Львівський національний університет імені Івана Франка
Біологічний факультет
Кафедра генетики та біотехнології

Назва курсу	Прикладна молекулярна філогенетика
Викладач (-і)	Богдан Омелянович Осташ
Профайл викладача	http://bioweb.lnu.edu.ua/employee/ostash-b-o
Контактний тел.	032 2394407
E-mail:	bohdan.ostash@lnu.edu.ua
Сторінка курсу на сайті кафедри	http://bioweb.lnu.edu.ua/course/applied_molecular_phylogenetics
Консультації	*Очні консультації*: II семестр (2020 р), щовівторка, 11:30-13:00 *Онлайн- консультації:* у форматі "питання-відповідь" через електронну пошту, в робочі дні тижня, з 10:00-16:00; очікуйте на мою відповідь не пізніше ніж за три доби з моменту надходження питання

1. Коротка анотація до курсу

Вибухоподібне зростання даних про первинну структуру генетичного матеріалу усіх форм життя відкриває нові можливості до виявлення родинних та еволюційних зв'язків між ними. Молекулярна філогенетика – наукова дисципліна, що вивчає такі зв'язки на основі даних про нуклеотидні чи амінокислотні послідовності. Сьогодні методи молекулярної філогенетики знайшли своє практичне застосування у низці галузей людської діяльності, що спонукало до розроблення цього курсу, який названо відповідно – "прикладна молекулярна філогенетика". Ця навчальна дисципліна охоплює сучасні уявлення та методологічні підходи до реконструкції філогенетичних зв'язків (побудови філогенетичних дерев) між видами чи окремими генами/білками й геномами на основі нуклеотидних чи амінокислотних послідовностей. В основі - сучасні концепції молекулярної еволюції, що пояснюють фізико-хімічну логіку розвитку живої матерії на Землі; методологічні підходи до кількісної оцінки філогенетичних відносин на основі молекулярно-генетичних даних. У першій частині курсу буде розглянуто теоретичні засади філогенії та їхнє застосування на рівні генетичних послідовностей. У другій частині, курсу на основі вибраних прикладів буде розглянуто прикладні аспекти філогенетичної реконструкції у галузях біотехнологій, систематики, охорони природи, епідеміології, медицини та криміналістики.

2. Мета та цілі курсу

Мета: сформувати у слухачів курсу систему знань про методи молекулярної філогенетики та множину прикладних проблем, до вивчення яких ці методи залучаються

Цілі: *а*) викласти молекулярно-біологічні та еволюційні засади, на яких ґрунтується філогенетичний аналіз; *б*) пояснити, чим керуються дослідники при виборі даних для філогенетичного аналізу; *в*) пояснити основні особливості функціонування та застосування різних алгоритмів філогенетичної реконструкції; *г*) окреслити напрями практичного використання філогенетичного аналізу

3. Формат курсу – очний

4. Результати навчання

Після курсу студент буде: *а*) розуміти сутність і відмінності різних алгоритмів філогенетичної реконструкції; *б*) створювати, відбирати і курувати масиви даних для філогенетичного аналізу; *в*) розуміти зміст і важливість, що їх несуть різні елементи філогенетичного дерева; *г*) вміти створювати філогенетичні дерева на основі різних масивів даних за допомогою різних програмних знарядь та веб-сервісів; *д*) розуміти значення і цінність філогенетичної реконструкції у таких галузях як ідентифікація видів, епідеміологія, криміналістика, біотехнологія.

5. Обсяг курсу

Вид заняття	**Загальна к-сть годин**
лекції	40
практичні	-
самостійна робота	140

6. Ознаки курсу:

Рік викладання	**Се-мест р**	**Спеці-альність**	**Курс (рік навчання)**	**Нормативний\ вибірковий**
2020	2	Бакалаври, біологія	4	Вибірковий

7. **Пререквізити.** Знання англійської мови на рівні, достатньому для перекладу наукових статей; прослухання загальних (бакалаврських) курсів генетики, біохімії, зоології та ботаніки**.** Розуміння базових математичних понять та теорії імовірностей та статистичного аналізу даних. Базові навички роботи з комп'ютером.

