

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА**

ННЦ « Інститут біології та медицини»
Кафедра вірусології

Завідувач кафедри проф. Ірина БУДЗАНІВСЬКА
Протокол № засідання кафедри
від “ ____ ” _____ 2024 р.

**МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ВПЛИВУ ОСНОВНИХ
ЕВОЛЮЦІЙНИХ ФАКТОРІВ НА ДИНАМІКУ ВІРУСНИХ
ПОПУЛЯЦІЙ**

Кваліфікаційна робота бакалавра
денної форми навчання
за спеціальністю біологія
Остапчук Іванна Олегівна
Науковий керівник від кафедри
Д.б.н., проф. Будзанівська І.Г.

Оцінка захисту дипломної роботи

Київ-2024р.

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ ТА ПОЗНАЧЕНЬ

ЗТ – зворотна транскриптаза;

μ , u – частота мутацій;

G – розмір геному;

N – загальний розмір популяції;

N_e – ефективний розмір популяції;

s – коефіцієнт відбору

DFE - розподіл фітнес-ефектів

ЗМІСТ

ВСТУП.....	1
РОЗДІЛ 1. Огляд літератури	4
1.1. Загальна характеристика факторів, що впливають на еволюцію вірусів.....	4
1.2. Швидкість мутацій і мутаційний тиск у нестатевих популяціях вірусів.....	6
1.3. Основні популяційно-генетичні процеси, що формують вірусне різноманіття.....	9
1.4. Математичні наслідки подібних спостережень.....	17
РОЗДІЛ 2. Матеріали та методи досліджень.....	20
2.1. Матеріали досліджень.....	20
2.2. Методи досліджень та постановка моделі.....	20
РОЗДІЛ 3. Результати досліджень та обговорення	29
ВИСНОВКИ.....	41
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	42

ВСТУП

Віруси розвиваються під впливом імунної системи хазяїна. Цей процес відомий як коєволюція, коли імунна система постійно піддає віруси селективному тиску, змушуючи їх адаптуватися для виживання. В результаті раніше інфіковані особи стають імунними до конкретного штаму вірусу, що створює бар'єр для його розповсюдження. У відповідь вірус змушений знаходити нових, чутливих господарів або мутувати, змінюючи свої антигенні властивості, щоб уникнути розпізнавання імунною системою (Moya et al., 2004). Математично цей процес можна описати підвищенням пристосованості у просторі, шляхом набуття нових мутацій і зміни ландшафтів пристосованості. Такі ландшафти включають в себе різні генетичні варіанти, які можуть надавати вірусу різний ступінь виживаності та репродуктивного успіху в залежності від селективного тиску, що на нього діє. Таким чином, еволюція вірусів є динамічним процесом, що відбувається на фоні постійної взаємодії з імунною системою хазяїна та іншими факторами довкілля.

Зазвичай наукові дослідження концентруються навколо вигідних для вірусу мутацій, що збільшують пристосованість штаму і сприяють його активному поширенню у популяції. Ці мутації, що надають вірусу переваги, називаються адаптивними або корисними мутаціями. Вони можуть покращувати здатність вірусу до інфікування нових господарів, уникати імунного відповіді або збільшувати його реплікаційну швидкість. Наприклад, мутації в поверхневих білках вірусу, таких як гемаглютинін у вірусу грипу, можуть дозволяти вірусу уникати розпізнавання антитілами, виробленими після попередніх інфекцій або вакцинацій (Frost & Volz, 2010). Проте, згідно з емпіричними даними, більшість мутацій є нейтральними або шкідливими. Нейтральні мутації не мають значного впливу на життєздатність вірусу і часто накопичуються випадково через генетичний дрейф (Pybus & Rambaut, 2009). Шкідливі мутації, навпаки,

знижують пристосованість вірусу, наприклад, погіршуючи його здатність до реплікації або збільшуючи його чутливість до імунної відповіді. Однак навіть шкідливі мутації можуть зберігатися в популяції, особливо якщо вони знаходяться у геномі разом з корисними мутаціями через феномен, відомий як генетичне тягнення (Sanjuán et al., 2010).

Хоча корисні мутації становлять важливу частку досліджень з вірусних адаптацій, не менш цікавим є розглянути ефект впливу ефективного розміру популяції на еволюційні процеси вірусів. Ефективний розмір популяції, який відображає кількість індивідів, що беруть участь у репродукції і передачі генетичного матеріалу, варіює між різними видами вірусів. Ефективний розмір популяції впливає на швидкість еволюції та акумуляції шкідливих мутацій. Важливою є роль ефективного розміру популяції у розумінні еволюційної динаміки вірусів, так як це дозволяє оцінювати потенційні шляхи розвитку вірусних інфекцій і передбачати можливі еволюційні траєкторії. Таким чином, вивчення впливу ефективного розміру популяції на віруси є ключовим для розробки стратегій контролю і профілактики вірусних захворювань (Gandon & Day, 2009).

У роботі буде проведений теоретичний аналіз гіпотетичних популяцій для побудови математичної моделі оцінки еволюції вірусів та визначення їх генеалогічної моделі. Це передбачає створення симуляцій, що відображатимуть різні сценарії еволюції вірусів, враховуючи змінні, такі як швидкість мутацій, ефективний розмір популяції, селективний тиск і вплив імунної відповіді господаря. Застосування математичних моделей дозволяє глибше зрозуміти механізми, які визначають еволюційну динаміку вірусів, та передбачати можливі шляхи їх розвитку (Drummond & Rambaut, 2007).

Отримавши інформацію про режим еволюції різних вірусів, з'явиться можливість більш ефективно оцінювати швидкість та характер

появи нових варіантів вірусів. Зокрема, можна буде передбачити, які мутації можуть стати домінуючими у популяції і як швидко вони поширюватимуться. Це важливо для розробки ефективних стратегій боротьби з вірусними інфекціями, таких як вакцини та антивірусні препарати, які можуть бути більш цілеспрямованими і адаптованими до конкретних штамів вірусу (Grenfell et al., 2004). Це дозволяє не лише розуміти поточну динаміку вірусних інфекцій, але й передбачати майбутні тенденції, що є критично важливим для глобальної охорони здоров'я і управління інфекційними хворобами.

Саме тому, метою роботи було використання математичної моделі Райта-Фішера в режимі шкідливих мутацій для оцінки впливу ефективного розміру популяції та коефіцієнта селекції на еволюцію безстатевих вірусних популяцій.

Завдання:

- 1) Створити і запустити математичну модель на базі теорії популяцій Райта-Фішера, що описує еволюцію вірусних популяцій у різних мутаційних режимах
- 2) Вирахувати та підібрати параметри з реальних даних вірусних популяцій
- 3) Перевірити модель на гіпотетичних вірусних популяціях декількох вибірок вірусів з різними ефективними розмірами популяції, режимами мутацій та селективним тиском для візуалізації темпу та динаміки еволюції

РОЗДІЛ 1. Огляд літератури

1.1. Загальна характеристика факторів, що впливають на еволюцію вірусів

Віруси мають дуже гетерогенні геномні структури та стратегії реплікації. Ці основні особливості корелюють зі схильністю вірусів до мутацій, а отже, з генетичною різноманітністю та еволюційністю вірусів. Швидкість мутацій вірусів залежить від складу геному (РНК або ДНК), розміру та структури (одно- або дволанцюжкові). Загалом, РНК-віруси продукують від 10^{-6} до 10^{-4} нових замін основ на нуклеотид на клітину, тоді як у ДНК-вірусів цей показник коливається від 10^{-8} до 10^{-6} (рис. 1). Крім того, одноланцюгові віруси мутують швидше, ніж дволанцюгові, і не виявлено зворотної кореляції між розміром геному та швидкістю мутації. Таким чином, у групі ДНК-вірусів одноланцюгові віруси і, ймовірно, деякі дволанцюгові віруси з малими геномами мутують швидше, ніж дволанцюгові віруси з великими геномами. Термін "квасівид" використовується для позначення високогетерогенних вірусних популяцій, в яких частота мутацій в популяції визначається швидкістю мутацій і пристосованістю (Holmes et al., 2002).

Концепція квазівиду застосовується в основному до РНК-вірусів через їх особливо високу швидкість мутацій. Вважається, що висока генетична різноманітність популяції відіграє ключову роль у здатності РНК-вірусів адаптуватися до зміни селективного тиску і проникати в нові організми. Дійсно, РНК-віруси включають велику кількість добре відомих патогенних мікроорганізмів, таких як ВІЛ-1, вірус гепатиту С, вірус грипу, еболавірус і зікавірус, а також основні збудники хвороб рослин і тварин. ДНК-віруси, навпаки, зазвичай менш різноманітні і еволюціонують

повільніше, хоча вони також включають такі важливі патогени, як віруси герпесу та віспи. Тим не менш, поділ на РНК- і ДНК-віруси є недосконалим, оскільки деякі РНК-віруси еволюціонують повільно, а деякі ДНК-віруси швидко еволюціонують, зокрема, одноланцюгові і малі дволанцюгові ДНК-віруси (Moуа et al., 2000, Wang X et al, 2024). Ця стаття зосереджена на джерелах генетичної різноманітності вірусів та інших еволюційних процесах на популяційному рівні (ми не розглядаємо довготривалу еволюцію вірусів на філогенетичному рівні). Обговорюються основні механізми, відповідальні за внесення нових мутацій, включаючи помилки реплікації, відсутність надійного зчитування, уникнення репарації, елементи, що генерують різноманітність, редагування вірусних геномів хазяїном та рекомбінації (Таблиця 1). Ми будемо розглядати тут як популяційне різноманіття змінюється під впливом природного відбору та генетичного дрейфу, і як це призводить до різних еволюційних процесів у вірусах.

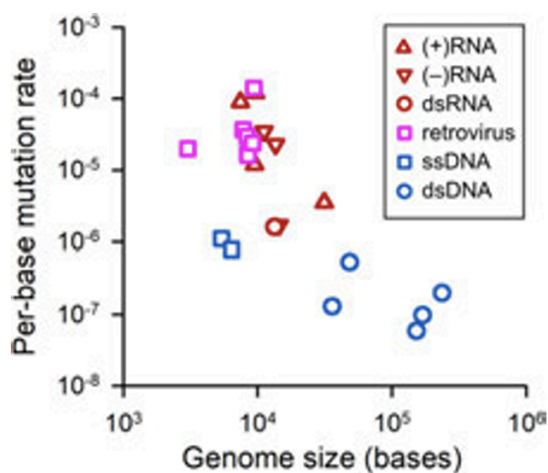


Рис 1. Частота мутацій у різних типах вірусів (взято з Sanjuán R, Domingo-Calap P. Genetic Diversity and Evolution of Viral Populations)

1.2. Швидкість мутацій і мутаційний тиск у нестатевих популяціях вірусів

Джерела генетичного різноманіття у вірусів досить різняться, але в загальній своїй тенденції це поява нових мутацій, асортативність і т.ін (Moysa et al., 2004).

Різні фактори впливають на ці параметри, вони підсумовані у таблиці 1

	<i>dsDNA</i>	<i>ssDNA</i>	<i>Para retro</i>	<i>Retro</i>	<i>dsRNA</i>	<i>(+)RNA</i>	<i>(-)RNA</i>
Lack of 3' exonuclease proofreading	-	+/-	+	+	+	+	+
Avoidance of post-replicative repair	-	+/-	+	+	+	+	+
Use of error-prone repair polymerases	+/-	+/-	-	-	-	-	-
Diversity-generating retro-elements	+/-	-	-	-	-	-	-
APOBEC hyper-mutation	+/-	+/-	+	+	-	-	-
ADAR hyper-mutation	-	-	-	-	+/-	+/-	+/-
Reassortment	+/-	+/-	-	-	+/-	+/-	+/-
Template switching	-	-	+	+	+	+	-
Recombinases	+/-	-	-	-	-	-	-

Таблиця 1. Джерела генетичного різноманіття семи основних класів вірусів (взято з Sanjuán R, Domingo-Calap P. Genetic Diversity and Evolution of Viral Populations)

Точність реплікації полімерази насамперед визначається її селективністю до правильної основи. Така селективність, як правило, становить 10^4 - 10^5 , і не має суттєвих відмінностей між РНК-залежними РНК-полімеразами, зворотними транскриптазами (ЗТ) та ДНК-полімеразами.

