

Введение в биоинформатику

Пржибельский Андрей

*Центр Алгоритмической Биотехнологии, СПбГУ
(бывшая Лаборатория Алгоритмической Биологии, СПбАУ)*

ОСНОВЫ

Что такое биоинформатика?

Что такое биоинформатика?

- Компьютеризация биологических исследований
- Анализ биологических данных
- Моделирование биологических процессов
- Объединение биологов и программистов

Центральная догма молекулярной биологии

ДНК



Транскрипция

РНК

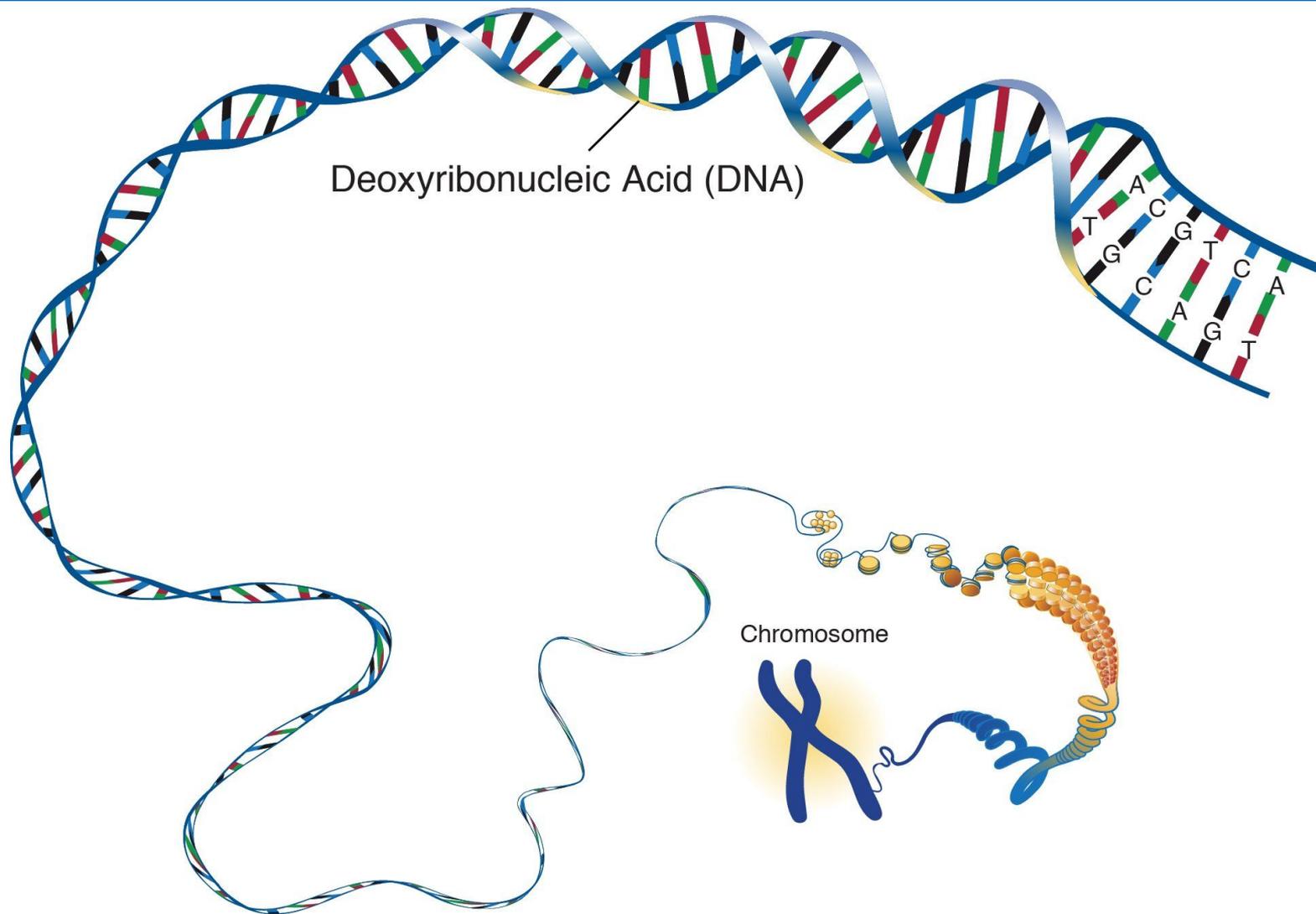


Трансляция

Белок

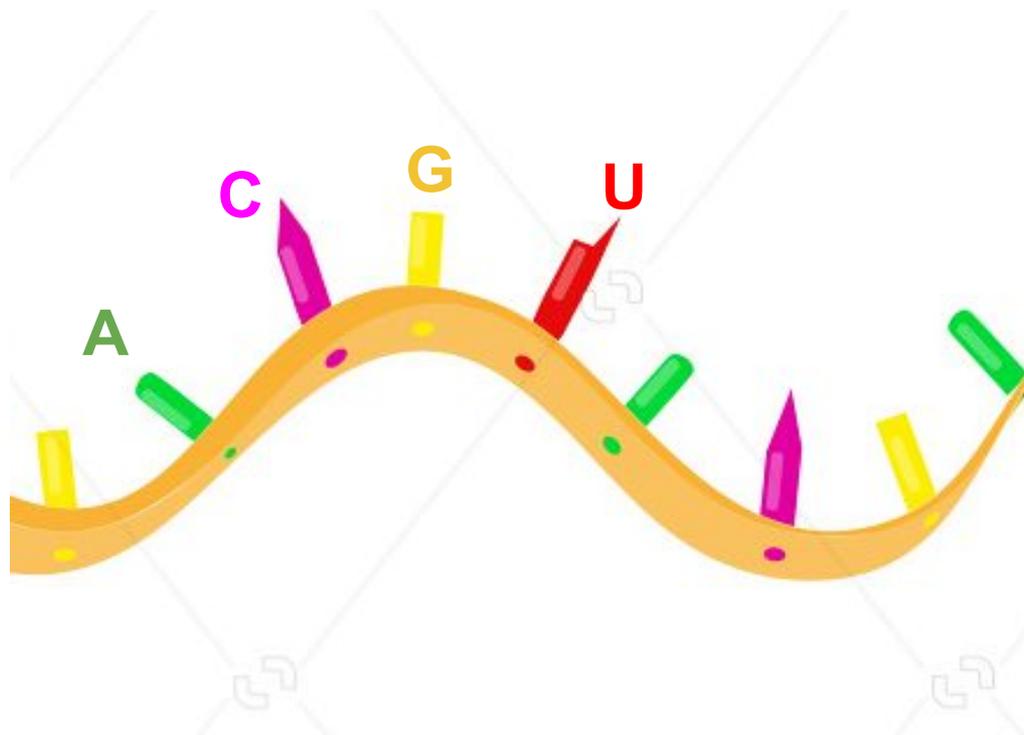


ДНК

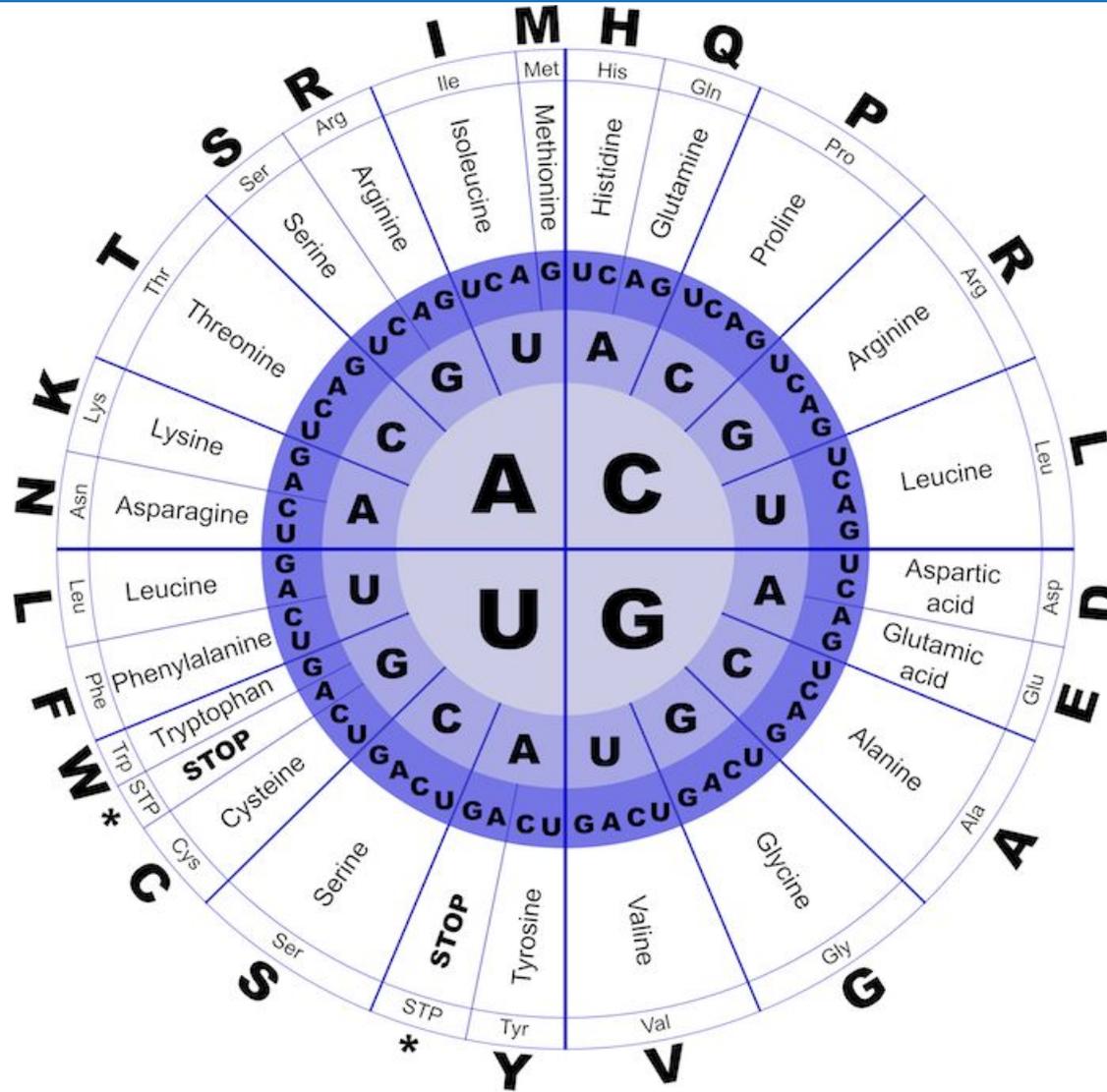


PHK

T → U



Белки



От ДНК к белкам

ДНК



Транскрипция

РНК



Трансляция

Белок



От ДНК к белкам

Тироцидин В1 (производится *Bacillus brevis*)

Val-Lys-Leu-Phe-Pro-Trp-Phe-Asn-Gln-Tyr

V K L F P W F N Q Y

От ДНК к белкам

Тироцидин В1 (производится *Bacillus brevis*)

Val-Lys-Leu-Phe-Pro-Trp-Phe-Asn-Gln-Tyr

V K L F P W F N Q Y

- Сколько различных последовательностей ДНК могут кодировать Тироцидин В1?

