412-14>
2. Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C., & de la Cruz, F. (2010). Mobility of plasmids. *Microbiology and Molecular Biology Reviews*, 74(3), 434–452.
<https://doi.org/10.1128/MMBR.00020-10>
3. Rozwandowicz, M., Brouwer, M. S. M., Fischer, J., Wagenaar, J. A., Gonzalez-Zorn, B., Guerra, B., Mevius, D. J., & Hordijk, J. (2018). Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. *Journal of Antimicrobial Chemotherapy*, 73(5), 1121–1137.
<https://doi.org/10.1093/jac/dkx488>
4. Arredondo-Alonso, S., Rogers, M. R. C., Braat, J. C., Verschuuren, T. D., Top, J., Corander, J., Willems, R. J. L., & Schürch, A. C. (2018). MOB-suite: Software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microbial Genomics*, 4(8), e000206
<https://doi.org/10.1099/mgen.0.000206>

5. Garcillán-Barcia, M. P., Francia, M. V., & de la Cruz, F. (2009). The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiology Reviews*, 33(3), 657–687.
<https://doi.org/10.1111/j.1574-6976.2009.00168.x>
6. Thomas, C. M., & Nielsen, K. M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews Microbiology*, 3(9), 711–721.
<https://doi.org/10.1038/nrmicro1234>

Практична робота 6: Структурний аналіз LTR-ретротранспозона ONSEN (ATCOPIA78) у геномі *Arabidopsis thaliana*

Мета роботи

Ознайомитися з будовою LTR-ретротранспозонів типу Ty1/copia та навчитися ідентифікувати й описувати копії конкретної родини ретротранспозона у референтному геномі *Arabidopsis thaliana*. Опанувати практичні підходи до оцінювання інтактності вставок (наявність обох LTR, внутрішньої ділянки та кодувального потенціалу) і порівняльної оцінки «віку» копій за ступенем дивергенції LTR.

Теоретична основа

LTR-ретротранспозони є мобільними генетичними елементами I класу, які поширюються в геномі за механізмом реплікативної транспозиції типу «copy-and-paste». На відміну від ДНК-транспозонів, вони не вирізаються з вихідного локусу у формі ДНК, а проходять стадію транскрипції в РНК, після чого їхня комплементарна ДНК синтезується зворотною транскриптазою та інтегрується в нову ділянку геному за участі інтегрази. Унаслідок такої інтеграції на межах вставки зазвичай формуються короткі цільові дуплікації, які є характерною ознакою завершеного інтеграційного акту.

Типова повнорозмірна копія LTR-ретротранспозона складається з двох довгих кінцевих повторів, 5'-LTR і 3'-LTR, які безпосередньо фланкують внутрішню ділянку елемента. У внутрішньому регіоні локалізуються одна або кілька відкритих рамок зчитування, що кодують білки GAG та POL; до складу POL входять протеаза, інтеграза, зворотна транскриптаза і RNase H. Для елементів надродини Ty1/copia, до якої належить ONSEN, характерне розташування доменів POL у порядку protease–integrase–reverse transcriptase–RNase H, що є важливою діагностичною ознакою цієї групи. Крім того, для життєвого циклу LTR-ретротранспозона істотними є primer binding site (PBS)

і polypurine tract (PPT), які беруть участь у ініціації та завершенні зворотної транскрипції.

Після інтеграції 5'- та 3'-LTR однієї копії спочатку є практично ідентичними, оскільки походять із одного РНК-шаблону. Згодом ці повтори накопичують нуклеотидні заміни відносно незалежно, тому ступінь їхньої дивергенції використовують як молекулярний індикатор відносного віку вставки. Саме тому порівняння LTR однієї копії є коректнішим для датування, ніж порівняння різних копій між собою: міжкопійні відмінності відображають не лише час після інтеграції, а й глибшу еволюційну історію родини, можливу субродинну диференціацію та нерівномірну швидкість накопичення мутацій. Оцінка **віку** за дивергенцією LTR є корисною, але наближеною. Формально її подають як

$$T=K/(2\mu), \text{ де}$$

K – нуклеотидна дивергенція між 5'- та 3'-LTR, а

μ – швидкість нейтральних замін на сайт за рік.

Однак така оцінка спирається на низку припущень, які не завжди виконуються: швидкість замін може відрізнятись між видами і навіть між ділянками геному; конверсія генів здатна штучно зменшувати дивергенцію між LTR і призводити до недооцінки віку; додаткові похибки можуть вносити множинні заміни, рекомбінаційні події, неточності збірки та неідеальне визначення меж елемента. Тому в навчальній роботі методологічно правильніше робити акцент передусім на відносному, а не абсолютному порівнянні віку вставок.

ONSEN, також позначуваний як ATCOPIA78, є LTR-ретротранспозоном типу Ty1/copia у *Arabidopsis thaliana*, який став модельною системою для дослідження стрес-індукованої мобілізації транспозонів у рослин. Для нього показано активацію під впливом теплового стресу, причому ця активація пов'язана з наявністю у LTR термочутливих регуляторних мотивів, що розпізнаються факторами теплового шоку. Унаслідок цього ONSSEN є не лише структурним об'єктом аналізу, а й функціонально релевантним елементом,

здатним впливати на транскрипцію прилеглих генів та, за певних умов, сприяти виникненню нових вставок. Для референтного геному *Arabidopsis thaliana*, зібраного на основі екотипу Col-0, у літературі описано лише обмежену кількість повнорозмірних копій ONSEN, що робить цю родину зручною для локус-орієнтованого структурного аналізу.

У цій роботі вихідною моделлю родини слугує консенсусна послідовність із бази Dfam. **Dfam** є відкритим ресурсом родин повторів, який зберігає для кожної родини множинні вирівнювання канонічних копій, профільні приховані марковські моделі та консенсусні послідовності. Важливо розуміти, що консенсус не є буквальним еквівалентом якоїсь однієї природної копії, а узагальненою моделлю родини, побудованою за сукупністю вирівняних екземплярів. Саме тому пошук копій ONSEN у геномі за консенсусом дає змогу виявляти як майже інтактні, так і помірно дивергентні представники родини. Dfam також підтримує завантаження консенсусної послідовності у форматі FASTA, що робить цей ресурс придатним для подальшого гомологічного пошуку.

Для картування потенційних копій ONSEN у геномі *Arabidopsis thaliana* доцільно використовувати BLASTn в Ensembl Plants, оскільки цей ресурс надає доступ до актуальної референтної збірки TAIR10 для екотипу Col-0 і дозволяє працювати безпосередньо з геномною ДНК. На цьому етапі слід інтерпретувати не лише статистичну значущість збігів, а й їхнє покриття та протяжність: повнорозмірні або майже повнорозмірні вирівнювання свідчать на користь інтактних чи близьких до інтактних копій, тоді як короткі або переривчасті збіги частіше відповідають фрагментам, деградованим вставкам або окремим LTR. Відбір кількох локусів із різним характером вирівнювання є методично виправданим, оскільки дозволяє порівняти інтактні й частково деградовані представники однієї родини в межах одного геному.

Структурне передбачення меж елемента виконується не за гомологією до консенсусу як такою, а за сукупністю архітектурних ознак, характерних для LTR-ретротранспозонів. Інструмент LTRharvest реалізує de novo-пошук

кандидатних елементів на основі пар подібних LTR, відстані між ними, наявності TSD і, за потреби, кінцевих мотивів. Його параметри – мінімальна та максимальна довжина LTR, мінімальна та максимальна внутрішня відстань, поріг подібності, допустимі межі для TSD і пошук мотивів – безпосередньо визначають баланс між чутливістю та специфічністю. Отже, LTRharvest дає не «остаточну істину» про елемент, а структурно обґрунтовану гіпотезу про його межі, яку надалі потрібно зіставити з довжиною внутрішньої ділянки, наявністю двох LTR, цільових дуплікацій і кодувального потенціалу.

У контексті цієї лабораторної поняття інтактності слід трактувати строго структурно-функціонально. Імовірно інтактною можна вважати таку копію ONSSEN, у якій виявляються обидва LTR, безперервна внутрішня ділянка очікуваної довжини, ознаки коректних меж інтеграції та довгі ORF, сумісні з наявністю gag/pol. Натомість відсутність одного з LTR, великі делеції, численні внутрішні розриви, передчасні стоп-кодони або фрагментація ORF свідчать про деградацію копії. Водночас навіть наявність довгих ORF не є прямим доказом функціональної автономності: ORFfinder лише виявляє відкриті рамки зчитування і повертає їхні координати та передбачений білковий продукт, але не встановлює експресію, трансляцію, доменну повноцінність, коректність процесингу поліпротеїну чи реальну здатність елемента до транспозиції.

Для порівняння 5'- та 3'-LTR доцільно використовувати глобальне вирівнювання, а не локальне, оскільки завдання полягає в оцінюванні повної парної подібності двох гомологічних послідовностей однакової функціональної природи. EMBOSS Needle реалізує саме глобальне вирівнювання за алгоритмом Нідлмана—Вунша і тому є коректним інструментом для обчислення відсотка ідентичності по всій довжині LTR. Під час інтерпретації цього показника слід пам'ятати, що висока ідентичність між 5'- та 3'-LTR свідчить про відносно недавню інтеграцію, тоді як нижча ідентичність вказує на триваліше незалежне накопичення замін; однак сам

відсоток ідентичності завжди потрібно оцінювати разом із якістю визначення меж LTR та можливими наслідками конверсії генів або рекомбінації.

Отже, **структурний аналіз ONSEN** у референтному геномі *Arabidopsis thaliana* поєднує три рівні інтерпретації.

- Перший рівень – гомологічний, коли за консенсусом Dfam ідентифікують локуси, споріднені з родиною ATCOPIA78.
- Другий рівень – архітектурний, коли за допомогою LTRharvest уточнюють межі повнорозмірної вставки, LTR і потенційних TSD.
- Третій рівень – функціонально-еволюційний, коли за ORF оцінюють збереженість кодувального потенціалу, а за подібністю 5'- та 3'-LTR роблять висновок про відносний вік копії.

