



# Методы секвенирования ДНК.

*Васильев Геннадий Владимирович*

Институт цитологии и генетики СО РАН



## Немного истории



1977 г. - Первый полный геном бактериофага Ф-Х174 (5386 нуклеотидов) секвенирован методом Shotgun

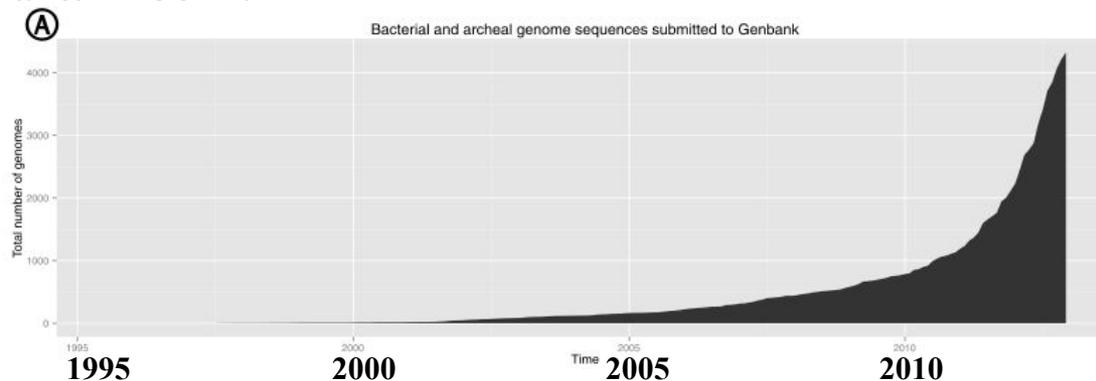
1995 г. – секвенирование генома первого свободноживущего организма – бактерии *Haemophilus influenzae*.

Проект «Геном человека» - 1990 – 2003 - 2006 гг., стоимость -15 млрд.\$.

Секвенировано 92,3% генома.



2010 г. – завершён проект «тысяча геномов», для каждого персонального генома человека секвенировано около 85% последовательности.





## «Расшифровка» генома



### Характерные размеры геномов

вирус папиллом человека - 8 т.п.о

*Escherichia coli* - 4900 т.п.о

*Saccharomyces cerevisiae* – 12 млн п.о.

*Arabidopsis thaliana* – 101 млн.п.о

*Triticum aestivum* – 16 млрд. п.о.

Человек – 3 млрд.п.о.



