

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ВИСОКИХ ТЕХНОЛОГІЙ

Завідувач кафедри молекулярної біотехнології та біоінформатики
проф. Нипорко Олексій Юрійович
Протокол № _____ засідання кафедри
від "___" _____ 2022 р.

Можливі шляхи походження *Drosophila simulans* України

Випускна кваліфікаційна робота магістра
студента спеціальності
091 Біологія
ОП “Біоінформатика і структурна біологія”
Грубіяна Назарія Олександровича

Науковий керівник від кафедри
доцент кафедри молекулярної
біотехнології та біоінформатики
к.б.н **Самофалова Дарія Олексіївна**

Робота виконана на базі кафедри загальної та медичної генетики
ННЦ “Інститут біології та медицини” Київського національного університету
імені Тараса Шевченка під керівництвом асистент, к.б.н. **С.В. Серги**

Оцінка захисту роботи

Київ – 2022 р.

АНОТАЦІЯ

Грубіян Н.О. Можливі шляхи походження *Drosophila simulans* України – Випускна кваліфікаційна робота магістра за спеціальністю 091 Біологія ОП “Біоінформатика і структурна біологія”.

У роботі було проаналізовано геноми 7 ізосамкових ліній *Drosophila simulans*, зібраних на території України. Проведено кладистичний аналіз відсеквенованих ліній для з’ясування походження місцевих популяцій серед популяцій світу. Встановлено, що місцеві популяції, які були зібрані в Одесі у 2015 та 2018 роках, мають африканське походження, що узгоджується із дослідженнями поширення *Drosophiliadae* у помірному кліматичному поясі. Було показано, що лінії *D. simulans* з Одеси інфіковані ендосимбіонтом *Wolbachia* штаму wRi. Отримані результати є першими стосовно особливостей геному даного виду з території України, а також його ендосимбіонту *Wolbachia*. Вони можуть бути використані в подальшому для дослідження поширення виду *D. simulans* у помірному кліматичному поясі, а також встановити механізми адаптації цього виду в умовах зміни клімату та глобального потепління.

Ключові слова: *Drosophila simulans*, *Wolbachia*, кластерний аналіз, діапауза, гаплотип, мітохондріальний геном.

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

мтДНК – мітохондріальна ДНК

COI – Cytochrome c oxidase subunit I (Цитохром с-оксидаза субодиниця 1)

SNP – Single Nucleotide Polymorphism (однонуклеотидний поліморфізм)

ЈН – Juvenile hormone (ювенільний гормон)

ЗМІСТ

ВСТУП	6
РОЗДІЛ 1. ОГЛЯД ЛІТЕРАТУРИ	8
1.1. Опис та характеристика виду <i>D. simulans</i>	8
1.1.1. Відкриття виду <i>D. simulans</i>	8
1.1.2. Спільні та відмінні риси між <i>D. melanogaster</i> та <i>D. simulans</i>	8
1.2. Характеристика гаплогруп <i>D. simulans</i>	10
1.3. Ідентифікація <i>D. simulans</i> на території України	12
1.4. Діапауза, як еволюційний механізм адаптації до помірних широт	13
1.4.1. Дослідження діапаузи у <i>D. melanogaster</i>	13
1.4.2. Дослідження діапаузи серед родини <i>Drosophilidae</i>	14
РОЗДІЛ 2. МАТЕРІАЛИ ТА МЕТОДИ ДОСЛІДЖЕННЯ	16
2.1. Збір та культивування ізосамкових ліній <i>D. simulans</i>	16
2.2. Отримання повногеномних сиквенсів та аналіз одонуклеотидних поліморфізмів	16
2.3. Аналіз мітохондріальної ДНК <i>D. simulans</i> м.Одеса	17
2.4. Аналіз повногеномних зчитувань штаму <i>Wolbachia</i> ізольованих ліній <i>D. simulans</i>	20
2.5. Аналіз продуктів генів діапаузи одеських ліній <i>D. simulans</i> та світу ...	21
РОЗДІЛ 3. РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ ТА ЇХ ОБГОВОРЕННЯ	25
3.1. Секвенування ліній <i>D. simulans</i> з м.Одеси	25
3.2. Кладистичний аналіз даних секвенування	25
3.3. Встановлення штаму ендосимбіонта <i>Wolbachia</i> у відсеквенованих лініях <i>D. simulans</i>	29

3.4. Дослідження продуктів генів діапаузи ліній зі світу та України.....	30
ВИСНОВКИ.....	33
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ	34
ДОДАТКИ.....	39

ВСТУП

Drosophila simulans - сиблінговий вид *Drosophila melanogaster*, який має низку спільних рис та займає подібну екологічну нішу [1]. Відомо, що *D. melanogaster* поширений далеко за межами тропічних і субтропічних регіонів, зокрема в помірному кліматичному поясі. Донедавна вважалось, що його сиблінговий вид *D. simulans* не заселяє північні широти, зокрема території Європи з помірним кліматом. Однак аналіз симпатричних популяцій обох видів в тропічних та помірних широтах засвідчив про можливість повного або часткового заміщення популяцій *D. melanogaster* на популяції *D. simulans* [2]. Існує гіпотеза, що швидке поширення цих двох видів стало можливим завдяки міжнародному трафіку фруктів, що забезпечило швидку колонізацію помірних широт цими двома видами. На сьогодні достеменно невідомий характер конкуренції між двома видами та шляхи поширення цих двох видів по світу. Вважалось, що основною еволюційною перевагою *D. melanogaster* над *D. simulans* є наявність зимової діапаузи. Дослідження Serga [3] показало, що *D. simulans* успішно колонізує помірні широти європейського континенту, зокрема території України, ймовірно завдяки наявності діапаузи. Згодом вперше була доведена можливість *D. simulans* реагувати на зміни температури шляхом входження в репродуктивну діапаузу у лабораторних умовах [4].

Однак, досі залишаються невідомими механізми та напрямки поширення *D. simulans* з місць з тропічним та субтропічним кліматом в Європу та, зокрема, на території України. Також на сьогодні невідомо які генетичні зміни зумовили фенотиповий прояв діапаузи у цього виду.

Метою роботи було встановити шляхи походження *D. simulans* на території України та визначити особливості генів, залучених у фенотиповий прояв діапаузи у даного виду.

Відповідно до мети поставлено такі **завдання**:

1. Визначити однонуклеотидні поліморфізми у складі відсеквенованих геномів ізосамкових ліній *D. simulans* на території України.
2. Встановити гаплотип особин *D. simulans* м.Одеса.
3. Встановити штам ендосимбіонта *Wolbachia* у відсеквенованих лініях *D. simulans* м.Одеса.
4. З'ясувати походження популяцій *D. simulans* м.Одеса.
5. Проаналізувати особливості послідовностей генів, залучених у формування діапаузи, у природніх популяціях *D. simulans* м.Одеса.

РОЗДІЛ 1. ОГЛЯД ЛІТЕРАТУРИ

1.1. Опис та характеристика виду *D. simulans*

1.1.1. Відкриття виду *D. simulans*

D. simulans, сиблінговий вид *D. melanogaster*, був вперше охарактеризований та описаний у 1919 році Альфредом Стертевантом, де зазначались відмінності між *D. simulans* і *D. melanogaster*. Стертевант проаналізував генетику цього виду, порівнявши його з *D. melanogaster* за допомогою гібридів [5]. Крім роботи Стертеванта, більшість робіт була присвячена аналізу таких гібридів. Серед цих робіт є стаття Bonnier [6] про видоутворення у *D. simulans*, Ephrussi [7] про хондрію яєчника і Kerkis [8] про розвиток гонад і кон'югацію хромосом, а також Muller [9] про хромосомні фактори, які спричиняють стерильність утворених гібридів. У кількох із цих досліджень наголошувалося, що, хоча ці два види схожі фенотипово та є тісно споріднені, вони можуть бути дуже розбіжними з високим рівнем міжвидової дивергенції.

1.1.2. Спільні та відмінні риси між *D. melanogaster* та *D. simulans*

Найперша робота з аналізу відмінностей була присвячена кількісним ознакам. Перші з цих досліджень Tantawy [10] описували відмінності термостійкості, спадковості, реакції на відбір для різних довжин крил, сезонної варіації довжини крил, генетичних відхилень морфологічних ознак між двома видами. На основі цієї роботи було висунуто припущення, що генетична мінливість кількісних ознак у *D. simulans* нижча.

Протягом 1970-х і 1980-х років кілька досліджень продемонстрували паралельні географічні варіації двох видів щодо морфометричних і фізіологічних ознак. У той же час розвиток електрофоретичного аналізу, відкрив нові можливості для досліджень відмінностей. Це призвело до різкого збільшення кількості досліджень, пов'язаних із порівняннями двох видів, проведених протягом 70 років. Ще раз було показано, що поліморфізм у *D. simulans* був нижчим, ніж у його сиблінгового виду [11].