Саме в такій багаторівневій логіці поняття «копія ONSEN» переходить від простого BLAST-збігу до біологічно змістовної характеристики конкретної вставки як імовірно інтактної, частково деградованої або фрагментованої.

Матеріали та методи

- Комп'ютер з доступом до Інтернету.
- Ensembl Plants (геномний браузер і BLAST для *Arabidopsis thaliana*):
https://plants.ensembl.org/Arabidopsis_thaliana/Tools/Blast
- Dfam (пошук та завантаження консенсусних послідовностей повторів):
<https://www.dfam.org/>
- Galaxy Europe (веб-платформа для запуску біоінформатичних інструментів): <https://usegalaxy.eu/>
- NCBI ORFfinder (пошук відкритих рамок зчитування):
<https://www.ncbi.nlm.nih.gov/orffinder/>
- EMBOSS Needle або інший інструмент глобального вирівнювання (доступний у Galaxy або онлайн):
https://www.ebi.ac.uk/jdispatcher/psa/emboss_needle

Завдання

- Отримати консенсусну нуклеотидну послідовність родини ONSEN (ATCOPIA78) з бази Dfam та зафіксувати її довжину.
- Знайти в геномі *Arabidopsis thaliana* локуси, гомологічні консенсусу ONSEN, використовуючи BLAST в Ensembl Plants; відібрати щонайменше 5 значущих збігів.
- Для трьох відібраних локусів експортувати геномні фрагменти з розширенням не менше ± 8 тис. п.н. від межі вирівнювання (щоб охопити повну вставку та фланкуючі ділянки).
- Виконати структурне передбачення LTR-ретротранспозонів на кожному фрагменті (LTRharvest або альтернативний інструмент у Galaxy) та визначити межі 5'- та 3'-LTR.
- Оцінити інтактність кожної копії: наявність двох LTR, цільових дуплікацій (TSD), внутрішньої ділянки та кодувального потенціалу (gag/pol; ORFfinder).
- Порівняти 5'- та 3'-LTR кожної копії (глобальне вирівнювання) і за відсотком ідентичності оцінити відносний вік вставок; за бажання – розрахувати орієнтовний час інтеграції за формулою: $T = K/(2\mu)$.
- Оформити результати у вигляді таблиці та сформулювати висновки про різноманіття копій ONSEN у геномі та їхній можливий функціональний стан.

Хід роботи

Підготовка референтної моделі родини ONSEN (ATCOPIA78)

1. У базі Dfam виконують пошук родини за назвами-синонімами **ONSEN** і **ATCOPIA78** та відбирають запис, що відповідає *Arabidopsis thaliana* і класифікується як **LTR-ретротранспозон типу Ty1/copia**.
2. Для обраного запису фіксують:

- назву родини;
 - accession або family ID;
 - класифікацію;
 - довжину консенсусної послідовності;
 - тип представленої моделі.
3. Консенсусну нуклеотидну послідовність завантажують у форматі **FASTA** і використовують як єдиний референтний запит для подальшого пошуку в геномі.
 4. У робочому журналі окремо зазначають, який саме запис Dfam було використано, щоб уникнути плутанини між синонімами родини та забезпечити відтворюваність аналізу.
 5. Довжину консенсусної послідовності заносять до протоколу як базову характеристику родини, з якою надалі порівнюють довжину вирівнювань і повноту знайдених копій.

Пошук гомологічних локусів ONSEN у геномі *Arabidopsis thaliana*

1. Пошук виконують у **Ensembl Plants** для ***Arabidopsis thaliana*** на референтній збірці **TAIR10**, використовуючи саме **BLASTN**, а не BLAT.
2. Як мішень обирають **геномну ДНК референтного геному**, а не транскриптом або білкові бази.
3. Для цієї лабораторної доцільно використовувати **немасковану геномну послідовність**, щоб не втрачати повторні та TE-подібні ділянки.
4. Як запит використовують повну консенсусну послідовність ONSEN у форматі FASTA.
5. Якщо результати виявляються надто фрагментованими, допускається повторний запуск із підвищеною чутливістю, але основний відбір локусів здійснюють на одному узгодженому режимі пошуку.
6. Із вихідного списку збігів відбирають **щонайменше 5 значущих локусів**.
7. Перевагу надають збігам, для яких характерні:
 - високе покриття запиту;

- статистично значущі E-value;
 - велика довжина вирівнювання;
 - чітка локалізація в одному геномному інтервалі.
8. Якщо один і той самий локус представлений кількома перекривними HSP, їх спочатку об'єднують на рівні локусу, а потім враховують як **один кандидат**, а не як кілька незалежних копій.
9. Для кожного відібраного локусу фіксують:
- хромосому;
 - початкову та кінцеву координати вирівнювання;
 - орієнтацію;
 - відсоток ідентичності;
 - довжину вирівнювання;
 - E-value.
10. На цьому етапі не роблять висновок про повнорозмірність елемента лише за BLAST-збігом; BLAST використовують для **локалізації кандидатних копій**, а не для остаточного структурного опису.

Відбір локусів для структурного аналізу

1. Із щонайменше 5 знайдених локусів для детального аналізу відбирають **3 репрезентативні ділянки**:
 - щонайменше 2 локуси з високим покриттям консенсусу;
 - 1 локус із частковим або очевидно деградованим збігом.
2. Для кожного з трьох локусів формують геномний інтервал, який охоплює не лише зону BLAST-вирівнювання, а й фланкуючі послідовності – **не менше ніж ± 8 тис. п.н.** від крайніх координат вирівнювання.
3. Якщо вирівнювання розташоване близько до краю попередньо обраного вікна або є підозра, що елемент довший за очікуване початкове вікно, інтервал розширюють доти, доки потенційна повна вставка разом із фланками не буде повністю охоплена.

4. Кожний експортований фрагмент зберігають окремим FASTA-файлом і називають так, щоб у назві були однозначно відображені:
 - хромосома;
 - координатний інтервал.
5. У журналі для кожного локусу окремо фіксують зв'язок між BLAST-координатами та ширшим експортованим інтервалом, щоб надалі не змішувати координати вирівнювання з координатами передбаченого LTR-елемента.

Структурне передбачення меж LTR-ретротранспозона

1. Структурне передбачення виконують у **Galaxy Europe** за допомогою актуально доступного інструмента для пошуку LTR-ретротранспозонів.
2. Якщо в поточному інтерфейсі відсутній окремий запуск **LTRharvest**, використовують доступний еквівалентний LTR-предиктор або зв'язку, у якій задіяні **LTRharvest/LTR_retriever**.
3. Для кожного з трьох FASTA-фрагментів виконують окремий аналіз, щоб результати по локусах не змішувалися.
4. Як стартові критерії використовують параметри, сумісні з повнорозмірним **LTR-ретротранспозоном родини соріа**:
 - мінімальна довжина LTR – близько 100 п.н.;
 - максимальна довжина LTR – до 2000 п.н.;
 - мінімальна подібність між LTR – близько 80%;
 - допустима відстань між LTR – у межах, що дозволяють охопити внутрішню ділянку типового соріа-елемента.
5. Якщо назви параметрів у поточному інтерфейсі відрізняються від класичних назв LTRharvest, добирають еквівалентні за змістом параметри:
 - межі довжини LTR;
 - мінімальну схожість LTR;
 - допустиму довжину внутрішньої ділянки;
 - загальну довжину елемента.

6. Із набору кандидатів для кожного локусу залишають той елемент, який:
 - найкраще узгоджується з BLAST-збігом ONSEN;
 - має дві LTR у правильній взаємній орієнтації;
 - містить внутрішню ділянку між ними.
7. Для обраного кандидата фіксують:
 - повні межі вставки;
 - координати 5'-LTR;
 - координати 3'-LTR;
 - довжину кожної LTR;
 - довжину внутрішньої ділянки;
 - за наявності – ознаки PBS, PPT або TSD.
8. Якщо жоден кандидат не відповідає очікуваній будові повнорозмірної сорія-вставки, локус далі аналізують як **частковий або деградований елемент**.

Оцінювання інтактності копій ONSEN

1. Для кожного з трьох локусів оцінюють:
 - чи присутні обидві LTR;
 - чи мають вони порівнянні довжини;
 - чи зберігається внутрішня ділянка між ними;
 - чи немає у вставки очевидних ознак деградації.
2. До ознак структурної інтактності відносять:
 - наявність 5'-LTR і 3'-LTR;
 - безперервну внутрішню ділянку;
 - симетрично розташовані межі вставки;
 - за можливості – короткі цільові дуплікації по краях.
3. До ознак деградації відносять:
 - великі внутрішні делеції;
 - втрату однієї з LTR;

- укорочення елемента;
 - сильну фрагментацію;
 - численні розриви;
 - явні порушення безперервності внутрішньої ділянки.
4. Якщо інструмент структурного передбачення не подає TSD автоматично, короткі фланкуючі послідовності безпосередньо поза межами 5'- та 3'-кінців порівнюють вручну.
 5. TSD фіксують лише тоді, коли дуплікація виглядає однозначною.
 6. За підсумком цього етапу кожен копію попередньо класифікують як:
 - імовірно інтактну повнорозмірну;
 - частково деградовану;
 - фрагментарну.

Оцінювання кодувального потенціалу внутрішньої ділянки

1. Для кожної копії окремо виділяють внутрішню ділянку між 5'-LTR і 3'-LTR та аналізують її в **ORFfinder**.
2. Аналіз виконують для обох ланцюгів або принаймні для того ланцюга, який узгоджується з орієнтацією елемента.
3. У протоколі фіксують:
 - кількість знайдених ORF;
 - координати ORF;
 - орієнтацію;
 - рамку зчитування;
 - довжину.
4. Особливу увагу приділяють довгим ORF, сумісним із **gag/pol-подібним кодувальним простором**.
5. ORF довжиною понад **300 амінокислот** розглядають як інформативні для непрямої оцінки кодувального потенціалу, але не трактують їх як прямий доказ функціональної автономності без додаткового доменного аналізу.

