

# Биоинформатика, ЛМШ-2025

Вопросы к зачёту

## Регламент

Для записи на зачёт вечером накануне будет **таблица в чате**. Приходите через 15-минутные интервалы, чтобы не терять много времени на ожидание. В аудитории одновременно двое учеников: один отвечает, другой готовится (не больше **15 минут**). Отвечающий отвечает и уходит, готовившийся садится отвечать, новый входит, берёт билет и садится готовиться.

В билете будет **2 вопроса** из раздела «вопросы», либо **1 вопрос и 1 задача** из раздела «задачи». При верном и полном (по решению преподавателя) ответе на билет Вы получаете оценку А. При желании поднять оценку выше предложенной по результатам ответа на билет (в т. ч. с А до А+), Вы получаете **дополнительные вопросы** (любые из списка по усмотрению преподавателя), на которые отвечаете **без подготовки**.

## Вопросы

1. Рентгеноструктурный анализ: физические основы, получаемые данные, ограничения метода.
2. Спектроскопия ядерного магнитного резонанса: физические основы, получаемые данные, достоинства и ограничения метода.
3. Криоэлектронная микроскопия: физические основы, получаемые данные, достоинства и ограничения метода.
4. Секвенирование ДНК по Сэнгеру – принцип метода, достоинства. Особенности и ограничения методов секвенирования ДНК первого поколения.
5. Секвенирование ДНК по технологии Illumina (этапы изоляции, амплификации и собственно секвенирования фрагментов). Принцип метода. Особенности, достоинства и ограничения методов секвенирования ДНК второго поколения.
6. SMRT- и нанопоровое секвенирование. Принципы методов. Особенности, достоинства и ограничения методов секвенирования ДНК третьего поколения.
7. Общая идея сборки последовательностей по перекрытиям чтений. Контиг, скэффолд, N50, L50, среднее покрытие – определения. Q-score – формула, значение.
8. Сборка генома с помощью графа перекрытий и графа де Брюина: постановки задачи, получаемая на выходе информация, достоинства, ограничения и применимость.
9. Проблема повторов при сборке геномов, возможные способы её преодоления. Парноконцевые и встречноконцевые чтения, объединение контигов в скэффолды.
10. Архивные, курируемые, автоматические базы данных: определения, примеры. Чему соответствует одна запись в базах данных GenBank, RefSeq, UniProtKB.
11. Строение Fasta-файла, правила записи последовательностей в формате Fasta.

12. Определение выравнивания. Эволюционное, функциональное, структурное, оптимальное выравнивание. Задача выравнивания последовательностей в биологии и её практическое решение.
13. Алгоритм Нидлмана-Вунша: входные данные, алгоритм, результат работы.
14. Множественные выравнивания. Причина применения эвристических алгоритмов. Идея прогрессивного выравнивания. Консервативные блоки.
15. Особенности алгоритма BLAST, ускоряющие поиск схожих последовательностей в базах данных (по сравнению с алгоритмом Смита-Вотермана).
16. E-value в BLAST: определение, значение, зависимость от длины запроса, размера базы данных, счёта выравнивания (теорема Карлина).
17. Составные части филогенетического дерева. Неукоренённое, неразрешённое дерево. Тривиальные и нетривиальные ветви. Скобочная форма записи деревьев.
18. Общая схема реконструкции филогенетического дерева. Дистанционные, символьные, прямые и переборные методы (с примерами). Бутстреп-анализ.

### **Задачи**

19. Даны две короткие последовательности. Выровняйте их алгоритмом Нидлмана-Вунша (нарисуйте и заполните граф, найдите оптимальное выравнивание и его счёт). Также даны бонус за совпадение букв, штраф за любое несовпадение и штраф за гэп (линейный).
20. Дана матрица расстояний для 3-5 последовательностей. Постройте по ней дерево методом UPGMA. Укажите на нём длины ветвей.