



Пакети прикладних програм для задач молекулярної біології
Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Другий (освітній)</i>
Галузь знань	16 – Хімічна та біоінженерія
Спеціальність	162 Біотехнології та біоінженерія
Освітня програма	<i>Біотехнологія</i>
Статус дисципліни	вибіркова
Форма навчання	Денна
Рік підготовки, семестр	5 курс, весняний семестр
Обсяг дисципліни	Загальна кількість 150 год.
Семестровий контроль/ контрольні заходи	Іспит/МКР,РР
Розклад занять	<i>Згідно розкладу лекції 1 год/тиждень, практика – 1 год. тиждень</i>
Мова викладання	Українська
Інформація про керівника курсу / викладачів	Лектор: ст. викладач, Кисляк Сергій Володимирович, bmk-ksv-fbmi@iit.kpi.ua Практичні: ст. викладач, Кисляк Сергій Володимирович, bmk-ksv-fbmi@iit.kpi.ua
Розміщення курсу	https://classroom.google.com/c/Mzg4NDA4MzIxNzcx?cjc=576om7e

Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Актуальність. З появою ефективних методів секвенування біологічних послідовностей визначається вектор розвитку сучасної обчислювальної молекулярної біології, особливістю якого є аналіз великих об'ємів даних. Експоненційне накопичення молекулярно-біологічної інформації, вимагає від дослідників вирішення однієї з основних проблем біоінформатики, що пов'язана з наявністю невеликої кількості описаних біологічних послідовностей, що зберігаються у базах даних (наприклад Uniprot, Genbank), у порівнянні з тими, що потребують повного анотування. Удосконалення та оптимізація основних алгоритмів та базових методів біоінформатики, а також максимальна інтеграція математичних статистичних обчислювальних методів та сучасних інформаційних технологій, дозволить ефективно вирішувати поставлені молекулярно-біологічні задачі. Кожний етап біоінформаційного аналізу, починаючи з секвенування, асемблювання, картування, ідентифікації кодуєчих ділянок тощо, пов'язаний з застосуванням певних алгоритмів та вимагає від дослідника вміння використовувати програмні продукти та online-сервіси, що дозволяють вирішити поставлені задачі.

Метою дисципліни є формування у студентів навичок роботи з популярними програмами та сервісами, що можуть бути застосовані для вирішення різноманітних задач обчислювальної молекулярної біології, таких як: проведення ресеквенування біологічних послідовностей; вивчення молекулярних основ генетичних захворювань; побудови філогенетичних дерев; проектування лікарських препаратів, ідентифікації генів. При цьому студенти отримують детальну інформацію щодо застосованих алгоритмів з можливістю їх подальшої реалізації та можливого удосконалення (оптимізації).

Основні завдання дисципліни.

Згідно з вимогами освітньо-професійної програми здобувачі вищої освіти повинні засвоїти компетентності, якими повинен оволодіти здобувач:

- ЗК1-Здатність проведення досліджень на відповідному рівні;
- ЗК2-Здатність до пошуку, оброблення та аналізу інформації з різних джерел;
- ЗК 5 Здатність вчитися і оволодівати сучасними знаннями;
- ФК2 - Здатність здійснювати пошук необхідної інформації в науковій і технічній літературі, базах даних та інших джерелах;

ФК3 -Здатність відбирати та аналізувати релевантні дані, у тому числі за допомогою сучасних методів аналізу даних і спеціалізованого програмного забезпечення;

Згідно з вимогами програми навчальної дисципліни студенти після засвоєння кредитного модуля мають продемонструвати такі результати навчання:

знання :

- основних концепцій аналізу біологічних текстів;
- методів оцінки якості даних секвенування;
- алгоритмів асемблювання геномів та оцінки якості отриманої збірки геному;
- молекулярної організації геномів та методів (алгоритмів) їх порівняльного аналізу;
- особливостей проведення ресеквенування геномів прокариот;
- методів розрахунку генетичних дистанцій;
- методів побудови філогенетичних дерев;
- алгоритмів пошуку генів,
- основних принципів та підходів щодо створення сучасних лікарських препаратів