6. Якщо у внутрішній ділянці є лише короткі ORF або ORF численно перериваються стоп-кодонами, це трактують як ознаку деградації кодувального потенціалу.
7. Якщо довгі ORF займають значну частину внутрішньої області та узгоджуються з очікуваною організацією соріа-елемента, копію описують як **структурно більш інтактну** порівняно з фрагментарними або ORF-дефектними копіями.

Порівняння 5'- та 3'-LTR і відносна оцінка віку вставок

1. Для кожної копії окремо експортують 5'-LTR і 3'-LTR за координатами, визначеними на етапі структурного передбачення.
2. Пари LTR вирівнюють глобально за допомогою **EMBOSS Needle** або еквівалентного інструмента глобального вирівнювання.
3. Для кожної пари фіксують:
 - довжину вирівнювання;
 - відсоток ідентичності;
 - кількість розбіжностей;
 - наявність інделів.
4. Вищу ідентичність між 5'- та 3'-LTR інтерпретують як ознаку **відносно молодшої вставки**, а нижчу – як ознаку **давнішої вставки**.
5. У цій лабораторній основний акцент роблять саме на **відносному порівнянні віку** між копіями, а не на абсолютному датуванні.
6. Якщо обчислюють орієнтовний час інтеграції за формулою $T = K/(2\mu)$, у звіті обов'язково вказують:
 - як саме було отримано K;
 - яке значення μ використано;
 - з якого джерела воно взят;
 - що така оцінка є наближеною.

7. Якщо обґрунтоване значення μ не задане або не процитоване окремо, абсолютний час інтеграції не розраховують і обмежуються ранжуванням копій за відносною «молодістю».

Уніфікація інтерпретації для трьох локусів

1. Після завершення структурного та ORF-аналізу результати трьох локусів зіставляють між собою не лише за BLAST-подібністю до консенсусу, а й за:
 - повнотою будови;
 - наявністю довгих ORF;
 - схожістю парних LTR.
2. Локус із двома добре визначеними LTR, суцільною внутрішньою ділянкою, наявними довгими ORF і дуже високою ідентичністю між 5'- та 3'-LTR розглядають як найкращого кандидата на **відносно молоду й структурно найбільш інтактну копію**.
3. Локус із двома LTR, але частково порушеним кодувальним потенціалом або зниженою ідентичністю LTR трактують як **старішу або частково деградовану копію**.
4. Локус без чітко визначених двох LTR, без повної внутрішньої ділянки або з короткими розірваними ORF класифікують як **фрагмент або деградований залишок елемента**.
5. На цьому етапі не змішують поняття «молодша копія» і «активна копія»: низька дивергенція LTR є аргументом на користь молодості вставки, але не є самостійним доказом сучасної транспозиційної активності.

Оформлення результатів

1. Результати оформлюють у підсумковій таблиці щонайменше для трьох локусів.

2. До таблиці включають такі поля:

- локус;
- хромосома;
- координати BLAST-збігу;
- координати повної вставки;
- загальна довжина елемента;
- довжина 5'-LTR;
- довжина 3'-LTR;
- довжина внутрішньої ділянки;
- відсоток ідентичності між 5'- та 3'-LTR;
- наявність або відсутність TSD;
- кількість ORF;
- довжина найдовшої ORF;
- висновок щодо інтактності.

3. Для кожного локусу додають короткий аналітичний коментар:

- повнорозмірна чи часткова копія;
- чи є ознаки збереженого кодувального потенціалу;
- чи виглядає копія відносно молодою або давнішою.

4. У підсумкових висновках окремо зазначають:

- чи є серед проаналізованих локусів структурно інтактні копії ONSSEN;
 - чи спостерігається різноманіття за довжиною, інтактністю та дивергенцією LTR;
 - яка копія виглядає наймолодшою;
 - яка копія виглядає найбільш деградованою.

5. Якщо отримані результати не узгоджуються з очікуванням щодо повнорозмірної копії, у висновку це фіксують прямо і пояснюють через:

- можливу деградацію елемента;
- обмеження структурного предиктора;
- особливості локального геномного контексту;

- розбіжності між анотаційними підходами.

Контрольні запитання

1. Поясніть різницю між ДНК-транспозонами та LTR-ретротранспозонами (молекулярний проміжний продукт, механізм копіювання).
2. Які структурні компоненти характерні для повнорозмірної копії LTR-ретротранспозона (LTR, PBS, PPT, gag/pol)?
3. Що таке цільові дуплікації (TSD) і чому вони утворюються під час інтеграції?
4. Що означає «інтактність» мобільного елемента у контексті геному? Які ознаки вказують на деградацію копії?
5. Чому для оцінки віку LTR-ретротранспозона порівнюють 5'- та 3'-LTR однієї копії, а не різні копії між собою?
6. Які обмеження має оцінка віку за формулою $T = K/(2\mu)$? Назвіть щонайменше три джерела похибки.
7. Яку інформацію дає ORFfinder у цій роботі? Чому наявність довгих ORF важлива, але не є прямим доказом активності?
8. Наведіть приклад, як вставка LTR-ретротранспозона може впливати на експресію сусідніх генів.

Рекомендована література

1. Cavrak, V. V., Lettner, N., Jamge, S., Kosarewicz, A., Bayer, L. M., & Scheid, O. M. (2014). How a retrotransposon exploits the plant's heat stress response for its activation. *PLoS Genetics*, 10(1), e1004115.
2. <https://doi.org/10.1371/journal.pgen.1004115>
3. Ellinghaus, D., Kurtz, S., & Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics*, 9, 18.
4. <https://doi.org/10.1186/1471-2105-9-18>

5. Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., & Paszkowski, J. (2011). An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature*, 472(7341), 115–119.
6. <https://doi.org/10.1038/nature09861>
9. Ito, H., Yoshida, T., Tsukahara, S., & Kawabe, A. (2013). Evolution of the ONSEN retrotransposon family activated upon heat stress in Brassicaceae. *Gene*, 518(2), 256–261.
10. <https://doi.org/10.1016/j.gene.2013.01.034>
11. Jedlicka, P., Lexa, M., & VanBuren, R. (2020). What can long terminal repeats tell us about the age of LTR retrotransposons, gene conversion and ectopic recombination? *Frontiers in Plant Science*, 11, 644.
12. <https://doi.org/10.3389/fpls.2020.00644>
13. Roquis, D., Robertson, M., Yu, L., Thieme, M., Julkowska, M., & Bucher, E. (2021). Genomic impact of stress-induced transposable element mobility in *Arabidopsis*. *Nucleic Acids Research*, 49(18), 10431–10447.
14. <https://doi.org/10.1093/nar/gkab828>

Практична робота 7: Аналіз поліморфних ретротранспозонних вставок (Alu, LINE-1, SVA) у геномі Homo sapiens за допомогою dbRIP та UCSC Genome Browser

Мета роботи

Навчитися знаходити та інтерпретувати поліморфні вставки мобільних генетичних елементів у геномі людини за допомогою геномного браузера UCSC. Опрацювати практику підключення трек-хабів, відбору кандидатних локусів, опису структурних ознак ретротранспозонних вставок і оцінювання їхнього поліморфізму (наявність/відсутність у популяціях, частота та генотипи), а також аналізу геномного контексту вставок.

Теоретичні відомості

У геномі людини активними або нещодавно активними залишаються три основні родини не-LTR-ретротранспозонів: **LINE-1 (L1)**, **Alu** та **SVA**. Із них лише L1 є автономним елементом, оскільки кодує білки, необхідні для власної ретротранспозиції, тоді як Alu і SVA є неавтономними та мобілізуються у trans за рахунок ферментативного апарату LINE-1. Усі три класи поширюються за механізмом «copy-and-paste» через РНК-проміжний продукт, а тому формують особливий різновид структурної мінливості геному – вставки мобільних елементів, або mobile element insertions (MEI).

Молекулярною основою інтеграції LINE-1 та залежних від нього елементів є target-primed reverse transcription, або TPRT. Під час цього процесу ендонуклеаза ORF2p LINE-1 розщеплює цільову ДНК, утворюючи 3'-гідроксильну групу, яка слугує праймером для зворотної транскрипції РНК-матриці безпосередньо в місці майбутньої вставки. Саме тому для L1-опосередкованих інтеграцій характерні короткі цільові дуплікації по обидва боки вставки, 3'-полі(А)-хвіст, локалізація поблизу послідовностей, подібних до консенсусного сайту розщеплення L1-ендонуклеази 5'-TTTT/AA-3', а також часті 5'-усічення; для частини L1-вставок і деяких похідних подій описано

також 5'-інверсії та 3'-трансдукції. Ці ознаки є важливими структурними «слідами» механізму інтеграції й тому входять до стандартного опису людських поліморфних ретротранспозонних вставок.

Кожен із трьох класів має власні характерні структурні риси.

Повнорозмірний людський **L1** має довжину близько 6 кб і містить 5'-UTR із внутрішнім промотором, дві відкриті рамки зчитування ORF1 та ORF2 і 3'-UTR із поліаденілювальним сигналом.

Alu – це короткий SINE-елемент довжиною приблизно 300 п.н., похідний від 7SL РНК, з внутрішніми А- та В-бокс-промоторами РНК-полімерази III та 3'-полі(А)-хвостом. **SVA** є складеним елементом, що включає 5'-гексамерні повтори, Alu-подібний модуль, GC-багатий VNTR-регіон, SINE-R-ділянку, похідну від HERV-K, і полі(А)-хвіст; канонічні **SVA** зазвичай мають довжину близько 2 кб, хоча реальні копії можуть істотно варіювати за розміром. Саме ці архітектурні відмінності лежать в основі класифікації вставок у ресурсах на кшталт dbRIP або RepeatMasker.

