

# СБОРКА И АНАЛИЗ ГЕНОМА

Краткий обзор курса

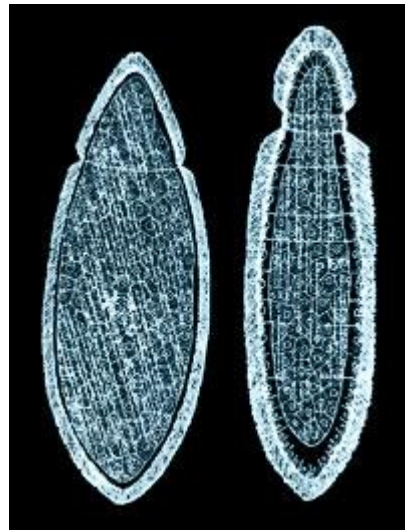


**ИЛС**  
ИнтерЛабСервис



# Цели работы

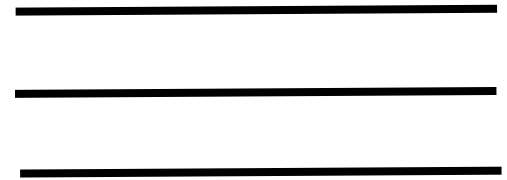
*Intoshia* - представитель типа Ортонектиды. Паразит морских беспозвоночных. Примитивное и необычное существо.



**Цель** - филогенетический анализ, то есть установление её родства с другими группами.

# Проблемы

- Человек (*Homo sapiens*) - геном известен
- Морской ёж (*Strongylocentrotus purpuratus*) - геном известен
- Гидра (*Hydra vulgaris*) - геном известен
- ...
- ...

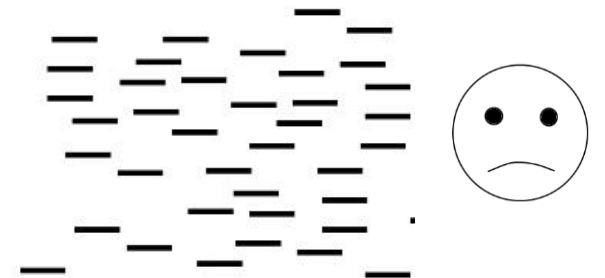


# Проблемы

- Человек (*Homo sapiens*) - геном известен
- Морской ёж (*Strongylocentrotus purpuratus*) - геном известен
- Гидра (*Hydra vulgaris*) - геном известен
- ...
- ...

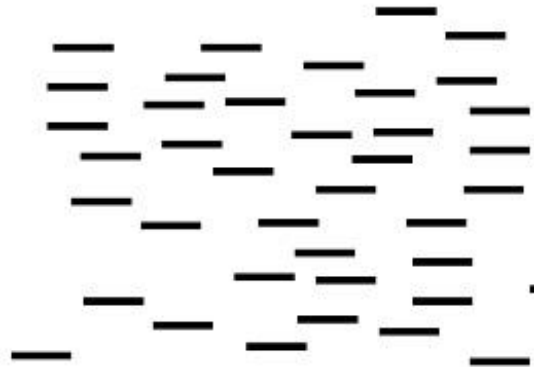


Intoshia - только риды



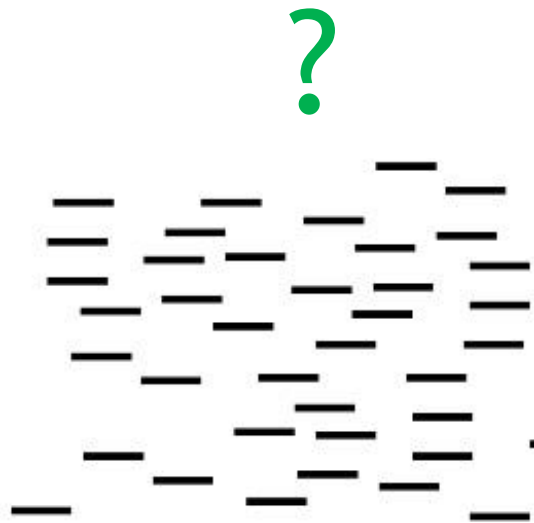
# Адаптеры в ридях

- Не все риды одинаково полезны

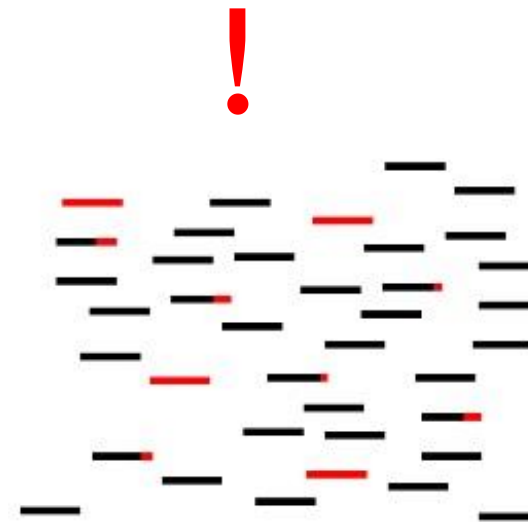


# Адаптеры в ридсах

- Не все ридсы одинаково полезны



На самом деле,  
идеального  
секвенирования не бывает



В ридсах встречаются  
адаптеры

# Адаптеры в ридсах

- Удаление адаптеров (а также удаление частей ридов, имеющих низкое качество по другим причинам) называется **ТРИММИНГ**
- Для удаления адаптеров - программа **Trimmomatic**
- Для контроля качества ридов - программа **FastQC**
- Контроль качества полезно делать **до** и **после** удаления адаптеров