Розеттский камень

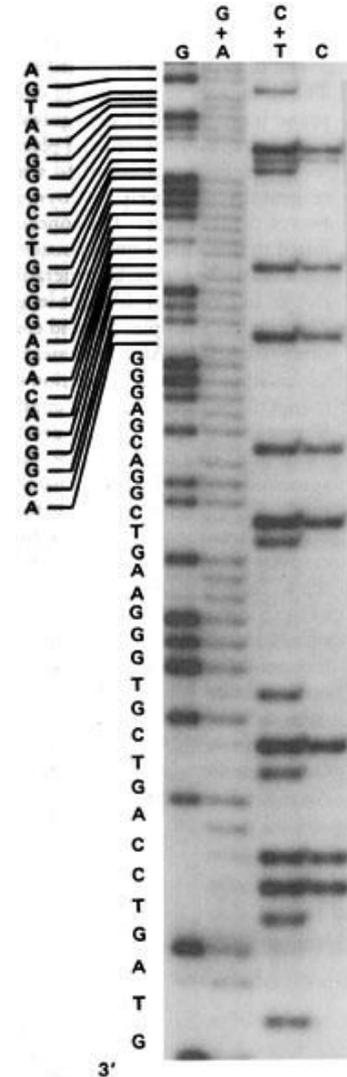
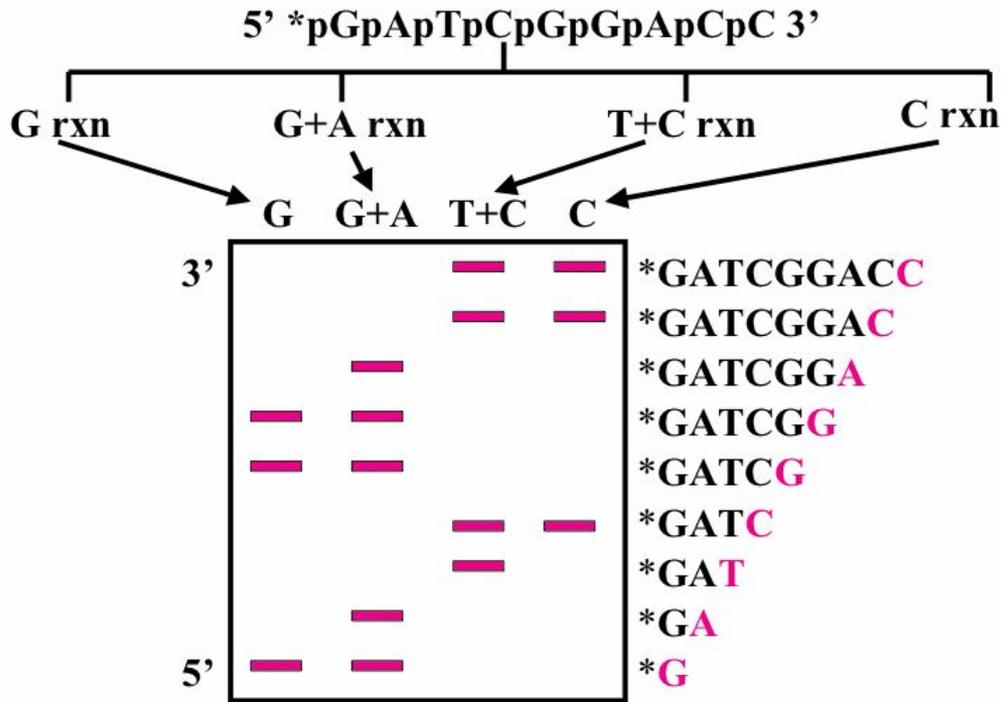


