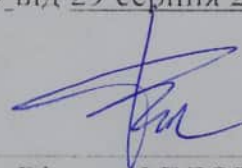


МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
Львівський національний університет імені Івана Франка  
Біологічний факультет  
Кафедра генетики та біотехнології

Затверджено  
на засіданні кафедри генетики та біотехнології  
(протокол № 19 від 29 серпня 2024 р.)

Завідувач кафедри. \_\_\_\_\_



Віктор ФЕДОРЕНКО

Силабус з навчальної дисципліни «Біоінформатика»  
що викладається в межах ОПП «Біохімія», «Біофізика», «Ботаніка», «Генетика»,  
«Зоологія», «Мікробіологія», «Фізіологія людини і тварин», «Фізіологія рослин»  
другого (магістерського) рівня вищої освіти для здобувачів зі спеціальності  
091 Біологія та біохімія

Львів 2024

<b>Назва курсу</b>	Біоінформатика
<b>Адреса викладання курсу</b>	вул. Грушевського 4, 79005 Львів
<b>Факультет та кафедра, за якою закріплена дисципліна</b>	біологічний факультет, кафедра генетики і біотехнології
<b>Галузь знань, шифр та назва спеціальності</b>	09 Біологія, 091 Біологія та біохімія
<b>Викладачі курсу</b>	Богдан Омелянович Остап Степан Іванович Тістечок
<b>Контактна інформація викладачів</b>	bohdan.ostash@lnu.edu.ua 032 2394407 stepan.tistechok@lnu.edu.ua <a href="http://bioweb.lnu.edu.ua/employee/ostash-b-o">http://bioweb.lnu.edu.ua/employee/ostash-b-o</a>
<b>Консультації по курсу відбуваються</b>	<i>Очні консультації:</i> II семестр (2025 р), щовівторка, 11:30-13:00 <i>Онлайн-консультації:</i> у форматі "питання-відповідь" через електронну пошту, в робочі дні тижня, з 10:00-16:00; очікуйте на відповідь не пізніше ніж за три доби з моменту надходження питання
<b>Сторінка курсу</b>	<a href="https://bioweb.lnu.edu.ua/course/bioinformatyka">https://bioweb.lnu.edu.ua/course/bioinformatyka</a>
<b>Інформація про курс</b>	Курс розроблено так, щоб магістри набули фахові знання з аналізу нуклеотидних та амінокислотних послідовностей (НАП) за допомогою спеціалізованого програмного забезпечення. У курсі розглядаються біоінформатичні бази даних, основні математичні моделі НАП, методи попарного і множинного вирівнювань, моделей на їхній основі, зокрема великі мовні моделі. Курс включає теоретичний матеріал у вигляді лекцій і виконання лабораторних занять.
<b>Коротка анотація курсу</b>	У 2024 року у базах даних є інформація про гени більше 400 000 організмів, або більш ніж $21 \times 10^{12}$ пар нуклеотидів. Це колосальна кількість інформації про генетичну основу життя, і її кількість невпинно зростає. Нуклеотидні послідовності не мають змісту доти, доки дослідник його не визначить, як-от здатність кодувати білок, бути промотором тощо. Систематичний аналіз усього масиву даних експериментальними методами нереалістичний як економічно так і технологічно. Це стимулювало розвиток методів комп'ютерного аналізу НАП, що допомогли упорядкувати дані у межах різних веб-сервісів. Найважливішим результатом синтезу молекулярної біології, геноміки, інформатики, комп'ютерних технологій і статистики – біоінформатики – стало те, що ця нова галузь перетворилася у самостійне знаряддя наукового відкриття. Курс "Біоінформатика" включає розділи, присвячені структурам баз даних: попарному і множинному вирівнюванню НАП; моделям, що узагальнюють множинні вирівнювання; основам філогенетичної реконструкції на основі; передбаченню функцій генів і пошукові мотивів у НАП; передбаченню тривимірних структур білків; аналізу молекул РНК.
<b>Мета та завдання курсу</b>	<b>Мета:</b> сформувати у слухачів курсу систему знань про основні бази даних нуклеотидних й амінокислотних послідовностей (НАП) та їхнього комп'ютерного аналізу, зокрема попарного і множинного вирівнювання, моделей на основі вирівнювань, філогенетичної реконструкції на основі НАП.