уміння:

- використовувати програмні пакети та веб-сервіси при проведенні ресеквенування біологічних послідовностей;
- отримувати множинне вирівнювання, визначати еволюційну відстань між послідовностями та отримувати ультраметричні та адитивні філогенетичні дерева, проводити бутстреп аналіз;
- використовувати програмні пакети та веб-сервіси з метою побудови філогенетичних дерев для отримання інформації щодо динаміки зараження на коронавірус (Sars-Cov-2) по різних країнам світу ;
- отримання навичок роботи з сучасними молекулярно-біологічними БД;
- працювати з геномними браузерями;
- проводити аналіз геномів еукаріот та визначати мутації, що призводять до генетичних захворювань;
- використовувати технології автоматизованого проектування лікарських препаратів in silico;

2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі)

Дисципліну в структурно-логічній схемі забезпечують такі дисципліни як «Інформаційні технології» та «Прикладна біоінформатика». Засвоєння дисципліни «Пакети прикладних програм для задач молекулярної біології» забезпечує успішне засвоєння курсу «Біоінформаційні бази даних».

3. Зміст навчальної дисципліни

Вступна частина.

Розділ 1 Аналіз біологічних послідовностей

Тема 1.1 Методи секвенування біологічних послідовностей

Тема 1.2 Алгоритми асемблювання біологічних послідовностей. Огляд сучасних геномних асемблерів.

Тема 1.3 Ресеквенування біологічних послідовностей.

Тема 1.4 Алгоритми вирівнювання на референсний геном.

Тема 1.5 Алгоритми та веб-сервіси пошуку генів про- та еукаріот.

Розділ 2 Молекулярна філогенетика

Тема 2.1 Філогенетичний аналіз біологічних послідовностей.

Тема 2.2 Методи визначення еволюційних дистанцій між амінокислотними та нуклеотидними послідовностями.

Тема 2.3 Дистанційні та символно-орієнтовані методи побудови ультраметричних та адитивних філогенетичних дерев.

Тема 3 Сучасні методи пошуку лікарських засобів.

Тема 3.1 Технології автоматизованого проектування лікарських препаратів *in silico*.

4. Навчальні матеріали та ресурси

Базова література:

1. Кисляк С.В., Настенок Є.А. Основи молекулярної біології та біоінформатики [Електронний ресурс] : навч. посіб. для студ. спеціальності 122 «Комп'ютерні науки та інформаційні технології спеціалізації «Інформаційні технології в біології та медицині» / С.В.Кисляк, Є.А.Настенок; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл, 2957 Кбайт). – Київ : КПІ ім. Ігоря Сікорського, 2018. – 95 с. <https://ela.kpi.ua/bitstream/123456789/27529/1/molbiolbioinformatics.pdf>
2. 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.
3. Огурцов А.Н. Основы биоинформатики: учеб.пособие.– Харьков: НТУ "ХПИ", 2013.- 400 с.
4. Ней М., Кумар С. Молекулярная эволюция и филогенетика. Киев: КВЦ, 2004. — 418 с.
5. Д.В. Ребриков, Д.О. Коростин, Е.С.Шубина, В.В. Ильинский. NGS: высокопроизводительное секвенирование.-2-е изд.-М.:БИНОМ. Лаборатория знаний, 2015.- 232с.
6. Хельтье Х.-Д. Молекулярное моделирование: теория и практика/ Х.-Д. Хельтье, В.Зиппль, Д.Роньян, Г.Фолькерс; пер.с англ.-2-е изд.-М.:БИНОМ. Лаборатория знаний, 2013.-319с.

Додаткова:

1. Хаубольд Б., Вие Т. Введение в вычислительную биологию: Эволюционный подход. Introduction To Computational Biology: An Evolutionary Approach. — Ижевск : РХД, 2011. — 456 с
2. Гасфилд, Дэн. Строки, деревья и последовательности в алгоритмах : Информатика и вычисл. биология / Дэн Гасфилд ; Пер. с англ. И.В. Романовского. - СПб. : Невский диалект : БХВ-Петербург, 2003 - 653 с.