# Секвенирование по Максому-Гилберту



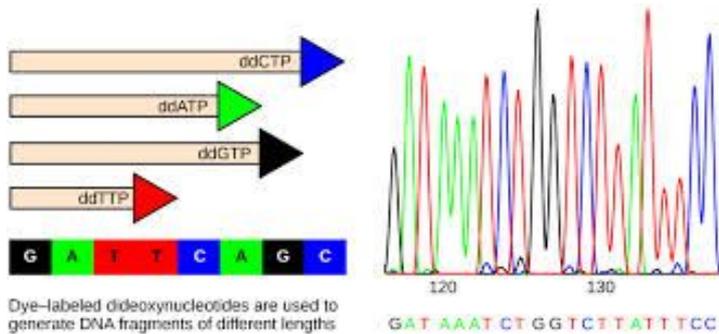
Химическая деградация ДНК

## Maxam-Gilbert sequencing

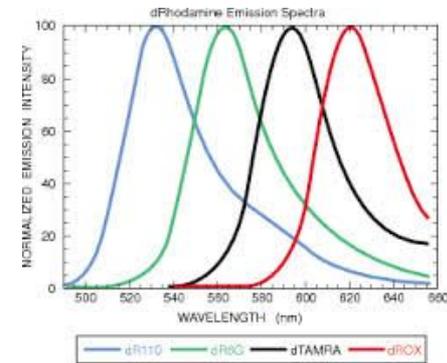




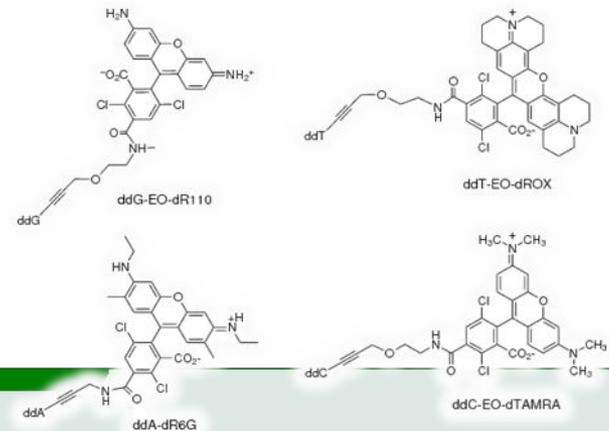
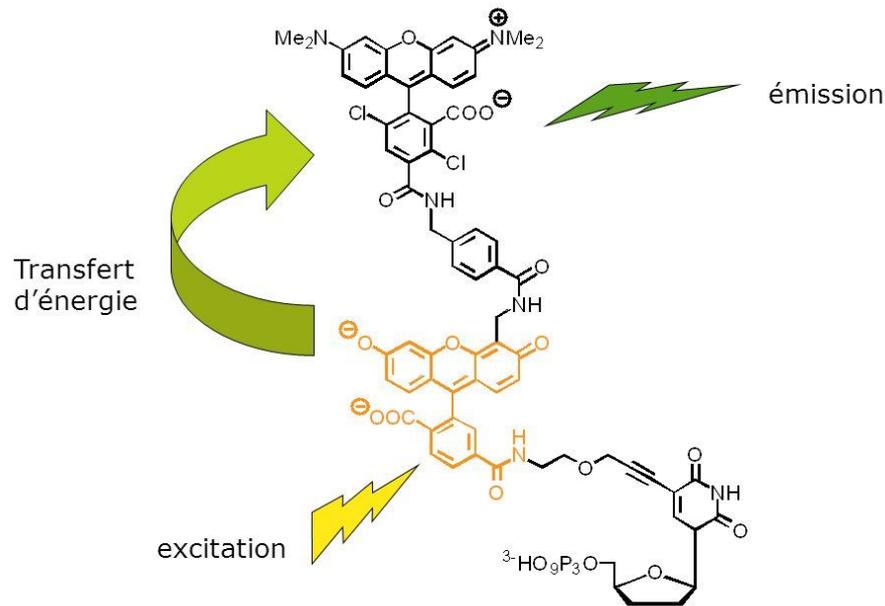
# Секвенирование методом Сэнгера



Реакция Сэнгера – не ПЦР, а многократный синтез с одной матрицы



## ddT-BigDye terminator





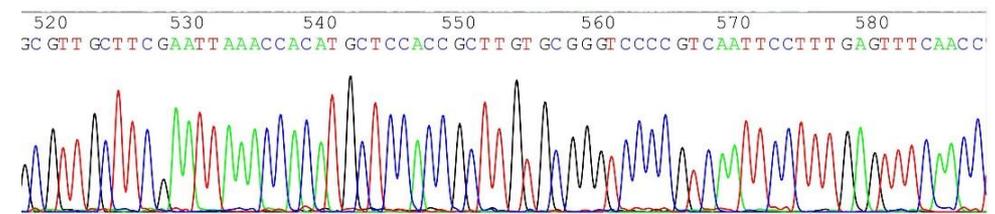
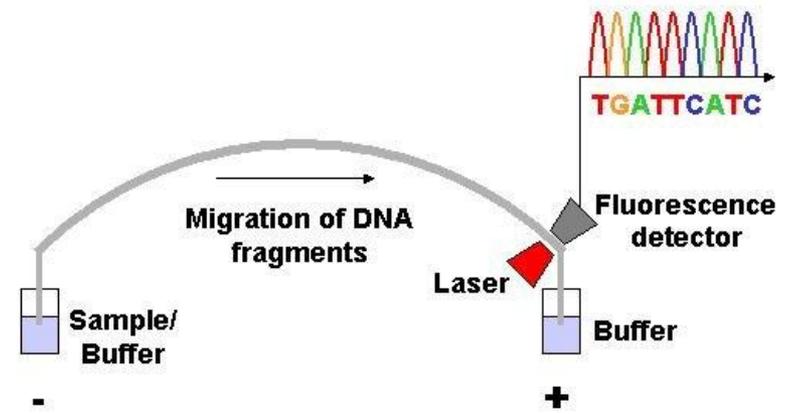
# Первое поколение секвенаторов - капиллярные