	<p><b>Завдання:</b> а) викласти концептуальні математичні та молекулярно-біологічні засади, на яких ґрунтується біоінформатика; б) ознайомити студентів з наявним арсеналом методів біоінформатики, їхніх можливостей і обмежень, останніх тенденцій розвитку дисципліни; в) сформулювати у студентів арсенал активних знань у галузі цієї дисципліни, тобто навчити їх формулювати наукові питання, відповіді на які можна шукати із залученням методів біоінформатики.</p>
<p><b>Література для вивчення дисципліни</b></p>	<p><b>Основна література:</b></p> <p><b>Базова:</b></p> <ol style="list-style-type: none"> <li>1. Остап Б.О. Біоінформатика: аналіз генетичних послідовностей. Електронний підручник. Видавничий центр ЛНУ ім. Івана Франка, 2022, 232 стор. ISBN 978-617-10-0729-1. Доступ онлайн: <a href="http://dspace.lnlibrary.lviv.ua/handle/123456789/169">http://dspace.lnlibrary.lviv.ua/handle/123456789/169</a></li> <li>2. Higgs PG, Attwood TK. Bioinformatics and Molecular Evolution. Blackwell Publishing, Oxford, 2005. 398 p. ISBN 1-4051-0683-2.</li> </ol> <p><b>Допоміжна:</b></p> <ol style="list-style-type: none"> <li>3. Durbin R, Eddy S, Krogh A, Mitchison G. Biological Sequence Analysis. Probabilistic Models of Proteins and Nucleic Acids. Cambridge University Press, Cambridge, 1998. 371 p. ISBN-13 978-0-521-62971-3</li> <li>4. Borodovsky M, Ekisheva S. Problems and Solutions in Biological Sequence Analysis. Cambridge University Press, Cambridge, 2006. 362 p. ISBN-13 978-0-521-61230-2</li> <li>5. Bioinformatics: a practical guide to the analysis of genes and proteins, 2<sup>nd</sup> Ed / AD Baxevanis, BFF Ouellette. – New York: John Wiley &amp; Sons, 2001. – 455 p.</li> <li>6. Allman ES, Rhodes JA. Mathematical Models in Biology. An Introduction. Cambridge University Press, Cambridge, 2003. 386 p.</li> <li>7. Pevsner J. Bioinformatics and functional genomics. 3<sup>rd</sup> edition. Wiley Blackwell, UK. – 2015- 1116 p. ISBN 978-1-118-58178-0.</li> </ol>
<p><b>Тривалість курсу</b></p>	<p>Один семестр</p>
<p><b>Обсяг курсу</b></p>	<p>Лекції – 32 год, лабораторні – 16 год, самостійна робота – 102 год; разом – 150 год (5 кредитів ECTS)</p>
<p><b>Очікувані результати навчання</b></p>	<p>Після курсу магістр буде знати:</p> <ul style="list-style-type: none"> <li>- основні бази даних НАП, принципи функціонування алгоритмів вирівнювання НАП, і спектр питань щодо структури, функції та еволюції НАП, на які можна шукати відповіді із застосуванням біоінформатичних веб-ресурсів і програм;</li> <li>- методи філогенетичної реконструкції на основі НАП;</li> </ul> <p><b>вміти:</b></p> <ul style="list-style-type: none"> <li>- користуватись базами даних, де зберігаються нуклеотидні та амінокислотні послідовності, інформація про структури геномів, їхню експресію (транскриптомні дані – генні чіпи і RNAseq) та відповідні протеоми;</li> <li>- порівнювати НАП з гомологічними послідовностями за допомогою методів попарного і множинного вирівнювань: оперувати паттернами і профілями;</li> <li>- виявляти регуляторні послідовності у геномах – повтори, паліндроми, консервативні мотиви; передбачати функціональність</li> </ul>