Інформаційні ресурси

1. <https://biomolecula.ru/>
2. https://www.youtube.com/playlist?list=PL-cKNuVAYAVsD-Ee5BSVCqpqkUXpkK_r
3. <https://ru.coursera.org/>
4. <http://ugene.net/ru/>
5. <https://www.ncbi.nlm.nih.gov/>
6. <https://www.bioinformatics.nl/emboss-explorer/>
7. <https://www.megasoftware.net/docs>
8. <http://exon.gatech.edu/GeneMark/>
9. <http://www.mathnet.ru/links/3daa19de3270eae061d3ba639f490d7c/kvant1757.pdf>
10. <https://www.illumina.com/>
11. <https://www.pacb.com/smrt-science/smrt-sequencing/>
12. <https://www.youtube.com/watch?v=sv9fFeSd3kE>

5. Методика опанування навчальної дисципліни (освітнього компонента)

№	Назва лекції та перелік основних питань
1	<p>Вступна частина. Предмет, цілі та задачі основні розділи біоінформатики. Історія розвитку біоінформатики. Предмет біоінформатики. Аналіз біологічних текстів. Основні розділи біоінформатики: структурна та функціональна геноміка, молекулярна філогенетика, клінічна, структурна та функціональна протеоміка.</p> <p>Розділ 1 Секвенування біологічних послідовностей</p> <p>Лекція 1. Методи секвенування біологічних послідовностей. Платформи для секвенування Illumina, Ion Torrent. Методи секвенування одиничних молекул Oxford Nanopore, Pacific biosciences.</p> <p>Література: базова [4], додаткова [1], інформаційні джерела [1,10-12]</p>
2	<p>Лекція 2. Алгоритми асемблювання біологічних послідовностей. Огляд сучасних асемблерів. Euler, Velvet, ALLPATHS, ABySS, SPAdes. Особливості асемблювання ридів на етапах: попереднього виправлення помилок, побудові графа де Брюїна; корекції графу; дозволу повторів; отримання контигів та скефолдів. ДНК-чипи. Гамільтонов та Ейлеров цикл. Алгоритми асемблювання ридів на основі графа де-Брюїна. Оцінка якості збірки геному за допомогою QUAST</p> <p>Література: базова [4], додаткова [1], інформаційні джерела [1,9]</p>
3	<p>Лекція 3. Ресеквенування біологічних послідовностей. Оцінка якості даних секвенування. Формат Fastq. Формати SAM, BAM. Алгоритми пошуку референсного геному для проведення ресеквенування. Сервіси BLAST та RNAmmer. Класичні алгоритми пошуку підрядку в рядку.</p> <p>Література: базова [1], додаткова [1-2], інформаційні джерела [5]</p>
4	<p>Лекція 4. Алгоритми вирівнювання на референсний геном. Вирівнювання біологічних послідовностей з афінним штрафом у контексті картування ридів. Перетворення Барроуза-Уїллера. Модифікований алгоритм вирівнювання, що враховує повтори. Вирівнювання з лінійною пам'яттю. Алгоритм Міллера-Майерса. Огляд пакету програм Ugene та Emboss.</p> <p>Література: базова [2-3], додаткова [1-3], інформаційні джерела [2-4]</p>
5	<p>Лекція 5. Профілі послідовностей та приховані марковські моделі. Алгоритм <i>Вітербі</i>.</p> <p>Література: базова [2], додаткова [1], інформаційні джерела [2,6,8]</p>
6	<p>Лекція 6. Forward-backward алгоритм. Апостеріорне декодування. Веб-сервіси пошуку генів у про- та еукаріот. GeneMarc та GeneScan. Ідентифікація CpG острівців за допомогою пакету програм Emboss.</p> <p>Література: базова [2], додаткова [1], інформаційні джерела [2,6,8]</p>
7	<p>Розділ 2 Молекулярна філогенетика</p> <p>Лекція 7. Молекулярні основи філогенетики. Еволюція нуклеотидних та амінокислотних послідовностей. Визначення еволюційних дистанцій між нуклеотидними та амінокислотними послідовностями. Пакет програм для проведення філогенетичного аналізу Mega</p> <p>Література: базова [3], додаткова [1], інформаційні джерела [1,7]</p>
8	<p>Лекція 8. Дистанційні та символно-орієнтовані методи побудови ультраметричних та адитивних філогенетичних дерев. Методи UPGMA, приєднання сусідів, Фітча-Марголіаша. Метод максимальної економії та найбільшої правдоподібності. Бутстреп-аналіз.</p> <p>Література: базова [3], додаткова [1], інформаційні джерела [1,7]</p>
9	<p>Лекція 9. Технології автоматизованого проектування лікарських препаратів in silico. Метод QSAR дослідження кількісних співвідношень «просторова структура - активність» ліганду, за умови невідомої просторової структури рецептору. Віртуальний</p>