96-capillary ABI 3730XL



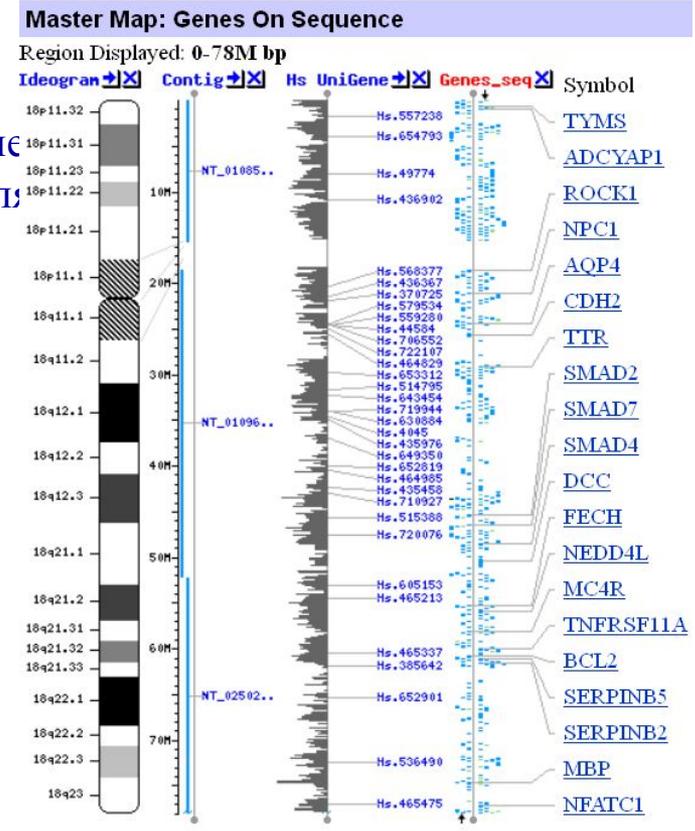
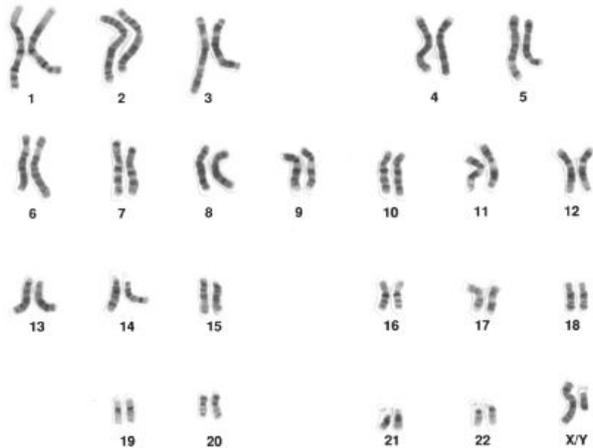
Нанофор 05 – российская модель





# Начало проекта по секвенированию эукариотического генома

- А) Выбор объекта
- Б) Подробная биологическая характеристика объекта, изучение его жизненного цикла, их особенностей, отбор варианта для получения и поддержания материала.
- В) Подробное кариотипирование.
- С) Выбор стратегии секвенирования