Поліморфізм вставки означає, що певний локус присутній у вигляді двох алельних станів: з інсерцією та без неї. На відміну від SNP або коротких інделів, MEI зазвичай відображає одиничну історичну подію інтеграції, тому такі варіанти часто розглядають як маркери identical-by-descent і як майже гомоплазійно-вільні маркери популяційної генетики. Водночас окремі поліморфні вставки мають не лише еволюційне, а й медичне значення, оскільки можуть порушувати структуру генів, змінювати сплайсинг, модулювати експресію або бути безпосередньою причиною моногенних захворювань. Саме тому аналіз частот, генотипів і популяційного розподілу ретротранспозонних вставок має одночасно популяційно-генетичне й біомедичне значення.

dbRIP є спеціалізованим ресурсом для документування людських поліморфних ретротранспозонних вставок, інтегрованим із анотаційною екосистемою UCSC Genome Browser. Офіційний опис бази вказує, що її метою

є збирання опублікованих локусів поліморфних вставок Alu, L1 і SVA, а сучасний dbRIP Track Hub для UCSC містить окремі підтреки Polymorphic Alu, Polymorphic L1 і Polymorphic SVA, доступні принаймні для hg19 та hg38. За даними сторінок прикладів і хабу, окремий запис може містити не лише координати та класифікацію елемента, а й фланкуючі послідовності, TSD, характеристику самого елемента, генний контекст, праймери, відомості про асоціацію із захворюванням і посилання на першоджерела. Отже, dbRIP слід розуміти не як «колер» нових вставок у конкретному геномі, а як курований каталог уже описаних локусів, які потім інтерпретуються у браузерному контексті UCSC.

UCSC Genome Browser є інтегративним графічним середовищем, у якому геномні координати поєднуються з багатьма шарами анотацій, зокрема генними моделями, повторами, CpG-островами, регуляторними сигналами та зовнішніми хабами. Сам браузер не формулює біологічних висновків автоматично, а надає можливість зіставляти незалежні джерела даних у межах одного інтерфейсу. Для зовнішніх наборів даних UCSC використовує track hubs – віддалені веб-доступні каталоги анотацій, що завантажуються в браузер за потребою; дані таких хабів можна переглядати так само, як і вбудовані треки, включати до збережених сесій і аналізувати через Table Browser. Table Browser, своєю чергою, забезпечує текстовий доступ до тих самих анотаційних таблиць, дозволяє відбирати записи за координатами або атрибутами, експортувати їх у табличних форматах і комбінувати результати між треками. Для навчального аналізу поліморфних вставок це принципово важливо, оскільки дає змогу спочатку відібрати кандидатні локуси масово, а потім перейти до поглибленої візуальної інтерпретації кожного з них.

Функціональна **інтерпретація вставки** вимагає урахування її геномного контексту. Якщо локус перекриває екзон, інтрон, промоторну область або розташований поблизу точки старту транскрипції, потенційний вплив на функцію гена є вищим, ніж у віддаленій міжгенній ділянці. RepeatMasker у UCSC дає змогу перевірити, чи узгоджується локальна

повторна анотація з типом вставки та чи не накладається вона на інші повтори або низькоскладні послідовності, що можуть ускладнювати трактування координат і меж елемента. Додаткові треки, зокрема генні моделі, CpG Islands і регуляторні анотації, дозволяють оцінити, чи може вставка бути пов'язана зі зміною регуляції експресії, зі сплайсинговими наслідками або, навпаки, найімовірніше має нейтральну локалізацію. Однак будь-який такий висновок у межах браузерного аналізу є інтерпретаційним і потребує чіткого розмежування між анотаційним збігом і експериментально доведеним функціональним ефектом.

Порівняння записів dbRIP з незалежними ресурсами структурних варіантів, такими як трек **gnomAD Structural Variants** v4.1 у UCSC, дає змогу оцінити відтворюваність локусу в іншому аналітичному середовищі, але не гарантує повного збігу позначень. За описом UCSC, трек gnomAD SV v4.1 відображає структурні варіанти розміром не менше 50 нуклеотидів із 63 046 неспоріднених геномів і за замовчуванням показує лише варіанти, що пройшли фільтр якості; серед типів подій окремо представлені й INS-вставки. Розбіжності між dbRIP і gnomAD можуть бути зумовлені різними вибірками, різною метою ресурсів, відмінностями алгоритмів виявлення, фільтрами якості, способом представлення брейкпоінтів і тим, що повторні та низькомаповні ділянки є складними для short-read-детекції. Огляди з аналізу mobile element variation підкреслюють, що short-read-підходи істотно залежать від мапованості, часто гірше відновлюють точну послідовність не-референсних копій і пропускають варіанти у складних повторних регіонах, тоді як long-read- або гаплотипно-асамблейні підходи значно краще розв'язують внутрішню структуру вставок, точні межі та вкладені варіації всередині самих мобільних елементів. Тому відсутність локусу в одному ресурсі не слід автоматично трактувати як його біологічну відсутність, якщо різняться технологія, аналітична схема й критерії представлення варіанта. Таким чином, аналіз поліморфних вставок Alu, L1 і SVA у цій лабораторній роботі поєднує два взаємодоповнювальні рівні. Перший рівень є молекулярно-

генетичним і стосується природи MEI як наслідку L1-опосередкованої ретротранспозиції з характерними структурними ознаками – TSD, полі(A)-хвостом, 5'-усіченням та інколи 5'-інверсією чи трансдукцією. Другий рівень є інформаційно-анотаційним і полягає в інтерпретації вже каталогізованих локусів у координатному просторі геному hg38 за допомогою UCSC Genome Browser, Track Hubs, Table Browser та суміжних треків. Саме поєднання цих двох рівнів дозволяє коректно переходити від простого факту наявності запису про вставку до науково обґрунтованого висновку про її тип, структурні особливості, популяційний поліморфізм і можливий функціональний контекст.

Матеріали та методи

- Комп'ютер з доступом до Інтернету.
- UCSC Genome Browser (збірка GRCh38/hg38): <https://genome.ucsc.edu/cgi-bin/hgGateway?db=hg38>
- Track Hub Registry / Track Data Hubs (підключення dbRIP_Hub): <https://genome.ucsc.edu/cgi-bin/hgHubConnect>
- dbRIP (опис бази та довідкові матеріали): <https://dbrip.brocku.ca/>
- Вбудовані треки UCSC для інтерпретації контексту: RefSeq Genes або GENCODE; RepeatMasker; CpG Islands (за потреби).
- Опційно: gnomAD Structural Variants v4.1 (для порівняння незалежного SV-ресурсу; доступний як трек UCSC для hg38).

Завдання

- Підключити трек-хаб dbRIP_Hub у UCSC Genome Browser (hg38) та увімкнути треки поліморфних вставок Alu, L1 та/або SVA.
- Відібрати щонайменше 5 поліморфних вставок одного типу (на вибір: Alu або L1 або SVA) у різних геномних контекстах.

- Для кожної вставки описати: координати (hg38), тип/підродину елемента, орієнтацію, довжину (якщо подано), ознаки інтеграції (TSD, полі(A), 5'-усічення – якщо вони наведені у деталях запису).
- Проаналізувати поліморфізм: наявність/відсутність у вибірці, частоти або генотипи (якщо подано), та сформулювати інтерпретацію отриманих показників.
- Оцінити функціональний контекст вставки (міжгенна/інтронна/екзонна, відстань до TSS, перекриття з відомими регуляторними ділянками – за треками UCSC).
- Для однієї обраної вставки виконати порівняння з альтернативним SV-ресурсом (наприклад, gnomAD SV) і пояснити можливі причини збігів/розбіжностей.

Хід роботи

Підготовка робочого середовища та вибір збірки геному

1. Для всіх етапів аналізу використовують збірку геному людини GRCh38/hg38 як єдину координатну систему.
2. На старті роботи в UCSC Genome Browser вмикають базові треки, необхідні для інтерпретації локусів:
 - NCBI RefSeq Genes або GENCODE Genes;
 - RepeatMasker;
 - за потреби – CpG Islands та доступні регуляторні треки.
3. Усі подальші координати, скриншоти, таблиці та висновки подають лише в нотації **hg38**, без змішування з hg19 або іншими збірками.
4. У робочому журналі окремо фіксують:
 - використану збірку;
 - набір активованих треків;
 - дату перегляду браузерних анотацій.

5. До початку відбору вставок задають однаковий підхід до масштабу перегляду локусів, щоб структурні та функціональні ознаки оцінювалися в порівнюваних вікнах.

Підключення dbRIP_Hub та вибір типу поліморфних вставок

1. У UCSC Track Data Hubs підключають dbRIP_Hub для збірки hg38.
2. Після підключення в браузері активують лише той підтрек, який відповідає обраному типу вставок:
 - Polymorphic Alu;
 - або Polymorphic L1;
 - або Polymorphic SVA.
3. Для однієї лабораторної роботи аналізують лише один тип вставок, щоб підсумкова вибірка була методично однорідною.
4. Решту підтреків dbRIP тимчасово приховують, щоб не змішувати різні класи мобільних елементів при візуальному перегляді.
5. Режим відображення для активного підтреку встановлюють таким, щоб були видимі окремі записи, а не лише сумарна щільність сигналу.
6. У журналі зазначають:
 - назву активного підтреку;
 - клас вставок;
 - чи використовувався додатково будь-який інший dbRIP-підтрек для перевірки контексту.

Отримання переліку кандидатних локусів

1. Для системного відбору записів використовують Table Browser UCSC, а не лише ручний перегляд окремих ділянок у графічному браузері.
2. У Table Browser задають:
 - геном Human;
 - збірку hg38;
 - відповідний підтрек dbRIP для обраного класу вставок.

