

Биоинформатика: цифровые основы живых организмов



Семинар, мастер-класс «Биоинформатика: (для учителей информатики, старшеклассников и студентов профильных вузов). Мастер-класс посвящен особенностям вычислительных методов, применяемых для получения биологической информации.



- В рамках реализации Соглашения 342 «Развитие проекта «Академический (научно-технологический) класс в московской школе» по мероприятию «Организация и руководство проектной и исследовательской деятельностью учащихся в области агробιοтехнологии растений и биоинформатики, формирование компетенций педагогов по этим направлениям»

Области исследования в биоинформатике

Биоинформатика - это совокупность математических методов для обработки различных биологических данных.

Геномная биоинформатика - методы компьютерного анализа в сравнительной геномике

Структурная биоинформатика - методы предсказания пространственной структуры белков и других молекул.

1. Геномная биоинформатика:

Изобретение метода секвенирования позволило определять последовательность нуклеотидов ДНК (и, чуть позже, РНК).

1977 г. Полная последовательность генома бактериофага ФХ174 - 5386nt

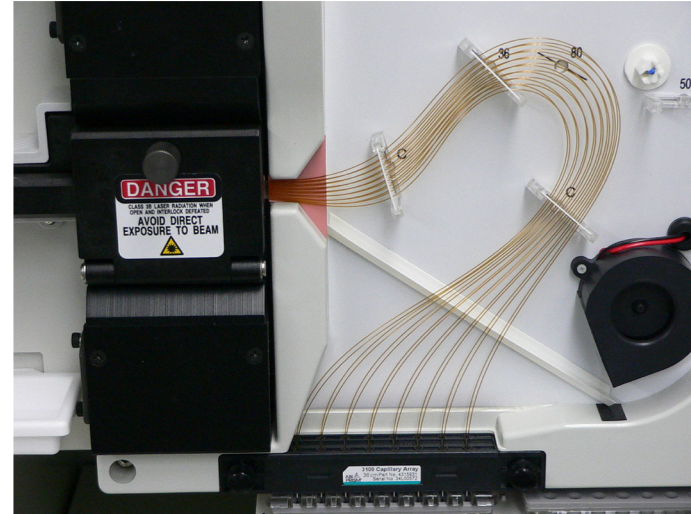
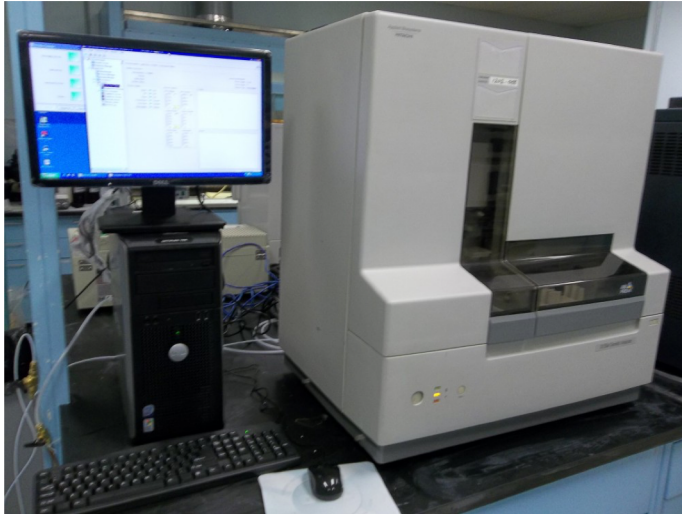
(Статья <https://www.nature.com/articles/265687a0>, нуклеотидная последовательность https://www.ncbi.nlm.nih.gov/nucleotide/NC_001422)

По состоянию на апрель 2020 г. в базе данных NCBI хранятся 1 484 079 258 последовательностей общей длиной в 8 203 903 249 287 nt

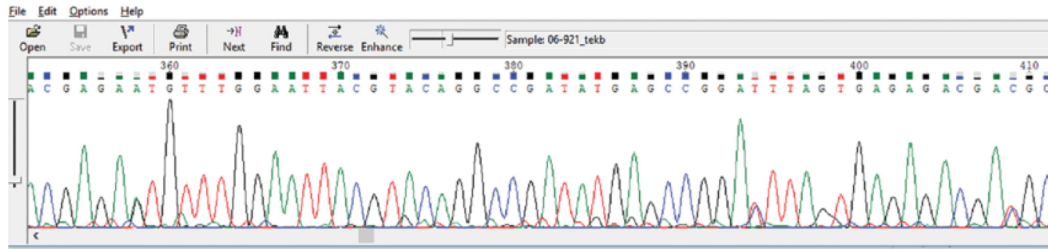
Такой объем данных невозможно обрабатывать вручную.