	<p>гена (білка) на основі аналізу консервативних доменів і каталітичних центрів:</p> <p>- користуватись методами передбачення і моделювання структури білків; здійснювати молекулярно-філогенетичний аналіз.</p>
<p><b>ЗК, ФК, ПРН</b></p>	<p>Курс розроблено так, щоб сформувати у студентів <u>загальні і фахові компетентності</u>:</p> <p>ЗК01. Здатність працювати у міжнародному контексті.</p> <p>ЗК02. Здатність використовувати інформаційні та комунікаційні технології.</p> <p>ЗК03. Здатність генерувати нові ідеї (креативність).</p> <p>ФК01. Здатність користуватися новітніми досягненнями біології, необхідними для професійної, дослідницької та/або інноваційної діяльності.</p> <p>ФК02. Здатність формулювати задачі моделювання, створювати моделі об'єктів і процесів на прикладі різних рівнів організації живого із використанням математичних методів й інформаційних технологій.</p> <p>ФК03. Здатність користуватися сучасними інформаційними технологіями та аналізувати інформацію в галузі біології і на межі предметних галузей.</p> <p>ФК06. Здатність прогнозувати напрямки розвитку сучасної біології на основі загального аналізу розвитку науки і технологій.</p> <p>ФК07. Здатність діагностувати стан біологічних систем за результатами дослідження організмів різних рівнів організації</p> <p>ФК10. Здатність використовувати результати наукового пошуку в практичній діяльності.</p> <p><u>Програмні результати навчання:</u></p> <p>ПР1. Володіти державною та іноземною мовами на рівні, достатньому для спілкування з професійних питань та презентації результатів власних досліджень.</p> <p>ПР2. Використовувати бібліотеки, інформаційні бази даних, інтернет ресурси для пошуку необхідної інформації.</p> <p>ПР4. Розв'язувати складні задачі в галузі біології, генерувати та оцінювати ідеї.</p> <p>ПР5. Аналізувати та оцінювати вплив досягнень біології на розвиток суспільства.</p> <p>ПР6. Аналізувати біологічні явища та процеси на молекулярному, клітинному, організменному, популяційно-видовому та біосферному рівнях з точки зору фундаментальних загальнонаукових знань, а також за використання спеціальних сучасних методів досліджень.</p> <p>ПР11. Проводити статистичну обробку, аналіз та узагальнення отриманих експериментальних даних із використанням програмних засобів та сучасних інформаційних технологій.</p> <p>ПР14. Дотримуватись норм академічної доброчесності під час навчання та</p>

	<p>провадження наукової діяльності, знати основні правові норми щодо захисту інтелектуальної власності.</p> <p>ПР16. Критично осмислювати теорії, принципи, методи з різних галузей біології для вирішення практичних задач і проблем.</p>
<b>Ключові слова</b>	Математичні моделі в біології, множинні і попарні вирівнювання, моделі Маркова, філогенетичні реконструкції, мовні моделі (AlphaFold), бази НАП.
<b>Формат курсу</b>	Очний або дистанційний
<b>Теми</b>	Наведено у таблиці внізу
<b>Підсумковий контроль, форма</b>	Іспит в кінці семестру. Усний.
<b>Пререквізити</b>	Знання англійської мови на рівні, достатньому для перекладу наукових статей; необхідні знання з основ генетики, біохімії, зоології та ботаніки. Розуміння базових математичних понять (логарифм, частка значень, експонента, відсоткові величини) та теорії імовірностей та статистичного аналізу даних. Базові навички роботи з комп'ютером. Наявність комп'ютера/смартфона з підключенням до інтернету (для лабораторних занять у випадку дистанційного формату навчання).
<b>Навчальні методи та техніки, які будуть використовуватися під час викладання курсу</b>	Лекції, пояснення, виконання завдань на лабораторних заняттях.
<b>Необхідне обладнання</b>	Персональний комп'ютер, загальноживані комп'ютерні програми і операційні системи, підключення до мережі Інтернет
<b>Критерії оцінювання (окремо для кожного виду навчальної діяльності)</b>	<p>Виконання завдань упродовж семестру/іспит – 50/50</p> <p>За змістом перших дев'яти 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 – 20 балів. Написання модуля обов'язкове.</p> <p>Ще 30 балів студент може набрати упродовж семестру за виконання шести контрольних завдань упродовж лабораторних занять (Бази даних, BLAST, MSA, WebLogo, Phylo1, Phylo2).</p> <p>Сумарний бал за модуль і за виконання контрольних завдань під час лабораторних занять має становити не менше 25, у такому співвідношенні: не менше 5 за модуль і не менше 20 за лабораторні.</p> <p>На іспиті – набір питань аналогічно до модуля: до термінів, питань і схем додаються тести. Письмова підготовка на протязі не більше 30 хв, далі усна відповідь. На іспит виноситься весь матеріал курсу</p> <p>Жодні форми порушення академічної доброчесності (<a href="https://www.lnu.edu.ua/wp-content/uploads/2019/06/reg_academic_virtue.pdf">https://www.lnu.edu.ua/wp-content/uploads/2019/06/reg_academic_virtue.pdf</a>) не толеруються.</p>
<b>Питання до екзамену</b>	<ol style="list-style-type: none"> <li>1. Моделі, що описують частоту зустрічності "слів" у генетичних послідовностях – Бернуллі і ланцюги Маркова.</li> <li>2. Консервативні і неконсервативні заміщення у попарних вирівнюваннях, способи їхнього оцінювання; підходи до уведення у вирівнювання прогалли</li> </ol>