От ДНК к белкам

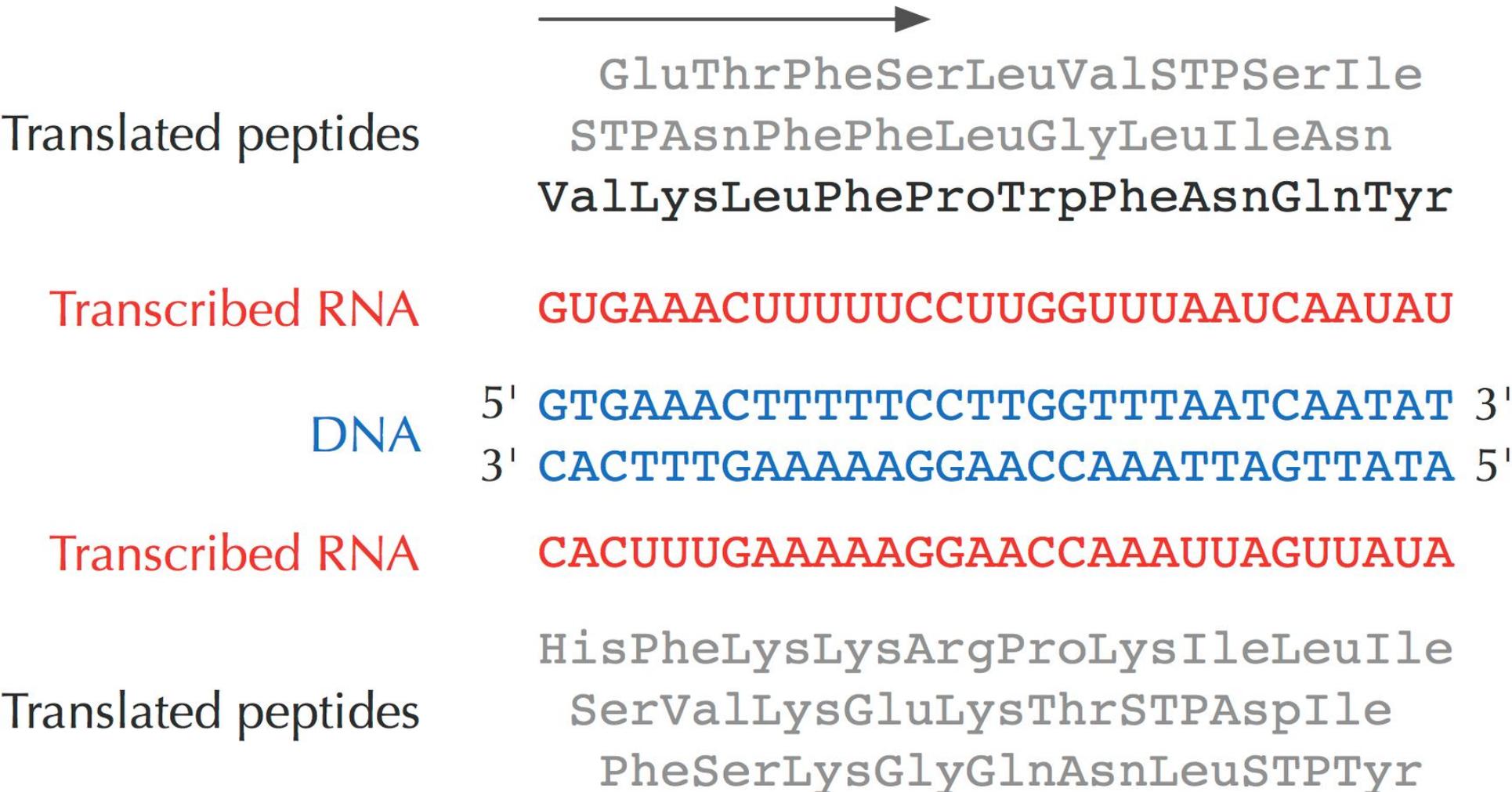
Тироцидин В1 (производится *Bacillus brevis*)

Val-Lys-Leu-Phe-Pro-Trp-Phe-Asn-Gln-Tyr

V K L F P W F N Q Y

- Сколько различных последовательностей белков может кодировать один участок ДНК?

От ДНК к белкам



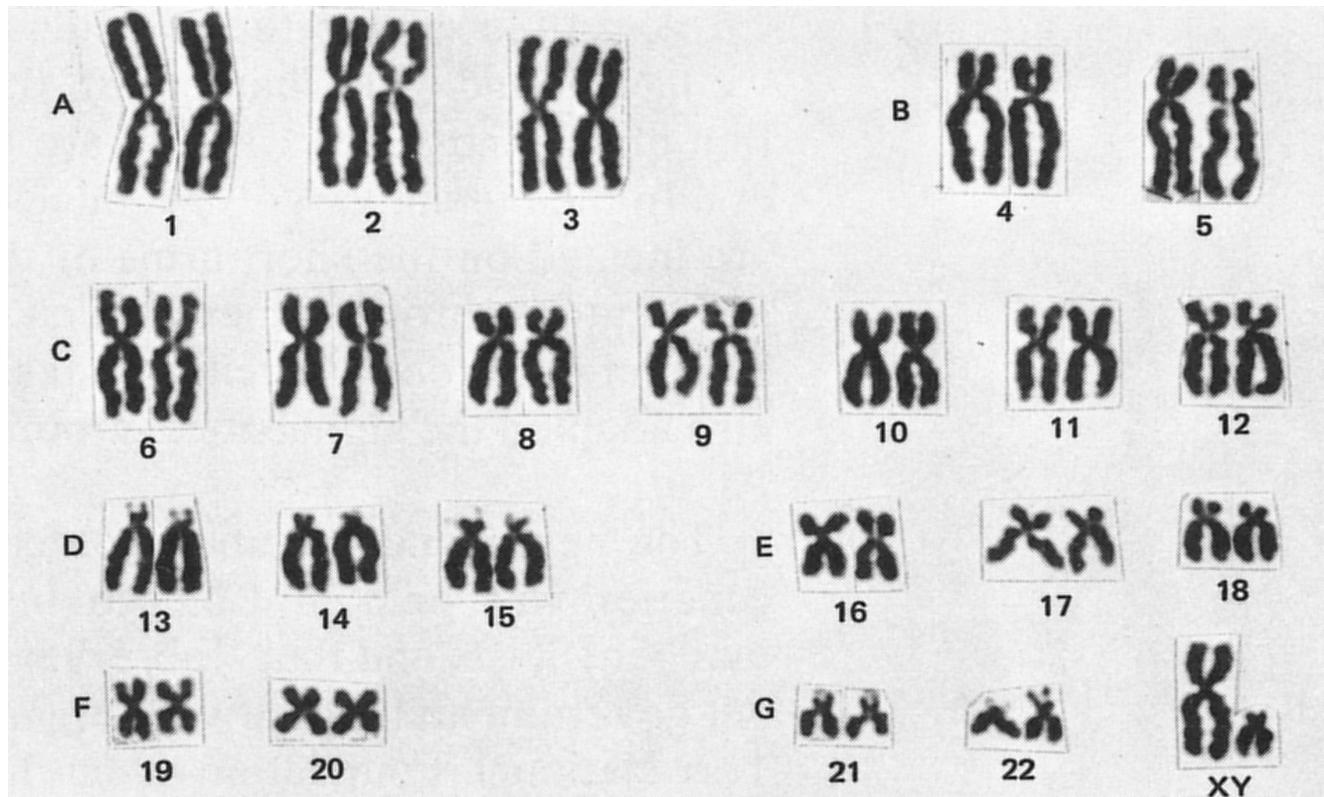
От ДНК к белкам

- У ДНК есть специальные участки — гены
 - Гены используются для создания белков
 - Остальная часть ДНК **раньше** считалась мусором
- Геном — совокупность всей информации закодированной в ДНК

Геном человека

Геном человека

- 23 пары хромосом



Геном человека

- 23 пары хромосом
- 3 миллиарда нуклеотидов
- ~22000 генов
- Гены составляют менее 2% от всей ДНК

Мутации

- Копия генома есть в каждой нашей клетке
- Клетки делятся, копируя геном
 - Во время копирования могут происходить ошибки — **мутации**
 - Большинство мутаций молчащие
 - Но некоторые влияют на здоровье
- Любое различие с эталонным геномом тоже называется **вариацией**

Вариации

- Замены

Эталонный геном:

....ACCCGCCAGGCAGGAC....

Исследуемый геном:

....ACCCGCCAT**GCAGGAC....**

Вариации

- Удаления

Эталонный геном:

....ACCCGCCAGGCAGGAC....

Исследуемый геном:

....ACCCGCCA_ GCAGGAC....

Вариации

- Вставки

Эталонный геном:

....ACCCGCCAGGCAGGAC....

Исследуемый геном:

....ACCCGCCAGAGGCAGGAC....

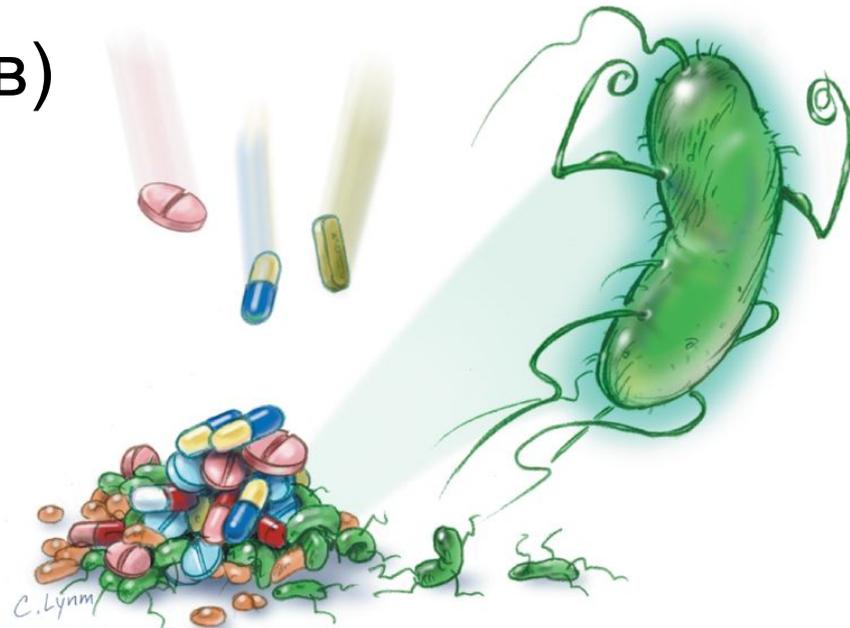
Применения биоинформатики

Пример №1

Поиск новых антибиотиков

Устойчивость к антибиотикам

- Антибиотики — естественные вещества, которые убивают бактерии или подавляют их размножение
- Люди используют антибиотики чаще, чем стоит (даже против вирусов)
- Некоторые бактерии все равно выживают



Устойчивость к антибиотикам

- Антибиотики — естественные вещества, которые убивают бактерии или подавляют их размножение
- Люди используют антибиотики чаще, чем стоит (даже против вирусов)
- Некоторые бактерии все равно выживают
- Бактерии становятся **устойчивыми**
- Необходимо найти новые антибиотики

Новые антибиотики

- Могут быть найдены в еще неизученных организмах
- Необходимо исследовать новые белки



Чтение белков

- Белок — строчка над алфавитом из 20 символов (аминокислот)
- Невозможно “прочитать” молекулу напрямую
- Для “чтения” используется масс-спектрометрия
- Пептид — короткий белок