Singh [12] узагальнив дані про просторову варіацію двох сиблінгових видів на основі всіх можливих ознак. Мета полягала в тому, щоб зробити висновок про історію та характер їхньої селекції. Як і в більшості попередніх порівнянь, було зазначено, що *D. simulans* був менш мінливим, ніж *D. melanogaster*, за винятком його нуклеотидного різноманіття. Очевидно, що в той час як два види демонструють приблизно однакову мінливість всередині популяції, *D. melanogaster* більш диверсифікований на рівні між популяціями. Однак, було показано, що в межах популяції генетична мінливість більша у *D. simulans* у випадку ядерної ДНК. Тому загальні тенденції не є незмінними, а детальний аналіз є ключовим у цих дослідженнях [1].

Як відомо, *D. simulans* має більш обмежене поширення [13]. Будучи поширеним у тропічних і субтропічних регіонах, він рідко зустрічається в помірному кліматі, тоді як його сиблінг, *D. melanogaster*, широко розповсюджений в усіх цих регіонах [14]. Було виявлено, що *D. melanogaster* краще переносить стресові умови, такі як екстремальна температура, голод або посуха [15]. Переваги місця яйцекладки, а також толерантність *D. melanogaster* до більш високих концентрацій етанолу можуть також відігравати значну роль у визначенні відмінностей між цими видами в їхньому географічному поширенні [3].

1.2. Характеристика гаплогруп *D. simulans*

Людський коменсал *D. simulans* є привабливим видом для вивчення мітонуклеарних взаємодій через високу різноманітність мітохондріальної ДНК, запропонованим відмінностям між гаплогрупами і невипадковим географічним розподілом гаплогруп.

Для характеристики географічних варіантів *D. simulans* використовується ідентифікація гаплотипу мітохондріальної ДНК разом із штамом внутрішньоклітинних симбіотичних бактерій, успадкованих від матері, роду *Wolbachia*. *Wolbachia* пов'язана з *Anaplasma marginale*, *Ehrlichia risticii* та *Rickettsia spp.* *Wolbachia* є облігатними мутуалістами нематод [16], і було показано, що 76% комах є ними інфіковані [17]. Така особливість робить цю групу бактерій найпоширенішим ендосимбіонтом на планеті. *Wolbachia* спричиняє ряд фенотипічних ефектів у комах і наземних ракоподібних, включаючи телітокозний партеноген [18], летальність для ембріонів чоловічої статі [19], а також цитоплазматичну несумісність у комарів та дрозофіл [20], [21]. Хоча ці фенотипи різноманітні, усі вони порушують репродуктивну систему господаря на користь бактерії. Здатність впливати на репродуктивну систему пояснює високу поширеність *Wolbachia* у світі безхребетних.

Серед трьох зареєстрованих мітохондріальних гаплогруп *D. simulans* (тобто siI, siII і siIII), siII є найбільш поширеним і єдиним присутнім у північній півкулі [20]. Таким чином, вважається, що саме гаплогрупа siII в першу чергу відповідає за процеси глобального розповсюдження *D. simulans*. Відомо, що ця гаплогрупа інфікована успадкованими від матері бактеріями роду *Wolbachia* штаму wRi [20]. Ці бактерії впливають на репродукцію господаря, і повідомляється, що мухи, заражені цим штамом, переважають в кількості неінфікованих особин або тих, хто інфікований іншими штамми бактерій [22].

Додаткове пояснення спостережуваного розподілу популяцій полягає в тому, що успадкована від матері цитоплазматична несумісність та збільшення пристосованості, викликане *Wolbachia*, значно вплинули на еволюцію мтДНК *D. simulans*. У найпростішому випадку несумісність виникає, коли неінфікована самка спарюється з інфікованим самцем, що призводить до зниження рівня висиджування яєць [23]. Несумісність, спричинена *D. simulans*, викликає збільшення частоти симбіонтів та зчепленого мітохондріального генотипу, успадкованого від матері. Встановлено, що і симбіонт, і господар впливають на вираження цитоплазматичної несумісності. У *D. simulans* виявлено шість штамів *Wolbachia*. Назви слідують за місцем або країною, де інфекція була вперше зафіксована. Неінфіковані лінії мух позначаються як w-. Було виявлено, що мухи з Нової Каледонії та Сейшельських островів подвійно інфіковані wHa і wMa_{No} [24], і було показано, що вони разом несумісні з wRi [25].

Ballard [20] у своєму дослідженні вказав основні штами *Wolbachia* та лінії *D. simulans*, якими вони були інфіковані (таблиця 1.1).

Таблиця 1.1 Лінії *D. simulans* і *Wolbachia*, які використовувались для ідентифікації штаму та ізоляту

Назва лінії	Штам <i>Wolbachia</i>	Гаплогрупа	Ізолят	Місце Збору
NC48	wHa & wMa	siI	wMa _{No}	Нова Каледонія, Океанія
HW00	wHa	siI		Гаваї, США
Coffs	wAu	siII		Коффс-Гарбор, Австралія
DSR	wRi	siII		Ріверсайд, США
C167	wRi	siII		Наньюкі, Кенія
MD199	wMa	siIII	wMa _{Ma}	Джофрривіль, Мадагаскар
KC9	wMa	siIII	wMa _{Ki}	Кіліманджаро, Танзанія

1.3. Ідентифікація *D. simulans* на території України

В Україні різні види дрозофіл вивчають з початку 20 століття. Проте повідомлення про поширення *D. simulans* в Україні є суперечливими. Ballard [20] повідомляє про виявлення *D. simulans* у Києві та Ялті. Зокрема у дослідженні Serga [3] було виявлено колонізацію *D. simulans* території України. У цьому дослідженні був оцінений зв'язок між відносною чисельністю *D. simulans* та його сиблінговим видом *D. melanogaster*; залежність між середньою місячною температурою протягом сезону розмноження та середньою зимовою температурою; мітохондріальну гаплогрупу *D. simulans* та штамп *Wolbachia*, що інфікує ці популяції для оцінки ймовірного географічного походження колонізаторів (рис.1.1).

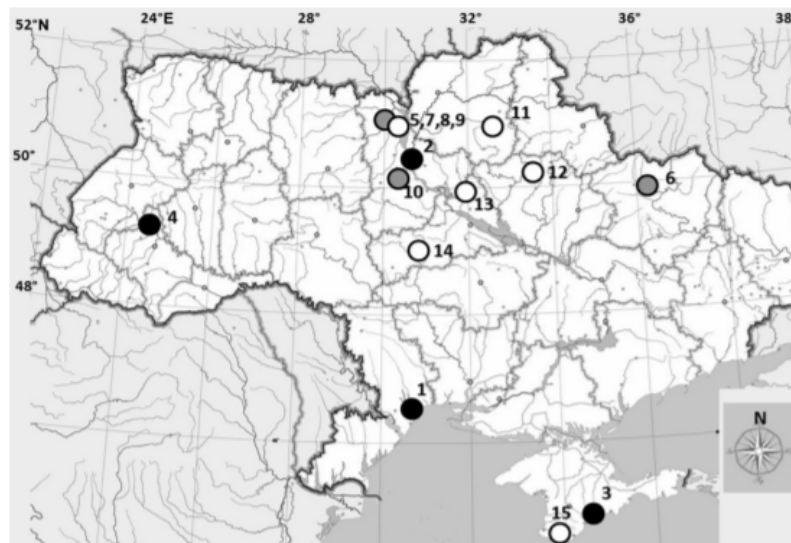


Рис.1.1 Поширення *D. simulans* на території України(2010-2014р.).

Чорні кола - місця із високою частотою виявлення. Сірі кола – місця із незначною кількістю виявлених особин (<3). Білі кола – місця спорадичного виявлення. Номер – локація збору [3]

1.4. Діапауза, як еволюційний механізм адаптації до помірних широт

1.4.1. Дослідження діапаузи у *D. melanogaster*

Організми, що проживають у помірному середовищі, розробили способи справлятися з ритмічними змінами в їхньому оточенні щоразу, коли Земля робить оберт навколо Сонця. До цих способів належить циркуляція температури, доступність їжі та води, тиск хижаків, доступність до укриття або гнізд — це лише деякі з факторів, пов'язаних зі зміною пори року.

Стратегія виживання для комах — діапауза, що дозволяє організмам вчасно уникнути несприятливих умов, а також не мігрувати на великі дистанції. У дикій природі діапауза характеризується фазою підготовки до діапаузи, фазою підтримки (зимівля) і фазою після діапаузи. Ці фази контролюються нейроендокринною системою і значно відрізняються рівнем метаболічної активності та експресії генів. Зокрема, *D. melanogaster* переживає дорослий репродуктивний стан спокою, викликаний зниженою температурою та скороченим світловим днем [26].