Варіанти полімераз зі зміненою селективністю та точністю були виявлені у широкого спектру РНК-вірусів. Вперше було показано, що послідовне перенесення в присутності аналога основи - рибавіріну - призводить до селекції полімерази з підвищеною точністю у поліовірусів. З того часу було ідентифіковано десятки мутантів з підвищеною точністю реплікації серед дуже різних сімейств РНК-вірусів (La Tourette & Garcia-Ruiz, 2022).

Еволюція варіантів з високою і низькою точністю у відповідь на селективний тиск, наприклад, хімічно-індукований мутагенез, демонструє, що відбір може оптимізувати реплікацію РНК-вірусів за точністю.

Однак, більшість варіантів вірності РНК-вірусів демонструють лише помірні відмінності у селективності до основ порівняно з диким типом, що свідчить про сильні структурні або біохімічні обмеження. Цікаво, що точність вірусних ДНК-полімераз, схоже, менш обмежена.

На додаток до внутрішньої точності полімерази, частота помилкових інкорпорацій основ залежить від таких факторів, як контекст послідовності, концентрація двовалентних катіонів та відносна кількість dNTP. Наприклад, прослизання полімерази часто відбувається в гомополімерних шаблонах, що призводить до виникнення вставок і делецій. У вірусу гепатиту С частота мутацій може сильно варіювати між послідовними основами у вірусному геномі, що свідчить про важливу роль контексту послідовності у визначенні точності реплікації. Вторинна структура РНК також може визначати швидкість, з якою виникають помилки реплікації (Rousseau et al, 2017).

Полімерази, що проявляють 3'-екзонуклеазну активність, можуть виправляти більшість нуклеотидних помилок. Однак ця активність, як правило, відсутня у РНК-полімеразах вірусів. Наразі доведено, що лише РНК-полімерази коронавірусу виконують 3'-екзонуклеазну коректуру. Відсутність коректури є основним фактором, що відповідає за вищі темпи спонтанних мутацій РНК-вірусів порівняно з ДНК-вірусами. Оскільки ЗТ також не мають 3'-екзонуклеазної активності, ретровіруси та параретровіруси мутують зі швидкістю, подібною до швидкості мутації більшості РНК-вірусів. Хоча реплікази ДНК-вірусів мають 3'-екзонуклеазну активність, певні амінокислотні заміни в їхніх полімеразах інактивують коректуру і призводять до появи сильного мутантного фенотипу, як це показано, наприклад, у бактеріофага Т4.

Цікаво, що в полімеразі T4 також існують амінокислотні заміни, які підвищують точність коректури до ста разів порівняно з диким типом. Нарешті, поліомавіруси, які використовують полімерази хазяїна для реплікації, можуть кодувати білки, що інактивують 3'-екзонуклеазний коректуючий домен полімерази хазяїна. Таким чином, уникнення коректури також може бути механізмом, за допомогою якого деякі ДНК-віруси збільшують різноманітність своєї популяції.

Роль постреплікативної репарації також є важливою у виникненні мутацій адже це високоефективна система усунення помилок реплікації та виправлення пошкоджень ДНК. Відсутність цього процесу може значно підвищити частоту мутацій. РНК-віруси не є субстратами клітинних систем репарації, тоді як деякі ДНК-віруси розвинули специфічні особливості, що дозволяють їм уникати репарації. Це стосується бактеріофага FX174 і, можливо, інших невеликих ДНК-фагів. У *Escherichia coli* специфічна репарація розриву ланцюга використовує статус метилювання мотивів послідовності GATC для розрізнення батьківського та дочірнього ланцюгів ДНК. Цікаво, що геном FX174 позбавлений мотивів GATC. Це запобігає метилюванню, а отже, постреплікативній репарації, що дозволяє фагу збільшувати швидкість мутацій порівняно з хазяїном.

Еукаріотичні віруси також можуть уникати репарації або порушувати регуляцію шляхів репарації для збільшення швидкості мутацій, хоча це не було детально досліджено. Відомо, що еукаріотичні віруси взаємодіють зі шляхами реагування на пошкодження ДНК. Наприклад, деякі невеликі ДНК-віруси активують ці шляхи для подовження фази де S клітинного циклу і сприяють їхній реплікації. Шляхи реагування на пошкодження ДНК включають схильні до помилок ДНК-полімерази для повторного синтезу вирізаних ланцюгів, що може сприяти підвищенню частоти вірусних мутацій. З іншого боку, деякі великі

ДНК-віруси кодують власну систему репарації ДНК. Наприклад, вірус африканської чуми свиней кодує ендонуклеазу, репаративну полімеразу (pol X) і АТФ-залежну ДНК-лігази. Pol X дуже схильна до помилок через низьку селективність до основ і відсутність 3'-екзонуклеазної активності. Використання pol X може сприяти відносно високому генетичному різноманіттю ізолятів вірусу африканської чуми свиней.

1.3. Основні популяційно-генетичні процеси, що формують вірусне різноманіття

Доля нових генетичних варіантів значною мірою визначається відбором. Популяційна частота нового мутанта матиме тенденцію до збільшення, випадкових коливань або зменшення залежно від того, чи має він сприятливий, нейтральний або згубний вплив на вірусну пристосованість, відповідно. Проте пристосованість певного алеля залежить від контексту, оскільки вона змінюється залежно від взаємодії між генотипом і фенотипом та генотипом і навколишнім середовищем. Значною мірою середовище існування вірусу диктується хазяїном, але воно також може модулюватися іншими членами вірусної популяції через конкуренцію за клітинні ресурси або кооперативну взаємодію (Crow, 1986, Johnson, 1999).

Частоти алелів також можуть змінюватися внаслідок випадкового генетичного дрейфу. Дрейф визначається ефективним розміром вірусної популяції, який, у свою чергу, залежить від просторової структури, "ефект пляшкового горлечка" у популяції під час передачі вірусу від хазяїна до хазяїна, динаміки вірусного росту тощо. Таким чином, як відбір, так і дрейф зумовлені численними демографічними та екологічними факторами,

що призводить до різноманітних популяційно-генетичних процесів (Рис. 2) (Sung et al., 2012).

Оскільки природні популяції, як правило, пристосовані до навколишнього середовища, більшість нових мутацій, що мають фітнес-ефект, є шкідливими, а отже, відбір прагне вилучити їх з популяції. Коли приплив генетичного різноманіття, спричинений мутаціями, дорівнює відтоку внаслідок відбору, популяція досягає рівноваги, яка називається балансом мутацій і відбору (Drake et al, 1998, Lewontin et al, 1974, Ohta, 1973).

У гаплоїдних популяціях популяційна частота певного шкідливого алеля при балансі мутацій і відбору просто дорівнює μ/s , де μ - швидкість мутацій, а s - коефіцієнт відбору. Коефіцієнт відбору коливається від нуля для нейтральних алелів до одиниці для повністю летальних алелів. Час досягнення цієї рівноваги обернено пропорційно залежить від s , причому рівновага досягається миттєво (за одне покоління).

У РНК-вірусів статистичний розподіл коефіцієнтів відбору нових мутацій було охарактеризовано за допомогою сайт-спрямованого мутагенезу. Це показало, що РНК-віруси демонструють дуже високі значення s (в середньому $s = 0,1$) порівняно з іншими мікроорганізмами. Отже, відбір проти шкідливих мутацій сильно діє на РНК-віруси (Moуа et al., 2004).

Це означає, що рівновага між мутаціями і відбором, як правило, досягається за кілька поколінь, і що в стані рівноваги більшість генетичних варіантів мають низьку популяційну частоту. Підтвердженням цього є те, що високоточне секвенування експериментальних популяцій поліовірусів за допомогою технології Illumina виявило велику кількість шкідливих варіантів з частотами, які неможливо виявити за допомогою звичайного секвенування наступного покоління. Ймовірним поясненням високих значень s -value у РНК-вірусів є те, що їхні геноми надзвичайно

компактні, з невеликою кількістю багатofункціональних генів, що часто перекриваються, багато з яких є життєво важливими. Ці особливості геному також стосуються невеликих ДНК-вірусів, а отже, середні значення s також дуже високі для ДНК-вірусів, таких як мікрівіруси (наприклад, фаг FX174) та іновіруси (наприклад, фаг M13).

Баланс мутацій та відбору досить часто намагаються пояснити теорією вірусних квазівидів. Це насамперед теорія балансу мутацій та відбору за надзвичайно високих темпів мутацій. Квазівиди спочатку вивчалися в рамках фізичної хімії як моделі динаміки самовідтворюваних послідовностей у гіпотетичному примітивному світі РНК.

Однак через високу швидкість мутацій, яку демонструють РНК-віруси, теорія згодом стала популярною серед вірусологів. Незважаючи на використання різної термінології, квазівидові та класичні моделі мутаційного відбору по суті мають справу з одним і тим же процесом і в значній мірі еквівалентні. Проте, квазівидові моделі пропонують більш чіткий підхід до динаміки мутантів, враховуючи швидкість реплікації кожної послідовності, а також ймовірність того, що кожна послідовність мутує в іншу конкретну послідовність. Таким чином визначається матриця переходу, власний вектор якої з найбільшим власним значенням визначає стабільний баланс мутацій та відбору. Цей підхід визначає пристосованість як властивість ансамблю послідовностей, що перебувають у рівновазі (квазівиду).

Як наслідок, пристосованість визначається не лише початковою послідовністю, але й її мутаційним оточенням. Це означає, що квазівид, який містить максимально пристосовану індивідуальну послідовність, може бути витіснений іншим квазівидом, який не містить найпристосованішої послідовності, але демонструє вищу середню пристосованість. Це явище називається виживанням найбільш пристосованих. Пов'язане з цим передбачення моделей квазівидів полягає

в тому, що при перевищенні певної швидкості мутацій найпристосованіша послідовність в межах даного квазівиду може бути витіснена власною хмарою шкідливих мутантів - процес, відомий як "катастрофа помилок". Як відбувається цей процес, залежить від топології ландшафту пристосованості. Початкові квазівидові моделі припускали ландшафт з єдиною послідовністю з високою пристосованістю (еталонною), в якому всі мутанти мали однакове значення пристосованості. У цій конкретній системі за порогом помилки всі мутанти досягають частоти, подібної до частоти еталонної послідовності, і, оскільки існує надзвичайно велика кількість таких можливих мутантів, фактична частота еталонної послідовності стає зникаюче малою (Holmes, 2002).

Теорія катастрофи помилок припускає, що вірусні популяції можуть бути ефективно знищені шляхом достатнього збільшення частоти вірусних мутацій. Обробка експериментальних популяцій РНК-вірусів хімічними мутагенами надала достатню підтримку можливості вимирання, спричиненого мутаціями, в культурах клітин і навіть на тваринних моделях. Однак моделі катастроф помилок розглядають лише придатність певних послідовностей відносно інших, і, таким чином, ігнорують той факт, що вимирання є абсолютним, демографічним процесом, який відбувається, коли базова репродуктивна кількість популяції падає нижче одиниці. Таким чином, ми припускаємо, що вимирання експериментальних популяцій РНК після обробки мутагенами краще визначати як процес летального мутагенезу. Незалежно від використовуваних моделей і механізмів, що лежать в основі вимирання, емпіричні дані переконливо свідчать на користь того, що РНК-віруси природно існують на межі вимирання, зумовленого мутаціями, оскільки вимирання може бути індуковане незначним збільшенням частоти мутацій. На противагу цьому, низька частота спонтанних мутацій у більшості

ДНК-вірусів робить їх менш схильними до летального мутагенезу або помилкової катастрофи.

Висока швидкість мутацій, яку демонструють РНК-віруси, скорочує час очікування, необхідний для появи селективно корисних мутацій, порівняно з іншими мікроорганізмами. Виникненню корисних мутацій також сприяє надзвичайно висока чисельність популяції РНК-вірусів, яка становить порядку мільярдів інфекційних частинок на одного інфікованого хазяїна. При швидкості мутацій 10^{-5} заміни на нуклеотид за раунд копіювання геному, реплікація популяції з 10^9 вірусних частинок повинна призвести до 10^4 мутацій, що наближається до загальної кількості можливих одноточкових мутацій в типовому геномі РНК-вірусу розміром 10 кб (Moуа et al., 2000).

Це робить РНК-віруси надзвичайно чутливими до адаптивного імунітету хазяїна, противірусних препаратів та інших швидкозмінних селективних впливів. Проте, незважаючи на сприятливі умови для появи корисних алелів, висока швидкість мутацій може перешкоджати їх закріпленню в популяції.