Масс-спектрометрия

G A S P V T C I L N D K Q E M H F R Y W
57 71 87 97 99 101 103 113 113 114 115 128 128 129 131 137 147 156 163 186

NQEL

484

Масс-спектрометрия

G A S P V T C I L N D K Q E M H F R Y W
57 71 87 97 99 101 103 113 113 114 115 128 128 129 131 137 147 156 163 186

NQEL

484

N NQ NQE

N-terminal

Масс-спектрометрия

G A S P V T C I L N D K Q E M H F R Y W
57 71 87 97 99 101 103 113 113 114 115 128 128 129 131 137 147 156 163 186

NQEL

484

N	NQ	NQE	QEL	EL	L
114	242	335	370	242	113
<i>N-terminal</i>			<i>C-terminal</i>		

Масс-спектрометрия

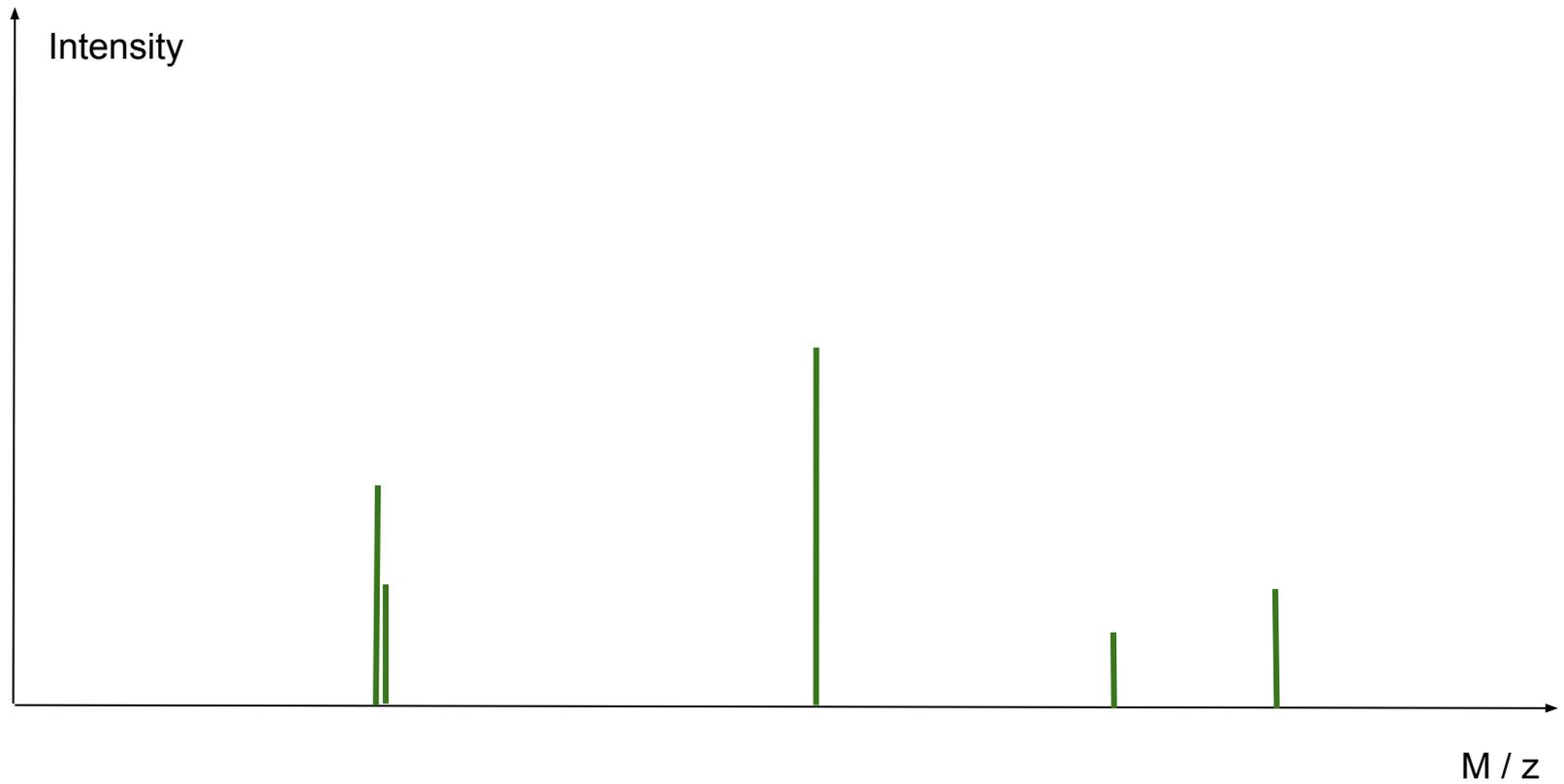
G A S P V T C I L N D K Q E M H F R Y W
57 71 87 97 99 101 103 113 113 114 115 128 128 129 131 137 147 156 163 186