## Стратегия секвенирования генома человека

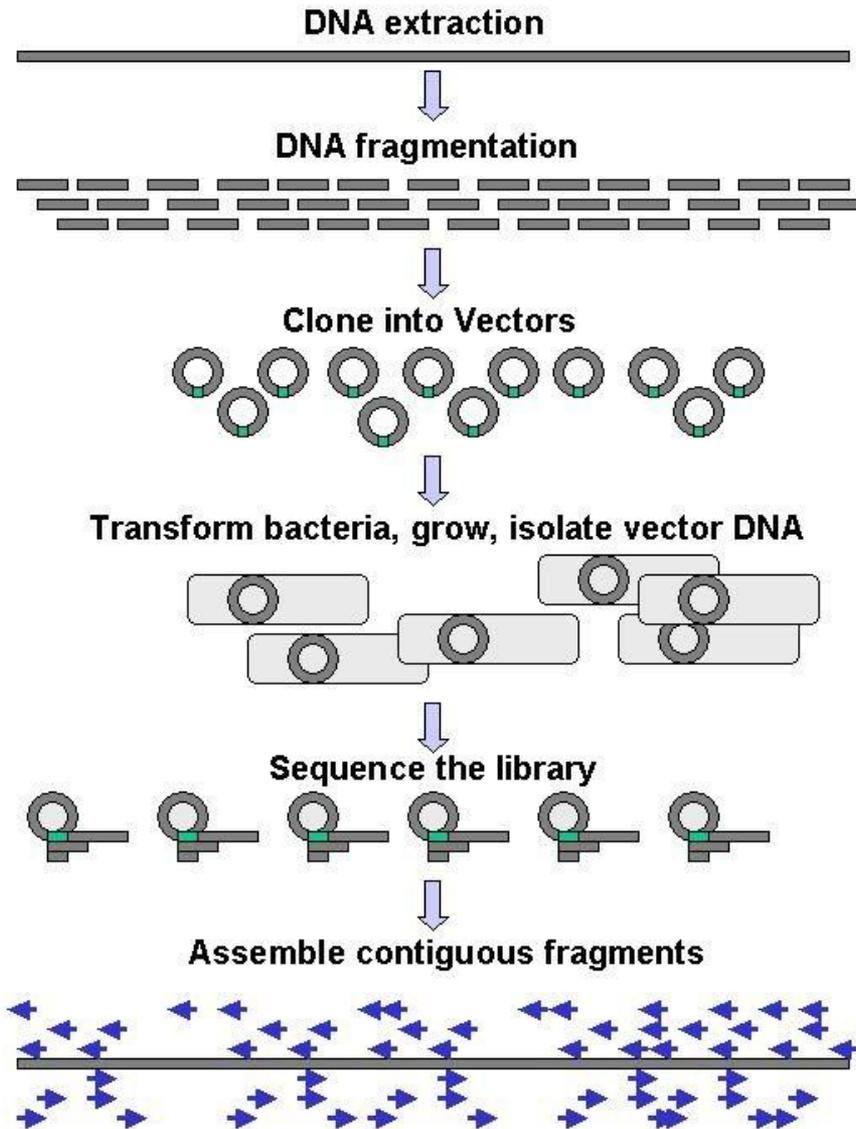


Проект «Геном человека» - 1990 – 2003 - 2006 гг.,  
стоимость – от 3 до 15 млрд.\$.  
Секвенировано 92,3% генома.

- 1) 23 хромосомы человека – 23 подпроекта по каждой
- 2) Субклонирование больших фрагментов генома в ВАС-библиотеках встройки 40 000 – 200 000 b.p. (YAC, дрожжевые – до 1 Mb, Cosmid, фаговые – до 45 kb). Суммарно получено 22,000 ВАС - клона
- 3) отдельный Shotgun для каждого клона ВАС (40 000 – 200 000 b.p.).



# Упрощённая схема процесса секвенирования генома методом Shotgun



Ферментативные методы фрагментации неприменимы для протяжённых районов.