3. Для первинного відбору локусів формують список щонайменше з 20 записів у межах однієї хромосоми або кількох обмежених інтервалів.
4. Вивантаження виконують у табличному форматі, придатному для подальшого ручного сортування.
5. Якщо в межах одного регіону записи явно кластеризовані або належать до подібного геномного оточення, вибірку розширюють на інші ділянки геному, щоб уникнути одноманітного контексту.
6. На цьому етапі не відбирають вставки випадково: попередній список потрібен для цілеспрямованого добору локусів із різними функціональними контекстами.
7. У робочому журналі фіксують:
 - межі регіону, з якого брали записи;
 - кількість попередньо переглянутих вставок;
 - критерії, за якими з цього списку відбирали фінальні 5 локусів.

Відбір п'яти вставок для детального аналізу

1. Із попереднього списку обирають 5 поліморфних вставок одного типу.
2. Вибірку формують так, щоб локуси відрізнялися геномним контекстом.
Бажано включити:
 - міжгенну вставку;
 - інтронну вставку;
 - вставку поблизу TSS або промоторної ділянки;
 - вставку в області з високою щільністю повторів;
 - вставку в гені з альтернативними транскриптами.
3. Якщо окремий локус одночасно відповідає кільком категоріям, його можна залишити у вибірці, але загалом набір має охоплювати різні типи контексту.
4. Для кожної кандидатної вставки перевіряють, чи відкривається сторінка деталей запису і чи містить вона достатній мінімум інформації для опису поліморфізму.

5. Якщо запис занадто бідний на анотації для виконання завдання, його замінюють іншим локусом до початку основного аналізу.
6. Остаточний набір із 5 вставок фіксують у чернетці таблиці ще до детального перегляду, щоб уникнути довільної заміни локусів уже після початку інтерпретації.

Опис базових структурних параметрів кожної вставки

1. Для кожної з 5 відібраних вставок відкривають сторінку деталей запису в dbRIP-треку та виписують:
 - ідентифікатор запису;
 - координати у форматі hg38;
 - тип мобільного елемента;
 - підродину;
 - орієнтацію;
 - довжину, якщо вона подана у записі.
2. Додатково перевіряють, чи наведені для конкретної вставки структурні ознаки інтеграції:
 - TSD;
 - полі(A)-хвіст;
 - 5'-усічення;
 - інші примітки щодо повноти вставки.
3. Якщо конкретний запис не містить частини цих полів, у таблиці це позначають як «не подано», а не замінюють припущенням.
4. Після цього локус звіряють із RepeatMasker, щоб перевірити:
 - чи відповідає локальна repeat-анотація заявленому типу вставки;
 - чи не перекриває запис інший транспозон;
 - чи не лежить вставка в низькоскладній або важко інтерпретованій повторній ділянці.

5. Якщо між записом dbRIP і локальною RepeatMasker-анотацією є розбіжність, її не виправляють самостійно, а описують у примітках як окреме спостереження.
6. Для кожної вставки формують короткий структурний підсумок: повна/усічена, добре охарактеризована/частково описана, проста/в складному повторному оточенні.

Аналіз поліморфізму вставок

1. Для кожної вставки зі сторінки деталей запису виписують усю доступну інформацію про поліморфізм:
 - наявність або відсутність у вибірці;
 - алельні підрахунки;
 - генотипи;
 - частоти в популяціях або підвибірках, якщо вони подані.
2. Якщо для запису наведено AC та AN, частоту вставки обчислюють як $AF = AC / AN$.
3. Якщо в записі подані генотипи, окремо оцінюють:
 - частку гетерозигот;
 - частку гомозигот за вставкою;
 - частку носіїв референсного стану без вставки.
4. Для кожної вставки формулюють коротку інтерпретацію отриманих показників:
 - рідкісна;
 - низькочастотна;
 - помірно поширена;
 - поширена у вибірці.
5. Якщо в записі подано кілька популяційних підрахунків, порівнюють їх між собою та відзначають, чи виглядає вставка рівномірно поширеною або популяційно зміщеною.

6. Якщо кількісні дані про поліморфізм відсутні, це прямо зазначають у таблиці і не підмінюють якісною оцінкою «ймовірно рідкісна» без джерела.
7. В інтерпретації враховують, що відмінності між записами можуть пояснюватися не лише реальною популяційною структурою, а й дизайном вибірки, методом виявлення вставок, фільтрами якості та неповнотою доступних полів.

Оцінювання геномного та функціонального контексту

1. Для кожної вставки визначають, у якому геномному контексті вона локалізована:
 - міжгенна ділянка;
 - інтрон;
 - екзон;
 - зона поблизу TSS;
 - область із перекриттям регуляторних елементів.
2. Для цього локус послідовно зіставляють із:
 - RefSeq Genes або GENCODE Genes;
 - RepeatMasker;
 - за потреби – CpG Islands та іншими доступними регуляторними треками.
3. Якщо вставка потрапляє в межі гена, окремо зазначають:
 - назву гена;
 - чи розташована вона в інтроні чи екзоні;
 - чи є в цій області альтернативні транскрипти.
4. Якщо вставка лежить поза межами гена, для інтерпретації визначають її відносне положення щодо найближчого TSS або найближчого анотованого гена.
5. Якщо вставка перекриває острівці CpG або регуляторний елемент, це не трактують автоматично як функціональний ефект, а формулюють як потенційно значущий контекст.

6. Для кожного локусу формують короткий висновок про можливий функціональний вплив:
 - імовірно нейтральний міжгенний контекст;
 - потенційний вплив на регуляцію експресії;
 - потенційний вплив на сплайсинг;
 - контекст складний і потребує обережної інтерпретації.
7. Усі такі висновки подають як аналітичні припущення, що впливають із геномного положення, а не як доведений біологічний ефект.

Порівняння однієї вставки з альтернативним SV-ресурсом

1. Для однієї з п'яти вставок виконують окреме порівняння з незалежним SV-ресурсом у UCSC, використовуючи трек gnomAD Structural Variants v4.1.
2. Перед порівнянням у треку gnomAD Structural Variants залишають відображення insertions (INS), щоб не змішувати вставки з делеціями, дуплікаціями чи іншими класами SV.
3. Для обраного локусу перевіряють:
 - чи присутній у цій області запис типу INS;
 - наскільки близькі координати;
 - чи узгоджується розмір варіанта;
 - чи доступні частотні показники для порівняння.
4. Якщо в gnomAD знаходиться відповідний INS-запис із близькими координатами, це описують як підтримку локусу незалежним ресурсом, але не як доказ повної тотожності варіантів.
5. Якщо відповідного запису в gnomAD немає, можливі причини розбіжності пояснюють через:
 - різний склад вибірок;
 - різні алгоритми виявлення;
 - різні критерії фільтрації;
 - неточність локалізації брейкпоінтів;
 - різне трактування одного й того самого сигналу як INS або іншого SV.

6. Якщо відповідний запис у gnomAD є, але його частота помітно відрізняється від dbRIP, це також фіксують окремо і пояснюють через відмінності джерел даних та принципів формування ресурсів.
7. У висновку до цього етапу не ставлять знак рівності між dbRIP-записом і gnomAD SV без явного збігу координатного інтервалу, типу варіанта та розумного узгодження за довжиною.

Оформлення висновків

1. Підсумкові результати подають у таблиці щонайменше для 5 вставок одного обраного типу.
2. До таблиці включають такі поля:
 - №;
 - координати (hg38);
 - тип;
 - підродина;
 - орієнтація;
 - довжина, bp;
 - геномний контекст;
 - TSD/полі(A)/5'-усічення;
 - AF або генотипи;
 - коротка примітка.
3. Для однієї з вставок у примітці окремо вказують результат порівняння з gnomAD Structural Variants v4.1.
4. Для щонайменше двох локусів до звіту додають:
 - оглядовий скриншот регіону;
 - збільшений скриншот рівня вставки.
5. Підписи до скриншотів мають містити:
 - координати локусу;
 - тип вставки;
 - перелік увімкнених треків, які використовувалися для інтерпретації.

6. У загальному висновку до лабораторної окремо формулюють:

- які геномні контексти трапилися серед обраних вставок;
- як змінювалися частоти або типи поліморфізму;
- наскільки добре структурні ознаки вставки були відображені в записах;
- які обмеження має аналіз, виконаний лише за браузерними анотаціями.

Контрольні запитання

1. Що таке поліморфізм вставок і чим він відрізняється від SNP/індел-поліморфізму?
2. Які родини ретротранспозонів у людини вважають активними або нещодавно активними? Наведіть приклади підродин.
3. Опишіть механізм TPRT та назвіть структурні «сліди» інтеграції (TSD, полі(A), 5'-усічення).
4. Чому LINE-1-вставки часто є неповними з 5'-кінця? Які наслідки це має для функціонального стану копії?
5. Які труднощі виникають під час виявлення MEI за даними short-read секвенування? Які переваги дають long-read підходи?
6. Що таке трек-хаб у UCSC Genome Browser і які задачі він спрощує в порівнянні з локальними файлами анотацій?
7. Назвіть щонайменше два можливі джерела розбіжностей між dbRIP і гномадівськими SV-анотаціями для однієї й тієї самої області.