скринінг та докінг. Алгоритми докінга біополімерів.
Література: базова [6]

Практичні заняття

Основні завдання циклу практичних занять:

- працювати з програмами та веб-сервісами для вирішення різноманітних задач молекулярної біології;
- орієнтуватись в базових алгоритмах геноміки;

1.	Практична робота 1 Оцінка та покращення якості даних секвенування. Програма FastQC, Trimmomatic. Література: базова [1,5], інформаційні ресурси [2,3].
2.	Практична робота 2 Асемблювання геному за допомогою Spades асемблера. Алгоритми асемблювання геномів. Оцінка якості отриманої збірки геному за допомогою Quast. Література: базова [1,5], інформаційні джерела [2,3]
3.	Практична робота 3 Пошук гомологічних послідовностей за допомогою Basic Local Alignment Search Tool. Перша частина практичної роботи пов'язана з пошуком референсного генома. Друга – з визначенням якості харчових продуктів. Література: базова [1,3], інформаційні джерела [1-3]
4.	Практична робота 4 Ресеквенування біологічних послідовностей Література: базова [1], інформаційні джерела [1-3]
5.	Практична робота 5 Пошук кодуєчих ділянок геному патогенного штаму E.coli Література: базова [1,3], інформаційні джерела [1-3,8]
6.	Практична робота 6 Філогенетичний аналіз біологічних послідовностей. Множинне вирівнювання філогенетичних маркерів. Розрахунок еволюційних дистанцій між амінокислотними послідовностями. Методи побудови філогенетичних дерев (UPGMA, Neighbor joining, Maximum likelihood). Статистична оцінка філогенетичного дерева, бутстреп-аналіз. Література: базова [4], додаткова [1] інформаційні джерела [2,7]
7.	Практична робота 7 Молекулярні основи генетичних захворювань Література: базова [1,2], інформаційні ресурси [2].
8.	Практична робота 8 Технології автоматизованого проектування лікарських препаратів in silico. Література: базова [1-3]
9.	Практична робота 9 МКР

6. Самостійна робота студента

Самостійна робота студента (33 години) по дисципліні включає підготовку до модульної контрольної роботи (4 години), виконання РР (15 годин), підготовка до іспиту (30 годин) та самостійне вивчення тем, перелік яких наводиться нижче (29 годин)