8. **Технічне й програмне забезпечення /обладнання.** Для кількох лекційних занять необхідно буде принаймні по одному ноутбуку на двох студентів для ознайомлення з наявними базами даних генетичних послідовностей та філогенетичних програм (за рахунок підключення до відкритої WiFi мережі Університету).

9. **Політики курсу.** Відвідування лекційної частини курсу вільне. Матеріали лекційного курсу (PowerPoint-презентації) буде надано електронною поштою усім студентам. Усі статті і матеріали, або гіперпосилання до них, що згадано нижче у схемі курсу (п. 10) – буде надано. Перша частина курсу (принципи та методи філогенетичної реконструкції) закінчується письмовим модулем. Написання модуля у визначений час обов'язкове, відсутність можлива лише за умови поважної причини, що має бути задокументовано (довідка про хворобу тощо). Протягом семестру, у межах часу, відведеного на самостійну роботу, всі виконують дослідницький мікропроект з філогенетичної реконструкції на основі

певного масиву нуклеотидних чи амінокислотних послідовностей, який отримає оцінку. Більше про систему оцінювання – див. нижче розділ **11**. Очікується, що студенти дотримуватимуться правил Академічної доброчесноті – див. http://www.lnu.edu.ua/wp-content/uploads/2019/06/reg_academic_virtue.pdf. Нульова толерантність (у вигляді недопуску до заліку) до плагіату, списування, хабарництва. Зниження оцінки при виявленні фактів несамостійного підготовлення завдань до практичних занять (нерозуміння підготовленої презентації, механічне використання перекладів, згенерованих автоматичними перекладачами тексту).

10. Схема курсу

Тиждень 1 (23 березня 2020)

Лекція 1 (2 год). Вступ. Знайомство з групою, з'ясування наукових інтересів групи. Структура, політика, оцінювання курсу. Чого навчиться студент під час цього курсу. Що таке філогенетика. Що спонукало розвиток молекулярної філогенетики? Припущення про спільний корінь всього живого. Поділ клітини, генеалогії. Видоутворення, філогенії. Що таке вид? – з точки зору тварини, рослини, бактерії. Механізми макро- та мікроеволюції. Природний добір і генетичний дрейф. *Матеріали* – презентація лекції amp_lecture1.pdf. *Література*: [1], розділ 7 (див. список наприкінці схеми курсу).

Самостійна робота 1 (9 год). Ознайомлення з життям і науковим доробком українсько-американського вченого-генетика Теодозія Добжанського – https://www.youtube.com/watch?v=TH2AC8fu34M. Сумісність біблійного та еволюційного вчень – погляд Т. Добжанського https://www.pbs.org/wgbh/evolution/library/10/2/text_pop/l_102_01.html

Лекція 2 (2 год). Філогенія – засадничі поняття. Визначення і моделі. Елементи дерева: нода (край, вузол), гілка (вісь), корінь, монофілетична група, парафілія, кінцеві таксони (листки дерева). Ієрархічність клад. Укорінення дерева. Морфологічні та фізіологічні ознаки як матеріал для філогенії. Синапоморфії, аутаморфії, плезіоморфії. Дендрограма, кладограма (топологічне дерево), філограма (метричне дерево). Масштаб дерева. Політомія. Яку інформацію містить і яку не містить філогенетичне дерево. Чи відображає філогенетичне дерево прогресивну еволюцію? Ротація клад. Призначення завдань для виконання мікропроектів. *Матеріали* – презентація лекції amp_lecture2.pdf. *Література*: [1], розділ 7; [3].

Тиждень 2 (30 березня)

Лекція 3 (2 год). Підходи до філогенії: від кістки до гена. Геном як "молекулярний палімпсест". Що таке ознака у філогенетиці. Оптимальні властивості ознаки. Стан ознаки, її філогенетичний сигнал. В чому переваги молекулярних маркерів над морфологічними чи фізіологічними? Один ген – одна ознака чи низка ознак? Дерево видів → дерево генів. Поняття гомології. Умови філогенії: однакова швидкість, єдиний і незмінний механізм змін, незалежність змін у різних частинах генів/білків. Концепція "молекулярного годинника". Тест Таджими для підтвердження чи спростування "молекулярного годинника" для пари послідовностей (у пакеті MEGA X). Виклики, що постають перед реконструкцією філогенії на основі молекулярних маркерів: різна еволюційна швидкість (для різних геномів, генів і частин гена), горизонтальне

перенесення генів, конвергентна та паралельна еволюція. Гомоплазія. П'ять етапів філогенетичної реконструкції: збір даних, множинне вирівнювання, вибір моделі еволюції, реконструкція дерева, оцінення дерева. *Матеріали* – презентація лекції amp_lecture3.pdf. *Література*: [1], розділ 7; [3].