NQEL

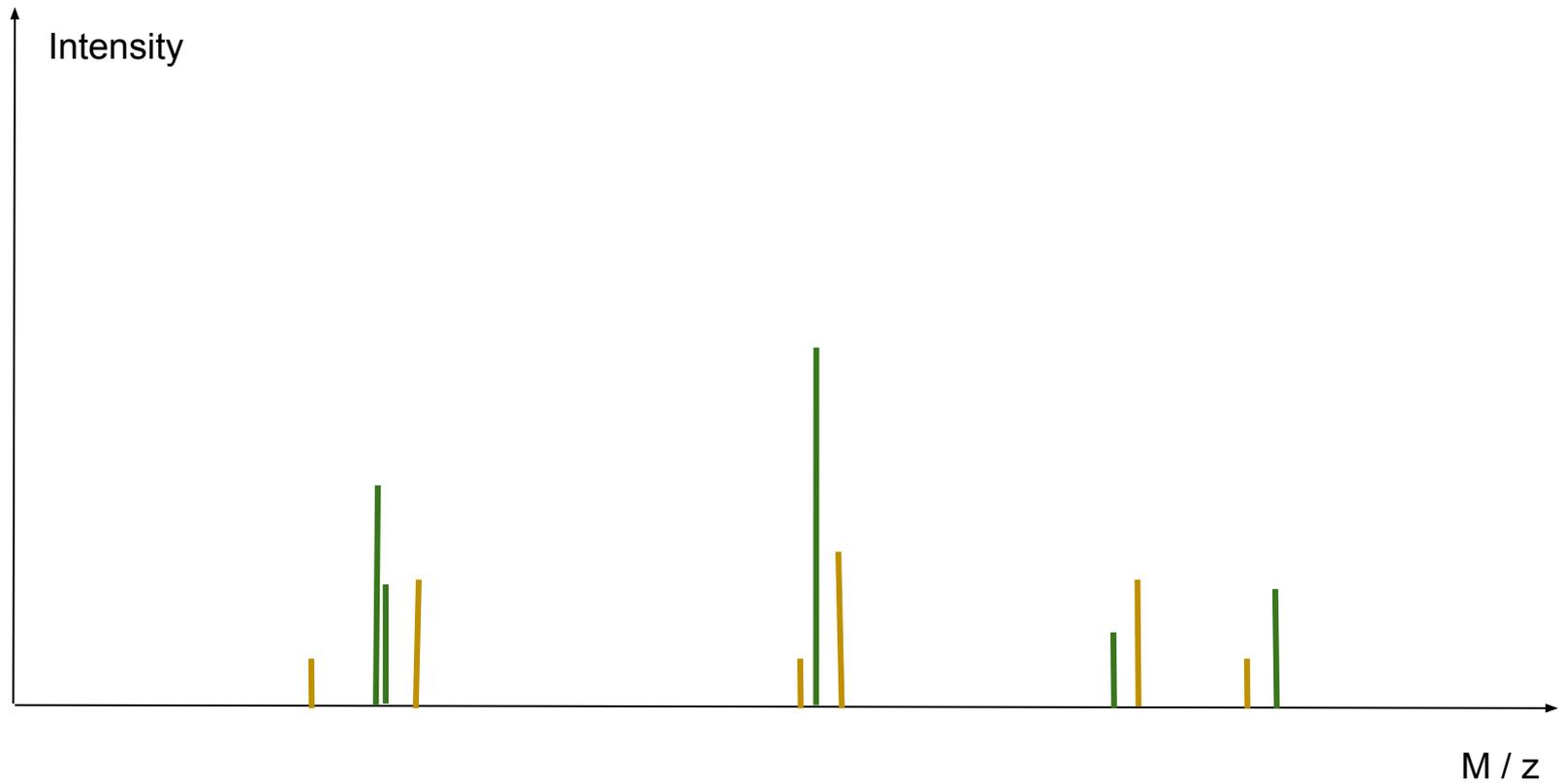
484

114 242 335 370 242 113

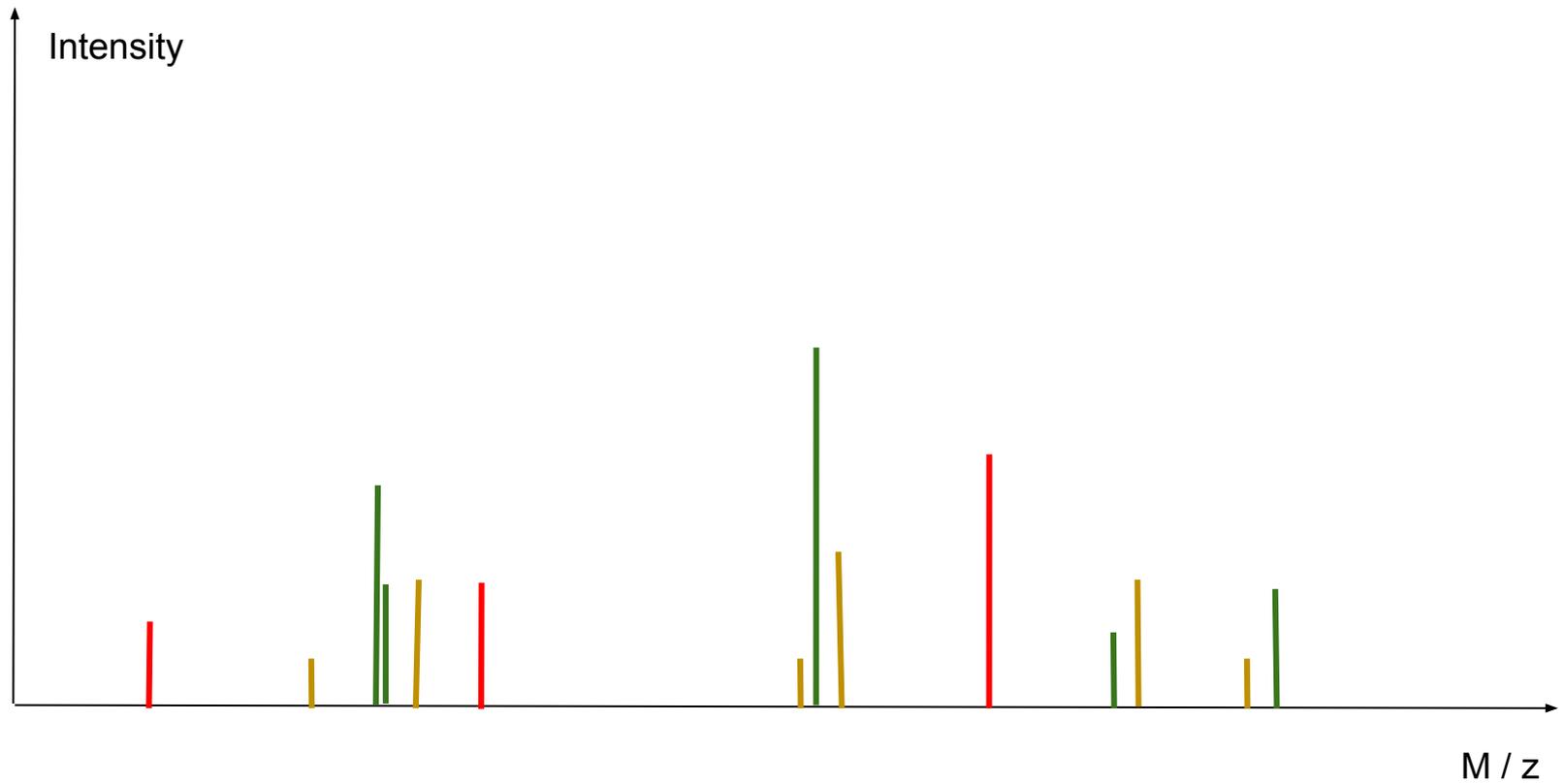
Масс-спектрометрия



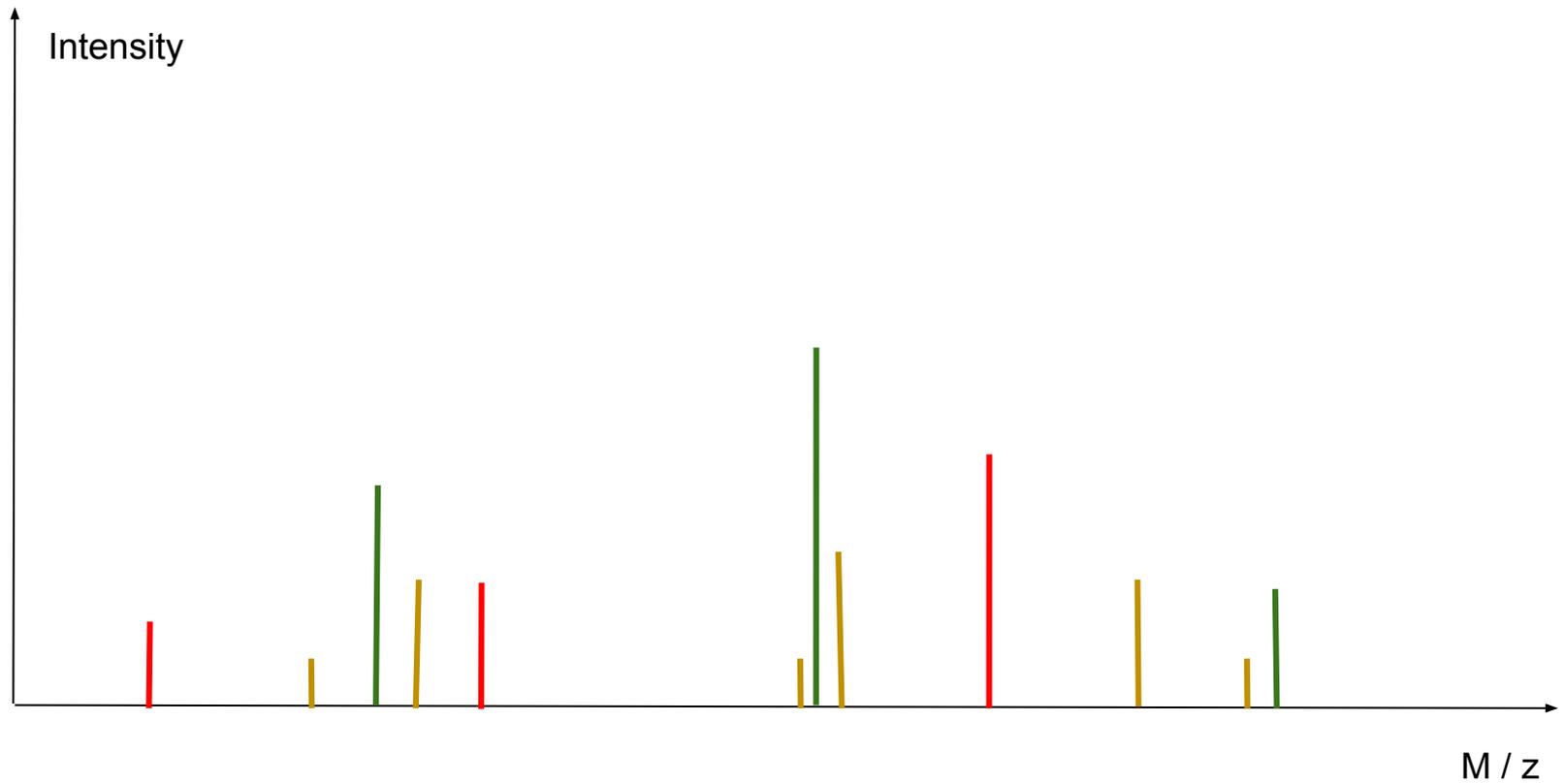
Масс-спектрометрия



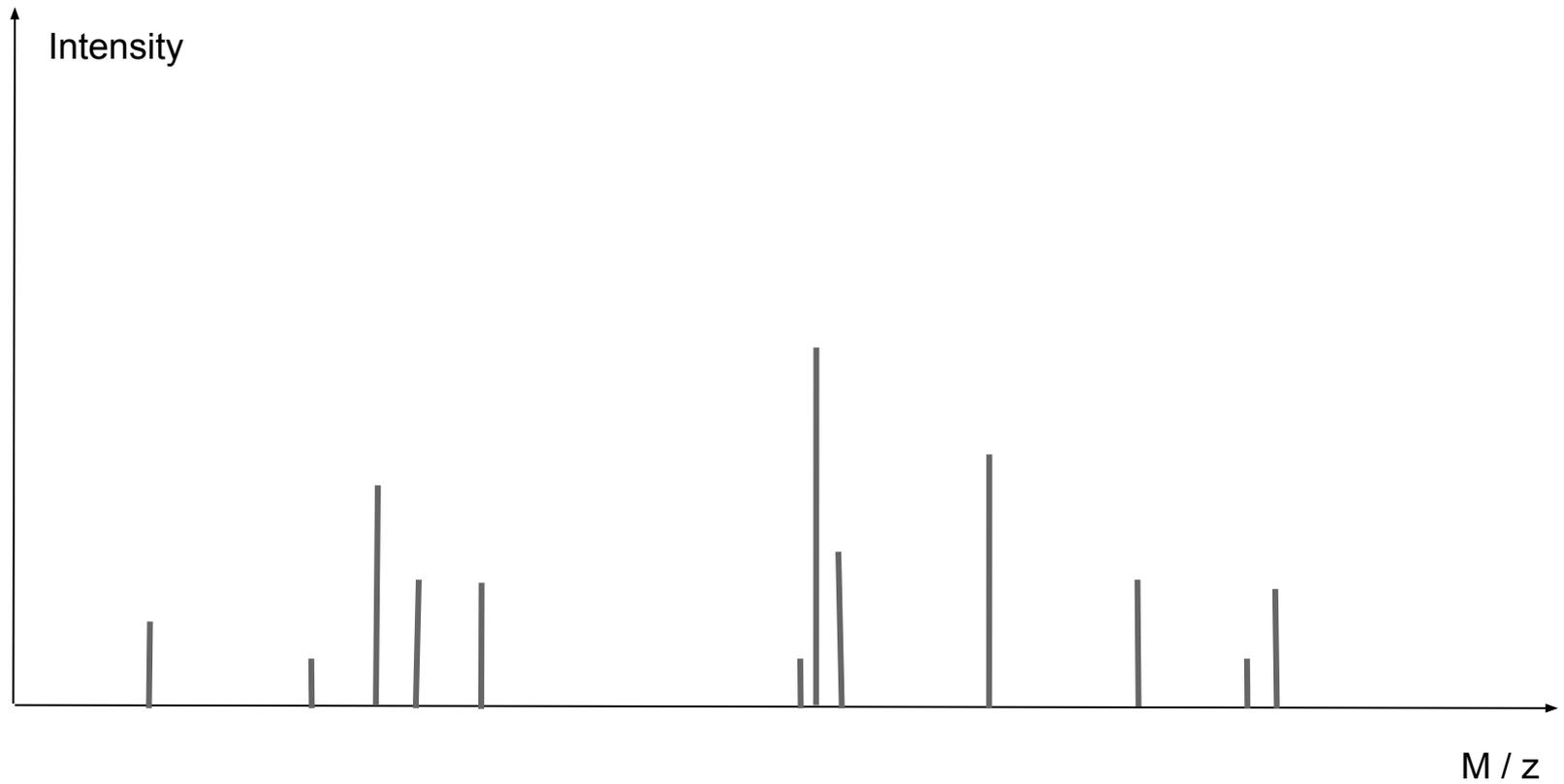
Масс-спектрометрия



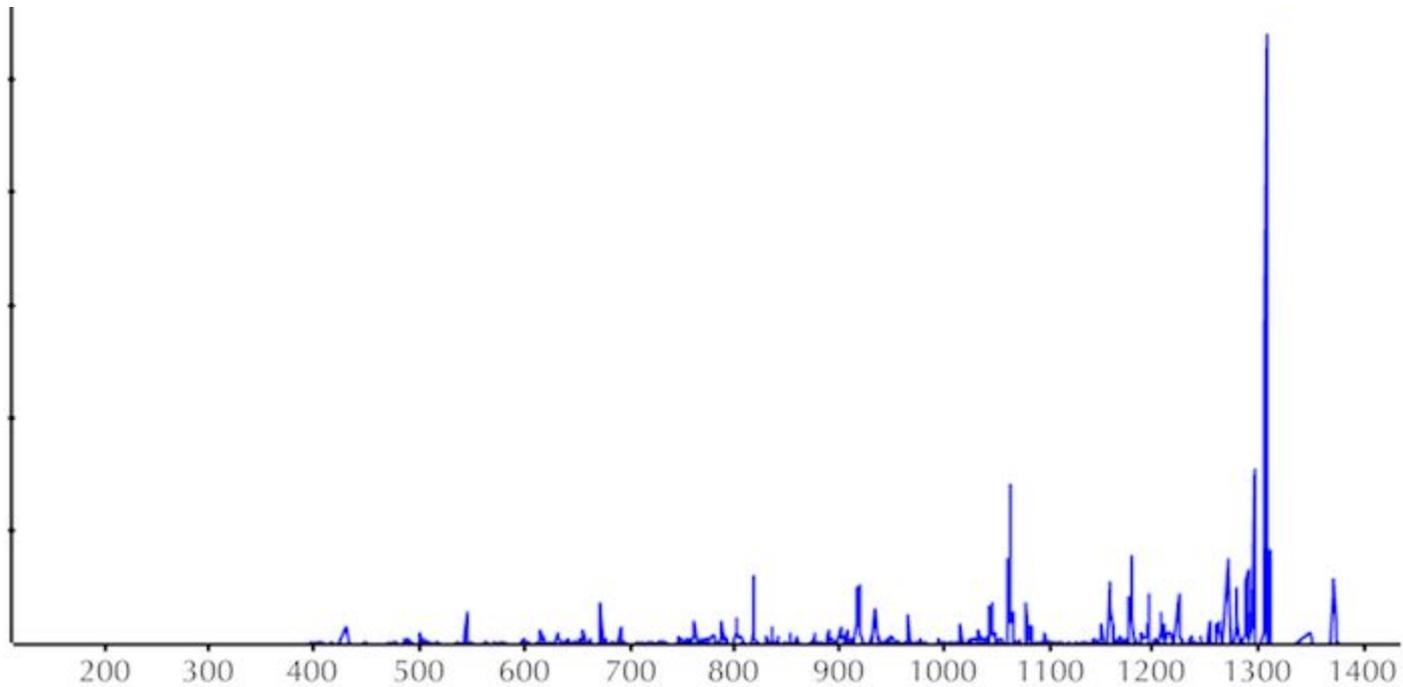
Масс-спектрометрия



Масс-спектрометрия



Масс-спектрометрия



Поиск с помощью базы данных

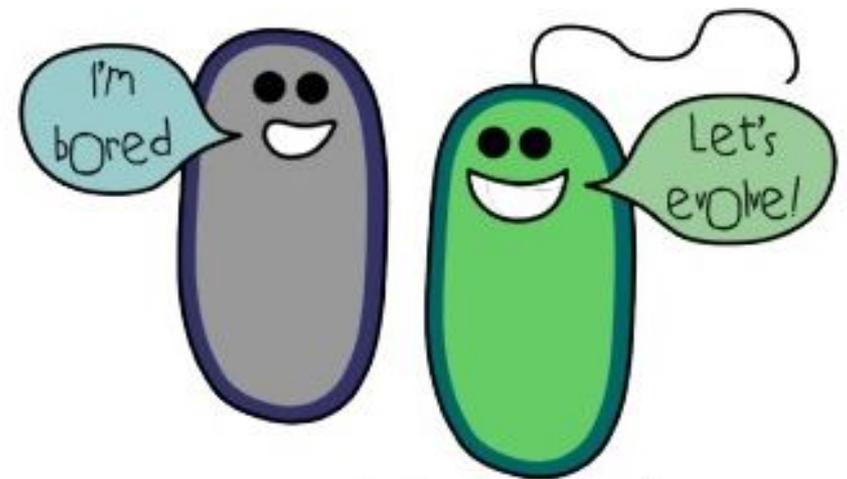
- Существует база данных известных белковых последовательностей
 - ~10К белков
- Цель — найти модификацию известного белка
 - 1М+ спектров
 - В каждом ~200 пиков

Пример №2

Изучение устойчивости к антибиотикам

Устойчивость к антибиотикам

- Люди используют антибиотики чаще, чем стоит (даже против вирусов)
- Некоторые бактерии все равно выживают
- Бактерии становятся **устойчивыми**

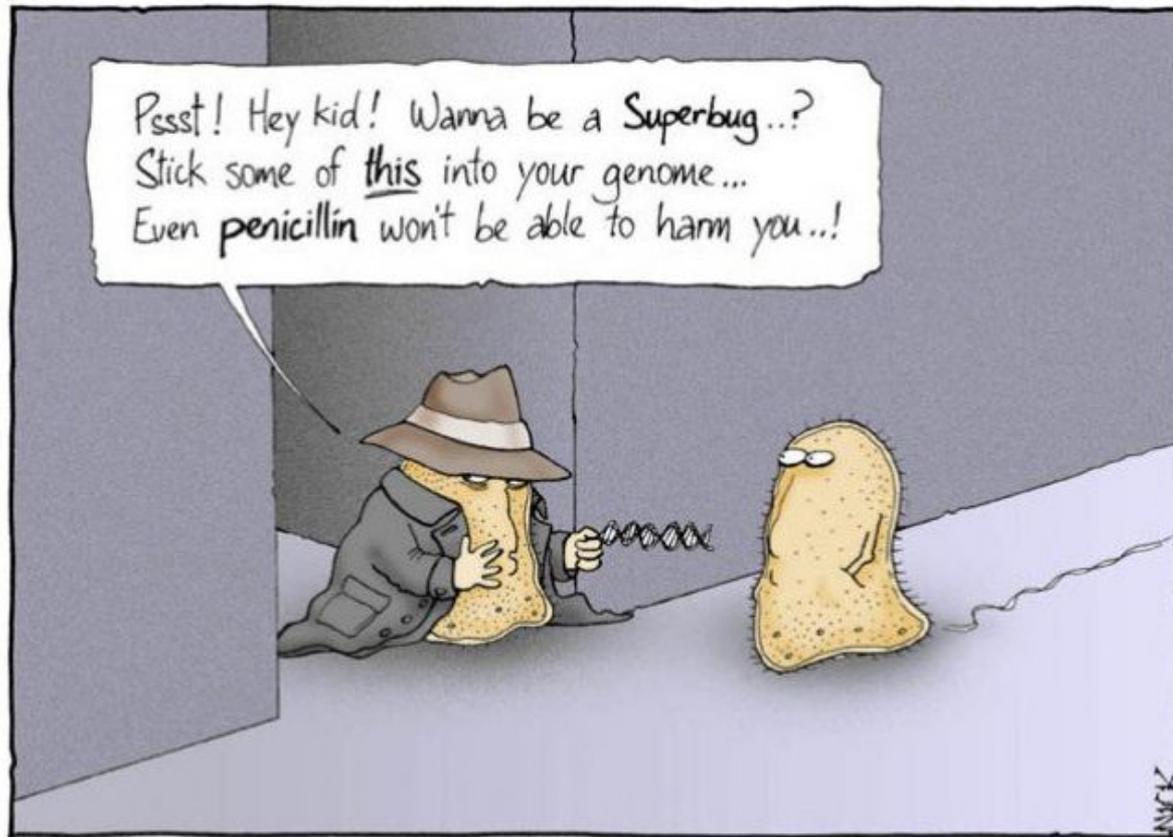


how multi-drug resistant bacteria came to be....