№	Назви тем і питань, що виносяться на самостійне опрацювання, та посилання на навчальну літературу	Кількість годин СРС
1.	Виконання розрахункової роботи: 1. Побудувати адитивні та ультраметричні філогенетичні дерева для просеквенованих геномів Sars-Cov-2 у різних країнах світу. 2. Відповідно до проведеного бутстреп тесту визначити найбільш вірогідну топологію дерева. Проаналізувати отриману інформацію та зробити висновок щодо динаміки зараження на коронавірус (Sars-Cov-2) по різних країнам світу. 3. Визначити країни, в яких Sars-Cov-2 потрапив з декількох різних джерел (країн) інфекції.	15
	<i>Теми:</i>	
2.	Методи секвенування наступного покоління Roche 454, Solid. Метод секвенування одиничних молекул Helicos. Література: базова [5], інформаційні ресурси [2,9]	5
3.	Алгоритми асемблювання ридів Сангера. Асемблер Selera Література: базова [5], інформаційні ресурси [2,9]	5
4.	Евристичний алгоритм множинного вирівнювання біологічних послідовностей ClustalW. Ітеративний алгоритм Muscle. Література: базова [2], інформаційні джерела [2]	7
5.	Методи побудови філогенетичних дерев Maximum parsimony, Minimum evolution Література: базова [3], додаткова [1], інформаційні джерела [1,7]	7
6.	Парне вирівнювання послідовностей за допомогою прихованих марковських моделей.	5
7.	<i>МКР</i>	4
8.	<i>Іспит</i>	30

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Правила відвідування занять. Відвідування лекцій, практичних занять, а також відсутність на них, не оцінюється. Студентам рекомендується відвідувати заняття, оскільки на них викладається теоретичний матеріал та розвиваються навички, необхідні для формування компетентностей, визначених стандартом освіти. Система оцінювання орієнтована на отримання балів за активність студента, а також виконання завдань, які здатні розвинути практичні уміння та навички. За об'єктивних причин (наприклад, хвороба, працевлаштування, міжнародне стажування тощо) навчання може відбутися у додатковий час за погодженням із керівником курсу. Захист індивідуальних завдань проводиться на практичних заняттях.

Політика щодо дедлайнів та перескладання. Роботи, які здаються із порушенням термінів без поважних причин, оцінюються з пониженням рейтингової оцінки. Викладачем визначаються дедлайни виконання розрахункової роботи та захистів 8 практичних робіт. Політика та принципи академічної доброчесності визначені у розділі 3 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше можна ознайомитись за посиланням: <https://kpi.ua/code>.

Правила поведінки на заняттях. Норми етичної поведінки студентів і працівників визначені у розділі 2 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше за посиланням: <https://kpi.ua/code>. Використання додаткових джерел інформації під час оцінювання знань заборонено (у т.ч. мобільних пристроїв).

8. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

Поточний контроль: виконання практичних робіт (24бали), виконання розрахункової роботи (20 балів) , МКР (16 балів). Докладніша інформація щодо поточного контролю та критеріїв оцінювання наведена в PCO з дисципліни (додаток 2)

Календарний контроль: проводиться двічі на семестр як моніторинг поточного стану виконання вимог силабусу.

Семестровий контроль: іспит. Загальна сума балів на іспиті – 40 балів. Докладніша інформація щодо проведення та оцінювання наведена в PCO з дисципліни (додаток 1).

Умови допуску до семестрового контролю: зарахування усіх практичних робіт, написання МКР, виконання та захист РР та семестровий рейтинг $R_c > 40$. Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

<i>Кількість балів</i>	<i>Оцінка</i>
100-95	Відмінно
94-85	Дуже добре
84-75	Добре
74-65	Задовільно
64-60	Достатньо
Менше 60	Незадовільно
Не виконані умови допуску	Не допущено

Робочу програму навчальної дисципліни «Пакети прикладних програм для задач молекулярної біології» складено старшим викладачем кафедри біоенергетики, біоінформатики та екобіотехнології Кисляком С. В.

Ухвалено кафедрою біоенергетики, біоінформатики та екобіотехнології (протокол №__від

Погоджено методичною комісією факультету (протокол № __ від

)

Рейтингова система оцінювання результатів навчання

Рейтинг студента складається з балів, що він отримує за:

- 1) виконання завдань на 8 практичних заняттях;
- 2) 1 модульну контрольну роботу;
- 3) виконання та захист РР.