D. melanogaster виник у Африці на південь від Сахари і згодом розширив своє середовище проживання завдяки зв'язку з людьми та сільським господарством [27]. Мухи колонізували помірні широти Євразії, а згодом Північну Америку та Австралію. Генотипи деяких популяцій легко входять у стан репродуктивного спокою у відповідь на низьку температуру та умови короткого дня, тоді як інші мають низьку або нульову схильність до діапаузи [28]. Сезонні популяції у високих широтах демонструють набагато більшу здатність до діапаузи, ніж мухи з субтропічних/тропічних популяцій низьких широт. Крім того, географічні варіації репродуктивного спокою у природніх популяціях *D. melanogaster* пов'язані з одонуклеотидними поліморфізмами в декількох локусах, що свідчить про те, що ознаки репродуктивного спокою генетично детерміновані [4].

Як і в інших видів комах, індукція репродуктивного спокою *D. melanogaster* пов'язана зі зниженням рівня ювенільного гормону (JH). Попереднє дослідження показало, що у виділеному СА (corpus allatum – місце синтезу ювенільного гормону) у комах, які перебували у стані репродуктивного стану спокою, був зменшений рівень біосинтезу JH [29]. При введенні аналога JH метопрену спостерігалось порушення репродуктивного стану спокою [30]. Фенотипи репродуктивного спокою також спостерігаються у самок мух, у яких синтез ферментів JH посилений або пригнічений. Наприклад, надмірна експресія гена, що кодує кислотну О-метилтрансферазу ювенільного гормону (*Jhamt*), значно знижує відсоток мух у стані спокою. І навпаки, СА-специфічна трансгенна дволанцюгова РНК-інтерференція *Jhamt* збільшує відсоток сплячих мух [31].

Репродуктивний стан спокою комах регулюється не тільки JH, але й багатьма іншими гормонами, включаючи інсуліноподібні пептиди та екдистероїди [26].

1.4.2. Дослідження діпаузи серед родини *Drosophilidae*

Окрім *D. melanogaster*, діпауза присутня у багатьох інших представників родини *Drosophilidae*. Так, наприклад, деякі з представників демонструють дуже сильну і стійку реакцію діпаузи. Більшість природних ліній *D. littoralis*, наприклад, показують чітку зміну фотоперіодичної реакції при короткому дні при 16°C [32]. Три види групи *D. virilis* (*littoralis*, *montana* і *ezoana*), демонструють подібну фотоперіодичну реакцію [33]. Zonato [4] спостерігав фенотип спокою і в інших тропічних видів дрозофіли (*D. yakuba*, *D. sechellia*, *D. erecta* та *D. ananassae*).

Останнім часом продовжуються дослідження щодо появи діпаузи у *D. simulans*. Так, Serga [3] висловила гіпотезу щодо наявності репродуктивної діпаузи у європейських популяціях *D. simulans*, зокрема на території України. Zonato [4] виявив, що популяції *D. simulans*, одна європейська та

одна африканська, демонструють значний рівень діпаузи через 12 днів. Крім того, спостерігався фотоперіодизм у європейських популяціях *D. simulans*. Оскільки *D. melanogaster* і *D. simulans* мають схожий фенотип спокою, це означає або паралельну еволюцію для однієї і тієї ж ознаки (конвергенція), або походження від спільного предка (тобто загальна реакція на стрес). Саме тому вивчення серед різних представників *Drosophilidae* діпаузи допоможе вирішити фундаментальні аспекти механізмів його дії, еволюції та появи даної ознаки.

РОЗДІЛ 2. МАТЕРІАЛИ ТА МЕТОДИ ДОСЛІДЖЕННЯ

2.1. Збір та культивування ізосамкових ліній *D. simulans*

Об'єкти дослідження - ізосамкові лінії *D. simulans*, започатковані у 2015 та 2018 роках у м.Одеса. Місце збору - це змішаний сад біля житла людини. Для приманювання мух використовували фруктові пивні (яблука) приманки, через 5 – 10 діб після їхнього встановлення здійснювали відлов за допомогою екстаустера та ентомологічного сачка. Після відлову встановлювали ізосамкові лінії, які культивували на стандартному середовищі (на 1 літр води 6 г агару, 15 г дріжджів, 50 г цукру, 55 г манної крупи та 4 мл пропіонової кислоти) при температурі 25 ± 1 °C та режимі освітлення 12 годин світло, 12 годин темрява (LD12:12).

2.2. Отримання повногеномних сиквенсів та аналіз однонуклеотидних поліморфізмів

З отриманих ізосамкових ліній виділяли тотальну ДНК за допомогою фенол-хлороформної екстракції. Повногеномне секвенування проводили за допомогою Illumina HiSeq2000.

У необроблених файлах fastq обрізали послідовності адаптерів Illumina за допомогою Cutadapt версії 1.2.1 [34]. Послідовності додатково обрізали за допомогою Sickle версії 1.2.00 з мінімальним показником якості 20 [35]. Послідовності, коротші за 20 пар основ., після обрізання були видалені. Непарні зчитування були виключені з аналізу.

Fastq файли відбирали для вирівнювання з референтним геномом *D. simulans* (Ідентифікаційний номер в базі даних Genome: ASM75419v2) за

допомогою bwa алгоритму [36]. Виклик одонуклеотидних поліморфізмів здійснювали за допомогою SAMtools [37] в командному рядку Linux:

```
// samtools mpileup -ugf ref.fna -b input.txt --output-tags DP, AD, ADF, ADR, SP, INFO/AD, INFO/ADF, INFO/ADR
```

та BCFtools:

```
// bcftools call -v -m -o file.vcf
```

Загальна методологія отримання повногеномних зчитувань та одонуклеотидних поліморфізмів представлена на рис.2.1.

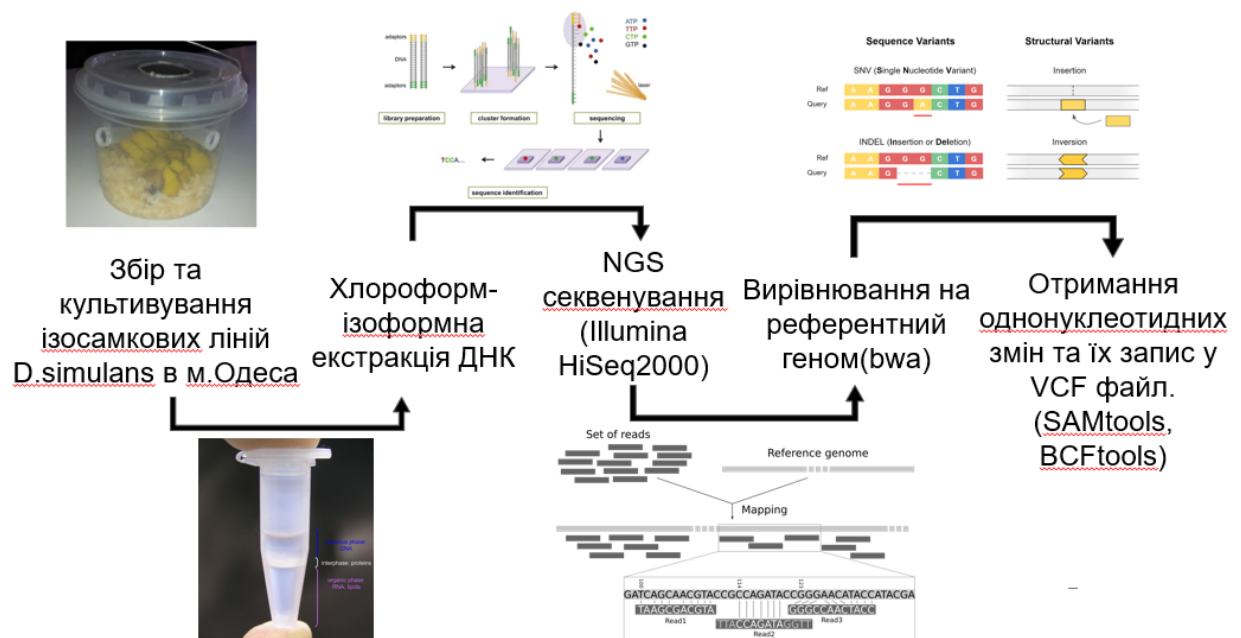


Рис. 2.1. Метод отримання повногеномних зчитувань та одонуклеотидних поліморфізмів ісосамкових ліній *D. simulans* з м.Одеси

2.3. Аналіз мітохондріальної ДНК *D. simulans* м.Одеса

Для отримання повногеномних сиквенсів мтДНК ліній *D. simulans* по всьому світу використовували бази даних PopSet [38]

(www.ncbi.nlm.nih.gov/popset/) з критеріями пошуку: “*D. Simulans* [porgn] AND mitochondrion [filter]”.