Что делает бактерию устойчивой?

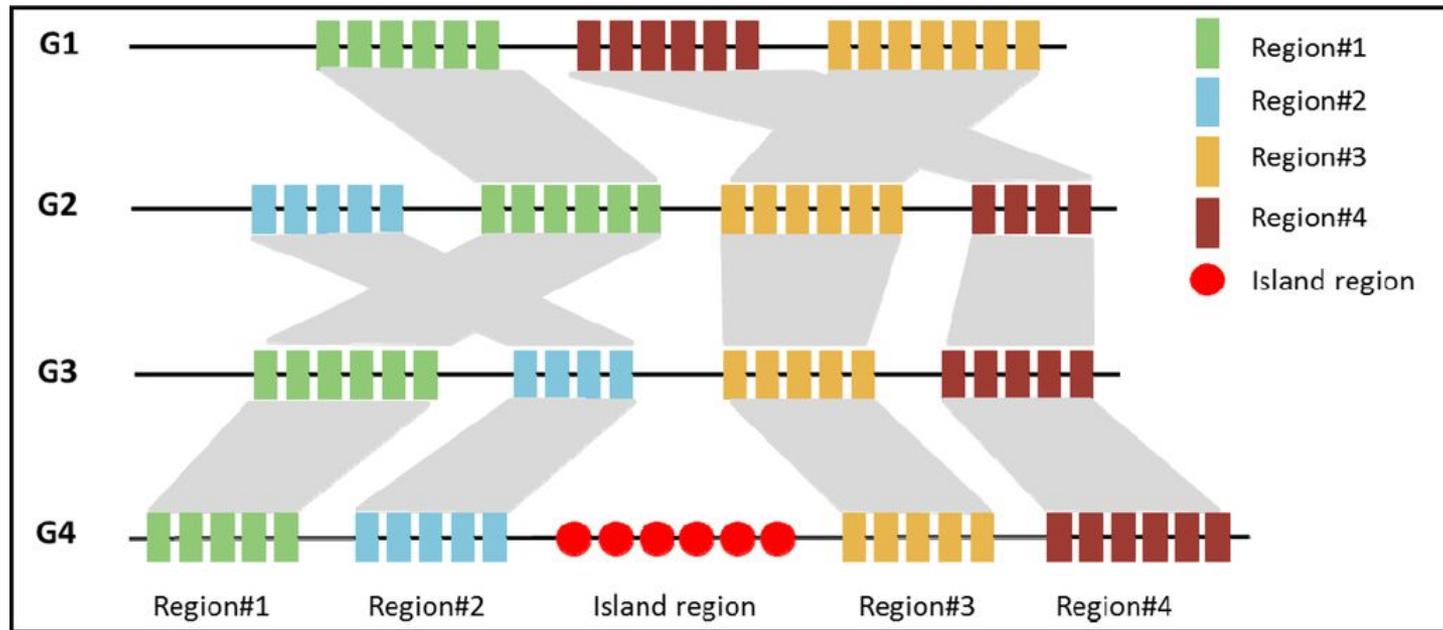
Что делает бактерию устойчивой?

В геном бактерии “встраивается”
новая последовательность



Как найти ген устойчивости?

- “Прочесть” геном обычной бактерии
- “Прочесть” геном устойчивой бактерии
- Сравнить два генома



Как прочесть геном?

Секвенирование ДНК

...AACCCGTACGTTTTGCAAACGACCGT...

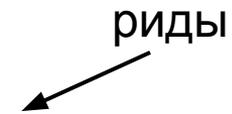
↑
геном

Секвенирование ДНК

- Секвенирование

GTACGTTTTGCA
GTTTTGCAAACG
CGTACGTTTTG
AACCCGTACGT AACGACCG
...AACCCGTACGTTTTGCAAACGACCGT...

ряды



Секвенирование ДНК

- Секвенирование

GTA GTTTTGCA
GTTTTGCAAACG

- Ошибки

CGTACGTTTTC

AACCCGTTTCGT

AACGACCG

...AACCCGTACGTTTTGCAAACGACCGT...



Секвенирование ДНК

- Секвенирование

GTA GTTTTGCA
GTTTTGCAAACG

- Ошибки

CGTACGTTTTC

AACCCGTTCGT AACGACCG

...AACCCGTACGTTTTGCAAACGACCGT...

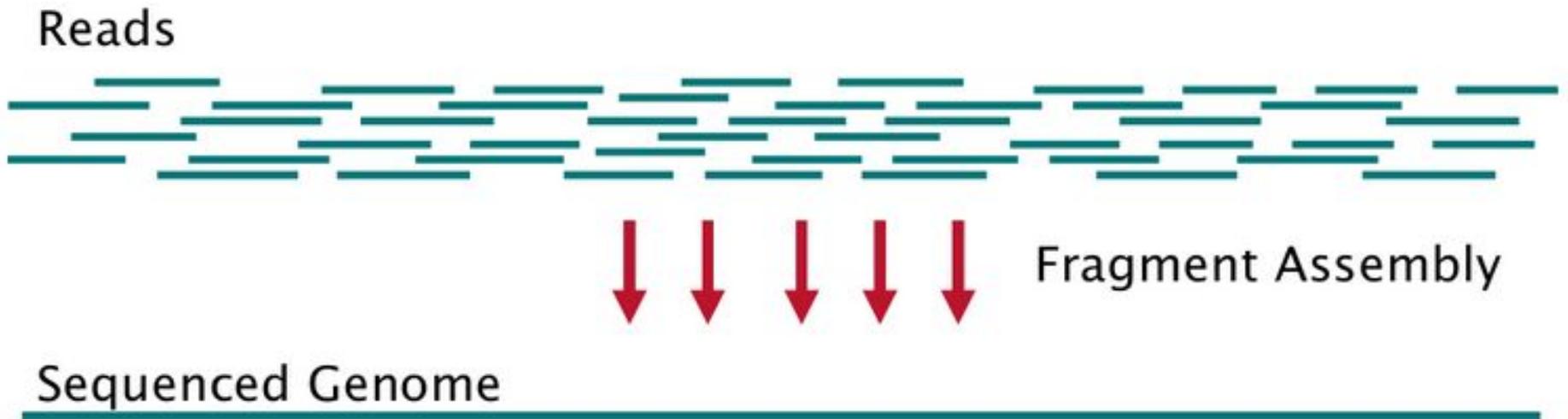
- Позиции ридов в геноме неизвестны и случайны

Достаточно ли ридов?

Достаточно ли ридов?

НЕТ

Сборка генома “с нуля”



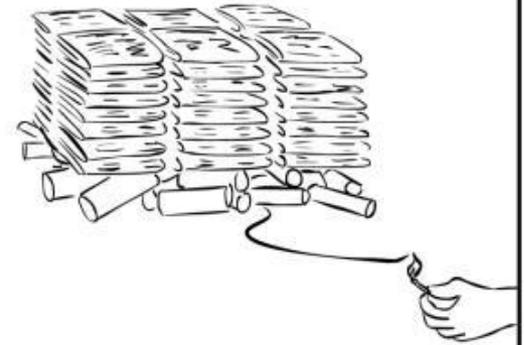
Сборка генома “с нуля”



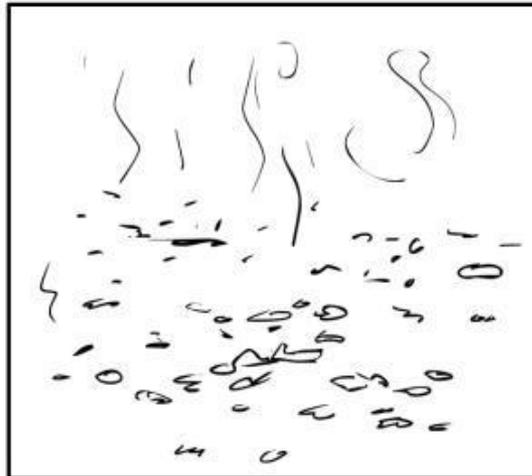
stack of NY Times, June 27, 2000



stack of NY Times, June 27, 2000
on a pile of dynamite



this is just hypothetical

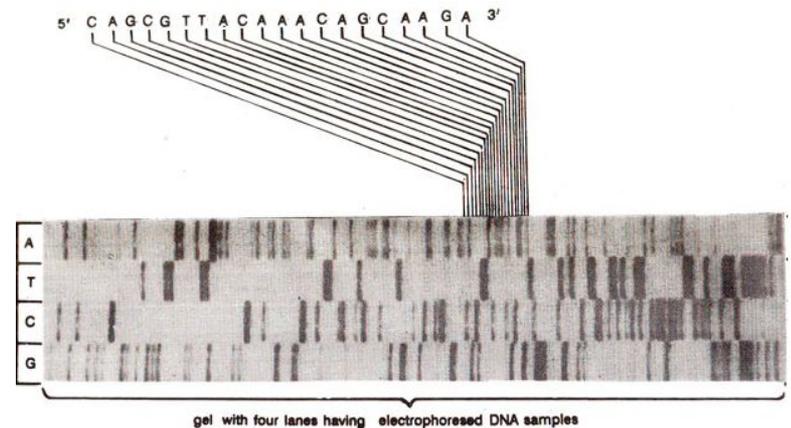


so, what did the June 27, 2000 NY
Times say?

Первые геномные проекты

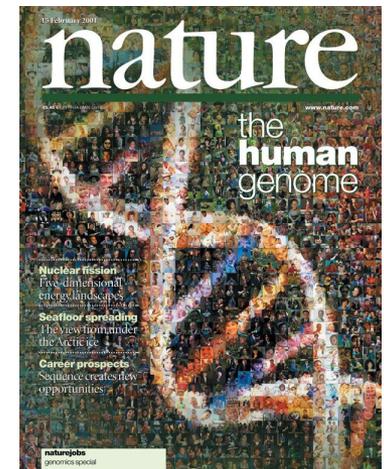
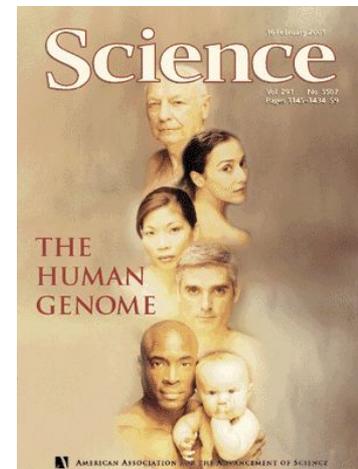
- Секвенирование по Сангеру

- Длинные риды (~900 bp)
- Огромная цена



- Геном человека

- 3 Mbp
- 3 миллиарда USD
- > 10 лет



Применения секвенирования

Применения секвенирования

- **Клинические исследования**
 - Диагностика болезней
 - Геномика рака
 - Персональная медицина
 - Микробиота
- **Фармацептика**
 - Разработка новых лекарств
 - Исследование устойчивости к лекарствам
- **Криминалистика**
- **Исследование древней ДНК**

Диагностика болезней

- Мутации могут влиять на склонность и устойчивость к различным заболеваниям
 - Устойчивость к ВИЧ/СПИД
 - Риски заболевания различными видами рака
 - Болезнь Паркинсона
 - Болезнь Альцгеймера
 - Астма
 - Псориаз
 - Диабет
 -
- 23andme.com