Рекомендована література

1. Wang, J., Song, L., Grover, D., Azrak, S., Batzer, M. A., & Liang, P. (2006). dbRIP: A highly integrated database of retrotransposon insertion polymorphisms in humans. *Human Mutation*, 27(4), 323–329. <https://doi.org/10.1002/humu.20307>
2. Stewart, C., Kural, D., Strömberg, M. P., Walker, J. A., Konkel, M. K., Stütz, A. M., Urban, A. E., Grubert, F., Lam, H. Y. K., Lee, W.-P., Busby, M., Indap,

- A. R., Garrison, E., Huff, C. D., Xing, J., Snyder, M. P., Jorde, L. B., Batzer, M. A., Korbel, J. O., & Marth, G. T. (2011). A comprehensive map of mobile element insertion polymorphisms in humans. *PLoS Genetics*, 7(8), e1002236. <https://doi.org/10.1371/journal.pgen.1002236>
3. Hancks, D. C., & Kazazian, H. H., Jr. (2016). Roles for retrotransposon insertions in human disease. *Mobile DNA*, 7, 9. <https://doi.org/10.1186/s13100-016-0065-9>
 4. Solyom, S., & Kazazian, H. H., Jr. (2012). Mobile elements in the human genome: Implications for disease. *Genome Medicine*, 4, 12. <https://doi.org/10.1186/gm311>
 5. Wang, H., Xing, J., Grover, D., Hedges, D. J., Han, K., Walker, J. A., & Batzer, M. A. (2005). SVA elements: A hominid-specific retroposon family. *Journal of Molecular Biology*, 354(4), 994–1007. <https://doi.org/10.1016/j.jmb.2005.09.085>
 6. Devine, S. E., Gardner, E. J., Steely, C. J., & Hancks, D. C. (2023). Emerging opportunities to study mobile element insertions and their source elements in human populations. *Mobile DNA*, 14, 25. <https://doi.org/10.1186/s13100-023-00321-0>

Практична робота 8: Біоінформатичний підбір праймерів для детекції Р-елементу у *Drosophila melanogaster*

Мета роботи

Навчитися підбирати праймери до транспозона Р (Р-елемент) та оцінювати їхню придатність для полімеразної ланцюгової реакції (ПЛР) за сукупністю критеріїв: специфічність у референсному геномі, параметри плавлення (T_m), GC-вміст, відсутність критичних вторинних структур, а також прогнозований розмір амплікона. За результатами роботи сформувані мінімум дві оптимальні пари праймерів: для внутрішньої детекції Р-елементу та (за бажанням) для детекції вставки за принципом «геном-мішень + Р-елемент».

Теоретичні основи

Р-елемент у *Drosophila melanogaster* є ДНК-транспозоном II класу, що переміщується за механізмом «cut-and-paste». Повнорозмірний елемент має довжину близько **2,9 кб**, кодує транспозазу, містить **31-п.н. термінальні інвертовані повтори**, а для транспозиції критичними є приблизно 150 п.н. на кожному кінці. Під час інтеграції Р-елемент формує 8-п.н. дуплікацію сайту-мішені. Саме ці особливості визначають логіку дизайну праймерів: для внутрішньої детекції доцільно обирати внутрішню, відносно консервативну ділянку елемента, а не крайові повтори; для junction-PCR один праймер розміщують у фланкуючій унікальній геномній послідовності, другий – усередині Р-елементу поблизу межі вставки.

Під час **підбору праймерів** оцінюють насамперед їх довжину, GC-вміст, температуру плавлення та очікуваний розмір амплікона. Для стандартної кінцевої ПЛР зазвичай найзручніші праймери довжиною близько 18–25 нт, із GC-вмістом у помірному діапазоні та близькими значеннями T_m у межах однієї пари; типовий робочий інтервал для дизайну становить приблизно 58–62 °C, а різниця T_m між прямим і зворотним праймерами має бути мінімальною. Для рутинної електрофоретичної детекції зазвичай обирають

амплікони помірної довжини, оскільки вони добре ампліфікуються і легко інтерпретуються. Саме ці параметри використовують Primer3 і Primer3Plus як базові критерії ранжування кандидатних праймерів.

Окремим критичним етапом є **аналіз вторинних структур** олігонуклеотидів. Внутрішні шпильки, self-dimer і hetero-dimer можуть знижувати ефективність ПЛР, але найбільш небезпечною є комплементарність на 3'-кінцях, оскільки вона створює умови для небажаного подовження праймерів один на одному. За рекомендаціями IDT, проблемними слід вважати структури з високою термостабільністю: для hairpin важливо, щоб її T_m була нижчою за умови реакції, а для self- та hetero-dimer суттєвим ризиком є ΔG близько -9 ккал/моль або більш негативне значення. Тому після первинного дизайну пари праймерів обов'язково фільтрують за показниками hairpin, self-dimer, hetero-dimer і 3'-комплементарності.

Primer3Plus є інструментом для генерації кандидатних праймерів із урахуванням T_m , довжини, GC-вмісту, розміру продукту, димеризації та ризику ектопічного праймінгу. Він дозволяє задавати включені та виключені ділянки, примусово обмежувати область пошуку, проєктувати праймери навколо цільового фрагмента або через межу з'єднання, а також використовувати бібліотеку mispriming, зокрема для *Drosophila*. Для задач, пов'язаних із мобільними елементами, це особливо важливо, оскільки повторні або низькокомплексні ділянки слід по можливості маскувати або виключати з дизайну ще до етапу перевірки специфічності.

Специфічність праймерів до Р-елемент потрібно перевіряти окремо для кожного праймера і для пари загалом. Для окремого праймера BLASTn придатний як аналіз короткого запиту; NCBI зазначає, що для коротких послідовностей параметри пошуку автоматично підлаштовуються. Для пари праймерів доцільніше використовувати Primer-BLAST, який поєднує дизайн Primer3 зі специфічною перевіркою на основі BLAST і може працювати як у режимі підбору нових праймерів, так і в режимі перевірки вже заданої пари.

При цьому NCBI рекомендує обирати відповідний організм і найменшу релевантну базу даних, оскільки це підвищує точність специфічної оцінки. Додаткову перевірку проводять за допомогою **in silico PCR**.

UCSC *In-Silico* PCR використовує BLAT-індекс і шукає потенційні амплікони, які утворює конкретна пара праймерів у заданій збірці геному; користувач може задавати максимальний розмір продукту, а також вимоги до точного та «доброго» збігу на 3'-кінцях праймерів. Водночас сам UCSC прямо застерігає, що інструмент не гарантує виявлення абсолютно всіх off-target-ампліконів і для дизайну праймерів у повторних регіонах його слід поєднувати з додатковою валідацією, зокрема Primer-BLAST. Це особливо актуально для Р-елемент, оскільки внутрішні праймери можуть ампліфікувати не один локус, а кілька копій або похідних елемента. Тому внутрішня ПЛР придатна насамперед для детекції наявності послідовностей Р-елемент, тоді як junction-PCR є методом локус-специфічного підтвердження конкретної вставки.

Матеріали та методи

- Комп'ютер з доступом до Інтернету.
- NCBI GenBank (отримання нуклеотидної послідовності):
<https://www.ncbi.nlm.nih.gov/>
- FlyBase або UCSC Genome Browser (перевірка контексту та координат):
<https://flybase.org/>; <https://genome.ucsc.edu/>
- Primer3Plus (дизайн праймерів): <https://www.primer3plus.com/> або локальний Primer3.
- NCBI Primer-BLAST (дизайн + перевірка специфічності):
<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>
- BLASTn (перевірка окремих праймерів як «short queries»):
<https://blast.ncbi.nlm.nih.gov/Blast.cgi>
- IDT OligoAnalyzer (оцінка димерів/шпильок та базових параметрів):
<https://www.idtdna.com/pages/tools/oligoanalyzer>

- UCSC In-Silico PCR (перевірка ампліконів у геномі dm6):
<https://genome.ucsc.edu/cgi-bin/hgPcr>
- MFEprimer (альтернатива для in silico PCR; веб або локально) – за наявності.

Завдання

- Отримати референсну нуклеотидну послідовність Р-елемент у форматі FASTA та зафіксувати її довжину.
- Визначити ділянку(и) для дизайну праймерів: (1) внутрішня консервативна ділянка елемента; (2) (опційно) межа «геном-мішень + Р-елемент» для детекції конкретної вставки.
- Згенерувати не менше 8 кандидатних пар праймерів (Primer3Plus) із заданими обмеженнями за Tm, GC-вмістом і довжиною амплікона.
- Відфільтрувати пари за вторинними структурами (шпильки, self/hetero-dimer) та за критеріями 3'-комплементарності.
- Перевірити специфічність праймерів у геномі *D. melanogaster* (Primer-BLAST/BLASTn + UCSC In-Silico PCR).
- Обрати мінімум дві найкращі пари праймерів і підготувати короткий протокол постановки ПЛР (температура відпалу, очікуваний розмір амплікона, інтерпретація).

Хід роботи

Вибір референтної послідовності Р-елемент

1. Для цієї роботи як референтну послідовність Р-елемент беруть не випадковий запис із нуклеотидної бази, а керовану канонічну послідовність родини з FlyBase.
2. У протоколі фіксують:
 - назву елемента;
 - FlyBase ID;

- джерело послідовності;
 - довжину повного елемента;
 - довжину термінальних інвертованих повторів.
3. Референтну послідовність зберігають у форматі FASTA як окремий файл, який далі використовують для дизайну внутрішніх праймерів.
 4. Якщо додатково використовують запис GenBank, його розглядають як допоміжне джерело, а не як основний еталон, якщо він не збігається з канонічною послідовністю родини.
 5. Для Р-елементу у *Drosophila melanogaster* у FlyBase подано канонічний елемент довжиною 2907 bp, термінальні повтори по 31 bp та TSD розміром 8 bp; саме ці параметри слід брати за основу під час планування дизайну праймерів.

Підготовка цільової ділянки для внутрішньої детекції

1. Для внутрішньої детекції обирають внутрішню ділянку Р-елементу, а не його крайові повтори.
2. Із дизайну виключають:
 - 5'- та 3'-кінцеві інвертовані повтори;
 - безпосередньо прикордонні зони вставки;
 - короткі низькокомплексні фрагменти;
 - ділянки з надмірно асиметричним GC-вмістом.
3. Перевагу надають відносно консервативному внутрішньому фрагменту, з якого можна одержати амплікон зручної довжини для стандартної ПЛР і подальшої електрофоретичної візуалізації.
4. На цьому етапі очікуваний амплікон визначають відносно самої канонічної послідовності Р-елементу, а не відносно геному dm6.
5. Такий підхід потрібний тому, що внутрішня детекція спрямована на наявність послідовності елемента як такої, а не на конкретний локус його інтеграції.