# Покрытие

- Покрытие (среднее) - сколько ридов в среднем ложится на каждый нуклеотид генома



**Проблема** - в генома есть места, не представленные ни одним ридом



# Покрытие

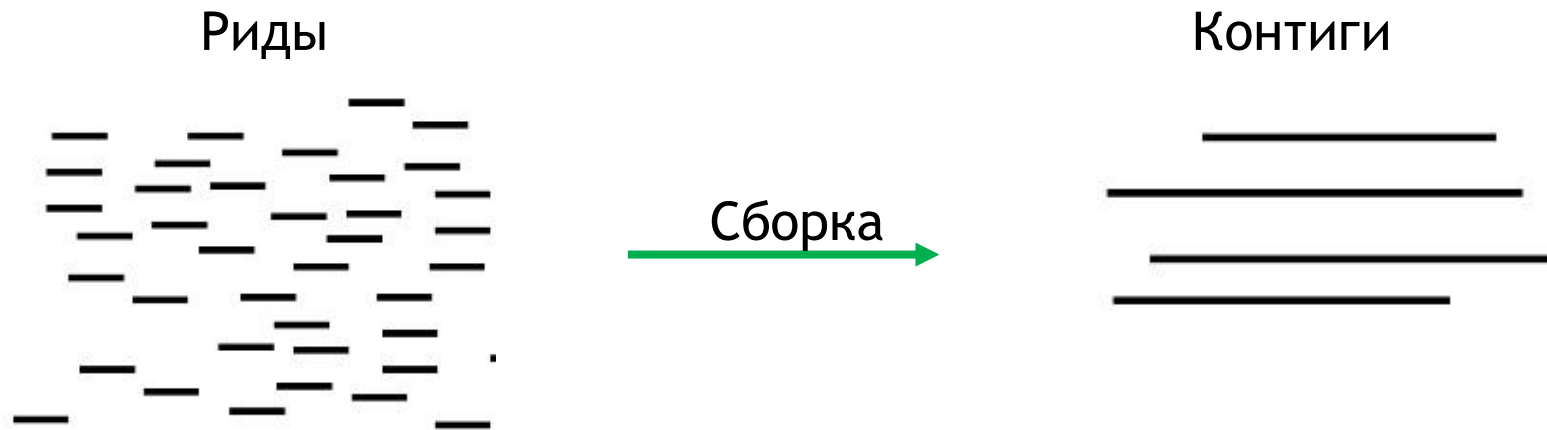
- Покрытие (среднее) - сколько ридов в среднем ложится на каждый нуклеотид генома



программа **Jellyfish** поможет вам построить похожий график и понять, какое же покрытие у *вашего* генома

# de novo сборка

- **Сборка** - процесс составления из ридов более длинных последовательностей - контигов.
- **Идеал** - собрать в один контиг весь геном
- Но такого почти никогда не бывает



Популярные программы - **Spades**, **Velvet**, **Ray** и многие другие (на самом деле, *лучших* буквально несколько)