Мутации, ассоциированные с болезнью

- **Эталонный геном**

CAAGTTTGAGGTAATGGATAGTGGGAATGAAAGTAGCATTATCCAACCGCCAAACTTAAATAGCA

- **Геном пациента**

CAAGTTTGTGGTAATGGAAAGTGGGAATGAAACTAGGATTATCCAACCGCCAAACTTAAATAGCA

Мутации, ассоциированные с болезнью

- **Эталонный геном**

CAAGTTTGAGGTAATGGATAGTGGGAATGAAAGTAGCATTATCCAACCGCCAAACTTAAATAGCA

- **Геном пациентов**

CAAGTTTGTGGTAATGGAAAGTGGGAATGAAACTAGGATTATCCAACCGCCAAACTTAAATAGCA

CTAGTTTGAGGAAATGGATAGTGGGAATGAAACTAGCATTATCCAACCGCGAAACTTAAATAGCA

CATGATTGAGGTAATGGATAGTGGGAATGAAACTAGCATTATCCAACGGCCAAACTTAAATAGCA

CAAGTTTGAGGTAATGGATAGTGGGAATGAAACTAGCATTATCGAACCGCCATACATAAATAGCA

CAAGTTTGAGGTAATGGATAGTGGGAATGAAACTAGCATTATCGAACCGCCAATCTTAAAAGCA

Мутации, ассоциированные с болезнью

- **Эталонный геном**

CAAGTTTGAGGTAATGGATAGTGGGAATGAAAGTAGCATTATCCAACCGCCAAACTTAAATAGCA

- **Геном пациентов**

CAAGTTTGTGGTAATGGAAGTGGGAATGAAAGTAGGATTATCCAACCGCCAAACTTAAATAGCA

CTAGTTTGAGGAATGGATAGTGGGAATGAAAGTAGCATTATCCAACCGCGAAACTTAAATAGCA

CATGATTGAGGTAATGGATAGTGGGAATGAAAGTAGCATTATCCAACGGCCAAACTTAAATAGCA

CAAGTTTGAGGTAATGGATAGTGGGAATGAAAGTAGCATTATCGAACCGCCATACATAAATAGCA

CAAGTTTGAGGTAATGGATAGTGGGAATGAAAGTAGCATTATCGAACCGCCAATCTTAAAAGCA

- **Геном здоровых людей**

CTAGTTTGTGGTAAAGGATAGTGGGAATGAAAGTAGCATTATCCAACCGCCAAACTAAAATAGCA

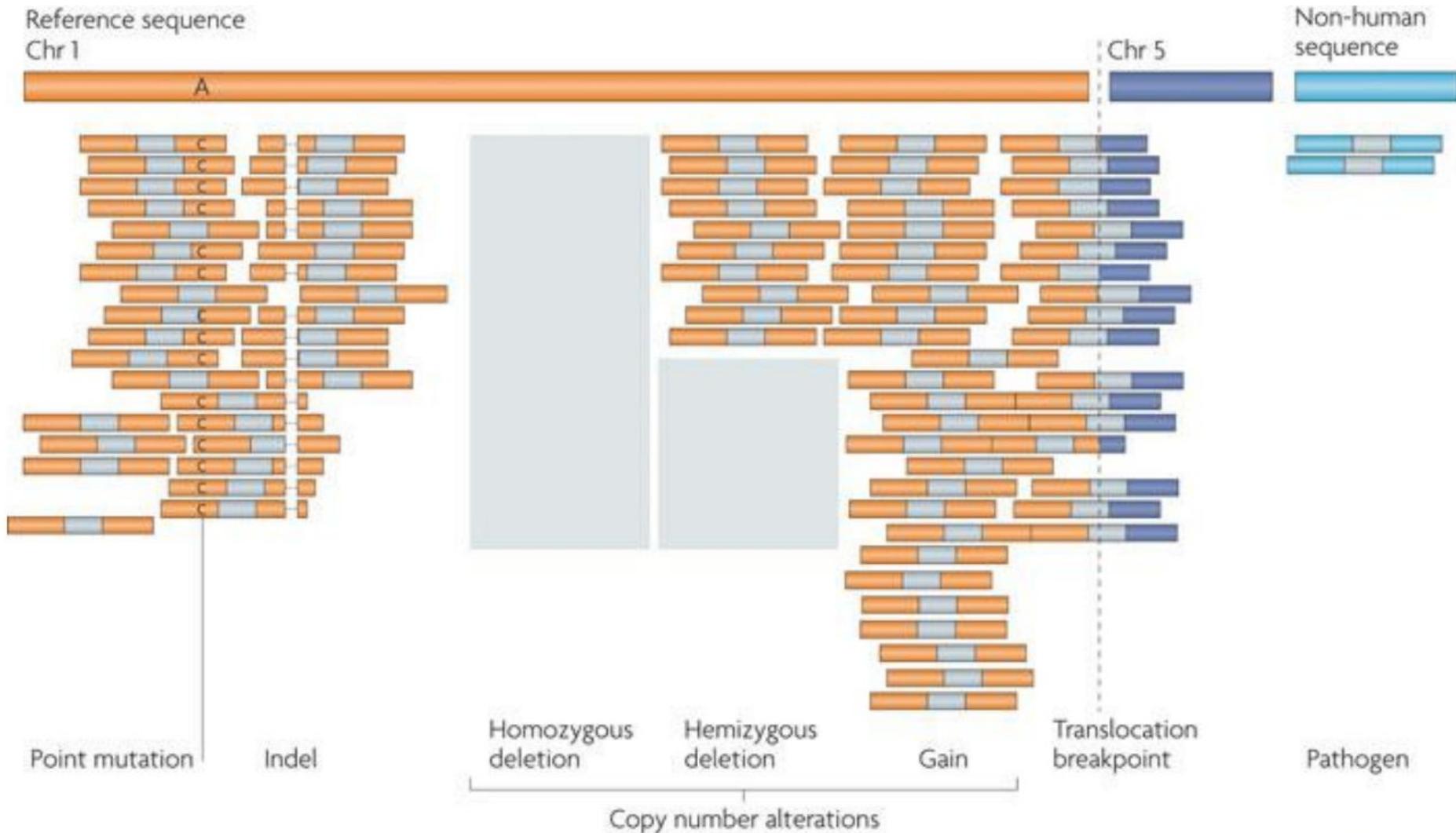
GAAGTTTGAGCSTAATGGATAGTGGGAATGAAAGTAGCATTATCCATCCGCCAAACATAAATAGCA

GAAGTTTGAGGTAATGCATAGTGGGAATGAAAGTAGCATTATCCTACCGCCAAACTTAAAAGCA

CAAGTATGAGGTTATGGATAGTGGGAATGAAAGTACCATTAACCAACGGCCAAACTTAAATAGCA

CAAGATTGAGGTAATGGATAGTGGGAATGAAAGTAGCATTAAACCAACCGCGAAACTTAAATACCA

Раковые мутации



Раковые мутации

Нормальный геном:

Chromosome 4



Chromosome 20

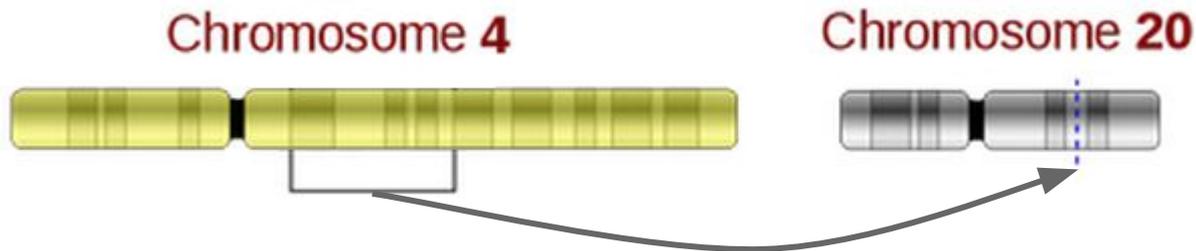


Раковые мутации

Нормальный геном:



Раковый геном:

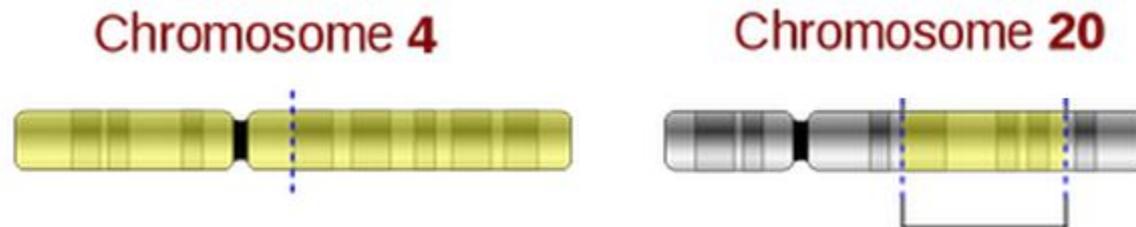


Раковые мутации

Нормальный геном:

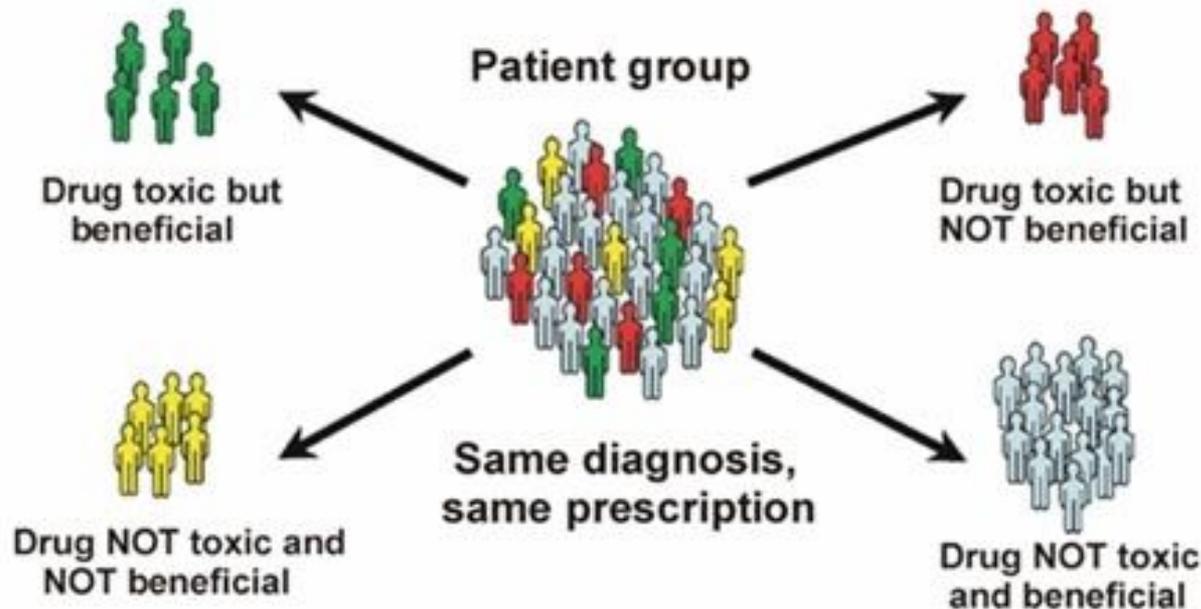


Раковый геном:



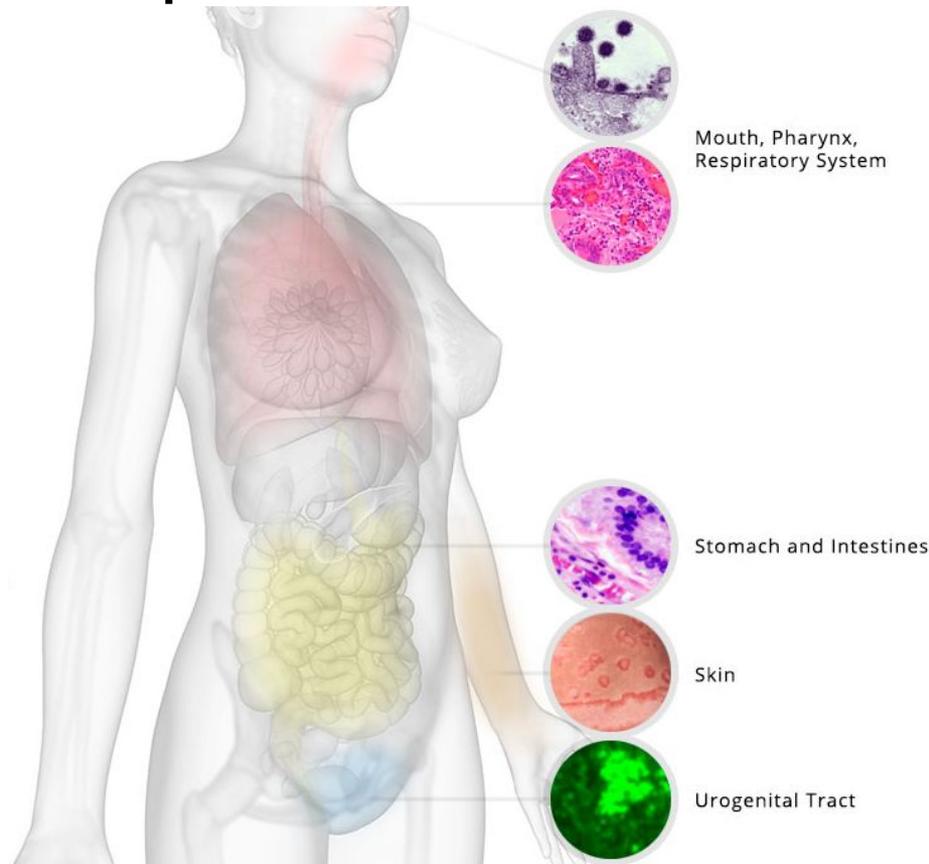
Персональная медицина

- У разных людей разная реакция на лекарства
- Назначать их стоит соответственно



Микробиом человека

- Человеческий организм содержит ~10000 типов бактерий



Микробиом человека

- Человеческий организм содержит ~10000 типов бактерий
- Изменения микробиоты могут влиять на здоровье
- Сложен для секвенирования

Как секвенировать микробиом?