Необходима комбинация библиотек с различным размером встоек.

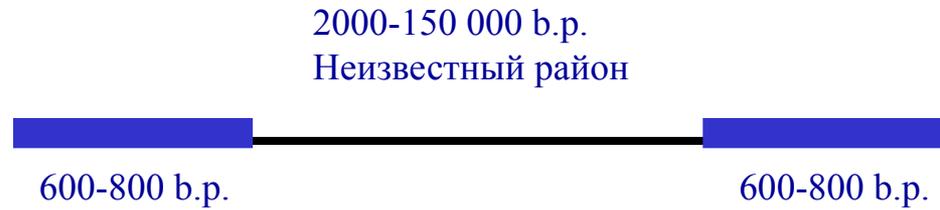
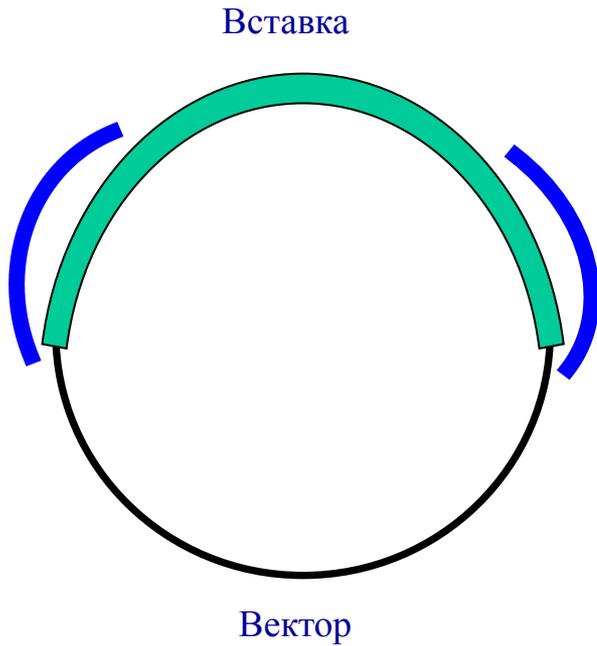
Необходим высокопрофессиональный биоинформатический анализ.



# Подробнее



Используемые векторы – ранее M13 phagemid  
Вентер – производная pBR322,  
Говорун - pCR2.1 (Invitrogen)  
Лейшмания - pUC18  
Для больших встроек (30-40 kb) – fosmid вектор pCC1FOS





## Whole genome shotgun



1998г. – старт проекта секвенирования первого индивидуального генома человека. Крейг Вентер, компания «Celera Genomics»

Использование метода Shotgun без предварительного деления генома на фракции

Преимущественно для бактерий и эукариот с малым геномом (простейшие, грибы).

Набор библиотек – обычно 2, 10, 50, редко до 150 kb

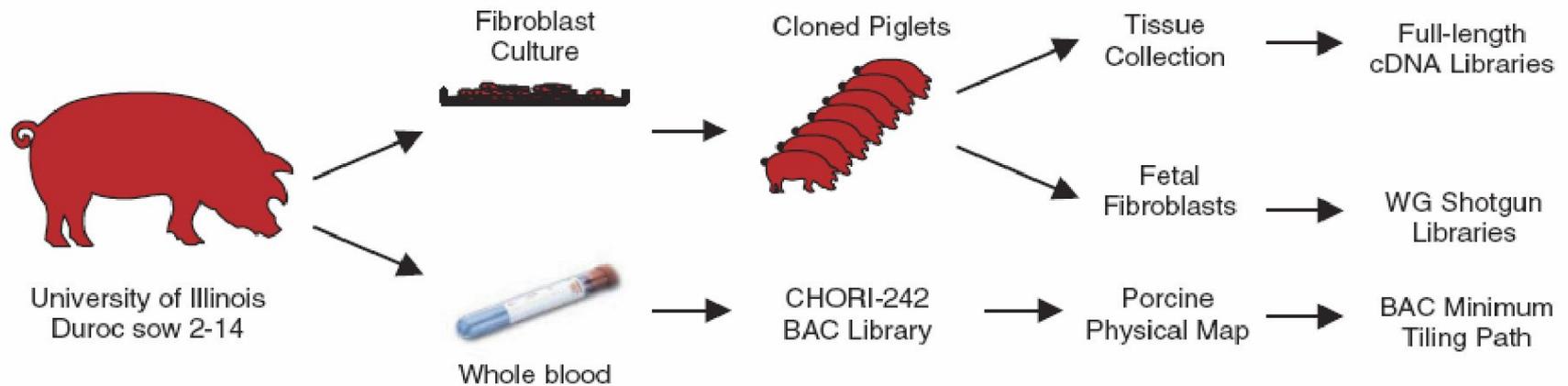
Геном Wenter'a – библиотеки со встройками 2 000 – 300 000 b.p.