# Ищем гены

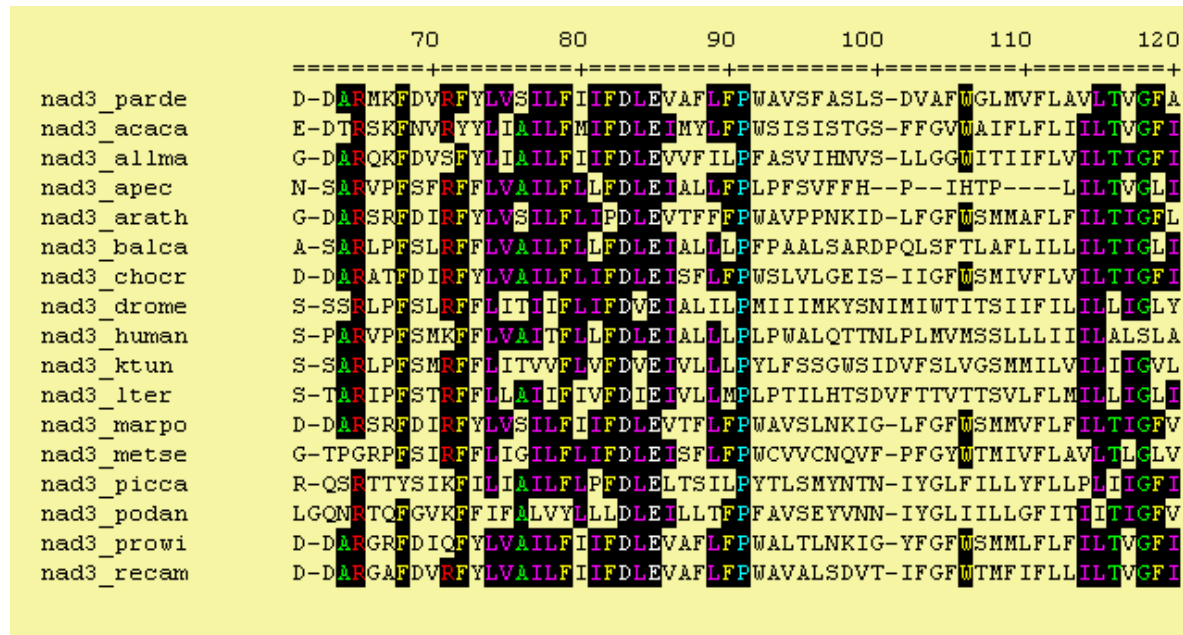
- Чтобы сравнивать *Intoshia* с другими видами, в получившихся контигах сначала нужно найти **гены**.
- Программа **CEGMA** существует для того, чтобы найти основные ~500 эукариотических генов.
- Затем на этих генах тренируется программа **Augustus** - она понимает, какие в этом геноме типичные параметры генов (использование кодонов, типичные длины интронов и т.п.) и на основе этого предсказывает **остальные** гены.

# Находим ортологи

- Мы получили гены. Но для сравнения нам ещё нужно поставить **соответствие** между генами разных видов. Иными словами, определить ортологи.
- Ортологи между разными геномами находит программа **OrthoMCL**

# Выравнивание ортологов

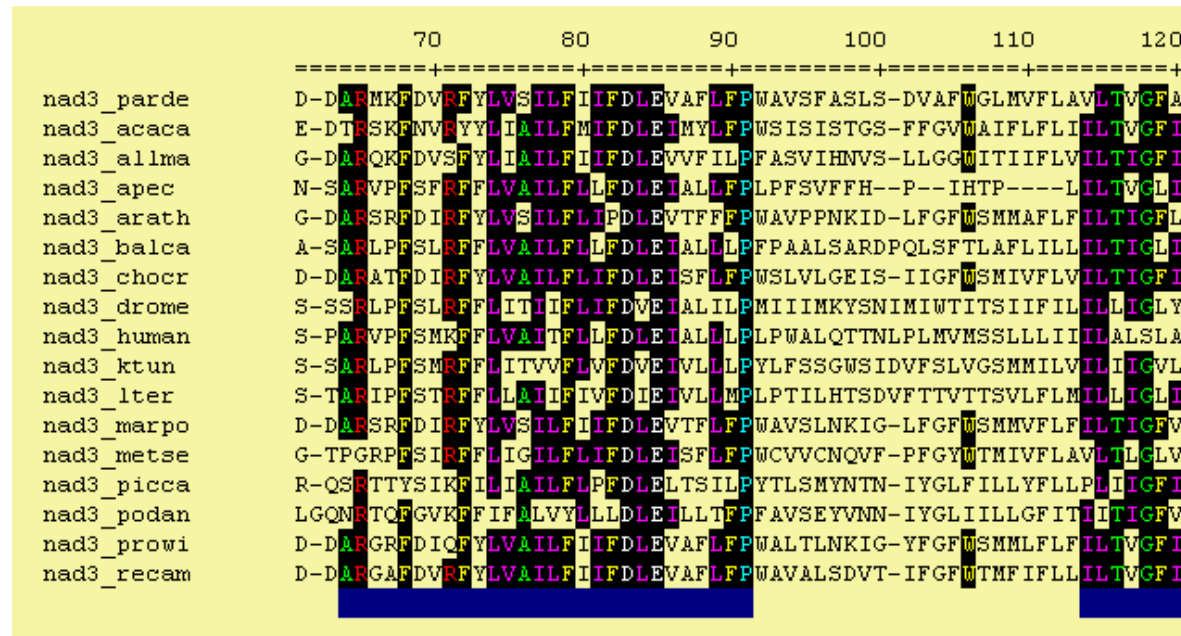
- Чтобы сравнить гены, нужно сначала провести их **выравнивание**



Это может делать, например, программа **MAFFT**

# Выравнивание ортологов

- После того, как вы сделали выравнивание, из него нужно удалить **места, не заслуживающие доверия**.



Синим  
отмечены  
места,  
которые  
стоит  
оставить

Это может делать, например, программа [trimAl](#)

# Построение дерева по выравниванию

- Методами, про которые вам рассказывали в других частях курса, вы наконец можете построить **дерево**, и предположить **эволюционную историю** необычного организма под названием **Intoshia**