Как секвенировать бактерию?

Как секвенировать бактерию?

Секвенирование требует большого количества генетического материала

Секвенирование бактерий

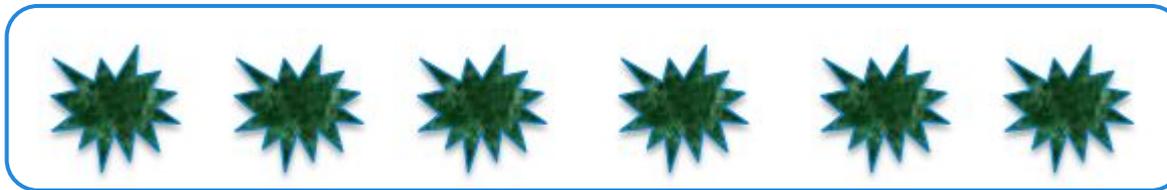
- Классическое секвенирование



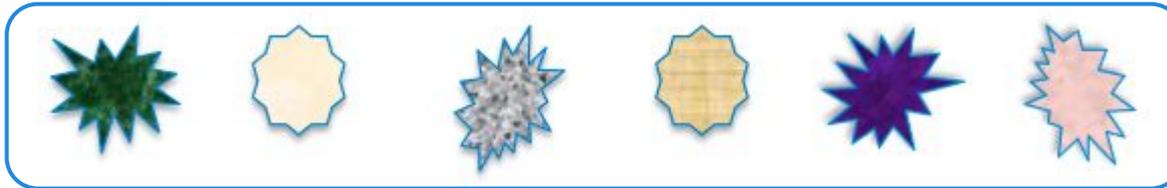
- >99% бактерий невозможно культивировать

Секвенирование бактерий

- Классическое секвенирование



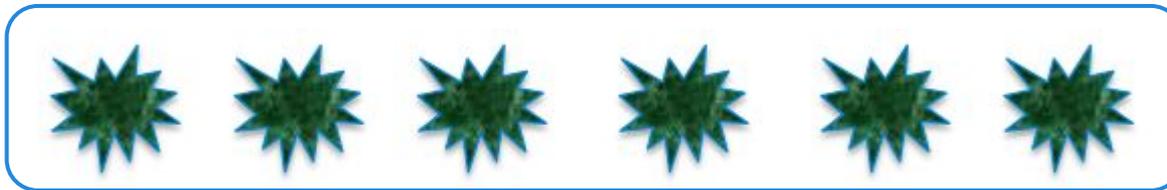
- Метагеномика



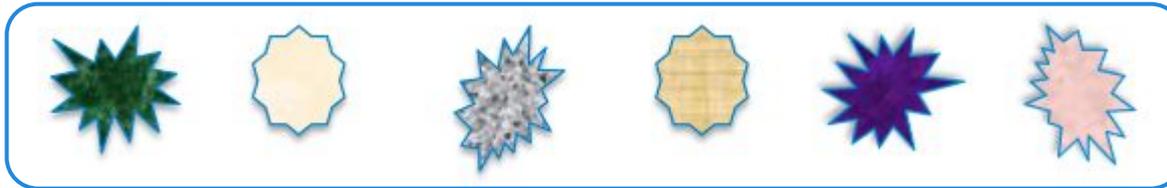
- Данные от сотен разных геномов в одном файле
- Трудно собрать и классифицировать

Секвенирование бактерий

- Классическое секвенирование



- Метагеномика



- По единичной клетке



- Необходима амплификация генома

Применения секвенирования

- Клинические исследования
 - Диагностика болезней
 - Геномика рака
 - Персональная медицина
 - Микробиота
- **Фрамацептика**
 - Разработка новых лекарств
 - Исследование устойчивости к лекарствам
- Криминалистика
- Исследование древней ДНК

Применения секвенирования

- Клинические исследования
 - Диагностика болезней
 - Геномика рака
 - Персональная медицина
 - Микробиота
- Фрамацептика
 - Разработка новых лекарств
 - Исследование устойчивости к лекарствам
- Криминалистика
- Исследование древней ДНК

Sequencing applications

- Экология
 - Метагеномика различных сред
- Сельское хозяйство
 - Большие геномы растений
- Геномика микробов
- Геномика грибов
- Эволюционные исследования

Пример №3

Изучение эволюции

Как сравнить два генома?

Как сравнить две книги?

- Два издания одного учебника

Как сравнить две книги?

- Два издания одного учебника
- Учебник для старших классов и для ВУЗа

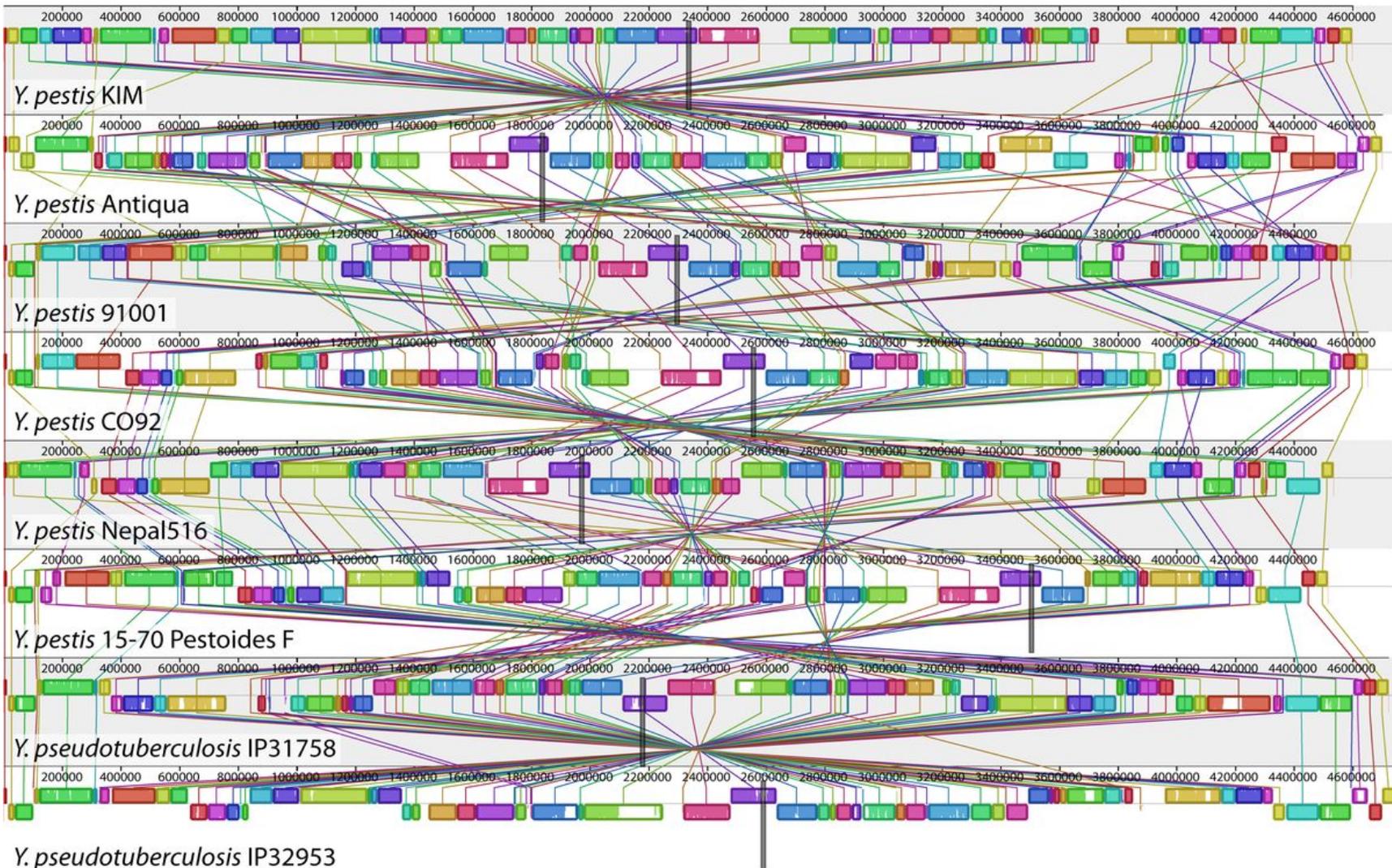
Как сравнить две книги?

- Два издания одного учебника
- Учебник для старших классов и для ВУЗа
- Учебник истории написанный при разных политических режимах

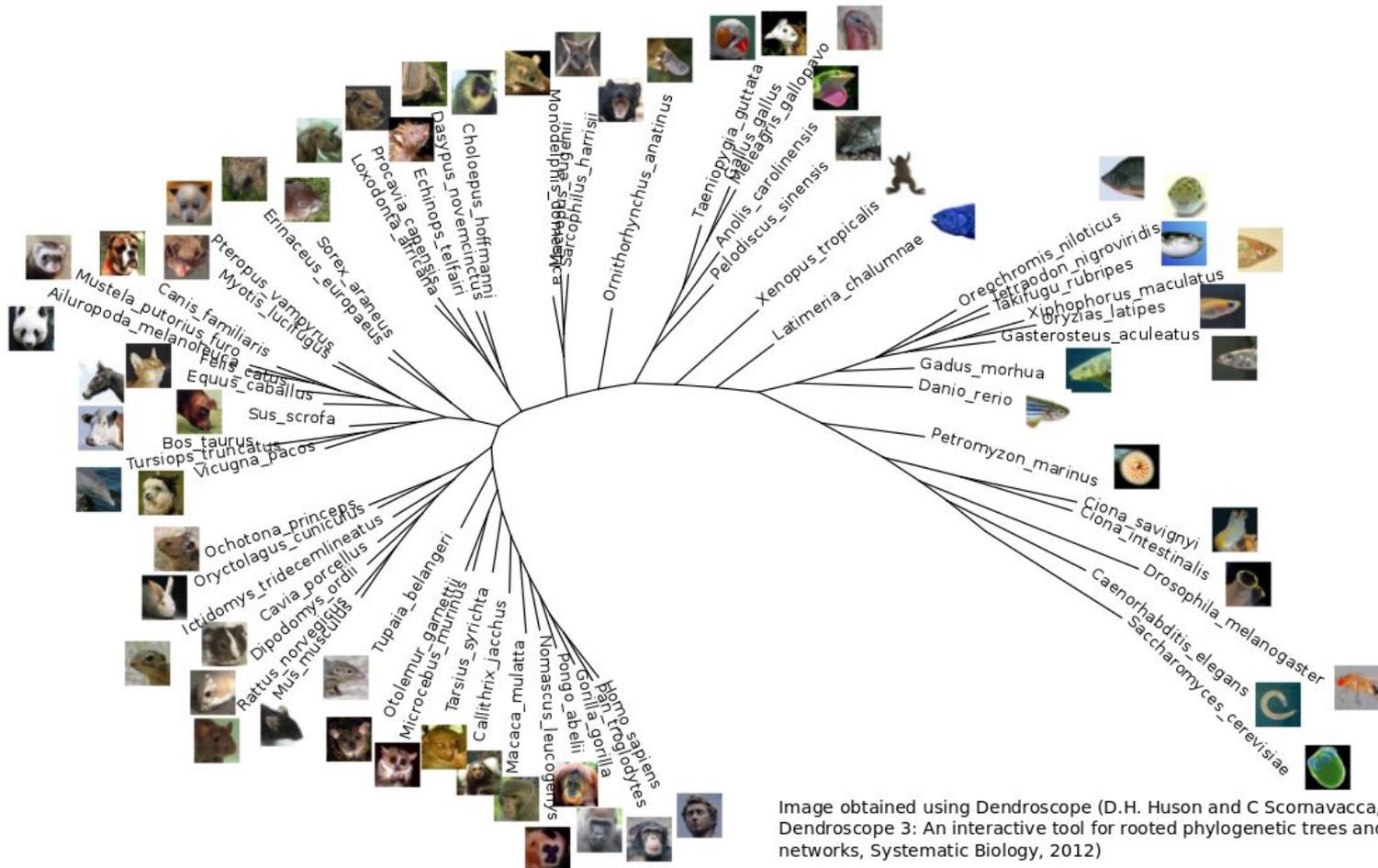
Как сравнить две книги?