3. PubMed – стратегії пошуку бібліографічних даних, значення ідентифікаторів статей
4. Структура флет-файлу GenBank
5. Національний центр біотехнологічної інформації США (NCBI) – структура і функції
6. Теорія прийнятних точкових мутацій (PAM) М. Дейгоф
7. Опишіть процес побудови позиційно-специфічних рахункових матриць (PSSM).
8. Алгоритми динамічного програмування у попарному вирівнюванні.
9. Основні терміни і теоретичні засади попарного вирівнювання
10. Опишіть етапи прогресивного принципу множинного вирівнювання
11. Моделі еволюції нуклеотидних послідовностей
12. Моделі Маркова в аналізі генетичних послідовностей
13. Способи опису складності НАП та районів з низькою складністю. Біологічне значення НАП низької складності.
14. Моделі заміщення в кодонних послідовностях
15. Опишіть основні елементи сторінки результатів BLAST
16. Підходи до уведення прогалин у попарні й множинні вирівнювання
17. Принципи побудови і використання прихованих моделей Маркова в аналізі НАП
18. Статистична оцінка попарних вирівнювань. Теорема Карліна-Альтшуля
19. Як розуміти унітарні й емпіричні системи обчислення рахунків вирівнювань з точки зору теорії інформації? Наведіть приклади. Теорема Альтшуля
20. У рівнянні з теореми Карліна-Альтшуля число очікування  $E$  прямо пропорційне простору пошуку (розміру бази даних) і обернено рахунку вирівнювання  $S$ . Тобто, у більших базах імовірність виявити випадкові збіги мала б зростати. Однак на практиці більші бази даних приводять до вирівнювань з дуже низьким  $E$ . Чому?
21. Які підходи використовують до оцінювання розривів (прогалин) у попарних вирівнюваннях?
22. Основні параметри опису попарного вирівнювання:  $S$ ,  $S'$ ,  $E$ ,  $P$ , ідейтичність  $ID$ , подібність  $SI$
23. НАП з точки зору інформації: інформаційний вміст, символічне повідомлення, частотний розподіл символів, імовірність зустрічності символів/слів у послідовності.
24. Параметри налаштування програми BLAST
25. PAM-матриці і способ їхньої побудови
26. Опишіть принципи побудови механістичних та емпіричних матриць заміщень: що відмінного і спільного між ними.
27. Принцип побудови матриць BLOSUM, їхні відмінні властивості (порівняно з PAM).
28. Відмінні і спільні риси паттернів, позиційно-специфічних матриць і прихованих моделей Маркова
29. Чому не існує не одна, а серія матриць PAM або BLOSUM? Як обчислюють матрицю PAM250? Для яких потреб слід використовувати матрицю PAM20?
30. Основні властивості моделей еволюції нуклеотидних

послідовностей.