Самостійна робота 2 (9 год). Як еволюція і філогенетика стосується мого повсякденного життя? – кожен зі студентів групи має вибрати і прочитати по одній статті зі списку, що є на веб-ресурсі університету Берклі: https://evolution.berkeley.edu/evolibrary/search/topics.php?topic_id=15. Будьте готові до трихвилинного виступу, що підсумовує прочитану статтю, і пояснити, яка роль філогенетики у розглянутому матеріалі. Завантаження програми MEGA X на комп'ютер (https://www.megasoftware.net/dload_mac_beta), ознайомлення з його базовими функціями на основі довідника користувача (https://www.megasoftware.net/docs)

Лекція 4 (2 год). Масив даних для філогенетичної реконструкції. Отримання нуклеотидних послідовностей: секвенування цілих геномів. Копіювання окремих ділянок генома для секвенування – стисло про метод ПЛР і підбір консервативних праймерів. Отримання інформації з баз даних: GenBank, ENA, Patric, BioCyc. Для білок-кодувальних генів – що аналізуємо: послідовність нуклеотидів, кодонів чи амінокислот? Збільшення філогенетичного сигналу – конкатеновані послідовності. Чи більше маркерів – краще? Поняття ортології та паралогії. Чим керуватися при виборі генів/білків для філогенетичної реконструкції? Вплив філогенетичної відстані (мілко – ДНК, глибоко – білки). Випадок №1: класифікація невідомої бактерії. Випадок №2: хітинази. Випадок №3: класифікація тритонів. Випадок №4: визначення, чи свіжовиділений ізолят роду Streptomyces належить до нового виду. *Матеріали* – презентація лекції amp_lecture4.pdf. *Література*: [1], розділ 7; [3].

Самостійна робота 3 (9 год). Ознайомлення з функціональними можливостями програми MEGA X. Знайомство з базами даних GenBank, Patric з метою пошуку послідовностей, що необхідні для виконання мікропроекту.

Тиждень 3 (6 квітня 2020)

Лекція 5 (2 год). Філогенетична реконструкція на основі гена 16S рРНК. Структура 16S рРНК: консервативні та варіабельні ділянки. Стратегії підбору праймерів. Філогенетичне дерево життя на основі 16S рРНК. Бази: RDP, Greengenes, SILVA, Gold. Універсальні 16S/18S праймери (див. https://doi.org/10.1371/journal.pone.0090053). Недоліки гена 16S рРНК як філогенетичного маркера: мікрогетерогенність набору генів 16S рРНК у межах генома; занадто слабкий філогенетичний сигнал, аби розрізняти види (див. т-ж https://www.drive5.com/usearch/manual/citation.html); недосконалість баз даних. Виклики, що постають при аналізі метагеномної ДНК. Операційні таксономічні одиниці (OTU; operating taxonomy units); α- і β-різноманітність метагеномних зразків. ПЛР-химери. Перехресне картування секвенованих ділянок висококонсервативних генів при аналізі метагеномних зразків (crosstalk). Програма UNOISE2. Філогенетичні маркери для реконструкції філогенії евкаріотів: 18S рРНК, ISSR (inter-simple sequence repeats), RAPD (randomly amplified polymorphic DNA), гени цитохромів, *rbcL*, ITS (internal transcribed sequences). Ідентифікація маркерів для близькоспоріднених видів – на прикладі групи *Bacillus pumilus*. *Матеріали* – презентація лекції amp_lecture5.pdf. *Література*: [1], розділ 9.