По-перше, за високої швидкості мутацій багато геномів несуть кілька шкідливих мутацій, і корисні мутації можуть не відібратися, якщо вони виникають на несприятливому генетичному фоні. З іншого боку, якщо корисний ефект мутації перевищує несприятливий ефект інших мутацій, присутніх у тому ж геномі, шкідливі мутації можуть закріпитися - процес, відомий як "генетичний автостоп" (Irwin et al., 2016, Smith et al., 1974). Ще однією проблемою в системах, що швидко мутують, є клональна інтерференція - процес, при якому в одній популяції співіснує кілька корисних мутацій, а відбирається лише найпристосованіший варіант, що призводить до вимирання інших корисних варіантів. Оскільки нижча швидкість мутацій збільшує час очікування появи корисних мутацій, але вища швидкість мутацій зменшує ефективність відбору корисних мутацій,

адаптація повинна бути максимальною для деякої проміжної, оптимальної швидкості мутацій. Подібний аргумент оптимальності може бути використаний і для нейтральної еволюції.

Аналіз емпіричних даних свідчить, що РНК-віруси реплікуються з майже оптимальною швидкістю мутацій (Lynch, 1999). Нарешті, клональна інтерференція, зустрічний відбір корисних алелів на несприятливому генетичному фоні та автостоп - все це можна усунути за допомогою рекомбінації, оскільки рекомбінація здатна руйнувати несприятливі генетичні комбінації і створювати нові (Desai and Fisher, 2007).

Пристосованість певного вірусу може змінюватися під впливом інших вірусів тієї ж популяції. Такі взаємодії можуть бути кооперативними або конкурентними, і найчастіше відбуваються між вірусами, що інфікують одну клітину, оскільки це дозволяє їм мати спільні білки. Коінфекція клітини незалежними віріонами вимагає високої локальної щільності вірусних частинок. Однак багато вірусів поширені як колективні інфекційні одиниці, що складаються з декількох частинок, що сприяє коінфекції навіть при низькій загальній щільності вірусів. Коли ситуація, в якій кожна клітина отримує кілька вірусних частинок, зберігається протягом кількох циклів інфікування, дефектні віруси мають тенденцію до вторгнення в популяцію. Дефектні віруси, утворені шляхом незаконної рекомбінації, не мають великих ділянок геному і можуть розмножуватися як гіперпаразити лише за рахунок повноцінних вірусних геномів, присутніх у тій самій клітині. Крім того, що дефектні віруси реплікуються швидше, ніж дикий тип, оскільки їхні геноми коротші, вони часто втручаються в деякі етапи інфекційного циклу. Таким чином, пристосованість як дикого типу, так і дефектних вірусів залежить від частоти. Вірусна інтерференція може спричинити різке падіння пристосованості, що дало підстави припустити, що дефектні інтерферуючі

віруси є можливими противірусними препаратами. Появі інтерферуючих вірусів також може сприяти хімічно індукований мутагенез у РНК-вірусів. З іншого боку, коінфекція клітин різними варіантами вірусу може сприяти кооперативній взаємодії вірусів. В принципі, два віруси, кожен з яких несе свій генетичний дефект, можуть взаємно підвищувати свою пристосованість, якщо вони обмінюються генетичними продуктами (генетична комплементация) (André et al., 2006). Також можливо, що кожен варіант оптимізував різні аспекти інфекційного циклу, і що такі варіанти досягають вищої пристосованості при коінфекції, ніж поодиноці. Такий тип кооперативної взаємодії був запропонований для вірусів кору та грипу і може сприяти підтримці вірусної генетичної різноманітності, оскільки пристосованість буде максимальною при проміжних частотах кожного варіанту в популяції. Таким чином, різноманітність може призвести до негативних або позитивних взаємодій між вірусами. У свою чергу, ці взаємодії впливають на різноманітність, змінюючи популяційну частоту певних варіантів (Harris et al., 2018).

Незважаючи на величезні розміри популяцій, віруси зазнають значного генетичного дрейфу. Це пов'язано з тим, що сила дрейфу залежить від ефективної чисельності популяції, а не від загальної чисельності. Демографічні “ефекти пляшкового горлечка” мають сильний і тривалий вплив на ефективну чисельність популяції, і такі події часто виникають під час передачі між хазяїнами майже у всіх вірусів (Miyashita et al., 2009).

Розмір популяції-засновника після передачі вірусу від одного хазяїна до іншого оцінюється за допомогою генетичних маркерів і, як правило, не перевищує 10 копій геному. Однак ця цифра може широко варіювати залежно від шляху передачі (аерозольний, фекально-оральний, статевий, через комах і т.д.). Крім того, вірусні інфекції демонструють розгалужену внутрішньоклітинну просторову структуру, спричинену обмеженням

переміщення вірусу між тканинами або органами (компаратменталізація органів), а також обмеженою дифузією віріонів у позаклітинному середовищі (локалізовані вогнища інфекції). Просторова структура також зменшує ефективну чисельність популяції, що сприяє дрейфу вірусу (Olofsson et al., 2023).

Безпосереднім наслідком дрейфу є втрата генетичного різноманіття, оскільки лише частина популяції робить внесок у наступне покоління. Довгостроковим наслідком є випадкове закріплення алелів у популяції. У дуже мінливих популяціях, таких як популяції РНК-вірусів, і при досить низьких розмірах популяції, дрейф може призвести до втрати послідовності дикого типу. За відсутності зворотної мутації або рекомбінації ця втрата є незворотною, а отже, призводить до накопичення випадкових мутацій в популяції, більшість з яких є нейтральними або шкідливими. Цей процес, відомий як храповик Мюллера, знижує середню пристосованість популяції і може поставити під загрозу виживання популяцій малих РНК-вірусів (Chao, 1990).

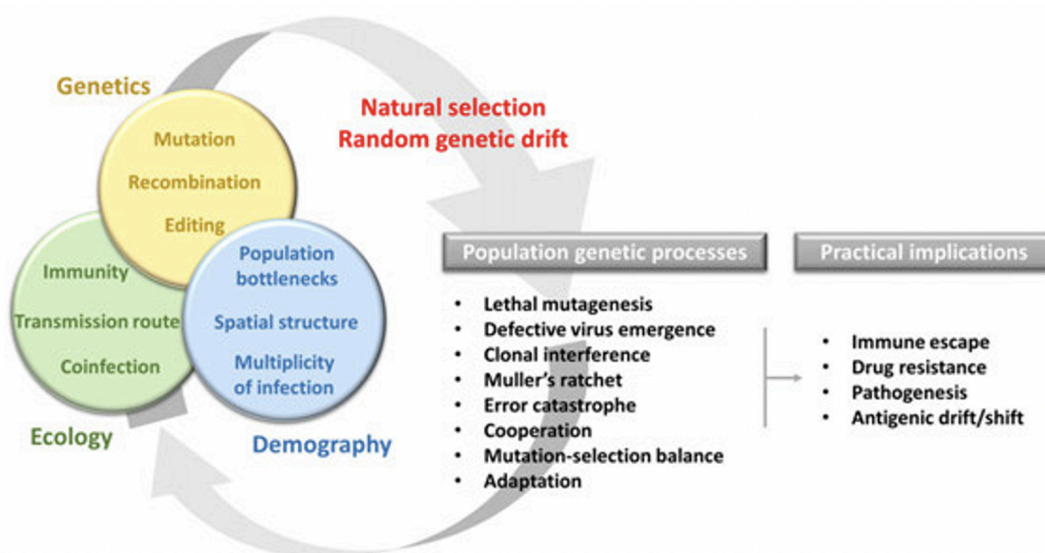


Рис 2. Генетичні, демографічні та екологічні процеси, що визначають різноманіття популяцій вірусів (узято з Rafael Sanjuán, Pilar Domingo-Calap, in Encyclopedia of Virology (Fourth Edition), 2021)

1.4. Математичні наслідки подібних спостережень

Для вірусів з дволанцюговою ДНК (длДНК) і прокаріотів залишається сильна підтримка гіпотези Дрейка, з масштабуванням частоти мутацій/сайтів/поколінь в 1,1 ступені від загального розміру геному, хоча очевидним залишається занепокоєння, що ця закономірність значною мірою залежить від включення геномів бактеріофагів.

Швидкість мутацій для РНК-вірусів більша, ніж для геномів длДНК порівняного розміру, але негативне масштабування зберігається (хоча це повністю залежить від однієї точки даних). Для прокаріотичних геномів, для яких доступні дані, помилка вибірки швидкості мутацій настільки велика, а діапазон варіації розміру геному настільки малий, що можна розрізнити лише слабе масштабування μ за допомогою G . Для більш чіткого вирішення цього питання бажано провести аналіз накопичення мутацій у всьому геномі, як це вже зроблено для деяких еукаріотів, для ряду прокаріотів, особливо тих, що мають екстремальні розміри геномів (Drake, 1991).

Часто стверджують, що швидкість мутацій оптимізується природним відбором для підвищення довготривалої швидкості адаптивних змін (Wilke et al. 2001; Andre and Godelle 2006; Denamur and Matic 2006). Логіка, що лежить в основі цієї точки зору, застосовується переважно до безстатевих популяцій, де корисні мутації залишаються назавжди пов'язаними з середовищем, в якому вони виникають (Johnson 1999a; Sniegowski et al. 2000). Однак спостереження про те, що більшість мутацій є шкідливими,

що зменшує середню пристосованість популяції і веде до небезпечного накопичення летальних мутацій змушує переглянути моделі, що розглядають лише адаптивну вигоду корисних мутацій, ігноруючи усі інші.

Що стосується прямої дії природного добору у вірусних популяціях, то зі зрозумілих причин особливу увагу привертає позитивний добір, пов'язаний зі стійкістю до противірусних препаратів або імунним ухиленням (Irwin et al., 2016a). Цей пошук адаптивних локусів, як правило, базується на виявленні закономірностей селективних прочисток (Maynard Smith and Haigh, 1974) у вірусних геномах завдяки ефекту генетичного автостоупу, спричиненого швидким зростанням частоти корисних мутацій у напрямку до закріплення. Однак, хоча цьому приділяється порівняно менше уваги у вірусологічній літературі, добре відомо, що переважна більшість нових мутацій, які впливають на пристосованість, у будь-якому організмі є шкідливими, а не корисними (Crow, 1993; Lynch et al., 1999; Bank et al., 2014b; Eyre-Walker and Keightley, 2007; Bank et al., 2014a).

Вибіркове видалення цих шкідливих варіантів за допомогою очищувального відбору сприятиме подальшому зменшенню ефективного розміру популяції, що значною мірою диктується швидкістю рекомбінації та силою відбору (Charlesworth et al., 1993; Charlesworth, 2013). Крім того, геномне зчеплення з таким частим надходженням шкідливих мутацій зменшує ймовірність фіксації інших зчеплених ділянок - в тому числі зменшує ймовірність адаптації (Hill and Robertson, 1966; Péron et al., 2017). Важливо, що так само, як збільшення частоти корисних мутацій може призвести до генетичного автостоупу пов'язаних з ними варіацій до закріплення в геномі (що призводить до ефекту селективної зачистки), так само і видалення набагато більшого внеску шкідливих мутацій може призвести до генетичного автостоупу пов'язаних з ними нейтральних варіантів до втрати. Цей ефект - втрата і зниження частоти через зчеплення

зі шкідливими варіантами - відомий як фоновий добір (Charlesworth et al., 1993).

Цей ефект може мати велике значення у вірусних популяціях з огляду на їхні кодуючі геноми (в яких більшість нових мутацій, як очікується, будуть шкідливими), особливо за відсутності рекомбінації. Хоча це ще належить ретельно вивчити у вірусних популяціях, попередні роботи виявили набагато більш домінуючу роль фонового відбору порівняно з селективними перевітками у формуванні геномної варіації у пацієнтів (Renzette et al., 2016). Кількісна оцінка відносного внеску шкідливих, нейтральних і корисних мутацій означає розуміння форми розподілу ефектів пристосованості (DFE).

Таким чином, вивчення ефектів різних режимів мутацій та ефективного розміру популяції, дасть змогу зробити остаточні висновки щодо способу еволюції вірусів, та механізми, що забезпечують встановлення рівноважного мутаційного режиму.

РОЗДІЛ 2. Матеріали та методи досліджень

2.1. Матеріали досліджень

Для запуску симуляцій було обрано розподіл ефективного розміру популяції вірусу згідно експериментальних даних, які зазвичай знаходяться в діапазоні 50-500 вірусних часток у популяції.