Завантажені файли FASTA перетворювали у VCF за допомогою інструменту SNP-sites [39] з використанням команди:

```
// snp-sites file.fas -v -o file_out.vcf
```

Файл VCF, що містить дані SNP одеських ліній, розділяли на окремі хромосоми за допомогою інструменту фільтрації SNPSift [40], а також фільтрували дані про наявні однонуклеотидні поліморфізми у складі мтДНК:

```
// java -jar SnpSift.jar filter "( CHROM = 'NC_005781.1' )" input.vcf > out.vcf
```

За допомогою програми GATK “VariantstoTable” перетворювали отримані VCF файли ліній світу в таблицю SNP-сайтів [41]. Матричну відстань розраховували за допомогою функції daisy в Rstudio[42]. Кластерну дендрограму створювали функцією hclust і будували в Rstudio. Для побудови кластерної дендрограми використовували метод повного зв’язку (Complete linkage method). Код програми для побудови кластерної дендрограми:

```
library(ggplot2)
library(ggtree)
library(cluster)
library(ape)
library(pegas)
library(readr)
mtDNAatab <- read_table("~/Drosophila/Haplotypemap/mergedmtDNA.txt")
mtDNAatab[] <- lapply(mtDNAatab, sub, pattern = "%.", replacement = "")
mtDNAatab[] <- lapply(mtDNAatab, sub, pattern = "\\*", replacement = NA)
mtDNAatab[] <- lapply(mtDNAatab, sub, pattern = "\\.", replacement = NA)
mtDNAatab$"POS" <- NULL
mtDNAatab$"REF" <- NULL
```

```

mtDNAtab <- na.omit(mtDNAtab)
mtDNAtab_transpose <- as.data.frame(t(mtDNAtab))
mtDNAtab_transpose[] <-
lapply(mtDNAtab_transpose[1:ncol(mtDNAtab_transpose)], as.factor)
mtDNAtab_transpose_dist<-daisy(mtDNAtab_transpose)
mtDNAtab_transpose_dist_clust <- hclust(mtDNAtab_transpose_dist)
pdf("/home/nazarii/Drosophila/Haplotypemap/clustering_Dsim.pdf",height=
8,width=10)
plot(mtDNAtab_transpose_dist_clust)
dev.off()

```

Для побудови мапи з відображенням розподілу відібраних ліній для аналізу записували дані про довжину та широту отриманих ліній в таблицю Excel. Мапу розподілу будували за допомогою середовища Rstudio. Код програми для побудови мапи в Rstudio:

```

library(ggmap)
library(ggplot2)
library(maptools)
library(maps)
library(readr)
library(ggrepel)
Dsim_metadata <- read_csv("D.simulans metadata.csv")
Dsim_metadata_df<- as.data.frame(Dsim_metadata)
world_map <- map_data("world")
ggplot() + geom_polygon(data = world_map, aes(x=long, y = lat, group =
group),fill = 'grey') +
coord_fixed(1.5)+geom_point(data = Dsim_metadata, aes(x =
Dsim_metadata$longitude, y = Dsim_metadata$latitude),col = 'black')+

```

```
geom_text_repel(aes(x = Dsim_metadata$longitude, y =
Dsim_metadata$latitude),label = Dsim_metadata$lineID, direction = "y",
nudge_x = 26, hjust = 0, segment.size = 0.3, segment.color = 'black', col =
'black') + theme_minimal()+scale_color_discrete()
```

Загальна методологія аналізу мітохондріальної ДНК представлена на рисунку 2.2.

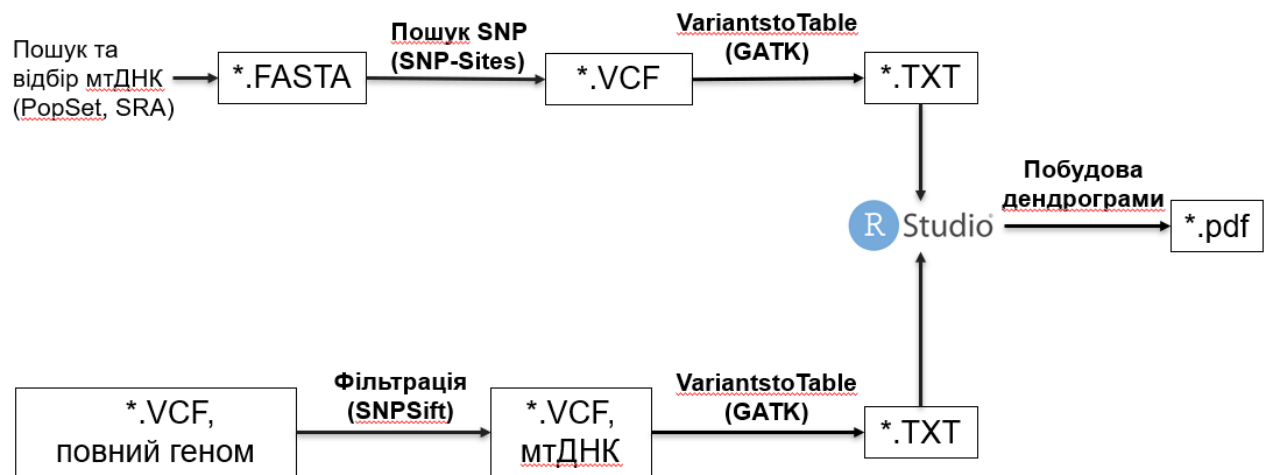


Рис. 2.2 Аналіз та пошук гаплотипу мітохондріальної ДНК ізосамкових ліній *D. simulans* з м.Одеса шляхом побудови кластерної дендрограми в

Rstudio

2.4. Аналіз повногеномних зчитувань штаму *Wolbachia* ізольованих ліній *D. simulans*

Генотип ендосимбіонту *Wolbachia* в складі секвенованих ліній *D. simulans* визначали шляхом мапування на відомі штами *Wolbachia*, отриманих з бази даних “Genome” [38] (www.ncbi.nlm.nih.gov/genome/). Візуалізацію результатів проводили за допомогою Integrative Genomics Viewer [43]. На основі даних візуалізації робили висновок про наявність

певного штаму ендосимбіонту *Wolbachia* в складі природніх популяцій *D. simulans* м. Одеса.

За допомогою SAMtools та VCFtools отримували однонуклеотидні поліморфізми у форматі VCF [37]. Далі визначали кількість однонуклеотидних поліморфізмів в складі відсеквенованого геному визначеного штаму *Wolbachia* та порівнювали з даними, отриманими в дослідженнях раніше.

2.5. Аналіз продуктів генів діпаузи одеських ліній *D. simulans* та світу

За допомогою програмного застосунку SNPEff [40] визначали типи однонуклеотидних замін командою в терміналі Linux:

```
// java -Xmx8g -jar snpEff.jar dsim_r2.02 input.vcf > out.vcf
```

Для отримання інформації щодо назви білкової послідовності, доменної організації, варіантів сплайсингу та його розташування всередині геному використовували бази даних UniProtKB[44] (<http://uniprot.org>) та FlyBase [45] (<http://flybase.org>). Зібрану інформацію записували у відповідну таблицю(додаток А).

Використовували відповідний bash скрипт для отримання продукту гена діпаузи *D. simulans*:

```
#!/bin/bash
echo Insert Name, ID, Variant of Sequence
read -r outname names varname
grep -F "$names" FullyAnnotatedGenomeOdesa.vcf > $names.vcf
java -Xmx6g -jar snpEff.jar ann -fastaProt $names.fa dsim_r2.02
$names.vcf>out_$names.vcf
grep -A1 "$varname" $names.fa | sed -e 's/--//g' > $outname.fasta
```

Вихідні файли з наявною інформацією про поліморфізми білкової послідовності об'єднували в один FASTA файл за допомогою середовища Rstudio [42]. Код програми для об'єднання білкових поліморфізмів в один файл FASTA:

```
library(ape)
library(seqinr)
library(dplyr)
library(stringr)
library(tidyr)
library(glue)

sequence_dir <- readline(prompt="Enter name of file: ")
seq_name <- readline(prompt = "Enter out sequence name:")
seq <- read.fasta(sequence_dir,'AA',whole.header = T,set.attributes = F)
seq_df <- as.data.frame(seq)
match <- NULL
for (i in colnames(seq_df))
{ match[[i]] = ifelse(seq_df[,1]==seq_df[,i],NA,seq_df[,i])}
match_df <- as.data.frame(match,col.names = 1:ncol(seq_df))
SNP <- unite(match_df,new,X1:ncol(match_df),na.rm = T)
SNP[SNP == ""] <- NA
fasta <- cbind(seq_df[1],SNP)
fasta$new[is.na(fasta$new)] <- fasta[,1][is.na(fasta$new)]
fasta_out<- fasta[2]
write.fasta(fasta_out, seq_name,glue("{seq_name}.fasta"))
```

У базі даних “Assembly” [46] (<https://www.ncbi.nlm.nih.gov/assembly/>) шукали та зберігали геномні послідовності ліній *D. simulans* світу. Для передбачення білкових послідовностей використовували референтний протеом лінії w501 *D. simulans* [47]. Для передбачення білкових