Лекція 6 (2 год). Множинне вирівнювання. Що таке вирівнювання? Попарне вирівнювання – збіги, незбіги, рахунок вирівнювання, штрафи за прогалини. Підходи до кількісного оцінювання попарних вирівнювань: нуклеотидні моделі, матриці PAM і BLOSUM для білків. Статистичне оцінення попарних вирівнювань. Знаряддя попарного вирівнювання – Fasta, BLAST. Множинне вирівнювання, чим відрізняється від попарного. Знаряддя для множинних вирівнювань на сайті https://toolkit.tuebingen.mpg.de/. Чому важливо використовувати різні програми множинного вирівнювання – на основі аналізу результатів вирівнювання за допомогою програм T-COFFEE та MUSCLE. Курування вирівняного масиву даних. *Матеріали* – презентація лекції amp_lecture6.pdf. *Література*: [1], розділи 3, 4, 6.

Самостійна робота 4 (9 год). Відбір 10-15 гомологічних послідовностей для потреб мікропроекту. Ознайомлення з можливостями пакету MEGA X щодо оцінки явища молекулярного годинника та моделі еволюції на основі зібраного студентом масиву даних. Форматування даних в MEGA X.

Тиждень 4 (13 квітня 2020)

Лекція 7 (2 год). Моделі заміщень нуклеотидних й амінокислотних залишків. Еволюційні моделі у філогенетиці і чому вони потрібні – на прикладі моделей Джакса-Кентора і K2P для нуклеотидних послідовностей. Моделі амінокислотних заміщень – LG, WAG, JTT. Гамма-розподіл для моделювання різних швидкостей заміщення у різних сайтах гена/білка. Знаряддя для вибору оптимальної моделі заміщення – вебсервер IQ Tree (http://iqtree.cibiv.univie.ac.at/), функція ProtTest в MEGA X. Критерії відбору – інформаційний критерій Акаїке (AIC), інформаційний критерій Байєза (BIC). Як вибір моделі впливає на остаточний результат? *Матеріали* – презентація лекції amp_lecture7.pdf. *Література*: [1], розділ 7.

Самостійна робота 5 (9 год). Побудова множинних вирівнювань на створеному масиві даних, з використанням онлайн-знарядь (Muscle, T-COFFEE) та програм у пакеті MEGA X. Порівняння результатів вирівнювань. Збереження результатів множинного вирівнювання.

Лекція 8 (2 год). Алгоритми філогенетичної реконструкції. Дистанційні і позиційні методи – визначення і відмінності. Дистанційні методи – UPGMA та NJ. Алгоритмічні особливості методу UPGMA (**U**nweighted **P**air-**G**roup **M**ethod with **A**rithmetic means). Усереднення відстаней та оцінка достовірності дерева. Алгоритм Фітча-Марголіяша (FM) – удосконалення UPGMA. Укорінення дерева в методах UPGMA та FM. Недоліки розглянутих вище алгоритмів. Ультраметричне дерево. Притягування довгих гілок. Алгоритм Neighbor Joining (NJ). Адитивність і неадитивність філогенетичних відстаней. Умова чотирьох точок. Переваги NJ над UPGMA. *Матеріали* – презентація лекції amp_lecture8.pdf. *Література*: [1], розділ 9; [2], розділ 5.

Тиждень 5 (27 квітня 2020) – Молодь і поступ біології

Лекція 9 (2 год). Позиційні методи філогенетичної реконструкції – максимальної ощадності. Принцип методу – відсутність еволюційних моделей, мутації рідкісні, найвірогідніший еволюційний шлях – той, що вимагає мінімум мутацій. Метод максимальної вірогідності (ML; maximum likelihood). Принцип методу – обчислення вірогідності того, що за обраної еволюційної моделі певне дерево приведе до заданої

нуклеотидної/амінокислотної послідовності. Недоліки ML. *Матеріали* – презентація лекції amp_lecture9.pdf. *Література*: [1], розділ 9; [2], розділ 5.

Самостійна робота 6 (9 год). Робота у програмному середовищі MEGA X. Філогенетична реконструкція на основі множинно вирівняних масивів даних, отриманих у межах попередньої самостійної роботи. Вплив вибору моделі еволюції та алгоритму реконструкції на топологію дерева – метод UPGMA.

Лекція 10 (2 год). Оцінка надійності філогенетичного дерева. Відтворення філогенії різними методами. Конгруентність дерев. Бутстреп-аналіз. Індекс aLRT (approximate Likelihood Ratio Test). Індекс CI. Формати збереження даних філогенетичної реконструкції. Формат Newick. Графічний рендеринг дерева – можливості phylogeny.fr. Програма FigTree. Колапсування клад. *Матеріали* – презентація лекції amp_lecture10.pdf. *Література*: [1], розділ 9; [3].

