



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 інженерія та біоінженерія
Спеціальність	162 Біотехнології та біоінженерія
Освітня програма	<i>Біотехнології</i>
Статус дисципліни	вибіркова
Форма навчання	Заочна
Рік підготовки, семестр	5 курс, весняний семестр
Обсяг дисципліни	Загальна кількість 150 год.
Семестровий контроль/ контрольні заходи	Іспит/МКР, ДКР
Розклад занять	<i>Згідно розкладу лекції 10 год, практика – 8 год.</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 програмні результати навчання:

- вміти обирати та застосовувати найбільш придатні методи моделювання та оптимізації при розробленні науково -технічних проєктів;
- вміти використовувати молекулярно-біологічні технології для створення та аналізу нових біологічних агентів;
- знати молекулярну організацію та регуляцію експресії генів, реплікації, рекомбінації та раперації, рестрикції та модифікації генетичного матеріалу у про- та еукаріот, стратегію створення рекомбінантних ДНК для цілеспрямованого конструювання біологічних агентів;
- вміти використовувати методи молекулярної біоінженерії для створення нових біологічних агентів.

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

Дисципліну в структурно-логічній схемі забезпечують такі дисципліни як «Інформаційні технології» та «Прикладна біоінформатика».

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., Настенко 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.Настенко; 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. Кисляк С.В., Голуб Н.Б., Дуган О.М., Аверьянова О.А. Моделювання молекулярної взаємодії [Електронний ресурс] : підручник для студ. спеціальності 162 «Біотехнології та біоінженерія» / С.В.Кисляк ; КПІ ім. Ігоря Сікорського. – Електронні текстові данні (1 файл 27167 Кбайт). – Київ: КПІ ім. Ігоря Сікорського, 2023 – 203с. <https://ela.kpi.ua/handle/123456789/55665>
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.

Інформаційні ресурси

1. <https://ru.coursera.org/>
2. <http://ugene.net/ru/>
3. <https://www.ncbi.nlm.nih.gov/>
4. <https://www.bioinformatics.nl/emboss-explorer/>
5. <https://www.megasoftware.net/docs>
6. <http://exon.gatech.edu/GeneMark/>
7. <https://www.illumina.com/>
8. <https://www.pacb.com/smrt-science/smrt-sequencing/>

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. Література: базова [3], інформаційні джерела [1]</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. 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 Література: базова [3], інформаційні джерела [1]</p>
2	<p>Лекція 2. 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. Література: базова [3], інформаційні джерела [1]</p>
3	<p>Лекція 3. 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. Профілі послідовностей та приховані марковські моделі. Алгоритм <i>Viterbi</i>. Forward-backward алгоритм. Апостеріорне декодування. Веб-сервіси пошуку генів у про- та еукаріот. GeneMarc та GeneScan. Ідентифікація CpG острівців за допомогою пакету програм Emboss. Література: базова [1,2,3], інформаційні джерела [2,4,5,6,8]</p>
4	<p>Розділ 2 Молекулярна філогенетика</p> <p>Лекція 4. Молекулярні основи філогенетики. Еволюція нуклеотидних та амінокислотних послідовностей. Визначення еволюційних дистанцій між нуклеотидними та амінокислотними послідовностями. Пакет програм для проведення філогенетичного аналізу Mega Література: базова [3], інформаційні джерела [1,7]</p>
5	<p>Лекція 5. 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. Література: базова [3], інформаційні джерела [1,7]</p>

Практичні заняття

Основні завдання циклу практичних занять:

- отримати навички роботи в терміналі операційної системи BioLinux
- 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. Асемблювання геному за допомогою Spades асемблера. Алгоритми асемблювання геномів. Оцінка якості отриманої збірки геному за допомогою Quast. Веб-сервіс Galaxy. Література: базова [1,5], інформаційні ресурси [2,3].
2.	Практична робота 2 Пошук гомологічних послідовностей за допомогою Basic Local Alignment Search Tool. Ресеквенування біологічних послідовностей. Література: базова [1,5], інформаційні джерела [2,3] Практична робота 3 Пошук кодуючих ділянок геному патогенного штаму E.coli Література: базова [1,3], інформаційні джерела [1-3]
3.	Практична робота 4 Філогенетичний аналіз біологічних послідовностей. Множинне вирівнювання філогенетичних маркерів. Кладограма, ультраметричне та адитивне філогенетичне дерево. Визначення нетривіальних гілок філогенетичного дерева. Newick формат. Топологія, формати запису та різновиди філогенетичних дерев. Дистанційні методи побудови ультра метричних філогенетичних дерев. Метод. UPGMA, метод трансформованої дистанції. Визначення довжин гілок філогенетичного дерева. Особливості побудови ультраметричного філогенетичного дерева. Література: базова [1,4], інформаційні джерела [1,2,7]
4.	МКР

6. Самостійна робота студента

Самостійна робота студента (132 години) по дисципліні включає підготовку до модульної контрольної роботи (4 години), виконання ДКР (10 годин), підготовку до аудиторних занять (16 годин), підготовку до іспиту (30 годин) та самостійне вивчення тем (72 годин)

№	Назви тем і питань, що виносяться на самостійне опрацювання, та посилання на навчальну літературу	Кількість годин СРС
1.	Виконання домашньої контрольної роботи: 1. За допомогою мережі Internet провести детальний аналіз літературних джерел щодо алгоритмів передбачення просторової структури РНК. Детально розглянути алгоритм Нусинів. 2. Для фрагменту послідовності, що кодує тРНК 10 нуклеотидів (для кожного студента повинна бути знайдена своя тРНК в геномному браузері ENSEMBL) за допомогою алгоритму Нусинів отримати вторинну структуру фрагменту тРНК. 3. При виконанні процедури зворотнього проходу необхідно врахувати хоча б одну додаткову альтернативну вторинну структуру, яка стубілізується за рахунок максимальної кількості водневих зв'язків.	10
	<i>Теми:</i>	
2.	Методи секвенування наступного покоління Roche 454, Solid. Метод секвенування одиничних молекул Helicos. Література: базова [5], інформаційні ресурси [2,9]	8
3.	Алгоритми асемблювання ридів Сангера. Асемблер Selera Література: базова [5], інформаційні ресурси [2,9]	8
4	Евристичний алгоритм множинного вирівнювання біологічних послідовностей ClustalW. Ітеративний алгоритм Muscle. Література: базова [2], інформаційні джерела [2]	8

5.	Дистанційні методи побудови адитивних філогенетичних дерев. Метод приєднання сусідів, метод мінімуму еволюції. Метод найменших квадратів. Особливості побудови адитивних філогенетичних дерев. Проведення тесту на адитивність. Література: базова [4-5], інформаційні джерела [2,7]	10
6	Дистанційні методи побудови адитивних філогенетичних дерев. Метод приєднання сусідів, метод мінімуму еволюції. Метод найменших квадратів. Особливості побудови адитивних філогенетичних дерев. Проведення тесту на адитивність. Література: базова [4-5], інформаційні джерела [2,7]	10
7	Методи побудови філогенетичних дерев Maximum parsimony, Minimum evolution Література: базова [3], додаткова [1], інформаційні джерела [1,7]	9
8	Парне вирівнювання біологічних послідовностей за допомогою прихованих марковських моделей. Література: базова [4]	9
9	Методи передбачення вторинної структури РНК. Алгоритми передбачення вторинної структури РНК, алгоритм Зукера Література: базова [4-5], інформаційні джерела [2,7]	10

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 на 0.5 балів. Викладачем визначаються дедлайни виконання ДКР та захистів 4 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 (36 балів), виконання ДКР (9 балів) , МКР (15 балів). Докладніша інформація щодо поточного контролю та критеріїв оцінювання наведена в PCO з дисципліни (додаток 2)

Календарний контроль: не проводиться.