послідовностей завантажених ліній *D. simulans* використовували команду програмного забезпечення GeneMark [48]:

```
// gmes_petap.pl --seq genome.fna --EP --dbep protein.faa --v -cores=8
max_intergenic 10000 --mask_penalty 0
```

Серед передбачених протеомів фільтрували продукти генів діпаузи та записували у окремий FASTA файл. Для міжвидового порівняння білків, залучених в появі діпаузи, завантажували з бази даних “Genome” [38] (www.ncbi.nlm.nih.gov/genome/) білкові послідовності відповідних представників Drosophilidae: *D. melanogaster*, *D. suzukii*, *D. virilis*, *D. ananassae.*, *D. yakuba*. Отримані протеоми *D. simulans* та інших представників Drosophilidae використовували для пошуку відповідних білкових послідовностей за допомогою інструменту grep:

```
#!/bin/bash
echo Insert desired Sequence and outfile name
read varname outname
grep -F "$varname" *.faa > $outname.fasta
sed -i 's/^/>/' $outname.fasta
sed -i 's/.faa:\n/' $outname.fasta
```

Отримані FASTA файли одеських та світових популяцій *D. simulans*, а також інших видів аналізували та порівнювали за допомогою програми MEGA X [49]. Для множинного вирівнювання використовували метод “ClustalW” із параметрами за замовчуванням. Для побудови філогенетичних дерев використовували метод максимальної правдоподібності та матричної моделі JTT[49]. Усі позиції, які містять прогалини та відсутні дані, були усунені. Отримані дані записували в таблицю Excel та порівнювали з лініями світу. Загальна методологія аналізу білкових послідовностей продемонстрована на рисунку 2.3.

Пошук геномів *D. simulans* та
референтного протеому в
базі даних Assembly
(NCBI)

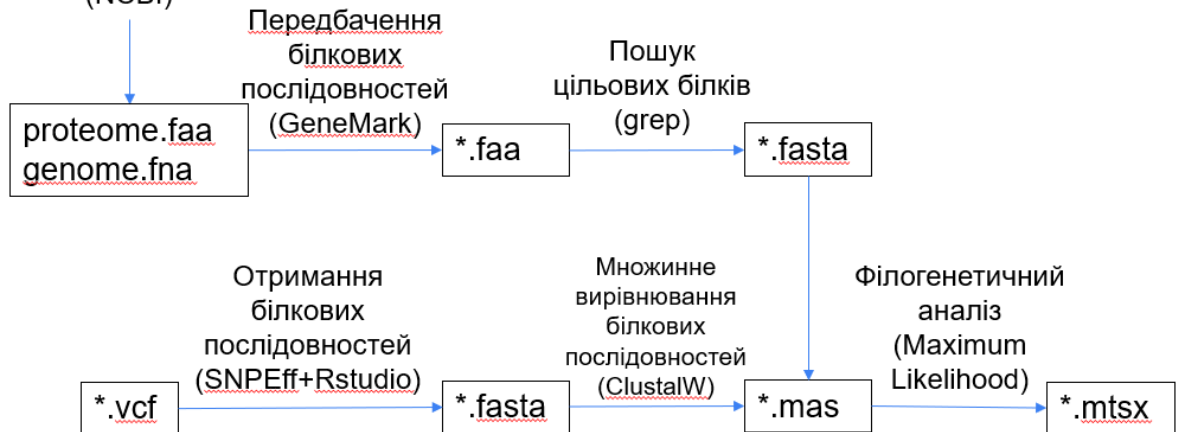


Рис. 2.3 Аналіз поліморфізмів продуктів генів діапаузи

РОЗДІЛ 3. РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ ТА ЇХ ОБГОВОРЕННЯ

3.1. Секвенування ліній *D. simulans* з м.Одеси

На основі зібраних попередньо 7 ізосамкових ліній *D. simulans* у м. Одеса (OD24, OD48, OD57, OD80 - 2015 року збору; Iry7, Iry8, Iry9 - 2018 року збору) було виділено їх тотальну ДНК. Дані секвенування, отримані за допомогою Illumina HiSeq2000 у лабораторії Andrea Betancourt (Ліверпульський університет, Велика Британія), були проаналізовані за представленими файлами, наданими лабораторією. Отримані повногеномні зчитування було відфільтровано та вирівняно на референтний геном лінії w501 *D. simulans* [47]. Було виявлено 3022829 SNP у відсеквенованих ліній відносно референтного геному з частотою мутацій 1 варіант на кожні 39 пар основ.

3.2. Кладистичний аналіз даних секвенування

З бази даних PopSet [38] було відібрано 32 мітохондріальних геноми *D. simulans* з різних регіонів світу (Америка, Африка, Океанія), лінії з яких відносяться до усіх трьох відомих гаплогруп *D. simulans*: siI, siII та siIII. Ідентифікаційний номер для пошуку відсеквенованих мітохондріальних геномів у базі даних PopSet: 8573373. Інформацію про локацію збору та назви ліній для аналізу зображені на рисунку 3.1. Отриманий архів нуклеотидних послідовностей мтДНК у форматі FASTA об'єднали і отримали файл для пошуку однонуклеотидних поліморфізмів серед мітохондріальних геномів ліній *D. simulans* світу. Отриманий файл перетворили на таблицю SNP сайтів.

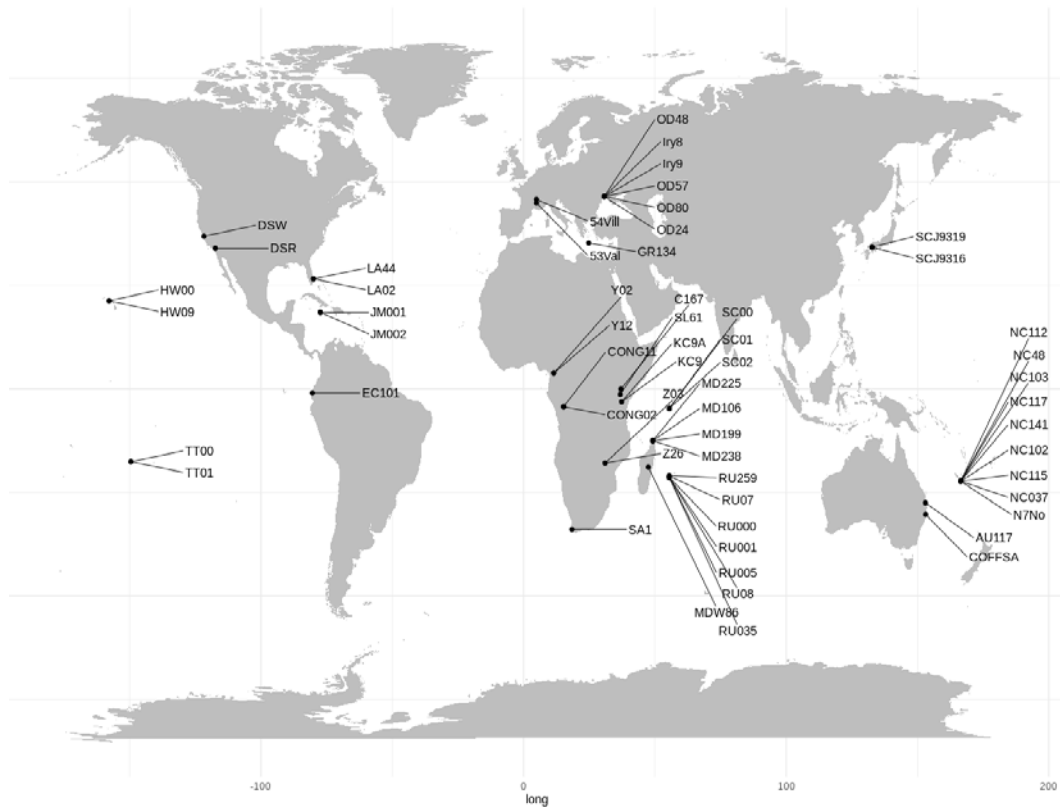


Рис.3.1 Відсеквеновані мітохондріальні геноми *D. simulans* разом із лініями з бази даних PopSet(NCBI), та їхні географічні координати.

З файлу VCF, у якому міститься інформація про одонуклеотидні поліморфізми ізосамкових ліній *D. simulans* з м. Одеси, отримали інформацію про SNP мітохондріальних геномів. Отриманий файл конвертували у SNP таблицю.

Дані про SNP ліній світу та популяцій з м. Одеси об'єднали в одну таблицю, що використовували для подальшого кладистичного аналізу. За допомогою методу повного зв'язку отримали кластерну дендрограму та встановили спорідненість ізосамкових ліній *D. simulans* з м.Одеса із лініями трьох гаплогруп *D. simulans*. Кластерний аналіз показав об'єднання ліній з України з лінією C167 та sm21 (Кенія, Африка, siII гаплогрупа), еволюційна відстань між якими була найменшою згідно кластерної дендрограми (рис. 3.2).