Самостійна робота 7 (9 год). Робота у програмному середовищі MEGA X. Філогенетична реконструкція на основі множинно вирівняних масивів даних, отриманих у межах попередньої самостійної роботи. Вплив вибору моделі еволюції та алгоритму реконструкції на топологію дерева – метод NJ.

Тиждень 6 (4 травня 2020 р)

Лекція 11 (2 год). Підсумок перших шести тижнів занять. Основні терміни, що стосуються філогенії. Етапи філогенетичної реконструкції. Значення моделі еволюції. Чи присутня вона у всіх методах філогенетичної реконструкції? Чи важливий вибір моделі еволюції? Яке значення має оцінка надійності топології дерева? *Матеріали* – презентація лекції amp_lecture11.pdf. *Література*: [1], розділ 9; [2], розділ 5; [3], [4].

Лекція 12 (2 год). Спеціалізовані філогенетичні сервіси. iTOL – interactive Tree of Life: https://itol.embl.de/. Включення метаданих. Датування часу дивергенції таксонів – TimeTree http://www.timetree.org/. Phylogeny.fr. *Матеріали* – презентація лекції amp_lecture12.pdf. *Література*: [1], розділ 9.

Самостійна робота 8 (9 год). Робота з базами даних SILVA і Greengenes. Порівняння схем класифікації бактерій роду Micrococcus у двох базах.

Тиждень 7 (11 травня 2020)

Лекція 13 (2 год). Мультилокусна філогенія (MLSA). Вимоги до маркерів. Базовий варіант – 16S, *atpD*, *gyrB*, *recA*, *rpoB*, *trpB*. Консервативні праймери. Протокол MLSA. Визначення рівня гомології ДНК на основі попарних відстаней між конкатенованими послідовностями (робота: DOI 10.1007/s10482-014-0348-4). Більше маркерів – PhyloPhLan. Ще більше маркерів – https://doi.org/10.1038/s41467-019-13443-4. Опис вибору маркерів для високо спорідненої групи *Bacillus pumilus* - doi:10.1371/journal. pone.0163098. Вибір маркерів для стрептоміцетів: робота Gupta et al. *Матеріали* – презентація лекції amp_lecture13.pdf. *Література*: [1], розділ 9.

Самостійна робота 9 (9 год). Робота в програмі MEGA X. Алгоритм ML. Статистична оцінка надійності топології філогенетичного дерева. Чи узгоджується дерево з концепцією молекулярного годинника?

Лекція 14 (2 год). Філогенія у межах одного виду/популяції – концептуальні відмінності від філогенії видів. Коалесцентна теорія. Фіксовані мутації між видами і поліморфізм у межах виду. Філогенетична реконструкція у вірусних популяціях, на прикладі вірусу імунодефіциту людини (HIV). Особливості біології HIV. Маркерні гени HIV. Філогенетична реконструкція HIV – глобальний рівень, між популяціями, у межах популяції, в одній особі. Про що свідчить топологія і довжина гілок дерева HIV? Практичне застосування філогенії HIV *Матеріали* – презентація лекції amp_lecture14.pdf. *Література*: [1], розділ 9.

Тиждень 8 (18 травня 2020)

Лекція 15 (2 год). Філогенетична реконструкція в епідеміологічних дослідженнях – на прикладі спалаху Burkholderia dolosa в хворих на муковісцидоз. Відстеження епідемії SARS. *Матеріали* – презентація лекції amp_lecture15.pdf. *Література*: [1], розділ 9.

Самостійна робота 10 (9 год). Графічний рендеринг філогенетичних дерев – FigTree, iTOL. Операційні середовища для роботи з великими масивами даних – Galaxy, Cyverse.

Лекція 16 (2 год). Філогенетична реконструкція в природоохоронних дослідженнях – виявлення походження китового м'яса на ринках Східної Азії. Криміналістична філогенетика. *Матеріали* – презентація лекції amp_lecture16.pdf. *Література*: [1], розділ 9.

Самостійна робота 11 (9 год). Ідентифікація виду на основі заданого масиву маркерів.