Підготовка шаблону для junction-PCR

1. Junction-PCR планують лише тоді, коли є конкретний локус вставки з відомими координатами та фланкуючою унікальною геномною послідовністю.
2. Для цього використовують:
 - координати вставки з FlyBase;
 - або координати з UCSC Genome Browser;
 - або координати зі stock- або insertion-ресурсу, якщо вони однозначно прив'язані до release 6.
3. Для junction-PCR один праймер розміщують у унікальній геномній фланкуючій ділянці, а другий – усередині Р-елемент біля його межі.
4. Для проєктування такої пари формують композитний шаблон, який містить:
 - частину фланкуючої геномної послідовності;
 - край Р-елемент;
 - саму межу з'єднання.
5. Якщо координати конкретної вставки відсутні, junction-PCR у цій лабораторній не конструюють, а як другу фінальну пару беруть альтернативну внутрішню пару до іншої ділянки Р-елемент.
6. Для роботи з координатами і фланками зручніше використовувати FlyBase JBrowse, оскільки FlyBase нині працює через JBrowse як основний геномний браузер для Drosophila.

Генерація кандидатних пар у Primer3Plus

1. Для дизайну праймерів використовують Primer3Plus у режимі підбору пар праймерів до заданого шаблону.
2. Як шаблон подають:
 - внутрішній фрагмент Р-елемент – для внутрішньої детекції;
 - або композитний фрагмент «геном + Р-елемент» – для junction-PCR.

3. Для цієї лабораторної встановлюють робочі обмеження:
 - довжина праймера – 18–25 нт;
 - T_m – 58–62 °C;
 - GC-вміст – 40–60%;
 - довжина продукту – 150–500 bp.
4. Якщо в шаблоні є ділянки, яких слід уникати, їх задають як excluded region, щоб Primer3Plus не розміщував у них праймери.
5. Генерують не менше 8 кандидатних пар і переносять їх у робочу таблицю.
6. Для кожної пари одразу фіксують:
 - послідовність forward-праймера;
 - послідовність reverse-праймера;
 - довжину кожного праймера;
 - T_m ;
 - GC-вміст;
 - очікуваний розмір продукту;
 - внутрішню позицію на шаблоні.
7. Primer3Plus у поточному інтерфейсі підтримує завантаження шаблонної послідовності, excluded/included regions і режим підбору primer pairs, тому цей етап залишається методично актуальним.

Первинна фільтрація кандидатних пар

1. Із початкового набору одразу відсіюють пари, у яких:
 - різниця T_m між праймерами перевищує робочий допуск;
 - довжина продукту виходить за межі заданого інтервалу;
 - один із праймерів має крайній GC-вміст;
 - на 3'-кінці є небажані довгі комплементарні мотиви.
2. Окремо позначають пари, у яких один праймер лежить занадто близько до межі виключеної або низькокомплексної ділянки.
3. Якщо кілька пар практично еквівалентні за базовими параметрами, перевагу надають тим, що:

- мають більш збалансовану Tm;
 - не закінчуються довгими мононуклеотидними траками;
 - формують продукт середньої, а не граничної довжини.
4. Після первинної фільтрації для термодинамічного аналізу залишають приблизно 4–6 найкращих пар.

Оцінка вторинних структур праймерів

1. Кожний праймер окремо перевіряють у IDT OligoAnalyzer.
2. Для кожного олігонуклеотиду аналізують:
 - hairpin;
 - self-dimer.
3. Для кожної пари додатково аналізують:
 - hetero-dimer;
 - особливо комплементарність на 3'-кінцях.
4. На цьому етапі відсіюють пари з вираженою 3'-димеризацією або з такими вторинними структурами, що можуть суттєво конкурувати з відпалом до цільової послідовності.
5. У робочій таблиці для кожної пари фіксують не лише факт «пройшла / не пройшла», а й коротку причину відсіву:
 - критичний self-dimer;
 - критичний hetero-dimer;
 - стабільна hairpin-структура;
 - небажана 3'-комплеметарність.
6. OligoAnalyzer у поточній версії підтримує розрахунок Tm, GC content, hairpin, self-dimer і hetero-dimer, тому його доцільно використовувати як основний інструмент термодинамічного фільтрування.

Перевірка окремих праймерів за допомогою BLASTn

1. Після термодинамічної фільтрації кожний окремий праймер перевіряють у BLASTn як короткий запит.

2. Мета цього етапу – не підтвердити цільову ампліфікацію, а виявити, чи має окремий праймер потенціал до небажаного зв'язування з іншими ділянками геному *D. melanogaster*.
3. Для коротких праймерів оцінюють насамперед:
 - збіги з повною або майже повною тотожністю;
 - наявність збігу на 3'-кінці;
 - кількість потенційних мішеней у геномі.
4. Якщо окремий праймер має велику кількість сильних збігів у геномі, таку пару переводять у групу підвищеного ризику, навіть якщо термодинамічні параметри були прийнятними.
5. Якщо праймер показує обмежену кількість збігів, але вони розташовані в повторному середовищі, це також позначають у примітках як фактор ризику для неспецифічної ампліфікації.
6. BLASTn у поточному інтерфейсі підтримує роботу з нуклеотидними запитами та є коректним інструментом для окремої перевірки коротких праймерів до етапу аналізу цілої пари.

Перевірка пар праймерів у Primer-BLAST

1. Для кожної відібраної пари виконують окрему перевірку у Primer-BLAST.
2. У Primer-BLAST можна працювати у двох режимах:
 - із шаблоном FASTA та автоматичним дизайном;
 - із уже готовими forward- і reverse-праймерами.
3. У цій лабораторній на етапі перевірки доцільно використовувати власні вже підібрані праймери, а Primer-BLAST застосовувати як інструмент оцінки специфічності.
4. Для внутрішньої детекції Р-елемент перевірку інтерпретують у двох площинах:
 - чи правильно пара дає очікуваний продукт на канонічній послідовності Р-елемент;

- чи не має пара потенціалу для небажаних off-target ампліконів у *D. melanogaster*.
5. Для junction-PCR перевіряють:
 - чи працює пара на композитному шаблоні межі вставки;
 - чи є геномний праймер достатньо унікальним у референсі.
 6. Якщо Primer-BLAST показує велику кількість альтернативних продуктів або кілька сильних off-target мішеней, пара не переходить до фінального відбору.
 7. Primer-BLAST нині офіційно поєднує Primer3 з перевіркою специфічності через BLAST, приймає FASTA-шаблон і дозволяє задавати власні праймери, тому цей етап є актуальним і методично виправданим.

Перевірка пар у UCSC *In-Silico* PCR

1. Для кожної фінальної або передфінальної пари виконують перевірку в UCSC In-Silico PCR на збірці dm6.
2. Цей крок використовують не як єдиний тест специфічності, а як додатковий фільтр для прогнозу можливих геномних ампліконів.
3. Для внутрішньої детекції Р-елемент результат інтерпретують так:
 - якщо пара дає продукт у dm6, це потенційний off-target і така пара небажана;
 - якщо пара не дає продукту в dm6, це узгоджується з відсутністю Р-елементу у референсному геномі та свідчить про відсутність очевидного off-target продукту в референсі.
4. Для junction-PCR dm6 використовують лише як контроль відсутності: за відсутності самої вставки в референсі така пара не повинна давати цільовий продукт на dm6.
5. Отже, очікуваний амплікон для:
 - внутрішньої детекції визначають на шаблоні Р-елемент;
 - junction-PCR визначають на композитному шаблоні межі вставки;

- dm6 використовують для виявлення небажаних або фонових ампліконів.
6. У робочій таблиці окремо фіксують:
 - кількість продуктів в *In-Silico* PCR;
 - їхні розміри;
 - інтерпретацію кожного результату.
 7. UCSC *In-Silico* PCR для *D. melanogaster* dm6 наразі доступний, вимагає праймери довжиною щонайменше 15 нт і прямо попереджає, що інструмент не гарантує виявлення всіх off-target сайтів, тому його слід використовувати разом із Primer-BLAST, а не замість нього. Крім того, Р-елементи відсутні в референсному геномі, що принципово впливає на інтерпретацію цього етапу.

Вибір двох фінальних пар праймерів

1. За результатами всіх фільтрів обирають мінімум дві найкращі пари.
2. Пара А має бути призначена для внутрішньої детекції Р-елемента.
3. Пара Б може бути:
 - альтернативною внутрішньою парою до іншої ділянки Р-елемента;
 - або junction-PCR парою, якщо координати конкретної вставки були відомі й коректно підготовлений композитний шаблон.
4. Для кожної фінальної пари подають:
 - послідовність forward-праймера;
 - послідовність reverse-праймера;
 - довжину праймерів;
 - Tm кожного праймера;
 - GC-вміст;
 - прогнозований розмір продукту;
 - підсумок щодо вторинних структур;
 - підсумок щодо специфічності.

5. Для кожної пари окремо вказують рекомендовану температуру відпалу як наближену стартову величину, яку оцінюють за нижчим T_m у парі з робочим пониженням приблизно на 3 °C.
6. Фінальне обґрунтування вибору має спиратися не на один параметр, а на сукупність критеріїв:
 - прийнятна T_m ;
 - збалансований GC-вміст;
 - відсутність критичних димерів;
 - відсутність небажаних продуктів у $dm6$;
 - коректний цільовий продукт на шаблоні Р-елемента або на junction-шаблоні.