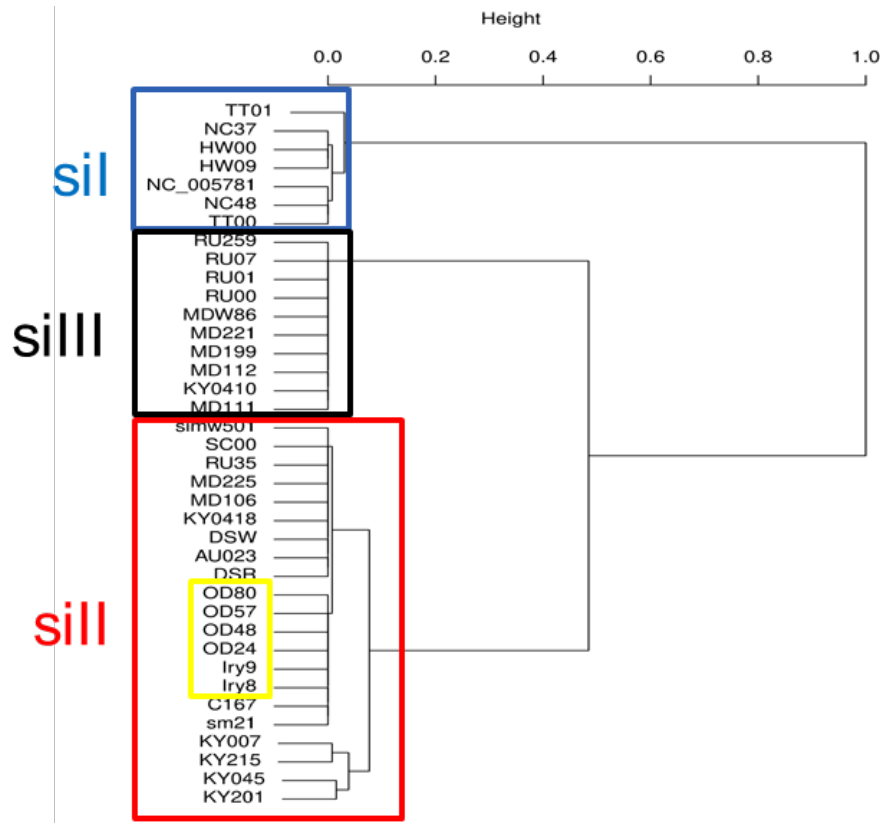


Рис.3.2 Кластерна дендрограма проаналізованих ізосамкових ліній *D. simulans*. Жовтим кольором виділені популяції *D. simulans* колекціоновані в м.Одеса. Синім кольором показані лінії, що належать до гаплотипу siI, чорним – до siIII, і червоним до siII відповідно.

За допомогою кластерної дендрограми отримали інформацію щодо генетичної спорідненості відсеквенованих ліній *D. simulans* м.Одеса. Встановлено генетичне споріднення із африканськими лініями C167 та sm21, що може свідчити про поширення проаналізованих ліній з субтропічного або тропічного регіону Африки в помірний кліматичний пояс, що з узгоджується із попередніми результатами поширення *D. simulans* на Євразійський континент [13]. Однак слід зазначити, що в Україну особини *D. simulans* могли потрапити і не безпосередньо з Африки, а через Європу, після колонізації африканськими представниками останньої. Проте для точного з'ясування цього питання необхідно порівняти отримані нами послідовності з геномами мух з Європи, які наразі відсутні у базах даних. Більше того, реальні шляхи

та терміни поширення даного виду на Євразійському континенті залишаються не вивченими та потребують подальших досліджень.

В результаті об'єднання в одну кладу одеських популяцій із лініями siII гаплотипу зробили висновок щодо приналежності відсеквенованих ліній *D. simulans* м.Одеса до siII гаплотипу, який є переважаючим у північній півкулі і представлений у помірних широтах. Дане дослідження узгоджується із отриманими даними Serga [3], де також вказали приналежність природніх популяцій *D. simulans* України до siII гаплотипу шляхом аналізу послідовностей гена Цитохром с-оксидази субодиниці I (COI). Однак аналіз послідовності короткого фрагменту одного гена для встановлення гаплотипу не є прецизійним, тому розробка методу для аналізу мітохондріального геному дозволила включити в аналіз всі наявні поліморфізми в його складі і з більшою точністю встановити конкретний гаплотип у відсеквенованих ліній. Крім того, розроблений метод можна застосовувати для знаходження гаплотипу не лише у відсеквенованих *D. simulans*, але й для аналізу гаплотипів інших представників *Drosophilidae*, що може значно розширити інформацію про шляхи колонізації нових територій. В майбутньому такий метод можна застосовувати і для аналізу зчитувань із пристроїв секвенування третього покоління, зокрема Oxford Nanopore. Такий метод був би зручним для польових досліджень, в умовах яких необхідно досліджувати багато природніх ліній одночасно, а культивування зібраних мух є неможливим чи сильно ускладненим. Крім того, комерційні інструменти від компанії Illumina для аналізу зчитувань та одонуклеотидних поліморфізмів на базі Illumina HiSeq2000 є дорогі, а використання розробленого методу значно зменшує вартість дослідження, оскільки всі інструменти є в безкоштовному доступі та легко інсталиються.

3.3. Встановлення штаму ендосимбіонта *Wolbachia* у відсеквенованих лініях *D. simulans*

Для генотипування *Wolbachia* у відсеквенованих зразках, використали 5 штамів (wHa, wNo, wRi, wAu, wMa), охарактеризованих раніше[20]. За допомогою повногеномного аналізу встановили наявність штаму wRi у всіх відсеквенованих зразках *D. simulans* (Рис.3.3). Загалом серед геномів *Wolbachia* одеських ліній було знайдено 22 однонуклеотидні поліморфізми та інсерції/делеції відносно референтного геному wRi.



Рис.3.3 Мапування відсеквенованих зразків на референтні штами ендосимбіонту *D. simulans* (на прикладі мапування з wRi та wNo).

Підтвердженням наявності гаплотипу siII в популяціях *D. simulans* м.Одеса є також встановлення wRi штаму ендосимбіонту *Wolbachia*, яким були інфіковані ізосамкові лінії мух. Як було вище зазначено, штам wRi є домінуючим серед siII гаплотипу і він здатен викликати цитоплазматичну несумісність у хазяїна. У дослідженні Ballard [20] було проаналізовано 2 лінії *D. simulans* з Ялти: UR8 та UR25. У цих лініях також був встановлений siII гаплотип мтДНК та були інфіковані wRi штамом. Згодом дослідження Serga [3] показало наявність штаму wRi у ізосамкових ліній *D. simulans* м.Одеса. Наше дослідження є підтвердженням поширеності штаму wRi на

євразійському континенті, зокрема на території України. Даний метод встановлення штаму є точнішим за електрофоретичне розділення ПЛР продуктів, оскільки запобігає отриманню хибно позитивних результатів і дозволяє порівнювати результати із декількома штамами одночасно.

3.4. Дослідження продуктів генів діапаузи ліній зі світу та України

Інформацію про гени, що впливають на розвиток діапаузи, було записано у таблицю Excel. До таблиці додали інформацію про назви генів, доменну організацію, та назви варіантів трансляції.

Для аналізу було відібрано 5 ліній з різних регіонів світу (dsgoth613,dsgoth31 — Франція, Європа; dssj27, dssj9 — Бразилія, Південна Америка; M252 — Мадагаскар, Африка) з бази даних “Assembly” [46] та завантажені повногеномні зчитування. За допомогою програми GeneMark [48] було передбачено білкові послідовності для всіх 5 аналізованих ліній. З бази даних “Genome” отримали протеоми інших представників Drosophilidae: *D. melanogaster*, *D. sukukii*, *D. virilis*, *D. ananassae*, *D. yakuba*. В результаті фільтрування передбачених протеомів *D. simulans* та представників Drosophilidae алгоритмом grep було знайдено білкові послідовності, пов’язані з діапаузою.

За допомогою SnpEff [40] та програмного коду Rstudio отримали дані білкові послідовності у відсеквенованих ізосамкових лініях *D. simulans* з м.Одеса. Зібрані дані використали для множинного вирівнювання білкових послідовностей та побудови кладограм в програмі MEGA X [49].

Загалом було отримано множинні вирівнювання для 33 генів, пов’язаних із діапаузою у *D. simulans*. Інформація про назву гена, назва варіанту трансляції, доменну організацію, та філогенетичну спорідненість представлена у таблиці 3.1.