Тиждень 9 (25 травня 2020)

Лекція 17 (2 год). Філогенетична реконструкція у дослідженні еволюції кластерів генів вторинного метаболізму актинобактерій. Реконструкція еволюції генів біосинтезу глікопептидних антибіотиків – еволюція виду, кластера, генів у межах кластера. Які висновки можна робити з аналізів такого типу. *Матеріали* – презентація лекції amp_lecture17.pdf. *Література*: [1], розділ 9.

Лекція 18 (2 год). Філогенетична реконструкція білкових родин – нові уявлення про еволюцію субстратної специфічності ферментів. Виявлення паралогії: порівняння дерева видів з деревом амінокислотних послідовностей. Реконструкція предкових послідовностей, біотехнологічне значення цього підходу. *Матеріали* – презентація лекції amp_lecture18.pdf. *Література*: [1], розділ 9.

Самостійна робота 12 (9 год). Ознайомлення з філогенетичними ресурсами NCBI – дерева, що супроводжують сервер BLAST. Як оцінюється надійність таких дерев?

Тиждень 10 (1 червня 2020)

Лекція 19 (2 год). Філогенетична реконструкція як метод виявлення горизонтального перенесення генів. Принципи методу, приклади застосування, зокрема у випадку еволюції стійкості до висихання (GRAS білки, стаття https://doi.org/10.1016/j.cell.2019.10.019). Методи виявлення рекомбінації і функції гена. Філогенетичні підходи і методи виявлення ознак природного добору. d_N/d_S. *Матеріали* – презентація лекції amp_lecture19.pdf. *Література*: [1], розділ 9.

Самостійна робота 13 (9год). Використання додаткових знарядь у програмі MEGA X: пошук ознак природного добору, часу дивергенції, молекулярного годинника, тест Таджими.

Лекція 20 (2 год). Методи філогенетичної реконструкції за межами біології: еволюція мови. Нейтральні і селективні процеси в еволюції мови – за і проти. Роль розміру популяції у швидкості набуття чи втрати слів у мові. *Матеріали* – презентація лекції amp_lecture20.pdf. *Література*: https://www.ncbi.nlm.nih.gov/pubmed/30898886; https://www.ncbi.nlm.nih.gov/pubmed/29755387

Самостійна робота 14 (9 год). Ознайомлення зі статтею Pagel et al. 2009; Nat Rev Genet.; doi:10.1038/nrg2560.

Література

1. Pevsner J. Bioinformatics and functional genomics. 3rd edition. Wiley Blackwell, London. – 2015- 1116 p. ISBN 978-1-118-58178-0.
2. Allman E.S., Rhodes J.A. Mathematical models in biology. An introduction. Cambridge University Press, Cambridge, 2004. 386 p. ISBN 978–0-511–07689–3. eBook.
3. Barton NH, Briggs DEG, Eisen JA, Goldstein DB, Patel NH. Chapter 26. Phylogenetic reconstruction. In: Evolution. CSHL Press, 2007, 833 p.
4. Yang Z. Computational Molecular Evolution. Oxford University Press, 2006. 374 p.

11. Система оцінювання та вимоги

Загальна система оцінювання курсу	участь в роботі впродовж семестру/залік - 50/50
Вимоги до письмової роботи (модуль)	За змістом перших десяти лекцій буде виконано поточний контроль знань у вигляді написання модуля. В модуль входять: визначення термінів (10 балів), два питання (по 6 б. кожне), одна схема чи таблиця, яку треба заповнити/зобразити (8 б.). Максимальна оцінка за модуль – 30 балів. Написання модуля обов'язкове.
Практичні заняття	Ще 20 балів студент може набрати упродовж семестру за виконання мікропроєкту. У 20 балів входять: створення масиву даних (5 балів), множинне вирівнювання і курування даних (5 балів), філогенетична реконструкція різними алгоритмами (5 балів) і оцінка надійності дерева (5 балів).
Умови допуску до підсумкового контролю	Написання модуля і виконання мікропроєкту (останній надсилаєте на пошту у вигляді PowerPoint-презентації (конвертованої в .pdf), що включає усі етапи виконання роботи, з необхідними поясненнями).
Залік	Набір питань аналогічно до модуля; до термінів, питань і схем додаються тести. Письмова підготовка на протязі не більше 30 хв, далі усна відповідь. На залік виноситься весь матеріал курсу