Були також обрані експериментальні дані з швидкості мутаційного тиску. Наприклад, оцінка ефективного розміру популяції вірусу грипу всередині хазяїна 32-72, а швидкість мутацій *in vivo* - $3,4 \times 10^{-6}$ на нуклеотид за покоління (Casper K et al., 2020).

2.2. Методи досліджень та постановка моделі

У ході аналізу генетичної та філогенетичної складової популяцій, ми працюємо з скінченними популяціями, що значно ускладнює підрахунки та апроксимування математичних моделей еволюції (які припускають що розмір популяції нескінченний – $N \sim \infty$).

Оцінка еволюції скінченних популяцій додає ключові фактори впливу – генетичний дрейф в результаті численних стохастичних сил. Наслідки генетичного дрейфу - випадкова і непередбачувана зміна частот генів, в результаті чого може відбутися випадкова фіксація того чи іншого алеля. Інші еволюційні сили, які присутні у популяціях (природний відбір, міграції та мутації), як правило, змінюють частоти генів певним чином, або підштовхнути їх до рівноважного значення та утримувати їх там.

Генетичний дрейф є єдиною стохастичною силою в еволюційному процесі. Спрощені еволюційні сили: мутацію та міграцію, можна детально дослідити, маючи лише середні значення та дисперсії частот генів (або альтернативно, враховуючи ймовірності ідентичності за походженням).

Для початку ми використовуємо найпростішу модель – гаплоїдну модель Райта-Фішера для дослідження ефектів появи шкідливих алелів та накопичення і зменшення пристосованості популяції (храповик Мюллера). Для цього ми використовуємо загальну формулу, що описує ймовірність популяції через t кількість поколінь від i копій алеля A досягти j копій. Початкова частота алеля $A - i/N$.

Модель нескінченної кількості алелів це модель яку ми можемо вивести із взаємодії генетичного дрейфу та мутацій в своїй основі має модель Райта-Фішера з фіксованою N кількістю індивідумів.

Модель припускає що усі алелі є еквівалентними для відбору – без ніяких відмінностей у пристосованості – кількість нових алелів є нескінченною, тому що кожна мутація веде до виникнення повністю нового алеля. Тому ми маємо нескінченне число можливих алелів (однаковий алель ніколи не з'являється двічі у різних мутаціях). Нам необхідно знати чи дві різні мутаційні події відбувалися у історії двох копій генів, щоб дізнатися чи це однакові/різні алелі – апроксимація ДНК послідовності, де більшість мутацій виникає у різних сайтах.

Ймовірність F_t , що 2 випадково обрані копії генів без заміщення з одної і тієї ж самої популяції є ідентичними за походженням становить:

$$F_{t+1} = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) F_t, \quad (1)$$

Якщо ми приймаємо, що після виникнення мутанту новий алель більше не вважається ідентичним за походженням до того з якого він утворився. Таким чином до вірогідності додається ще одна умова – жоден із 2 обраних генів не має бути мутантом. Кожна копія у поколінні $t+1$ має однакову і незалежну вірогідність u бути мутантом. Тому наше рівняння має бути помножене на вірогідність того що жодна з 2 копій не є мутантом.

$$F_{t+1} = (1 - u)^2 \left[\frac{1}{2N} + \left(1 - \frac{1}{2N}\right) F_t \right] \quad (2)$$

Ми можемо вирішити це рівняння для певного рівноважного частотного стану (адже значення буде завжди проміжним і ніколи за умови скінченного розміру популяції та ненульовій частоті мутацій не буде 0 або 1).

$$F_{t+1} = F_t = F \quad (3)$$

В результаті ми отримуємо апроксимацію:

$$F \simeq \frac{1}{1 + 4Nu} \quad (4)$$

Апроксимація до рівноважних розподілів

Коли мутації постійно з'являються і змінюють розподіл алелів у популяції, немає визначеного поняття фіксації алеля. Частоти генів продовжують змінюватися невизначену кількість разів у обох напрямках. Якщо мутація відбувається на стадії розмноження у життєвому циклі, і якщо природний відбір відбувається на стадії життєздатності життєвого циклу, до того, як N дорослих буде випадково відібрано для виживання популяції, що залежить від частотно-залежного відбору, тоді ми можемо обчислити ймовірності переходу приблизно так само, як раніше.

Вірогідність переходу у наступне покоління $P(j|i)$, тобто вірогідність що буде j копій алеля A у наступному поколінні, якщо ми маємо i копій серед $2N$ генів у дорослих особин, ми можемо отримати за допомогою:

- 1) Підрахунок частот генів $p = i/(2N)$ дорослих особин у теперішньому поколінні.
- 2) Пропорції A та a серед нескінченно великої кількості гамет, що продукуються цими дорослими особами буде p і $1-p$ відповідно.

- 3) Якщо частка u A буде мутувати до a і частка v a буде мутувати до A , після мутаційного процесу частота гамет A буде $p' = (1-u)p + v(1-p)$
- 4) Якщо частка m гамет буде замінена мігруючими гаметами, частота яких A буде фіксована частотою p_1 , тоді після міграції частота гамет A буде змінена до $p' = (1-m)p + mp_1$
- 5) Підрахунок частот генотипів серед новозапліднених зигот буде відбуватися за умови припущення випадкової комбінації гамет, що продукує пропорції Харді-Вайндберга за нової частоти генів p'
- 6) Використовуємо природний відбір. Домножуємо частоту кожного генотипу на його фітнес і ділимо на середній фітнес популяції. Це дає нам співвідношення генотипів серед тих хто вижив після природного добору. Припустимо що це P, Q, R для AA, Aa, aa .
- 7) Вірогідність, що серед N дорослих зигот, що вижили є k AA , l Aa , і $N-k-l$ aa це триніomialна ймовірність розподілу

$$\frac{N!}{k! l! (N - k - l)!} P^k Q^l R^{N-k-l}$$

- 8) Аби підрахувати ймовірність знайти j кількість A генів серед N особин, треба сумувати усі ймовірності серед усіх комбінацій k і l що мають $2k+l = j$

Наші перехідні вірогідності будуть передбачати один тип життєвого циклу. Як тільки ми маємо ймовірність переходу, рівноважний розподіл частот генів (або числа генів) дається

$$f_j = \sum_{i=0}^{2N} f_i P(j|i), \quad j = 0, 1, \dots, 2N \quad (5)$$

Де f_i це рівноважна ймовірність що тут знаходяться i A генів у популяції.

Якщо ми дамо еволюціонувати популяції достатньо довго поколінь, для того щоб популяція втратила усі ефекти своєї початкової частоти генів, в такому випадку f_i це вірогідність того що його частота генів $i/(2N)$.

Дифузійна апроксимація

Хороша дифузійна апроксимація може бути виведена з рівноважного розподілу.

$$f(p) = \sum_{\Delta p} f(p - \Delta p) P_{p-\Delta p}(\Delta p) \quad (6)$$

$$\phi(p) \delta p \simeq \sum_{\Delta p} \phi(p - \Delta p) P_{p-\Delta p}(\Delta p) \delta p \quad (7)$$

Після апроксимації функції неперервною функцією густини, апроксимуємо кожен з функцій з першими членами ряду Тейлора.

$$\phi(p) \simeq \sum_{\Delta p} \left[\phi(p) - \Delta p \phi'(p) + \frac{(\Delta p)^2}{2} \phi''(p) \right] \left[P_p(\Delta p) - \Delta p P'_p(\Delta p) + \frac{(\Delta p)^2}{2} P''_p(\Delta p) \right], \quad (8)$$

$$M(p) = \mathbb{E}(\Delta p) = \sum_{\Delta p} P_p(\Delta p)$$

$$V(p) = \mathbb{E}((\Delta p)^2) = \sum_{\Delta p} P_p(\Delta p) (\Delta p)^2 \quad (9, 10)$$

Ми використовуємо ці співвідношення, аби замінити змінні і отримуємо FKPP

$$M'(p) = dM(p)/dp = \sum_{\Delta p} P'_p(\Delta p) \Delta p$$

$$V'(p) = dV(p)/dp = \sum_{\Delta p} P'_p(\Delta p) (\Delta p)^2$$

$$V''(p) = d^2V(p)/dp^2 = \sum_{\Delta p} P''_p(\Delta p) (\Delta p)^2 \quad (11)$$

$$0 \simeq -\frac{d}{dp}[M(p)\phi(p)] + \frac{1}{2} \frac{d^2}{dp^2}[V(p)\phi(p)] \quad (12)$$

Рішення цього рівняння

$$C \simeq -M(p)\phi(p) + \frac{1}{2} \frac{d}{dp}[V(p)\phi(p)]. \quad (13)$$

C виявляється нулем, в результаті ми маємо диференційне рівняння 1 порядку.

$$\frac{2M(p)}{V(p)} = \frac{1}{f(p)} \frac{df(p)}{dp} = \frac{d \ln f(p)}{dp}. \quad (14)$$

При інтегруванні

$$\phi(p) = \frac{K}{V(p)} \exp \left[2 \int_c^p M(x)/V(x) dx \right]. \quad (15)$$

Константа K – параметр нормалізації

Якщо ми впроваджуємо параметр мутацій, ми пам'ятаємо що детерміністична зміна генної частоти це

$$p_{t+1} = (1 - u) p_t + v (1 - p_t) \quad (16)$$

$$M(p) = \mathbb{E}(\Delta p) = -u p + v (1 - p) \quad (17)$$

Таким чином підрахунок виконується

$$\frac{2M(x)}{V(x)} = \frac{-2ux + 2v(1 - x)}{x(1 - x)/2N_e} = \frac{-4N_e u}{1 - x} + \frac{4N_e v}{x}, \quad (18)$$

$$\phi(p) = K p^{4N_e v - 1} (1 - p)^{4N_e u - 1} \quad (19) \text{ (Felsenstein, 1984)}$$

Відбір проти дрейфу – загальна формула

Коли ми спостерігаємо дію природного добору який відбувається на 2 алелях і постійних індексах пристосованості - відносно проста формула для рівноважного розподілу може бути виведена.

$$M(p) = \frac{p(1-p)}{2\bar{w}} \frac{d\bar{w}}{dp} + v(1-p) - up. \quad (20)$$

$$V(p) = \frac{p(1-p)}{2N_e}. \quad (21)$$

$$2 \frac{M(x)}{V(x)} = \frac{2N_e}{\bar{w}} \frac{d\bar{w}}{dx} + \frac{4N_e v}{x} - \frac{4N_e u}{1-x}. \quad (22)$$

Ми можемо отримати таким чином рівноважний розподіл

$$\phi(p) = K p^{4N_e v - 1} (1-p)^{4N_e u - 1} \bar{w}^{2N_e}, \quad (23)$$

Симуляція моделі Райта-Фішера з накопиченням шкідливих мутацій у популяції

Ми моделюємо еволюцію розподілу пристосованості безстатевої популяції як хвилю, що рухається у просторі пристосованості. Ми розглядаємо шкідливі мутації, що описуються їхнім коефіцієнтом відбору і припущенням нескінченно великої кількості сайтів мутації.

Ми досліджуємо поведінку хвилі пристосованості і розподілу популяції у кожному поколінні зі зміною коефіцієнту відбору та вплив ефективного розміру популяції.

Початкова модель була мотивована представленням еволюції безстатевої популяції як фронту, що поширюється вздовж осі пристосованості. Її детально вивчали (Brunet et Derrida, 2006, 2007, 2008, 2012, 2013).

Популяція представлена у вигляді набору дійсних чисел, розташованих уздовж осі пристосованості. Кожна особина залишає фіксовану кількість нащадків, які можуть відрізнятися від батьків. Після цього розмір популяції відновлюється, залишаючи N крайніх правих особин у наступному поколінні.

Псевдокод симуляції (Райт-Фішер, N - фіксоване у кожному поколінні)

1. Популяція

Особина репрезентована як число, що визначає кількість корисних мутацій що є у особини. Тому, популяція записана у вигляді (N) numpy array.

2. Мутації

Кожна особина отримує нові мутації (збільшує кожного разу на 1) з вірогідністю μ .

3. Відбір

Наступне покоління формується як мультиноміальна вибірка з особин з заміною з попереднього покоління з вірогідностями пропорційними пристосованостями цих особин: де k і j кількість мутацій, що є у особин – розмір популяції постійний.