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 РГР 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 (силабус):

Складено старшим викладачем кафедри біоенергетики, біоінформатики та екобіотехнології Кисляком С. В.

Ухвалено кафедрою біоенергетики, біоінформатики та екобіотехнології (протокол №18 від 25.05.2023).

Погоджено Методичною комісією факультету (протокол № 11 від 26.06.2023)

Рейтингова система оцінювання результатів навчання

Рейтинг студента складається з балів, що він отримує за:

- 1) виконання завдань на 4 практичних заняттях;
- 2) 1 модульну контрольну роботу;
- 3) виконання та захист ДКР.

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

№	Вид контролю	Бал	Кількість	Сума балів
1	Завдання на практичному занятті	9	4	36
	ваговий бал			
	якість виконання та захист ¹	0-2		
2	Модульна контрольна робота	15	1	15
	ваговий бал			
	якість виконання та захист ²	0-15		
3	ДКР	9	1	9
	Якість виконання ³	0-9		
	Всього			60

¹ – Якість виконання завдання на практичному занятті:

- завдання виконане повністю, звіт надано своєчасно - 9 балів;
- робота містить несуттєві помилки або не повністю виконане завдання, звіт надано своєчасно - 7-8 балів;
- робота містить суттєві помилки або не повністю виконане завдання, звіт надано своєчасно – 5-6 балів;
- робота не зарахована, звіт надано не своєчасно з суттєвими помилками - 0 балів.

² – Якість виконання контрольної роботи:

- повна розкрита відповідь - 12 - 15 балів;
- помилка в одному завданні або неповна відповідь в двох завданнях - 10 - 12 балів;
- помилка в одному завданні або неповна відповідь на усі завдання - 9 - 10 балів;
- робота не зарахована - 0 - 8 балів;

³ – Якість виконання ДКР та захист:

- захист роботи (захист включає знання з лекційного матеріалу) - 3 бали;
- правильно виконана робота – 6 балів;
- робота виконана з помилками в одному завданні - 4 - 5 балів;
- робота виконана з помилками у двох завданнях - 3 бали;
- робота не може бути зарахована, якщо сумарна кількість балів з урахуванням захисту та виконання завдань ДКР не перевищує 3 балів.

Штрафні бали:

№		
1	Несвоєчасна здача виконаного завдання з практичного заняття	-0.5 б

Розрахунок шкали (R) рейтингу

Сума вагових балів контрольних заходів протягом семестру складає:

$$R_c = 36 + 15 + 9 = 60 \text{ балів}$$

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 ДКР 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 балів x 3 тер питання + 16 x 1 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

Питання до модульної контрольної роботи

1. Мутації. Основні види точкових мутацій. Транзиції та трансверсії. Не синонімічні та синонімічні заміни.
2. Особливості еволюції нуклеотидних послідовностей (первинні, вторинні та зворотні заміни, характеристика консервативних ділянок, конвергенція та дивергенція)
3. Концепція молекулярного годинника та нейтральна теорія еволюції.
4. Характеристика основних етапів філогенетичного аналізу біологічних послідовностей.