Підготовка короткого протоколу постановки ПЛР

1. Для кожної з двох фінальних пар формують короткий протокол постановки ПЛР.
2. У протокол включають:
 - назву пари;
 - ціль детекції;
 - послідовності праймерів;
 - рекомендовану температуру відпалу;
 - очікуваний розмір амплікона;
 - тип шаблону, на якому очікується продукт.
3. Для внутрішньої детекції окремо прописують, що:
 - позитивний контроль має містити Р-елемент;
 - негативний контроль без шаблону є обов'язковим;
 - відсутність продукту на $dm6$ не є помилкою, а відповідає логіці off-target перевірки в референсі.
4. Для junction-PCR окремо прописують, що:
 - позитивний результат можливий лише за наявності конкретної вставки;

- відсутність продукту на dm6 очікувана, якщо вставки в референсі немає;
 - інтерпретація залежить від того, чи використовувався геном із відповідним локусом вставки.
5. У завершальному висновку коротко зазначають, які обмеження залишаються навіть після біоінформатичної перевірки:
- повторна природа мобільного елемента;
 - можливість прихованих off-target сайтів;
 - залежність від якості анотації й референсної збірки;
 - відмінність між референсним геномом і реальною дослідною лінією.

Оформлення результатів

1. Результати подають у підсумковій таблиці для всіх кандидатних і окремо для двох фінальних пар.
2. Для кандидатних пар у таблиці доцільно включити такі поля:
 - №;
 - Forward primer;
 - Reverse primer;
 - довжина;
 - Tm;
 - GC%;
 - розмір продукту;
 - hairpin;
 - self-dimer;
 - hetero-dimer;
 - BLASTn-оцінка;
 - Primer-BLAST;
 - UCSC In-Silico PCR;
 - рішення.
3. Для фінальних пар окремо додають короткий аналітичний коментар:

- для чого призначена пара;
 - чому вона залишена;
 - яке її головне обмеження.
4. У підсумковому висновку формулюють:
- яку пару рекомендовано як основну для внутрішньої детекції;
 - яку пару рекомендовано як резервну або як junction-PCR;
 - чому саме ці пари найкраще відповідають поставленій меті.

Контрольні запитання

1. Чому під час дизайну праймерів важливо контролювати різницю T_m між прямим і зворотним праймерами?
2. Що таке 3'-димер праймерів і чому він критичніший за внутрішню комплементарність у середині праймера?
3. У чому відмінність між перевіркою специфічності окремого праймера (BLASTn) та пари праймерів (in silico PCR)?
4. Чому праймери до мобільних елементів можуть давати кілька ампліконів у геномі? Як це врахувати під час інтерпретації результатів?
5. За яких умов доцільно використовувати junction-PCR і які дані потрібні для її коректного планування?

Рекомендована література

1. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Research*, 40(15), e115. <https://doi.org/10.1093/nar/gks596>
2. Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13, 134. <https://doi.org/10.1186/1471-2105-13-134>

3. Kaminker, J. S., Bergman, C. M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D. A., Lewis, S. E., Rubin, G. M., Ashburner, M., & Celniker, S. E. (2002). The transposable elements of the *Drosophila melanogaster* euchromatin: A genomics perspective. *Genome Biology*, 3(12), research0084. <https://doi.org/10.1186/gb-2002-3-12-research0084>
4. Castro, J. P., & Carareto, C. M. A. (2004). *Drosophila melanogaster* P transposable elements. *Genetics and Molecular Biology*, 27(1), 148–157.
5. Daniels, S. B., Peterson, K. R., Strausbaugh, L. D., Kidwell, M. G., & Chovnick, A. (1990). Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics*, 124(2), 339–355.
6. Bergman, C. M., Haddrill, P. R., & Bettencourt, B. R. (2017). Genomic analysis of P elements in natural populations of *Drosophila melanogaster*. *PeerJ*, 5, e3824. <https://doi.org/10.7717/peerj.3824>

Додаток 1.

Назви та джерела еталонних послідовностей мобільних генетичних елементів *Drosophila melanogaster*

P-елемент_Dmel DNA-транспозон

Довжина: 2907 bp

Джерело: FlyBase Natural Transposon Report

Посилання: <https://flybase.org/reports/FBte0000037.html>

hobo_Dmel DNA-транспозон

Довжина: 2959 bp

Джерело: GenBank M69216.1

Посилання: <https://www.ncbi.nlm.nih.gov/nuccore/M69216.1?report=fasta>

coria_Dmel LTR-ретротранспозон

Довжина: 5146 bp

Джерело: ENA X04456.2

Посилання: <https://www.ebi.ac.uk/ena/browser/view/X04456>

gypsy_Dmel LTR-ретротранспозон

Довжина: 7469 bp

Джерело: GenBank M12927.1

Посилання: <https://www.ncbi.nlm.nih.gov/nuccore/M12927.1?report=fasta>

412_Dmel LTR-ретротранспозон

Довжина: 7566 bp

Джерело: FlyBase Natural Transposon Report

Посилання: <https://flybase.org/reports/FBte0000007.html>

roo_Dmel LTR-ретротранспозон

Довжина: 9092 bp

Джерело: FlyBase Natural Transposon Report

Посилання: <https://flybase.org/reports/FBte0000100.html>

Додаток 2

Рекомендовані бактеріальні геноми для аналізу профагів (із кодами доступу).

Escherichia coli O157:H7 str. Sakai

GenBank: BA000007.3 (хромосома, ~5.50 Mb).

Відомий патогенний штам STEC; має розвинений профаговий репертуар, зокрема Stx-кодуючі профаги.

Salmonella enterica subsp. *enterica* serovar Typhimurium str. LT2

RefSeq: NC_003197.2 (хромосома, ~4.86 Mb).

Класичний лабораторний штам; містить кілька профагів, зокрема Gifsy-1 та Gifsy-2.

Staphylococcus aureus subsp. *aureus* str. N315

GenBank: BA000018.3 (хромосома, ~2.81 Mb).

Референтний MRSA-штам; містить профагові регіони різного ступеня повноти, також наявні плазмідні реплікони.

Bacillus subtilis subsp. *subtilis* str. 168

RefSeq: NC_000964.3 (хромосома, ~4.22 Mb).

Лабораторний штам; має інтегрований профаг SP β (~134 kb).

Pseudomonas aeruginosa PAO1

RefSeq: NC_002516.2 (хромосома, ~6.26 Mb).

Референтний штам; містить інтегрований профаг Pf4 та інші профагоподібні/криптичні ділянки.

Додаток 3

Приклади зібраних геномів для аналізу плазмід

Klebsiella pneumoniae (клон ST258) – штам NJST258_1 (GenBank: CP006923–CP006928). Цей клінічний ізолят містить 1 хромосому і 5 плазмід. Серед них є велика плазміда, асоційована з карбапенемазою KPC, а також менші плазмідні елементи різних типів. Цей штам придатний для аналізу складного мультиплазмідного профілю у клінічно значущих ентеробактерій.

- Хромосома: <https://www.ncbi.nlm.nih.gov/nuccore/CP006923.1?report=fasta>
- Плазміда 1: <https://www.ncbi.nlm.nih.gov/nuccore/CP006924.1?report=fasta>
- Плазміда 2: <https://www.ncbi.nlm.nih.gov/nuccore/CP006925.1?report=fasta>
- Плазміда 3: <https://www.ncbi.nlm.nih.gov/nuccore/CP006926.1?report=fasta>
- Плазміда 4: <https://www.ncbi.nlm.nih.gov/nuccore/CP006927.1?report=fasta>
- Плазміда 5: <https://www.ncbi.nlm.nih.gov/nuccore/CP006928.1?report=fasta>

Staphylococcus aureus (MRSA) – штам USA300_FPR3757 (NC_007790–NC_007793). Має 1 хромосому і 3 плазмиди. До складу плазмідного набору входять мала криптична плазміда, плазміда з геном тетрациклін-резистентності та більша плазміда з додатковими детермінантами резистентності. Цей приклад зручний для аналізу плазмід грампозитивних бактерій, де виявляються гер-типи, а не типові Inc-групи ентеробактерій.

- Хромосома: https://www.ncbi.nlm.nih.gov/nuccore/NC_007793.1?report=fasta
- Плазміда 1: https://www.ncbi.nlm.nih.gov/nuccore/NC_007790.1?report=fasta
- Плазміда 2: https://www.ncbi.nlm.nih.gov/nuccore/NC_007791.1?report=fasta
- Плазміда 3: https://www.ncbi.nlm.nih.gov/nuccore/NC_007792.1?report=fasta

Escherichia coli (ентеротоксигенний) – штам H10407 (GenBank/ENA: FN649414–FN649418). Цей діарейний штам має 1 хромосому і 4 плазмиди: дві великі вірулентні плазмиди та дві малі плазмиди типу Col. Очікувано для цього штаму можуть виявлятися реплікони RepFIIA, ColE1 та ColE2. Цей приклад є

придатним для аналізу поєднання великих і малих плазмід в одному ентеробактеріальному ізоляті.

- Хромосома: <https://www.ebi.ac.uk/ena/browser/view/FN649414>
- Плазміда 1: <https://www.ebi.ac.uk/ena/browser/view/FN649415>
- Плазміда 2: <https://www.ebi.ac.uk/ena/browser/view/FN649416>
- Плазміда 3: <https://www.ebi.ac.uk/ena/browser/view/FN649417>
- Плазміда 4: <https://www.ebi.ac.uk/ena/browser/view/FN649418>

Enterococcus faecalis – штам V583 (AE016830–AE016833). Містить 1 хромосому і 3 плазмиди. Цей клінічний ванкоміцин-резистентний ізолят є класичним прикладом геному з кількома плазмідами різного розміру у грампозитивної бактерії. Може використовуватися для аналізу плазмідних реплікативних систем типу Rep у ентерококів.

- Хромосома: <https://www.ncbi.nlm.nih.gov/nuccore/AE016830.1?report=fasta>
- Плазміда 1: <https://www.ncbi.nlm.nih.gov/nuccore/AE016831.1?report=fasta>
- Плазміда 2: <https://www.ncbi.nlm.nih.gov/nuccore/AE016832.1?report=fasta>
- Плазміда 3: <https://www.ncbi.nlm.nih.gov/nuccore/AE016833.1?report=fasta>