4. Родовід

Кожна особина має список предків: індекси особин, що обрані як батьки з попередніх поколінь. Список оновлюється кожне покоління. Тобто, родовід зберігається у вигляді $(N * \text{кількість поколінь})$ -array. Кожна особина має список нащадків і додатковий список предків: індекси особин, що обрані як батьки з попередніх поколінь. Список оновлюється кожне покоління. Потім, базуючись на цьому списку підраховуються генеалогії: 3 ситуації спадковості – кожна обраховується незалежно і послідовно. В результаті ми отримуємо попарні відстані у часі між особинами, у який інформація про родовід зберігається у $(N-1)$ -array.

Наступним кроком, підраховуються середні $\langle T2 \rangle$, $\langle T3 \rangle$, $\langle T4 \rangle$. Для $\langle T2 \rangle$ це всі можливі максимальні часи парних комбінацій, для $\langle T3 \rangle$ всі можливі максимальні трійні комбінації. Час до найбільш нещодавнього спільного предка двох особин це максимальний індекс пари у списку часів цих відстаней (Brunet, Derrida et al., 2006–2012).

РОЗДІЛ 3. Результати досліджень та обговорення

3.1. Оцінка моделі Райт-Фішера для використання опису еволюційної динаміки вірусних популяцій з появою і накопиченням шкідливих мутацій

Модель коалесценції (злиття нащадків у предкові лінії походження алелів) є результатом ретельного аналізу моделі Райта-Фішера. Вона аналізує частоту і розподіл подій коалесценції в історії еволюції. Ці події наближаються до експоненціального розподілу з найвищою швидкістю в недавньому минулому. Кількість ліній, що виживають в одному поколінні, також збільшують свою частоту квадратично. Оскільки кількість вцілілих ліній завжди зменшується в міру віддалення в минуле, це фактично посилює ефект експоненціального розподілу. У випадку моделі Райт-Фішера є досить сильний недолік - ця модель, однак, є неповною, оскільки вона припускає постійну чисельність популяції, що часто не відповідає дійсності. Подальший аналіз показує, що ймовірність події коалесценції обернено пропорційна до величини популяції.

З таким розширенням коалесцентна модель все ще має обмеження, але можна усунути більшість з них зі схожими змінами. Це дозволяє врахувати такі речі як вплив природного відбору.

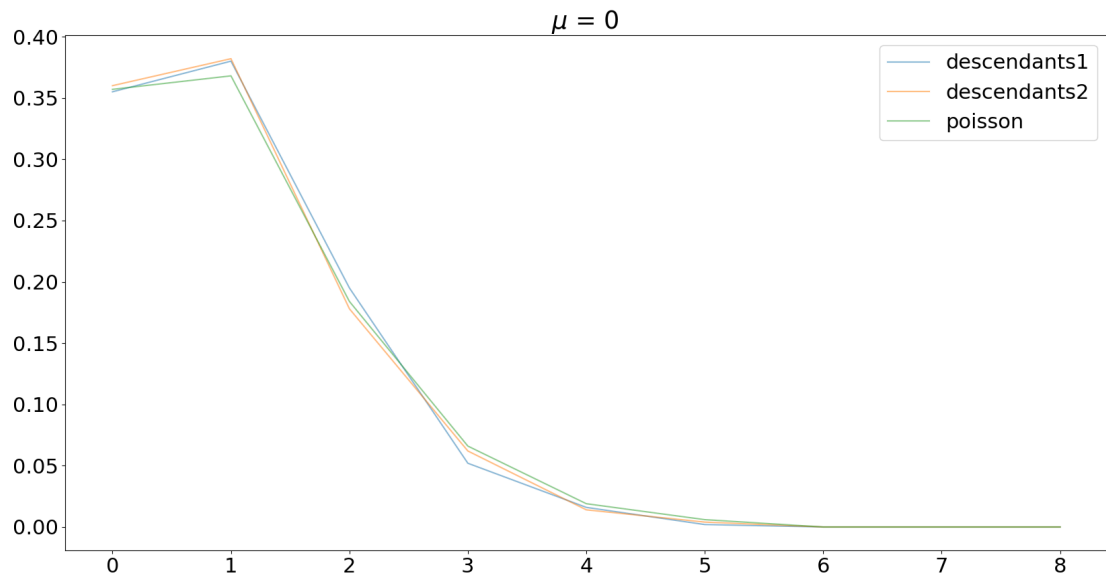
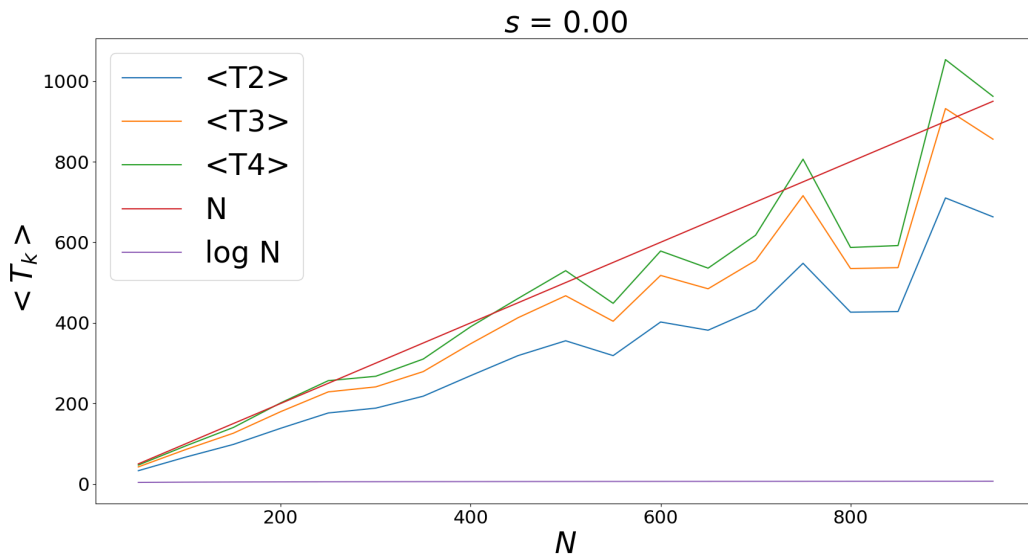


Рис 1. Розподіл нащадків при моделі Райт-Фішера

Розподіл кількості нащадків буде різним для різних видів, але простий (хоча і довільний) вибір полягає в тому, що кількість нащадків, які вижили на предка, підпорядковується пуассонівському розподілу (рис. 1). Розподіл Пуассона має лише один параметр, μ , який визначає як середнє значення, так і дисперсію.

А. коефіцієнт відбору $s = 0$, відсутність відбору



В. коефіцієнт відбору $s = 0.1$, помірний відбір

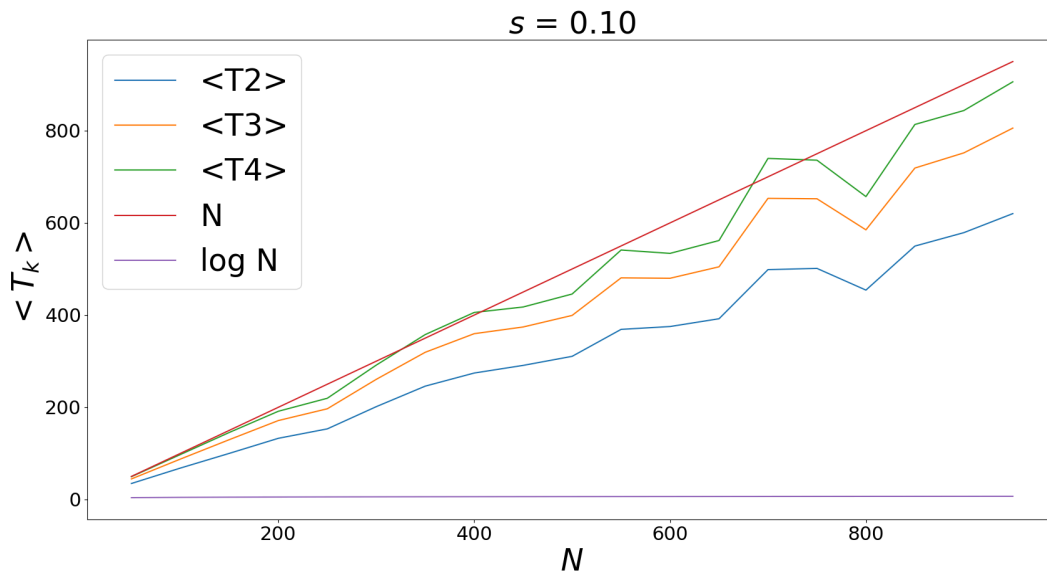


Рис 2. Значення різних $\langle T_k \rangle$ (часів коалесценції) залежно від числа особин у популяції з різними коефіцієнтами відбору

У нестатевих популяціях всі локуси мають однакову генеалогічну історію, і доля лінії залежить від пристосованості всього геному. Якщо пристосованість залежить від великої кількості поліморфних локусів з порівнянними ефектами, то розподіл пристосованості в популяції буде приблизно гауссівським. Древа коалесценції прямо пропорційно залежать від розміру популяції (Рис.2).

У малих популяціях з $N\sigma \ll 1$ (σ -дисперсія пристосованості) домінують нейтральні процеси коалесценції (неспадкові коливання чисельності нащадків, відомі як генетичний дрейф).

Без відбору ($s = 0$) (Рис 2А): Час коалесценції зростає з розміром популяції, оскільки генетичний дрейф є єдиним чинником, що впливає на злиття ліній. У великій популяції цей процес є повільним.

З відбором ($s = 0.1$) (Рис 2В): Відбір зменшує час коалесценції незалежно від розміру популяції, але ефект відбору стає більш вираженим у великій популяції, де він значно пришвидшує коалесценцію, зменшуючи генетичну дивергенцію швидше, ніж це відбувалося б лише під впливом генетичного дрейфу.

Отже, час коалесценції є важливим індикатором генетичної структури популяції, і його залежність від розміру популяції та коефіцієнта відбору допомагає зрозуміти динаміку генетичної різноманітності в еволюційному процесі.

3.2. Оцінка впливу ефективного розміру популяції

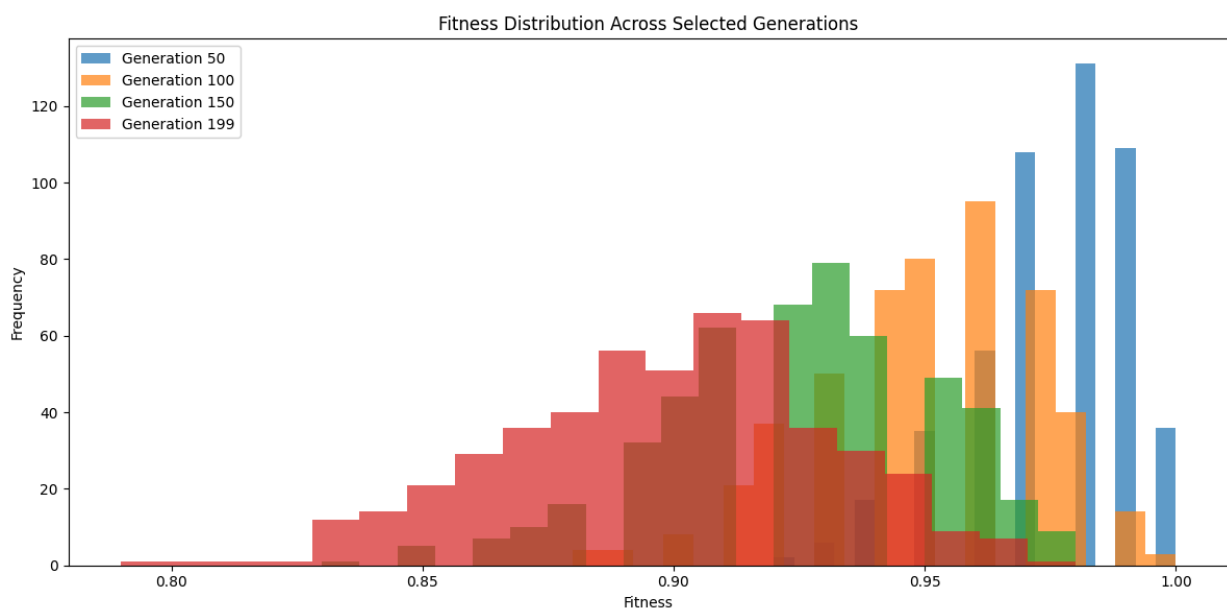
Природною відправною точкою будь-якого еволюційного аналізу є розгляд ефектів генетичного дрейфу. Ця стохастична еволюційна сила зазвичай описується в термінах ефективного розміру популяції (N_e) - тобто ідеалізованого розміру популяції, яка зазнає спостережуваного рівня генетичного дрейфу. Він протиставляється загальній чисельності популяції (N), тобто поточній кількості особин у популяції. Існує багато причин, чому N_e може сильно зменшитися порівняно з N , включаючи зміни чисельності популяції з плином часу, а також сам природний відбір, як буде описано далі. Хоча відомо, що чисельність вірусних популяцій часто є надзвичайно великою, важливо розуміти, що чисельність ефективних вірусних популяцій є набагато меншою через основні шляхи передачі та динаміку інфекції, як описано нижче, при цьому N_e оцінюється лише в межах від сотень до тисяч (Hughes, 2009; Miyashita and Kishino, 2010; Renzette et al., 2013).