Таблиця 3.1 Аналіз продуктів генів, пов'язаних із діапаузою, у відсеквенованих ліній *D. simulans* м.Одеса

Назва гена	Домен	Мутація (HGVS-номенклатура)	Подібні мутації у інших лініях <i>D. simulans</i> та <i>Drosophilidae</i>
<i>dock</i>	SH3(3),SH2	Ser192Asn	Dssj27
<i>chico</i>	PH	Ser36Ala	Dssj27, Dsgoth613, Dsgoth31, M252
<i>Pdk1</i>	S_TKc	Pro425Leu	Dsgoth613
<i>Pdk1</i>	PH_3	Leu762Trp	унікальна
<i>Pi3K92E</i>	PI3Kc	Leu831Phe	унікальна
<i>InR</i>	FU	Ile576Val	унікальна
<i>InR</i>	Fibronectin тип 3	Gly1079Ser	M252
<i>Cat</i>	Catalase	Ile97Leu	Dssj9
<i>Cat</i>	Catalase	Ile297Val	M252, <i>D. melanogaster</i>
<i>CG8630</i>	FA_desaturase	Leu342Ile	Dsgoth613
<i>CG17928</i>	Cyt-b5	Asp56Asn	Dssj9
<i>CG17928</i>	Cyt-b5	Gly97Ser	Dssj9, M252
<i>Eip71CD</i>	PMSR	Val153Ile	<i>D. ananassae</i>
<i>Eip71CD</i>	EGF like	Leu819Met	унікальна

Однією з особливостей відсеквенованих ізосамкових ліній *D. simulans* є здатність переживати суворі помірні широти шляхом входження в репродуктивну діапаузу. Саме тому розробка методу аналізу генів, пов'язаних з діапаузою, надала б корисну інформацію про еволюцію цього виду, шляхи його переміщення та стрімкої колонізації помірного клімату. Як було зазначено, наявність діапаузи часто є генетично детермінованою ознакою, що залежить від одонуклеотидних змін всередині генів, пов'язаних із накопиченням та зменшенням біосинтезу ювенільного гормону. Зокрема, було отримано продукти генів, що впливають на перебіг діапаузи, а також вивчено склад поліморфізмів всередині доменів, які є основними структурними одиницями білків та еволюційно консервативними. Серед 33 проаналізованих генів відібрали ті, які мають мутації всередині доменів та відрізняються від референтного геному *D. simulans*. Серед проаналізованих

генів особливу увагу мають ті, мутації яких є унікальними або містять подібні мутації з мухами, які мають здатність переживати суворі кліматичні умови шляхом входження в репродуктивну діапаузу (*D. melanogaster*, *D. ananassae*). В результаті було виявлено 14 поліморфізмів які здатні потенційно впливати на перебіг діапаузи у *D. simulans* (таблиця 3.1). Серед них це ген *Eip71CD*, який має мутацію у PMSR домені разом із *D. ananassae*. Ген *Cat* має спільну мутацію з *D. melanogaster* у каталазному домені. Унікальними є мутації в генах *Pdk1* (PH_3 домен), *Pi3K92E* (PI3Kс домен), *InR* (FU домен), та *Eip71CD* (EGF-like домен). Отримані результати можна використати в подальшому для дослідження цих мутацій, а також дослідити їхній вплив на входження ізосамкових ліній *D. simulans* з м.Одеси в стан репродуктивної діапаузи.

ВИСНОВКИ

1. У складі відсеквенованих геномів ізосамкових ліній *D. simulans* на території України виявлено 3022829 однонуклеотидних поліморфізмів.
2. У 7 лініях одеської популяції *D. simulans* наявний штам wRi, що належить до групи гаплотипів siII.
3. Всі відсеквеновані лінії інфіковані wRi штамом ендосимбіонта *Wolbachia*. Виявлено 22 однонуклеотидні поліморфізми відносно референтного геному wRi *Wolbachia*.
4. Кладистичний аналіз мітохондріальних геномів *D. simulans* з популяції м.Одеса засвідчив про імовірне походження цих ліній з Африки.
5. Серед проаналізованих 33 продуктів генів, пов'язаних із діапаузою, встановлено 14 поліморфізмів, відмінних від референтного геному *D. simulans*.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] P. Capy and P. Gibert, “*Drosophila melanogaster*, *Drosophila simulans*: so similar yet so different,” *Drosophila melanogaster, Drosophila simulans: So Similar, So Different*, pp. 5–16, 2004.
- [2] T. K. Watanabe, Y. Inoue, and M. Watada, “Adaptation of *Drosophila simulans* in Japan,” *The Japanese Journal of Genetics*, vol. 59, no. 3, pp. 225–235, 1984.
- [3] S. V. Serga, O. M. Maistrenko, A. I. Rozhok, T. A. Mousseau, and I. A. Kozeretska, “Colonization of a temperate-zone region by the fruit fly *Drosophila simulans* (Diptera: Drosophilidae),” *Canadian Journal of Zoology*, vol. 93, no. 10, pp. 799–804, 2015.
- [4] V. Zonato, L. Collins, M. Pegoraro, E. Tauber, and C. P. Kyriacou, “Is diapause an ancient adaptation in *Drosophila*?,” *J Insect Physiol*, vol. 98, pp. 267–274, 2017.
- [5] A. H. Sturtevant, “Genetic studies on *Drosophila simulans*. I. Introduction. Hybrids with *Drosophila melanogaster*,” *Genetics*, vol. 5, no. 5, p. 488, 1920.
- [6] G. Bonnier, “Contributions to the knowledge of intra- and inter-specific relationships in *Drosophila*,” *Acta Zoologica*, vol. 5, no. 1- 2, pp. 1–122, 1924.
- [7] B. Ephrussi, “Sur le chondriome ovarien des *Drosophila melanogaster* et *simulans*,” *CR Seanc. Soc. Biol., Paris*, vol. 92, pp. 778–780, 1925.
- [8] J. Kerkis, “Chromosome conjugation in hybrids between *Drosophila melanogaster* and *Drosophila simulans*,” *The American Naturalist*, vol. 70, no. 726, pp. 81–86, 1936.
- [9] H. J. Muller, “Recessive genes causing interspecific sterility and other disharmonies between *Drosophila melanogaster* and *simulans*,” *Genetics*, vol. 27, p. 157, 1942.

[10] A. O. Tantawy and M. H. Soliman, "Studies on natural populations of *Drosophila* VI. competition between *Drosophila melanogaster* and *Drosophila simulans*," *Evolution (N Y)*, pp. 34–40, 1967.

[11] P. Hyytia, P. Capy, J. R. David, and R. S. Singh, "Enzymatic and quantitative variation in European and African populations of *Drosophila simulans*," *Heredity (Edinb)*, vol. 54, no. 2, pp. 209–217, 1985.

[12] R. S. Singh and A. D. Long, "Geographic variation in *Drosophila*: from molecules to morphology and back," *Trends in Ecology & Evolution*, vol. 7, no. 10, pp. 340–345, 1992.

[13] D. Lachaise and J.-F. Silvain, "How two Afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster*-*D. simulans* palaeogeographic riddle," *Drosophila melanogaster, Drosophila simulans: so similar, so different*, pp. 17–39, 2004.

[14] J. R. David *et al.*, "Comparative life histories and ecophysiology of *Drosophila melanogaster* and *D. simulans*," *Drosophila melanogaster, Drosophila simulans: So Similar, So Different*, pp. 151–163, 2004.

[15] A. A. Hoffmann and M. Watson, "Geographical variation in the acclimation responses of *Drosophila* to temperature extremes," *The American Naturalist*, vol. 142, pp. S93–S113, 1993.

[16] C. Bandi, T. J. C. Anderson, C. Genchi, and M. L. Blaxter, "Phylogeny of *Wolbachia* in filarial nematodes," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 265, no. 1413, pp. 2407–2413, 1998.

[17] A. Jeyaprakash and M. A. Hoy, "Long PCR improves *Wolbachia* DNA amplification: wsp sequences found in 76% of sixty- three arthropod species," *Insect Mol Biol*, vol. 9, no. 4, pp. 393–405, 2000.

[18] J. A. J. Breeuwer and J. H. Werren, "Microorganisms associated with chromosome destruction and reproductive isolation between two insect species," *Nature*, vol. 346, no. 6284, pp. 558–560, 1990.

[19] G. D. D. Hurst and M. E. N. Majerus, “Why do maternally inherited microorganisms kill males?,” *Heredity (Edinb)*, vol. 71, no. 1, pp. 81–95, 1993.

[20] J. W. O. Ballard, “Sequential evolution of a symbiont inferred from the host: *Wolbachia* and *Drosophila simulans*,” *Molecular Biology and Evolution*, vol. 21, no. 3, pp. 428–442, 2004.

[21] J. H. Yen and A. R. Barr, “The etiological agent of cytoplasmic incompatibility in *Culex pipiens*,” *J Invertebr Pathol*, vol. 22, no. 2, pp. 242–250, 1973.