5. Множинне вирівнювання біологічних послідовностей. Алгоритм прогресивного множинного вирівнювання. Алгоритм Clustal в модифікації W. Оцінка якості вирівняних послідовностей. Ітеративні алгоритми множинного вирівнювання біологічних послідовностей. Алгоритм Muscle.
6. Опис консенсусного рядка програми у множинному вирівнюванні. Дати визначення: домен, мотив, патерн та сайт білкової послідовності. Правила створення запитів у БД Prosite.
7. Спостережені, істинні та розрахункові еволюційні дистанції між послідовностями. Р-дистанція. Обмеження та недоліки використання Р-дистанції.
8. Визначення генетичних відстаней між нуклеотидними послідовностями. Модель Джукса-Кантора, Кімури, Таджими-Неї.
9. Визначення еволюційних дистанцій між амінокислотними послідовностями.
10. Філогенетичні дерева. Топологія, види дерев, ультраметричні та адитивні філогенетичні дерева.
11. Методи побудови філогенетичних дерев. Загальна характеристика дистанційних методів побудови дерев філогенезу.
12. Метод неваженого попарного групування із середнім арифметичним (UPGMA).
13. Метод трансформованої дистанції.
14. Метод приєднання сусідів (Neighbor-Joining).
15. Метод Фітча – Марголіаша
16. Метод максимальної економії (Maximum parsimony)
17. Статистична оцінка достовірності вузлів філогенетичного дерева. Бутстреп-аналіз.
18. Основні елементи вторинної структури РНК. Алгоритми передбачення вторинної структури РНК. Алгоритм Нуссінов. Алгоритм Зукера.
19. Технологія секвенування першого покоління. Метод Сангера.
20. Ампліфікація при секвенуванні біологічних послідовностей. Етапи проведення, температурний профіль ПЦР. Застосування ПЦР в медицині.
21. Метод Illumina. Особливості проведення мостової ампліфікації (bridge amplification). Переваги та недоліки секвенування за допомогою Illumina. Чому рідки Illumina мають довжину не більше 300 нуклеотидів?
22. Особливості секвенування одиничних молекул у реальному часі (SMRT). Платформи Pacific Biosciences та Oxford Nanopore їх переваги та недоліки.
23. Оцінка якості рідів. Шкала якості Phred 33 та Phred 64. Формат FASTQ. Контроль якості даних та їх покращення.
24. Особливості асемблювання (збірки) геномів про- та еукаріот. Ейлерів та Гамільтонів цикли в графах. Для 5-6 к-мерів довільної довжини побудувати граф OLC (Overlap Layout Consensus) та граф де Брейна та відновити послідовність нуклеотидів фрагменту геному.
25. Особливості побудови парного графу де Брейна.
26. Алгоритм реалізації переходу від циклічних до лінійних послідовностей на етапі збірки «reads».
27. Вирівнювання біологічних послідовностей: глобальне, локальне, точкова матриця Dot plot.
28. Міра подібності біологічних послідовностей. Відстань Хеммінга та Левенштайна. Алгоритм Вагнера-Фішера.
29. Операції редагування. Види операцій редагування та штрафів за видалення (делеції).
30. Схеми оцінок нуклеотидних та амінокислотних замінів. Матриці PAM та BLOSUM.

31. Алгоритм глобального вирівнювання Нідмана – Вунша. Алгоритм локального вирівнювання.
32. Модифіковані алгоритми парного вирівнювання.
33. Парне вирівнювання біологічних послідовностей за допомогою прихованих марковських моделей.

Перелік питань, які виносяться на семестровий контроль:

1. Вирішення задачі пошуку підрядка у рядку для задач біоінформатики. Алгоритми пошуку підрядка у рядку. Наївний алгоритм пошуку.
2. Алгоритми пошуку підрядка у рядку. Основний препроцесинг підрядка.
3. Z-функція. Обчислення Z-функції за лінійний час. Z-алгоритм.
4. Алгоритм Кнута-Морріса-Пратта. Префікс функція. Обчислення префікс функції за лінійний час.
5. Алгоритм Бойєра-Мура. Препроцесинг для правила поганого символу та правила гарного суфікса.
6. Використання дерев ключових слів (префіксне дерево, бір) для вирішення задачі пошуку підрядка в рядку. Алгоритм Ахо-Карасік.
7. Загальна характеристика суфіксних дерев. Розв'язання задачі пошуку найбільшого загального підрядка для узагальненого суфіксного дерева. Визначення повторів та унікальних підрядків за допомогою суфіксних дерев та суфіксних масивів.
8. Мутації. Основні види точкових мутацій. Транзиції та трансверсії. Не синонімічні та синонімічні заміни.
9. Особливості еволюції нуклеотидних послідовностей (первинні, вторинні та зворотні заміни, характеристика консервативних ділянок, конвергенція та дивергенція)
10. Концепція молекулярного годинника та нейтральна теорія еволюції.
11. Характеристика основних етапів філогенетичного аналізу біологічних послідовностей.
12. Множинне вирівнювання біологічних послідовностей. Алгоритм прогресивного множинного вирівнювання. Алгоритм Clustal в модифікації W. Оцінка якості вирівняних послідовностей. Ітеративні алгоритми множинного вирівнювання біологічних послідовностей. Алгоритм Muscle.
13. Опис консенсусного рядка програми у множинному вирівнюванні. Дати визначення: домен, мотив, патерн та сайт білкової послідовності. Правила створення запитів у БД Prosite.
14. Спостережені, істинні та розрахункові еволюційні дистанції між послідовностями. P-дистанція. Обмеження та недоліки використання P-дистанції.
15. Визначення генетичних відстаней між нуклеотидними послідовностями. Модель Джукса-Кантора, Кімури, Таджими-Неї.
16. Визначення еволюційних дистанцій між амінокислотними послідовностями.
17. Філогенетичні дерева. Топологія, види дерев, ультраметричні та адитивні філогенетичні дерева.
18. Методи побудови філогенетичних дерев. Загальна характеристика дистанційних методів побудови дерев філогенезу.
19. Метод неваженого попарного групування із середнім арифметичним (UPGMA).
20. Метод трансформованої дистанції.
21. Метод приєднання сусідів (Neighbor-Joining).
22. Метод Фітча – Марголіаша
23. Метод максимальної економії (Maximum parsimony)
24. Статистична оцінка достовірності вузлів філогенетичного дерева. Бутстреп-аналіз.

25. Основні елементи вторинної структури РНК. Алгоритми передбачення вторинної структури РНК. Алгоритм Нуссінов. Алгоритм Зукера.
26. Технологія секвенування першого покоління. Метод Сангера.
27. Ампліфікація при секвенуванні біологічних послідовностей. Етапи проведення, температурний профіль ПЦР. Застосування ПЦР в медицині.
28. Метод Illumina. Особливості проведення мостової ампліфікації (bridge amplification). Переваги та недоліки секвенування за допомогою Illumina. Чому риди Illumina мають довжину не більше 300 нуклеотидів?
29. Особливості секвенування одиничних молекул у реальному часі (SMRT). Платформи Pacific Biosciences та Oxford Nanopore їх переваги та недоліки.
30. Оцінка якості рідів. Шкала якості Phred 33 та Phred 64. Формат FASTQ. Контроль якості даних та їх покращення.
31. Особливості асемблювання (збірки) геномів про- та еукаріот. Ейлерів та Гамільтонів цикли в графах. Для 5-6 к-мерів довільної довжини побудувати граф OLC (Overlap Layout Consensus) та граф де Брейна та відновити послідовність нуклеотидів фрагменту геному.
32. Особливості побудови парного графу де Брейна.
33. Алгоритм реалізації переходу від циклічних до лінійних послідовностей на етапі збірки «reads».
34. Вирівнювання біологічних послідовностей: глобальне, локальне, точкова матриця Dot plot.
35. Міра подібності біологічних послідовностей. Відстань Хеммінга та Левенштайна. Алгоритм Вагнера-Фішера.
36. Операції редагування. Види операцій редагування та штрафів за видалення (делеції).
37. Схеми оцінок нуклеотидних та амінокислотних замін. Матриці PAM та BLOSUM.
38. Алгоритм глобального вирівнювання Нідлмана – Вунша. Алгоритм локального вирівнювання.
39. Модифіковані алгоритми парного вирівнювання.
40. Парне вирівнювання біологічних послідовностей за допомогою прихованих марковських моделей.