Це важливо для генетики вірусних популяцій, оскільки вказує на верхню межу ефективності природного добору, враховуючи, що ця ефективність диктується N_e . Це також слугує нагадуванням про надзвичайно важливу роль генетичного дрейфу у визначенні результатів еволюції та про невиправданий, як правило, адаптаційно-центричний погляд на еволюцію вірусних геномів, який все ще поширений у певних частинах вірусологічної спільноти.

Таким чином, перш ніж розглядати тему відбору, необхідно спочатку розглянути нейтральну популяційну динаміку. У випадку вірусної популяції, відібраної від пацієнта, ця динаміка, ймовірно, включає сильний “ефект пляшкового горлечка” в популяції (тобто тимчасове зменшення розміру популяції), пов'язане з первинною інфекцією - ці зменшення можуть бути досить серйозними, причому деякі первинні інфекції спричиняються дуже малою кількістю одиниць вірусу, що інфікують.

Після такого первинного інфікування вірусна популяція часто вступає у фазу швидкого зростання чисельності, проте платить за це ціну зменшеного генетичного різноманіття.

A. Розмір популяції = 50



В. Розмір популяції = 500

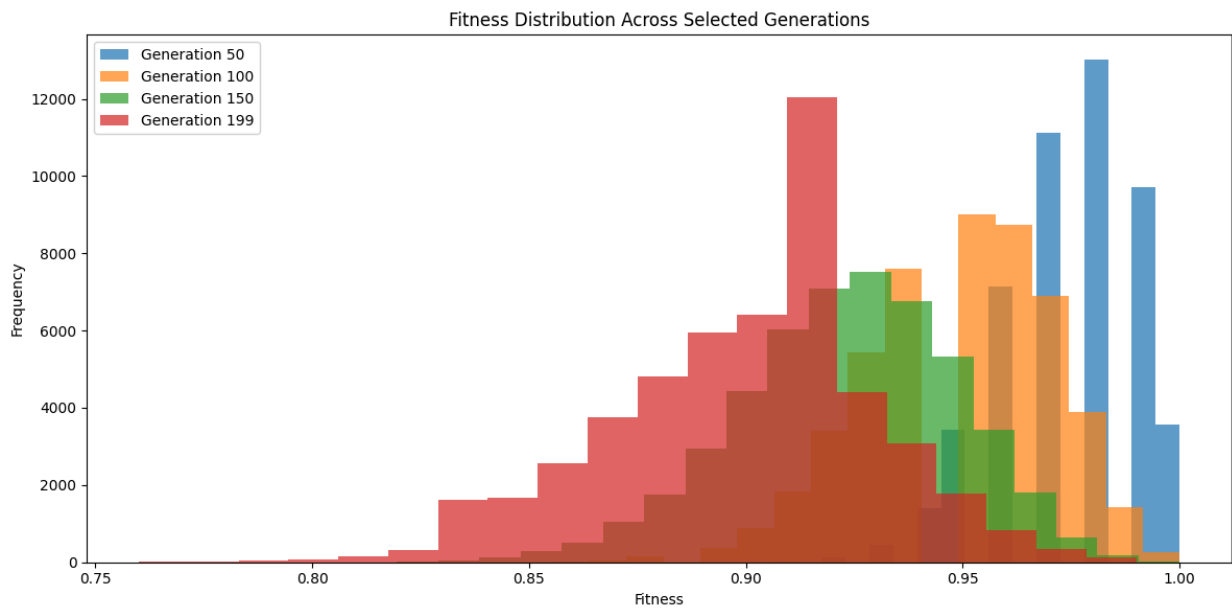


Рис.3. Розподіл пристосованості популяції через обрані покоління (різні кольори означають вибірку популяції в межах 1 покоління)

У випадку використання експериментально обчисленої швидкості мутацій для грипу можна побачити як пристосованість популяцій падає з збільшенням поколінь, що пройшли. При цьому розподіл стає більш плоским з збільшенням поколінь за маленького розміру популяції, тобто спостерігається більша варіабельність та гетерогенність популяції. Маленькі популяції особливо вразливі до ефектів генетичного дрейфу, який може призвести до випадкових змін у частоті алелів, не пов'язаних з їх адаптивною цінністю. З часом це спричиняє зниження середньої пристосованості, оскільки навіть нейтральні або шкідливі мутації можуть фіксуватися у популяції. Як наслідок, генетичний ландшафт популяції стає більш неоднорідним, що відображається у більшій дисперсії значень пристосованості серед різних індивідумів (Рис 3).

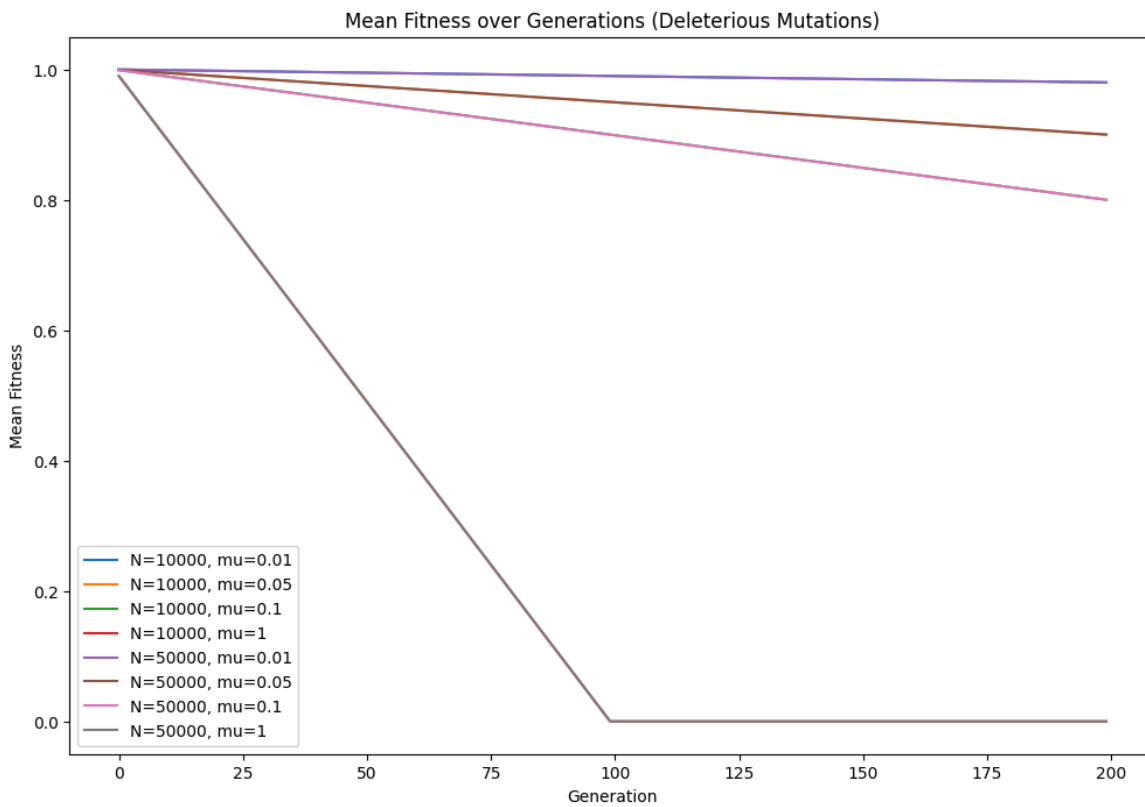
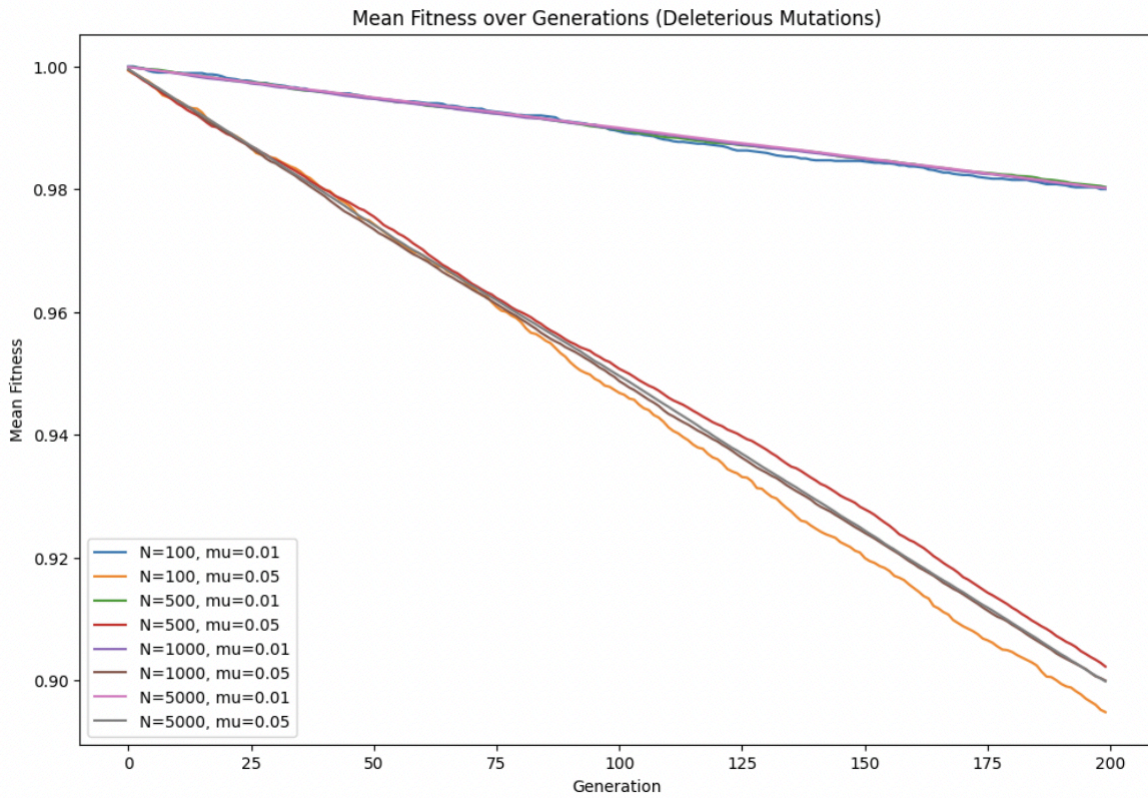


Рис.4. Розподіл пристосованості з часом поколінь за умови різних мутаційних режимів та різних ефективних розмірів популяцій

Шкідливі мутації виникають з досить високою швидкістю у більшості організмів, можливо, навіть по одній на гамету (Crow and Simmons 1983; Bell 1988a; Кондрашов 1988; Charlesworth et al. 1990; Houle et al. 1992). Багато зусиль було витрачено на те, щоб визначити, чи можуть популяції витримати цей постійний натиск мутацій без суттєвої втрати індивідуальної пристосованості (Haldane 1937; Kimura et al. 1963; Crow and Kimura 1979; Кондрашов 1988; Charlesworth 1990). За таких умов, популяції, які несуть більше мутаційне навантаження, повинні також мати вищий ризик вимирання. Мюллер (Muller, 1964) вказував на особливу гостроту цієї проблеми для невеликих нестатевих популяцій. Якщо припустити, що зворотні мутації рідкісні, то за відсутності рекомбінації жодна особина ніколи не зможе дати потомство з меншою кількістю шкідливих мутацій, ніж вона сама. Завжди існує ймовірність того, що випадково клас особин з найнижчою пристосованістю не дасть нащадків у якомусь поколінні. Після того, як цей клас особин буде втрачено, очікується, що другий за пристосованістю клас зрештою зазнає тієї ж долі, і так далі. Фельзенштейн (Felsenstein, 1974) назвав це явище храповиком Мюллера.