31. Який біологічний зміст несе уведення розривів (прогалів) у попарні вирівнювання? Чи можна переставляти місцями позиції вирівнювання, і якщо так – то який біологічний процес відображає така маніпуляція НАП
32. Принципи дотплот-аналізу НАП
33. Що спонукало дослідників до розробки емпіричних підходів до оцінки вирівнювань амінокислотних послідовностей? Чому емпіричні підходи не набули поширення для нуклеотидних послідовностей?
34. Спільне і відмінне в попарному і множинному вирівнюваннях
35. Стисло опишіть моделі, які створюють на основі множинного вирівнювання НАП. Що в них відмінного?
36. Еволюційні засади попарного вирівнювання. Основні терміни.
37. Поясніть, що таке позиційно-незалежні та позиційно-специфічні рахунки вирівнювання, наведіть приклади їхнього використання в аналізі НАП
38. Концепція псевдорахунків у біоінформатиці – приклади її використання
39. опишіть відомі вам способи множинного вирівнювання НАП
40. Принцип функціонування алгоритму пошуку оптимального глобального вирівнювання (Нідельмана-Ванча)
41. Відмінність між алгоритмами глобального і локального вирівнювання НАП
42. За якими ознаками певну модель чи процес можна віднести до Марковського?
43. Яку роль відіграють константи  $\lambda$  й  $K$  в рівнянні для обчислення числа  $E$ ?
44. Етапи побудови позиційно-специфічної рахункової матриці
45. Моделі оцінки частот символів/слів у НАП
46. Вичерпні та евристичні підходи до попарного вирівнювання НАП.
47. Порівняйте позиційно-специфічні матриці й приховані моделі Маркова як методи опису множинних вирівнювань.
48. ДНК- і білкові логотипи – принципи побудови, та інформація, яку він містить
49. У чому полягає складність побудови множинних вирівнювань? Які є способи оцінки їхньої якості?
50. Принцип функціонування алгоритму BLAST.
51. опишіть, як зміна налаштувань програми BLAST впливає на результат пошуку гомологів?
52. Прогресивний та ітеративний принципи множинного вирівнювання
53. Які фактори ускладнюють безпосереднє використання даних про частоти появи мутацій у ДНК для коректного опису її еволюції?
54. Основні етапи побудови і використання позиційно-специфічних вагових матриць
55. Можливості й обмеження методів попарного і множинного вирівнювання для виявлення гомологічних НАП
56. Які дані містить сторінка результатів BLASTP?
57. Основні етапи пошуку оптимального локального вирівнювання за

	<p>алгоритмом Сміта-Уотермана</p> <p>58. Еволюційні засади філогенії</p> <p>59. Основні поняття у галузі філогенетичної реконструкції</p> <p>60. Яку інформацію містить філогенетичне дерево, а яку – не містить?</p> <p>61. Основні елементи філогенетичного дерева, взаємозв'язок між ними</p> <p>62. Вибір даних для філогенетичної реконструкції</p> <p>63. Що таке модель еволюції у філогенетичній реконструкції і яке її значення?</p> <p>64. Опишіть основні етапи філогенетичної реконструкції</p> <p>65. Методи статистичної оцінки філогенетичних дерев</p> <p>66. Підхід до філогенетичної реконструкції: з'єднання сусідів (NJ)</p> <p>67. Підхід до філогенетичної реконструкції: максимальної ощадності (MP)</p> <p>68. Підхід до філогенетичної реконструкції: максимальної вірогідності (ML)</p> <p>69. Чому дерево-провідник з прогресивних підходів до множинного вирівнювання не є філогенетичним деревом?</p> <p>70. У чому полягає суть курування вихідного множинного вирівнювання, що передусє філогенетичній реконструкції?</p> <p>71. На чому ґрунтується суть пошуку мотивів ДНК?</p> <p>72. Принцип функціонування алгоритму максимізації очікування (EM)?</p> <p>73. Принцип функціонування програми MEME</p> <p>74. Поясніть поняття родина, фолд, клас білка (за системою SCOP)</p> <p>75. Поясніть принципи пошуку гомологів за допомогою програм BLAST й HHPred</p> <p>76. Чому порівняння первинних і третинних структур генетичних послідовностей дає різний результат?</p> <p>77. Практичне використання філогенетичного аналізу</p> <p>78. В чому полягає принцип зважування генетичних послідовностей у моделях множинних вирівнювань?</p> <p>79. Моделі заміщення в нуклеотидних і кодонних послідовностях, їхні особливості порівняно з моделями амінокислотних заміщень</p>
<b>Опитування</b>	Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу

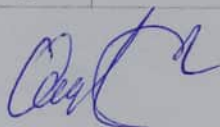
Таблиця

Схема курсу «Біоінформатика». Форма навчання - денна

Тиж-день	Тема занять (перелік питань)	Форма діяльності та обсяг годин	Додаткова література / ресурс для виконання завдань (за потреби)	Термін виконання
1	Вступ. Засади біоінформатики. Бази даних.	Лекції – 2 год. Лабораторні – 2 год. Самостійна робота – 24 год	<a href="http://rosalind.info/problems/list-view/">http://rosalind.info/problems/list-view/</a>	1 тиждень
2,3	Математичні	Лекції – 4	<a href="https://www.bv-brc.org/">https://www.bv-brc.org/</a>	2 тижні

	моделі	год. Лабораторні – 2 год. Самостійна робота – 15		
4,5,6	Попарне вирівнювання НАП	Лекції – 6 год. Лабораторні – 2 год. Самостійна робота – 15		3 тижні
7	Множинне вирівнювання НАП	Лекції – 2 год. Лабораторні – 2 год. Самостійна робота – 5		1 тиждень
8, 9	Моделі на основі множ. вирівнювань	Лекції – 4 год. Лабораторні – 2 год. Самостійна робота – 15	<a href="https://toolkit.tuebingen.mpg.de/tools/hhpred">https://toolkit.tuebingen.mpg.de/tools/hhpred</a>	2 тижні
10, 11, 12, 13	Філогенетична реконструкція	Лекції – 8 год. Лабораторні – 2 год. Самостійна робота – 38	<a href="https://phylogeny.fr">https://phylogeny.fr</a> <a href="https://evolution.berkeley.edu/evolibrary/search/topics.php?topic_id=15">https://evolution.berkeley.edu/evolibrary/search/topics.php?topic_id=15</a> .	4 тижні
14	Пошук паттернів та операторів	Лекції – 2 год.		1 тиждень
15	Бази й аналіз білкових структур	Лекції – 2 год. Лабораторні – 2 год. Самостійна робота – 2	<a href="https://meme-suite.org/meme/">https://meme-suite.org/meme/</a>	1 тиждень
16	Бази й аналіз структур РНК	Лекції – 2 год. Лабораторні – 2 год. Самостійна робота – 2		1 тиждень

Автор



Богдан ОСТАШ

"Погоджено"

Голова методичної ради  
біологічного факультету  
Віталій ГОНЧАРЕНКО

"28" "08" 2024 р.

Гарант ОПІ «Біохімія»  
Наталія СИБІРНА

28. 08. 2024 р.

Гарант ОПІ «Біофізика»  
Марта БУРА

28. 08. 2024 р.

Гарант ОПІ «Ботаніка»

Анастасія ОДІНЦОВА  
28. 08. 2024 р.

Гарант ОПІ «Генетика»

Наталія ГОЛУБ  
28. 08. 2024 р.

Гарант ОПІ «Зоологія»

Андрій БОКОТЕЙ  
28. 08. 2024 р.

Гарант ОПІ «Мікробіологія»

Світлана ГНАТУШ  
28. 08. 2024 р.

Гарант ОПІ «Фізіологія людини і тварин»

Оксана ІККЕРТ  
28. 08. 2024 р.

Гарант ОПІ «Фізіологія рослин»

Наталія РОМАНІУК  
28. 08. 2024 р.