- Два издания одного учебника
- Учебник для старших классов и для ВУЗа
- Учебник истории написанный при разных политических режимах
- Учебники по двум

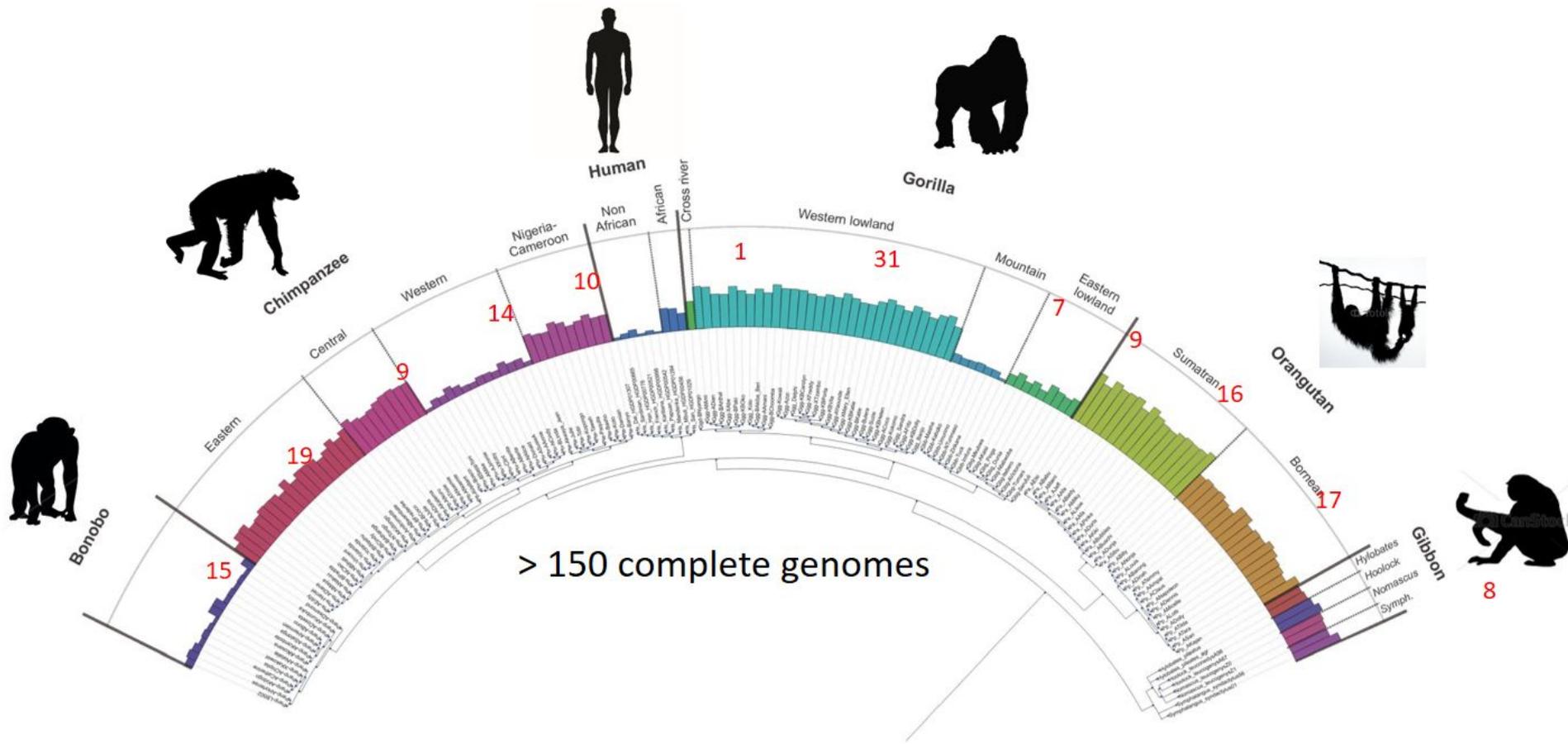
Сравнительная геномика



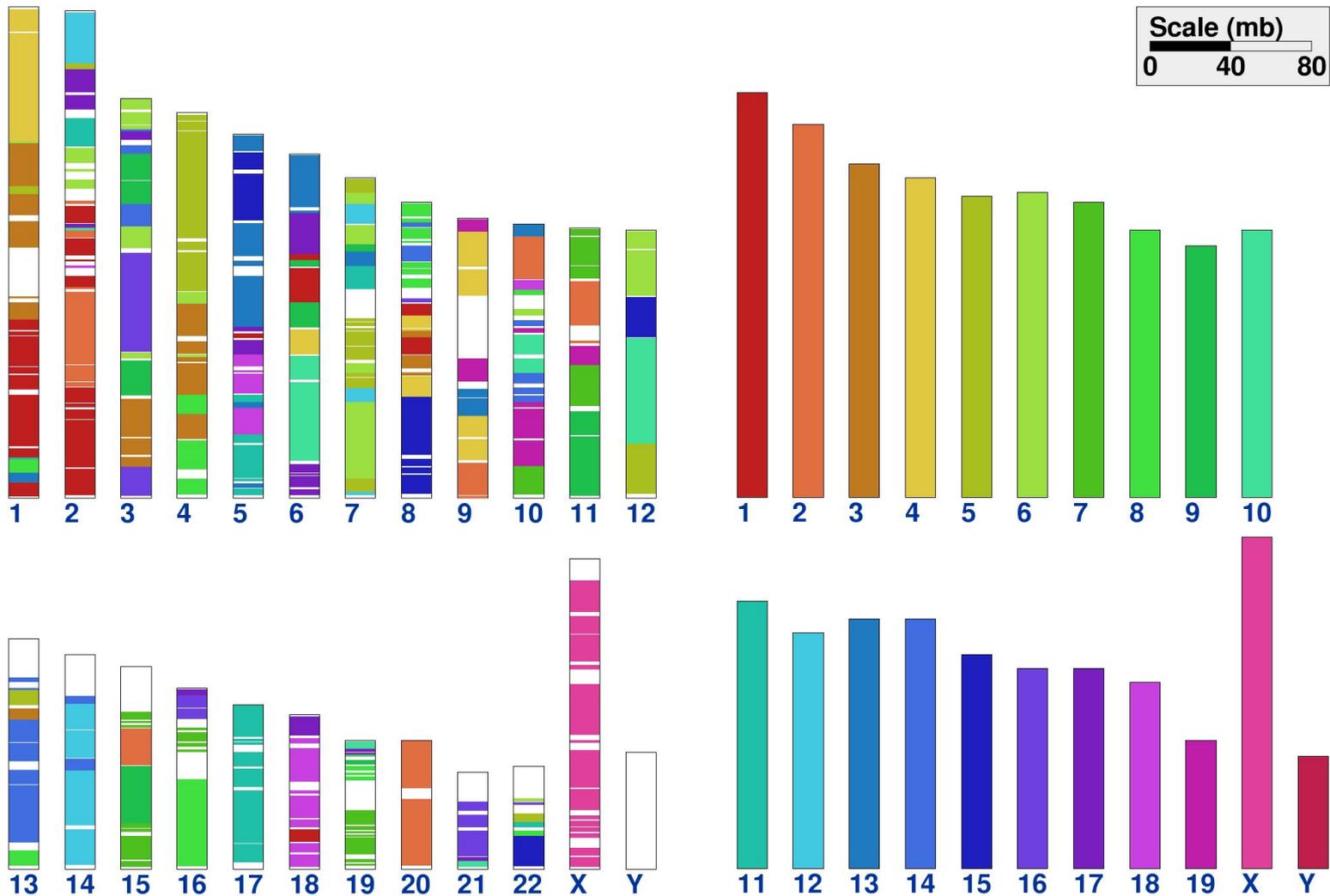
Филогенетика



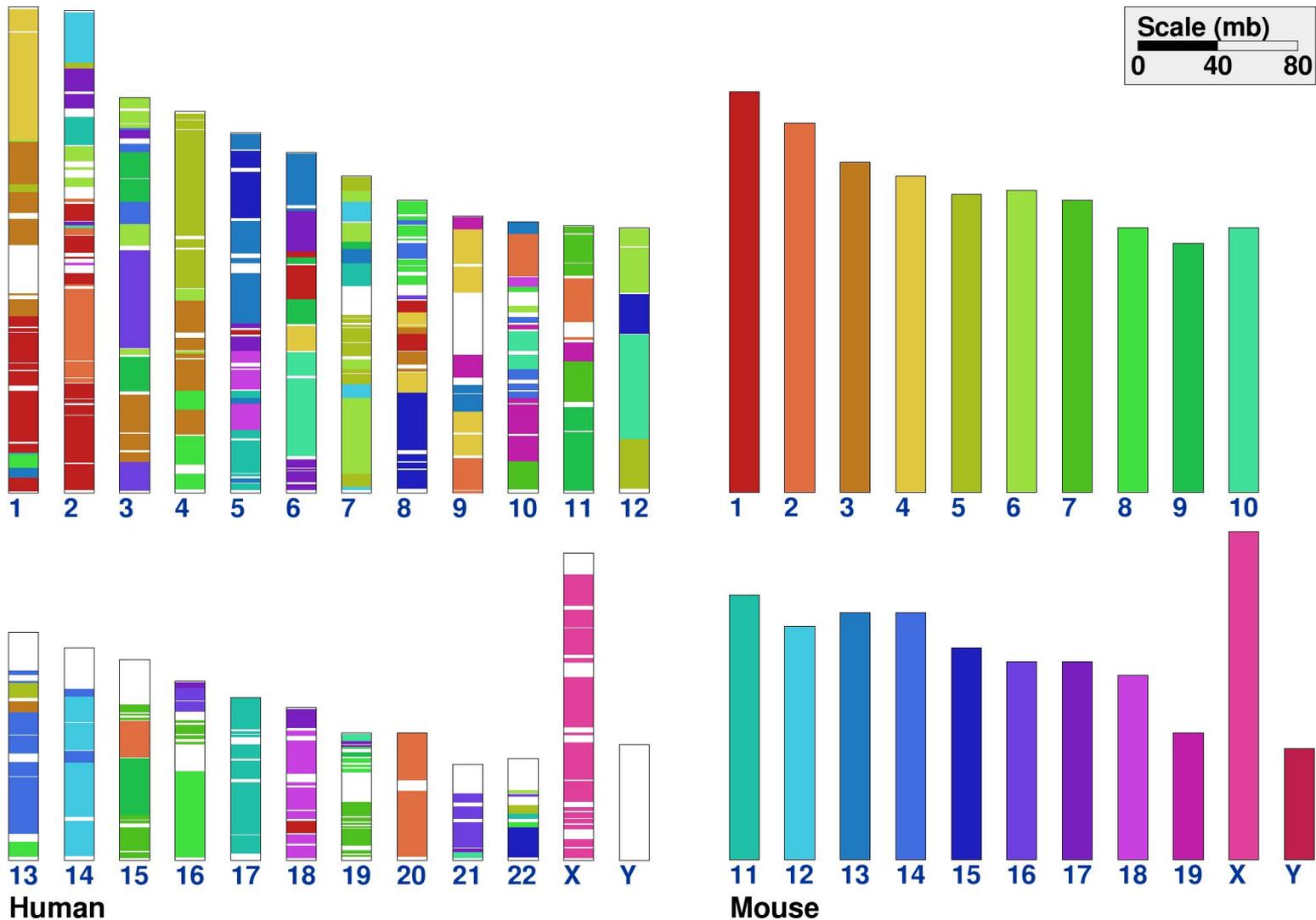
Филогенетика



Хромосомная карта



Хромосомная карта



Спасибо за внимание!

Вопросы?

ap@bioinf.spbau.ru