Можемо спостерігати, що, зменшуючи середню абсолютну пристосованість особин, накопичення мутацій врешті-решт призведе до зменшення чисельності популяції і згодом вимирання популяції (Рис.4). Але в меншій популяції випадковість відіграє більшу роль, тому подальші шкідливі мутації накопичуються з дещо більшою ймовірністю, спричиняючи дещо швидше зниження середньої пристосованості та чисельності популяції. Таким чином, очікується, що зниження пристосованості відбуватиметься з прискоренням - явище, яке ми називаємо мутаційним таненням. За умови порушення балансу між

відбором та мутаційним режимом, перевищення впливу навантаження мутацій, $\mu=1$, популяція вимирає вже через 100 поколінь.

Зворотні мутації не можуть зупинити храповик Мюллера, оскільки очікується, що швидкість зворотних мутацій менша, ніж швидкість прямих; хоча будь-яка мутація може змінити послідовність гена (і, можливо, зашкодити йому), лише зворотна мутація може відновити початкову послідовність. Шкідливі мутації, однак, можуть також компенсуватися мутаціями, які не змінюють ген до його початкової послідовності. Якщо такі компенсаторні мутації є досить поширеними, храповик Мюллера можна зупинити, цей випадок можна розглянути у подальших дослідженнях, але відносна частота зворотних, компенсаторних і корисних мутацій залишається емпіричним питанням. У нещодавніх експериментах такі компенсуючі мутації не зупиняли храповик Мюллера, тому було би важливо дослідити це питання.

Так як РНК-віруси мають високу швидкість спонтанних мутацій (частота мутацій ДНК становить від 10^{-9} до 10^{-10} помилок на реплікацію нуклеотидів, частота помилок для РНК становить від 10^{-4} до 10^{-5}), то проблема накопичення мутацій є дуже актуальною, відповідно постає питання вірогідності фіксації шкідливих мутацій у популяції. Це відбувається тому, що кожен цикл реплікації супроводжується значною кількістю помилок, більшість з яких є нейтральними або шкідливими.

Іншим важливим фактором є взаємодія між мутаціями. У випадках, коли шкідливі мутації з'являються у генах, які мають значну функціональну важливість, вони можуть впливати на здатність вірусу до інфекції та реплікації. Висока швидкість мутацій у РНК-вірусів означає, що навіть корисні мутації можуть швидко бути "перекриті" шкідливими, що значно ускладнює адаптацію популяції.

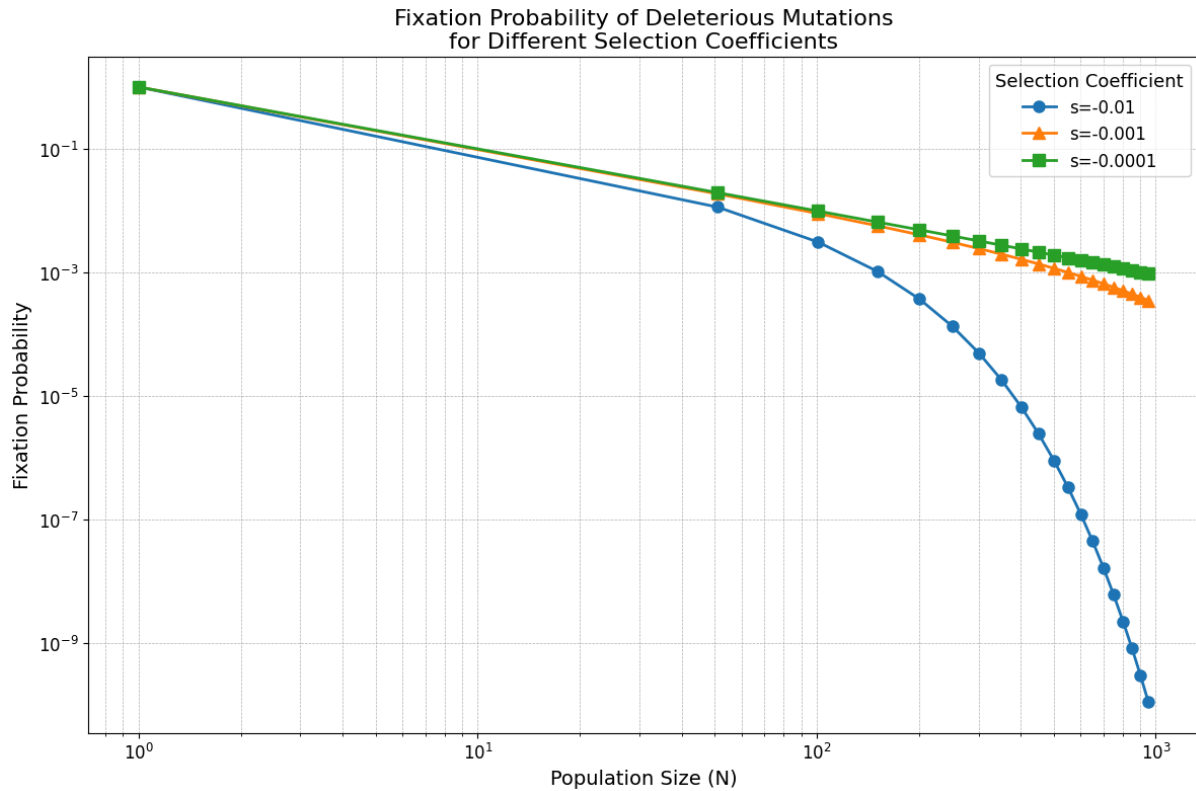


Рис.5. Оцінка вірогідності фіксації шкідливих мутацій з різними коефіцієнтами селекції $s=-0.01/0.001/0.0001$

У даному аналізі була прийнята теорія Кімури про врахування внеску майже нейтральних мутацій в еволюційний процес (Ohta 1973, 1992; Kimura 1983), що полягає в прийнятті спектру майже нейтральних мутацій - від злегка шкідливих до злегка корисних, причому нейтральність тієї чи іншої зміни залежить від розміру популяції; еволюційні траєкторії складаються з балансу між злегка шкідливими і злегка корисними заміщеннями. Тому фіксації шкідливих мутацій будуть частими, оскільки більша ймовірність закріплення компенсує їх відносну шкоду для пристосованості (Gillespie 1994).

В рамках нейтральної моделі розподіл s серед нових мутацій може бути бімодальним, з режимами, зосередженими навколо дуже шкідливих і нейтральних мутацій. Під час адаптивних епізодів розподіл матиме три режими, з невеликим додатковим режимом, зосередженим навколо

сприятливих мутацій. Оскільки сильно шкідливі мутації мають зникаюче малу ймовірність закріплення в популяції, більшість замін (тобто закріплених мутацій) будуть слабо шкідливими або нейтральними. У цьому випадку розподіл s між заміщеннями буде унімодальним і зосередженим навколо слабо-шкідливих та нейтральних мутацій. Під час адаптивного епізоду природний добір швидко призведе до закріплення багатьох позитивно відібраних мутацій. У цьому випадку розподіл замін буде бімодальним, з режимами, зосередженими навколо слабо шкідливих і вигідних замін.

Отже, зафіксуватися і збільшитися у частоті можуть ще і алелі, які є слабо шкідливими, тому не несуть прямого зменшення у пристосованості, проте кумулятивний ефект багатьох таких мутацій може призвести до підвищеного мутаційного тиску (Рис. 5). Ми можемо спостерігати що за умови маленького ефективного розміру популяції, що є випадком для вірусів, вірогідність фіксації слабо шкідливих мутацій є доволі високою.

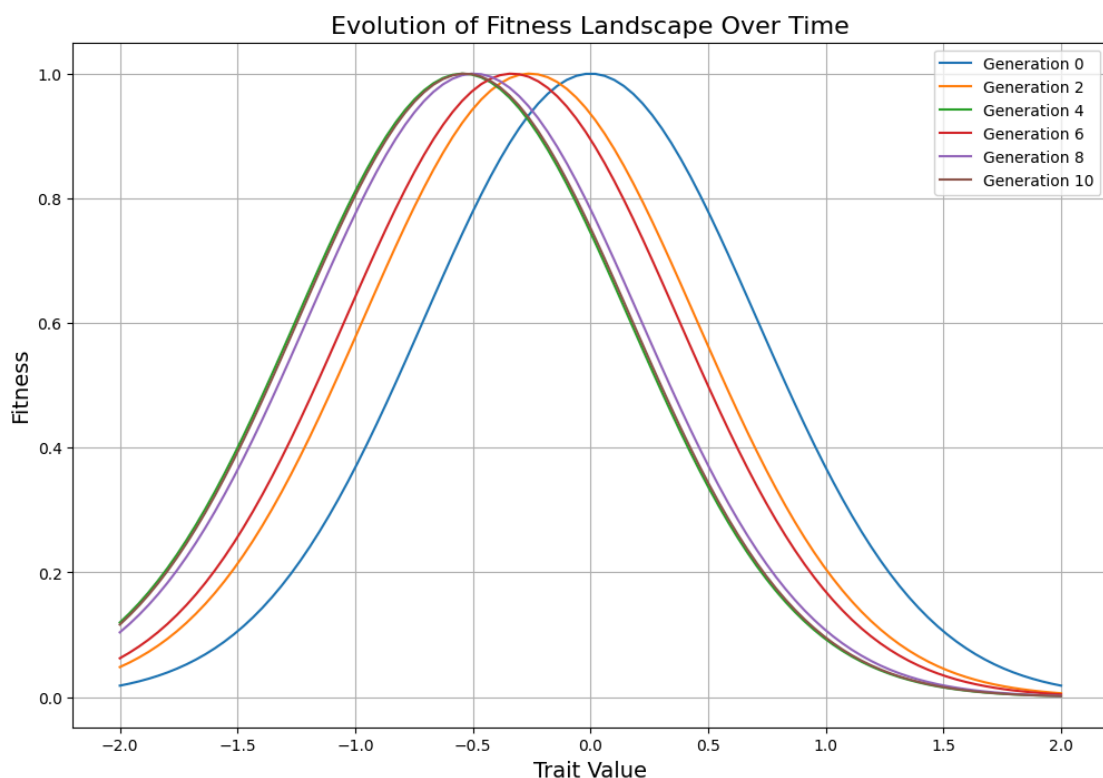


Рис 6. Еволюція ландшафту пристосованості популяції у просторі або репрезентація у форматі fitness wave

Швидкість адаптації популяції, як очікується, буде залежати від швидкості мутацій і від розподілу фітнес-ефектів (DFE) корисних мутацій, що виникають, оскільки наявність навіть декількох мутацій з великим ефектом може мати великий вплив на те, як швидко популяція рухається до свого фітнес-оптимуму.

З іншого боку, DFE шкідливих мутацій, зокрема частка слабо шкідливих мутацій, визначає очікуване дрейфове навантаження популяції - зниження пристосованості внаслідок численних шкідливих мутацій з малим ефектом, які окремо достатньо близькі до нейтральних, щоб іноді уникати відбору, але разом можуть мати значний вплив на пристосованість (Рис.6).

DFE нових мутацій впливає на багато еволюційних закономірностей, таких як очікуваний ступінь паралельної еволюції, еволюційний потенціал і здатність популяції реагувати на нові умови та підтримання варіації кількісних ознак. Таким чином, розуміння DFE мутацій є ключовою частиною нашого розуміння процесу еволюції.