Система рейтингових (вагових) балів занять і рейтингових оцінок за видами контролю

№	Вид контролю	Бал	Кількість	Сума балів
1	Завдання на практичному занятті	3	8	24
	ваговий бал			
	якість виконання та захист ¹	0-3		
2	Модульна контрольна робота	16	1	16
	ваговий бал			
	якість виконання та захист ²	0-16		
3	РГР	20	1	20
	Якість виконання ³	0-20		
	Всього			60

¹ - Якість виконання завдання на практичному занятті:

- завдання виконане повністю, звіт надано своєчасно - 3 бали;
- робота містить несуттєві помилки або не повністю виконане завдання, звіт надано своєчасно - 2 бали;
- робота не зарахована, звіт надано не своєчасно з суттєвими помилками - 0 балів.

² - Якість виконання контрольної роботи:

- повна розкрита відповідь - 12 - 16 балів;
- помилка в одному завданні або неповна відповідь в двох завданнях - 10 - 12 балів;
- помилка в одному завданні або неповна відповідь на усі завдання - 8 - 10 балів;
- помилка в двох завданнях - 6 - 8 балів;
- робота не зарахована - 0 балів.

³ - Якість виконання РР та захист:

- захист роботи (захист включає знання з лекційного матеріалу) - 5 балів;
- правильно виконана робота - 12 - 15 балів;
- робота виконана з помилками в одному завданні - 10 - 12 балів;
- робота виконана з помилками у двох завданнях - 8 - 10 балів;
- робота не зарахована - 0 балів.

Штрафні бали:

№		
1	Несвоєчасна здача виконаного завдання з практичного заняття	-1 б

Сума як штрафних балів не має перевищувати 0,1 R_c = 60 балів x 0,1 = 6 балів

Розрахунок шкали (R) рейтингу

Сума вагових балів контрольних заходів протягом семестру складає:

$$R_c = 20 + 16 + 24 = 60 \text{ балів}$$

Форма атестації передбачена у вигляді іспиту з сумарною кількістю балів 40, тому загальна рейтингова шкала з дисципліни складає: $R_c = 60 + 40 = 100$ балів. Необхідною умовою допуску до семестрової атестації є зарахування всіх практичних робіт, написання МКР, виконання та захист РР та семестровий рейтинг $R_c > 40$.

Екзаменаційна робота (Виходячи з розміру шкали $R_E = 40$ балів)

- практична задача – ваговий бал 16;
- 3 теоретичних завдання – ваговий бал 8.

Максимальна кількість балів за екзаменаційну роботу дорівнює $8 \text{ балів} \times 3 \text{ тер питання} + 16 \times 1 \text{ практичне завдання} = 40$ балів.

Критерій оцінювання теоретичного завдання

«Відмінно», відповідь правильна (не менше 90% потрібної інформації)	- 8 балів
«Добре», є несуттєві помилки у відповіді (не менше 75% потрібної інформації)	- 6-7 балів
«Задовільно», є недоліки у відповіді та помилки (не менше 60% потрібної інформації).	- 5 балів
«Незадовільно», відповідь відсутня	- 0 балів

Критерій оцінювання практичної задачі

«Відмінно», відповідь правильна (не менше 90% потрібної інформації)	- 16-15 балів
«Добре», є несуттєві помилки у відповіді (не менше 75% потрібної інформації)	- 14-12 балів
«Задовільно», є недоліки у відповіді та певні помилки (не менше 60% потрібної інформації).	- 11-10 балів
«Незадовільно», відповідь відсутня або не відповідає вимогам до «Задовільно»	- 0 балів

Для отримання студентом відповідних оцінок (ECTS та традиційних) його рейтингова оцінка **RD** розраховується згідно з таблицею:

RD = $r_c + r_e$	Оцінка ECTS	Традиційна оцінка
95...100	A	відмінно
85...94	B	добре
75...84	C	
65...74	D	
64...60	E	задовільно
$RD \leq 60$	Fx	не задовільно
$r_c < 30$ або не виконані інші умови допуску до заліку	F	не допущений