[22] A. O. Belousov and I. A. Kozeretskaia, “Symbiotic bacteria, which modify reproduction processes of *Drosophila melanogaster*,” *Mikrobiolohichnyi Zhurnal (Kiev, Ukraine: 1993)*, vol. 73, no. 2, pp. 43–52, 2011.

[23] M. Turelli and A. A. Hoffmann, “Cytoplasmic incompatibility in *Drosophila simulans*: dynamics and parameter estimates from natural populations.,” *Genetics*, vol. 140, no. 4, pp. 1319–1338, 1995.

[24] F. Rousset and M. Solignac, “Evolution of single and double *Wolbachia* symbioses during speciation in the *Drosophila simulans* complex,” *Proceedings of the National Academy of Sciences*, vol. 92, no. 14, pp. 6389–6393, 1995.

[25] H. Mercot, B. Llorente, M. Jacques, A. Atlan, and C. Montchamp-Moreau, “Variability within the Seychelles cytoplasmic incompatibility system in *Drosophila simulans*.,” *Genetics*, vol. 141, no. 3, pp. 1015–1023, 1995.

[26] Y. Kurogi, Y. Mizuno, E. Imura, and R. Niwa, “Neuroendocrine Regulation of Reproductive Dormancy in the Fruit Fly *Drosophila melanogaster*: A Review of Juvenile Hormone-Dependent Regulation,” *Frontiers in Ecology and Evolution*, vol. 9. Frontiers Media S.A., Sep. 23, 2021. doi: 10.3389/fevo.2021.715029.

[27] T. Flatt, “Life-history evolution and the genetics of fitness components in *Drosophila melanogaster*,” *Genetics*, vol. 214, no. 1, pp. 3–48, 2020.

- [28] D. K. Fabian *et al.*, “Spatially varying selection shapes life history clines among populations of *Drosophila melanogaster* from sub-Saharan Africa,” *Journal of Evolutionary Biology*, vol. 28, no. 4, pp. 826–840, 2015.
- [29] D. S. Saunders, D. S. Richard, S. W. Applebaum, M. Ma, and L. I. Gilbert, “Photoperiodic diapause in *Drosophila melanogaster* involves a block to the juvenile hormone regulation of ovarian maturation,” *Gen Comp Endocrinol*, vol. 79, no. 2, pp. 174–184, 1990.
- [30] M. Tatar and C.-M. Yin, “Slow aging during insect reproductive diapause: why butterflies, grasshoppers and flies are like worms,” *Exp Gerontol*, vol. 36, no. 4–6, pp. 723–738, 2001.
- [31] N. Ojima, Y. Hara, H. Ito, and D. Yamamoto, “Genetic dissection of stress-induced reproductive arrest in *Drosophila melanogaster* females,” *PLoS Genet*, vol. 14, no. 6, p. e1007434, 2018.
- [32] P. Lankinen, “Geographical variation in circadian eclosion rhythm and photoperiodic adult diapause in *Drosophila littoralis*,” *Journal of Comparative Physiology A*, vol. 159, no. 1, pp. 123–142, 1986.
- [33] T. S. Salminen and A. Hoikkala, “Effect of temperature on the duration of sensitive period and on the number of photoperiodic cycles required for the induction of reproductive diapause in *Drosophila montana*,” *J Insect Physiol*, vol. 59, no. 4, pp. 450–457, 2013.
- [34] M. Martin, “Cutadapt removes adapter sequences from high-throughput sequencing reads,” *EMBnet. journal*, vol. 17, no. 1, pp. 10–12, 2011.
- [35] Joshi NA and Fass JN, “Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software].” 2011.
- [36] H. Li and R. Durbin, “Fast and accurate long-read alignment with Burrows–Wheeler transform,” *Bioinformatics*, vol. 26, no. 5, pp. 589–595, 2010.
- [37] P. Danecek *et al.*, “Twelve years of SAMtools and BCFtools,” *Gigascience*, vol. 10, no. 2, p. giab008, 2021.
- [38] E. W. Sayers *et al.*, “Database resources of the national center for biotechnology information,” *Nucleic Acids Res*, vol. 49, no. D1, p. D10, 2021.

- [39] A. J. Page *et al.*, “SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments,” *Microb Genom*, vol. 2, no. 4, 2016.
- [40] P. Cingolani *et al.*, “A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3,” *Fly (Austin)*, vol. 6, no. 2, pp. 80–92, 2012.
- [41] A. McKenna *et al.*, “The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data,” *Genome Res*, vol. 20, no. 9, pp. 1297–1303, 2010.
- [42] Rs. Team, “RStudio: integrated development for R. RStudio, PBC, Boston, MA. 2020.” 2021.
- [43] J. T. Robinson *et al.*, “Integrative genomics viewer,” *Nat Biotechnol*, vol. 29, no. 1, pp. 24–26, Jan. 2011, doi: 10.1038/nbt.1754.
- [44] T. U. Consortium, “UniProt: the universal protein knowledgebase in 2021,” *Nucleic Acids Research*, vol. 49, no. D1, pp. D480–D489, Jan. 2021, doi: 10.1093/nar/gkaa1100.
- [45] “FlyBase: The *Drosophila* Database,” *Nucleic Acids Research*, vol. 24, no. 1, pp. 53–56, Jan. 1996, doi: 10.1093/nar/24.1.53.
- [46] P. A. Kitts *et al.*, “Assembly: a resource for assembled genomes at NCBI,” *Nucleic Acids Res*, vol. 44, no. D1, pp. D73–D80, 2016.
- [47] T. T. Hu, M. B. Eisen, K. R. Thornton, and P. Andolfatto, “A second-generation assembly of the *Drosophila simulans* genome provides new insights into patterns of lineage-specific divergence,” *Genome Res*, vol. 23, no. 1, pp. 89–98, 2013.
- [48] T. Brůna, A. Lomsadze, and M. Borodovsky, “GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins,” *NAR Genom Bioinform*, vol. 2, no. 2, p. lqaa026, 2020.
- [49] S. Kumar, G. Stecher, M. Li, C. Knyaz, and K. Tamura, “MEGA X: molecular evolutionary genetics analysis across computing platforms,” *Mol Biol Evol*, vol. 35, no. 6, p. 1547, 2018.

ДОДАТКИ

Додаток А

Таблиця 2.1. Гени, пов'язані з діапаузою, що були включені в аналіз

Назва гена у <i>D. melanogaster</i>	Назва гена у <i>D. simulans</i>	Хромосома	Назва варіанту сплайсингу	Домен
tim	tim	2L	FBtr0222661	timeless
dock	GD23050	2L	FBtr0357181	SH3(3),SH2
chico	GD22290	2L	FBtr0222200	PH,IRS
Pi3K21B	GD23016	2L	FBtr0222926	SH2(2), PI3K_P85_iSH2
melt	GD13996	3L	FBtr0355345	PH
Ilp-4	GD12891	3L	FBtr0212801	IIIGF domain
Ilp-8	GD12429	3L	FBtr0212339	-
Pdk1	GD13545	3L	FBtr0365517	S_TKc, PH_3
RpL8	GD13692	3L	FBtr0213602	Ribosomal_L2, Ribosomal_L2_C
Pi3K92E (Dp110)	GD20048	3R	FBtr0358541	PI3K_p85B, PI3K_rbd, PI3K_C2, PI3Ka, PI3Kc
cpo	GD19192	3R	FBtr0359615	RRM domain
foxo	GD18943	3R	FBtr0348901	FH
InR	GD19971	3R	FBtr0219881	FU,Fibronectin type 3 domain
Act42A	GD10342	2R	FBtr0210301	Actin
Arr1	GD24126	2L	FBtr0224037	Arrestin_C
Cat	GD14770,Cat	3L	FBtr0214680	Catalase
CG6785	GD22139	2L	FBtr0222052	-
CG8630	GD18893	3R	FBtr0218803	FA_desaturase
CG15531	GD17223	3R	FBtr0217144	-
CG17928	GD24082	2L	FBtr0358633	Cyt-b5,FA_desaturase
Droj2	GD18905	3R	FBtr0220413	Tryp_SPc
drpr	GD13447	3L	FBtr0352698	EGF
dy	GD15971,	X	-	Zona pellucida (ZP) domain
Eip71CD	GD14549	3L	FBtr0214460	PMSR
fru	GD20167	3R	FBtr0354475	BTB
Gale	GD13563	3L	FBtr0352865	adh_short domain
Hsc70	CG4147	X	FBtr0351859	HSP70 domain
ken	GD13771	2R	FBtr0349594	BTB
LanA	GD13136	3L	FBtr0354948	LamNT, EGF_Lam, EGF like
Mhc	GD24080	2L	FBtr0352496	MYS domain,IQ domain
ninaD	GD18923	3R	FBtr0356665	-
vri	GD23334	2L	FBtr0223244	BRLZ domain
Thor	GD23237	2L	-	eIF_4EBP domain