Геномная биоинформатика: секвенирование

I поколение: 1 (16, 96) фрагмент ДНК длиной до 1000 нт.

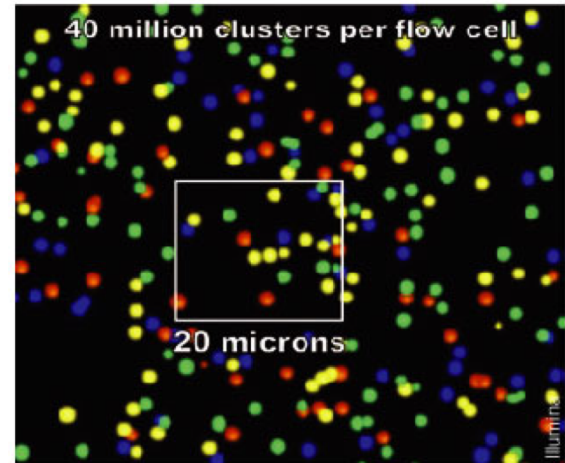


06-921_tekb_kkt_plate6_4Jan07_2007-01-09_G10.ab1 - Chromas



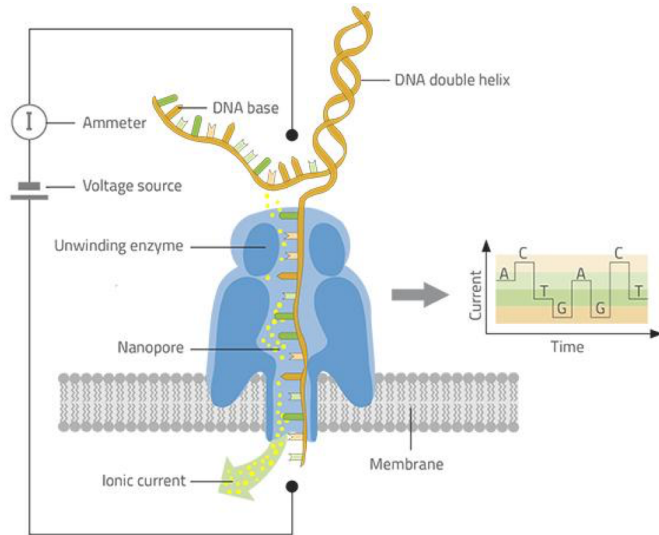
Геномная биоинформатика: секвенирование

II поколение: Одновременное чтение большого числа ($4 \cdot 10^6 \dots 10^9$) относительно коротких (75-300 нт) последовательностей ДНК (“shotgun sequencing”)



Геномная биоинформатика: секвенирование

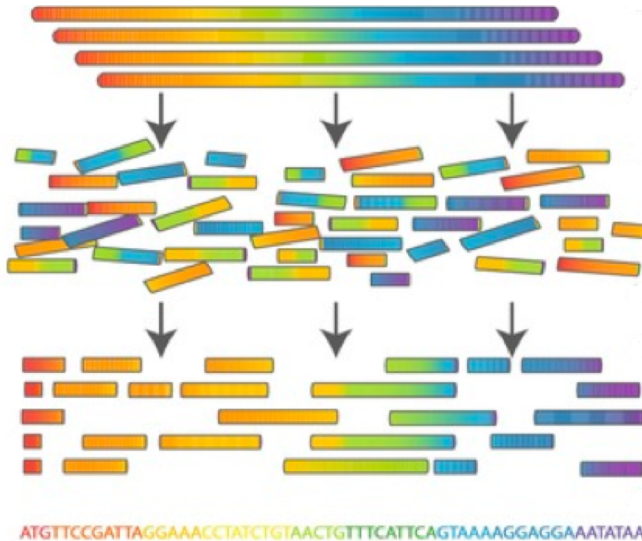
III поколение: Одновременное чтение сотен относительно длинных (до 10^6 нт) последовательностей ДНК.



Геномная биоинформатика: применение полученных данных

III поколение: Одновременное чтение сотен относительно длинных (до 10^6 нт) последовательностей ДНК.

Сборка генома/метагенома.



Исходные молекулы ДНК

Фрагментированная ДНК

Прочитанные фрагменты ДНК