ВИСНОВКИ

- 1) Показано, що більшість популяційно генетичних висновків ґрунтується на моделі Райта-Фішера, де існує припущення про пуассонівський розподіл нащадків, що є вагомим недоліком, що необхідно враховувати при вивченні вірусних популяцій, які є дуже варіабельними в кількості нащадків.
- 2) При аналізі генеалогій вірусних популяцій було порівняно відсутність відбору та переважання дрейфу з присутністю селекції, що демонструє вплив відбору на зменшення відстані між останнім спільним предком вірусних частинок у популяції.
- 3) Враховуючи високі ефекти генетичного дрейфу у вірусних популяціях через особливості життєвого циклу та як наслідок низький N_e , умова збереження популяцій швидко мутуючих вірусних геномів це підтримання балансу між коефіцієнтом селекції та швидкості мутацій. Було визначено, що якщо цей ліміт $N_e\mu \approx 1$ перевищується, це призводить до вимирання популяції та може бути потенційним таргетом для розробки нових терапевтичних препаратів.
- 4) Зафіксуватися і збільшитися у частоті можуть алелі, які є слабко шкідливими, тому не несуть прямого зменшення у пристосованості, проте сумарний ефект багатьох таких мутацій може призвести до підвищеного мутаційного тиску.
- 5) Адаптація вірусів до нових середовищ та виникнення нових штамів не є єдиною умовою для спостережуваної динаміки вірусів. Більшість моделей беруть до уваги лише позитивні мутації і їх швидко фіксацію, таким чином переоцінюючи можливість вірусних популяцій підтримувати певну швидкість мутацій, що і перевищує реальні цифри.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. André, J.B. and Godelle, B., 2006. Within-host evolution and virulence in microparasites. *Journal of Theoretical Biology*, 241(2), pp.402-409. doi: 10.1016/j.jtbi.2005.12.011. PMID: 16442566.
2. Bank, C., Ewing, G.B., Ferrer-Admetlla, A., Foll, M. and Jensen, J.D., 2014. Thinking too positive? Revisiting current methods of population genetic selection inference. *Trends in Genetics*, 30(12), pp.540-546. doi: 10.1016/j.tig.2014.09.010. PMID: 25438719.
3. Brunet, É. and Derrida, B., 2012. How genealogies are affected by the speed of evolution. *Philosophical Magazine*, 92(1-3), pp.255-271.
4. Brunet, É. and Derrida, B., 2013. Genealogies in simple models of evolution. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(1), p.P01006.
5. Brunet, É., Derrida, B., Mueller, A.H. and Munier, S., 2006. Noisy traveling waves: effect of selection on genealogies. *Europhysics Letters*, 76(1), pp.1-7.
6. Brunet, É., Derrida, B., Mueller, A.H. and Munier, S., 2007. Effect of selection on ancestry: An exactly soluble case and its phenomenological generalization. *Physical Review E*, 76(4), p.041104.
7. Brunet, É., Rouzine, I.M. and Wilke, C.O., 2008. The stochastic edge in adaptive evolution. *Genetics*, 179(1), pp.603–620.
8. Charlesworth, B., 2013. Stabilizing selection, purifying selection, and mutational bias in finite populations. *Genetics*, 194(4), pp.955-71. doi: 10.1534/genetics.113.151555. PMID: 23709636; PMCID: PMC3730922.
9. Charlesworth, B., Morgan, M.T. and Charlesworth, D., 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics*, 134(4), pp.1289-1303. doi: 10.1093/genetics/134.4.1289. PMID: 8375663; PMCID: PMC1205596.

10. Chao, L., 1990. Fitness of RNA virus decreased by Muller's ratchet. *Nature*, 348(6300), pp.454-455. doi: 10.1038/348454a0. PMID: 2247152.
11. Crow, J.F., 1993. How much do we know about spontaneous human mutation rates? *Environmental and Molecular Mutagenesis*, 21(2), pp.122-129.
12. Crow, J.F., 1986. Basic concepts in population, quantitative, and evolutionary genetics. New York: W.H. Freeman.
13. Desai, M.M. and Fisher, D.S., 2007. Beneficial mutation-selection balance and the effect of linkage on positive selection. *Genetics*, 176(3), pp.1759–1798. doi: 10.1534/genetics.106.067678.
14. Denamur, E. and Matic, I., 2006. Evolution of mutation rates in bacteria. *Molecular Microbiology*, 60(4), pp.820-827. doi: 10.1111/j.1365-2958.2006.05150.x. PMID: 16677295.
15. Drake, J.W., 1991. A constant rate of spontaneous mutation in DNA based microbes. *Proceedings of the National Academy of Sciences*, 88(16), pp.7160–7164. doi: 10.1073/pnas.88.16.7160.
16. Drake, J.W., Charlesworth, B., Charlesworth, D. and Crow, J.F., 1998. Rates of spontaneous mutation. *Genetics*, 148(4), pp.1667–1686.
17. Eyre-Walker, A. and Keightley, P.D., 2007. The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, 8(8), pp.610-618. doi: 10.1038/nrg2146. PMID: 17637733.
18. Gabriel, W., Lynch, M. and Bürger, R., 1993. Muller's ratchet and mutational meltdowns. *Evolution*, 47(6), pp.1744-1757. doi: 10.1111/j.1558-5646.1993.tb01266.x. PMID: 28567994.
19. Harris, R.B., Sackman, A. and Jensen, J.D., 2018. On the unfounded enthusiasm for soft selective sweeps II: examining recent evidence from humans, flies, and viruses. *PLOS Genetics*, 14, p.e1007859.
20. Hardy, G.H., 1908. Mendelian proportions in a mixed population. *Yale Journal of Biology and Medicine*, 76(2), pp.79–80.

21. Hill, W.G. and Robertson, A., 1966. The effect of linkage on limits to artificial selection. *Genetical Research*, 8(3), pp.269-294. PMID: 5980116.
22. Irwin, K.K., Laurent, S., Matuszewski, S., Vuilleumier, S., Ormond, L., Shim, H., Bank, C. and Jensen, J.D., 2016. On the importance of skewed offspring distributions and background selection in virus population genetics. *Heredity*, 117(6), pp.393-399. doi: 10.1038/hdy.2016.58. PMID: 27649621; PMCID: PMC5117839.
23. Johnson, T., 1999. Beneficial mutations, hitchhiking and the evolution of mutation rates in sexual populations. *Genetics*, 151(4), pp.1621-1631. doi: 10.1093/genetics/151.4.1621. PMID: 10101182; PMCID: PMC1460574.
24. Kimura, M., 1983. *The neutral theory of molecular evolution*. New York: Cambridge University Press.
25. La Tourrette, K. and Garcia-Ruiz, H., 2022. Determinants of Virus Variation, Evolution, and Host Adaptation. *Pathogens*, 11(9), Article 1039. doi: 10.3390/pathogens11091039.
26. Lewontin, R.C., 1974. *The genetic basis of evolutionary change*. New York: Columbia University Press.
27. Lynch, M., Blanchard, J., Houle, D., Kibota, T., Schultz, S., Vassilieva, L. and Willis, J., 1999. Perspective: Spontaneous deleterious mutation. *Evolution*, 53(3), pp.645-663. doi: 10.1111/j.1558-5646.1999.tb05361.x. PMID: 28565627.
28. Miyashita, S. and Kishino, H., 2010. Estimation of the size of genetic bottlenecks in cell-to-cell movement of soil-borne wheat mosaic virus and the possible role of the bottlenecks in speeding up selection of variations in trans-acting genes or elements. *Journal of Virology*, 84(4), pp.1828-1837. doi: 10.1128/JVI.01890-09. PMID: 19955302; PMCID: PMC2812364.

29. Moya, A., Holmes, E.C. and González-Candelas, F., 2004. The population genetics and evolutionary epidemiology of RNA viruses. *Nature Reviews Microbiology*, 2(4), pp.279-288. doi: 10.1038/nrmicro863. PMID: 15031727; PMCID: PMC7096949.
30. Ohta, T., 1973. Slightly deleterious mutant substitutions in evolution. *Nature*, 246(5428), pp.96–98. doi: 10.1038/246096a0.
31. Olofsson, P., Chipkin, L., Daileda, R.C. and Azevedo, R.B.R., 2023. Mutational meltdown in asexual populations doomed to extinction. *Journal of Mathematical Biology*, 87(6), p.88. doi: 10.1007/s00285-023-02019-y. PMID: 37994999.
32. Penisson, S., Singh, T., Sniegowski, P. and Gerrish, P., 2017. Dynamics and Fate of Beneficial Mutations Under Lineage Contamination by Linked Deleterious Mutations. *Genetics*, 205(3), pp.1305-1318.
33. Renzette, N., Gibson, L., Jensen, J.D. and Kowalik, T.F., 2014. Human cytomegalovirus intrahost evolution—a new avenue for understanding and controlling herpesvirus infections. *Current Opinion in Virology*, 8, pp.109-115.
34. Rousseau, E., Moury, B., Mailleret, L., Senoussi, R., Palloix, A. and Simon, V., 2017. Estimating virus effective population size and selection without neutral markers. *PLOS Pathogens*, 13(11), p.e1006702. doi: 10.1371/journal.ppat.1006702.
35. Sanjuán, R. and Domingo-Calap, P., 2021. Genetic Diversity and Evolution of Viral Populations. In: D.H. Bamford and M. Zuckerman, eds., *Encyclopedia of Virology (Fourth Edition)*, Academic Press, pp.53-61. ISBN 9780128145166. doi: 10.1016/B978-0-12-809633-8.20958-8.
36. Smith, J.M. and Haigh, J., 1974. The hitch-hiking effect of a favourable gene. *Genetical Research*, 23(1), pp.23-35. PMID: 4407212.

37. Sniegowski, P.D. and Gerrish, P.J., 2010. Beneficial mutations and the dynamics of adaptation in asexual populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1544), pp.1255–1263. doi: 10.1098/rstb.2009.0290.
38. Sniegowski, P.D., Gerrish, P.J., Johnson, T. and Shaver, A., 2000. The evolution of mutation rates: separating causes from consequences. *Bioessays*, 22(12), pp.1057-1066.
39. Sung, W., Ackerman, M.S., Miller, S.F., Doak, T.G. and Lynch, M., 2012. Drift-barrier hypothesis and mutation-rate evolution. *Proceedings of the National Academy of Sciences*, 109(45), pp.18488–18492. doi: 10.1073/pnas.1216223109.
40. Wang, X. and Feng, X., 2024. Challenges in estimating effective population sizes from metagenome-assembled genomes. *Frontiers in Microbiology*, 14, p.1331583. doi: 10.3389/fmicb.2023.1331583. PMID: 38249456; PMCID: PMC10797056.
41. Wilke, C.O. and Adami, C., 2003. Evolution of mutational robustness. *Mutation Research*, 522(1–2), pp.3–11. doi: 10.1016/S0027-5107(02)00307-X.
42. Wilke, C.O., Wang, J.L., Ofria, C., Lenski, R.E. and Adami, C., 2001. Evolution of digital organisms at high mutation rate leads to survival of the flattest. *Nature*, 412, pp.331–333. doi: 10.1038/35085569.
43. Benton, M.L., Abraham, A., LaBella, A.L., Abbot, P., Rokas, A. and Capra, J.A., 2021. The influence of evolutionary history on human health and disease. *Nature Reviews Genetics*, 22(5), pp.269-283. doi: 10.1038/s41576-020-00308-z. PMID: 33536697.
44. Dutta, A., Lanham, A., Ma, Z.S. and Merz, C., 2020. Virus mutation rates: Examination of recent data sets and methodology correction. *Journal of Genetics and Genomics*, 47(2), pp.63-72. doi: 10.1016/j.jgg.2019.12.001. PMID: 32057697.

45. Feder, A.F., Petrov, D.A. and Bergland, A.O., 2012. LDx: Estimation of Linkage Disequilibrium from High-Throughput Pooled Resequencing Data. *PLoS One*, 7(5), p.e48588. doi: 10.1371/journal.pone.0048588. PMID: 23112800; PMCID: PMC3484048.
46. Grueber, C.E., Wallis, G.P. and Jamieson, I.G., 2013. Genetic drift outweighs natural selection at toll-like receptor (TLR) immunity loci in a re-introduced population of a threatened species. *Molecular Ecology*, 22(17), pp.4470-4482. doi: 10.1111/mec.12370. PMID: 23909864.
47. Holmes, E.C., Dudas, G., Rambaut, A. and Andersen, K.G., 2016. The evolution of Ebola virus: Insights from the 2013-2016 epidemic. *Nature*, 538(7624), pp.193-200. doi: 10.1038/nature19790. PMID: 27734858.
48. Martin, G. and Lenormand, T., 2015. The fitness effect of mutations across environments: a survey in light of fitness landscape models. *Evolution*, 69(6), pp.1439-1454. doi: 10.1111/evo.12688. PMID: 26079746.
49. Seaton, D.D., McInnes, L. and Tollervey, D., 2020. Evolution of rapid trait loss: genomics of evolutionary time series following experimental shifts to extreme environments. *PLoS Genetics*, 16(2), p.e1008538. doi: 10.1371/journal.pgen.1008538. PMID: 32023289.
50. Wertheim, J.O. and Kosakovsky Pond, S.L., 2011. Purifying selection can obscure the ancient age of viral lineages. *Molecular Biology and Evolution*, 28(12), pp.3355-3365. doi: 10.1093/molbev/msr170. PMID: 21705368.