Два варианта организации работы – только shotgun с покрытием по геному 6-7 х. Бык – 6х, собака – 7.6х, лошадь – 6.8х, слон 7х *etc.*

Сочетание shotgun и pair-mate библиотек пиросеквенатора Roche.  
Кошка – 2х, кролик – 2х, тупайя – 2х, серый лемур 2х *etc.*



# Схема организации геномного проекта на примере генома свиньи (2.6 Gb)

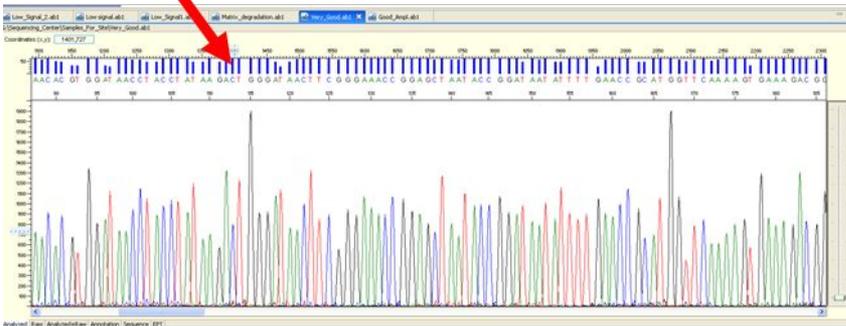


- А) Секвенирование транскриптома
- Б) Секвенирование Shotgun-библиотек с расчётным 3х покрытием, использование библиотек с встройками 3 kb, 10 kb и 50 kb
- В) Создание и секвенирование ВАС-библиотек с расчётным 3х покрытием
- С) Комбинирование данных при биоинформатическом анализе

Schook *et al.*,  
2005



# Phred quality score

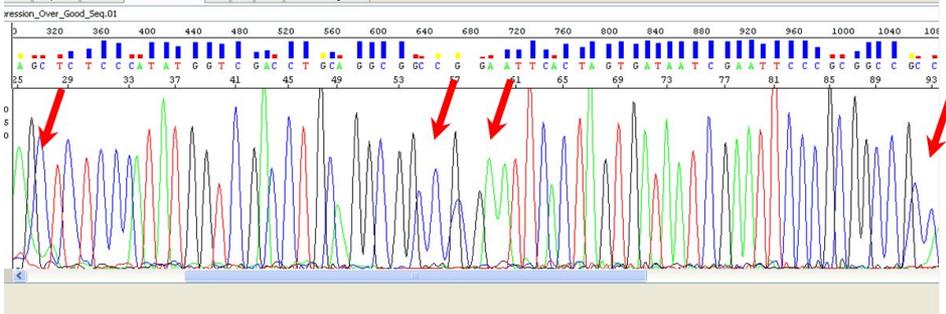


$$Q = -10 \log_{10} P_i$$

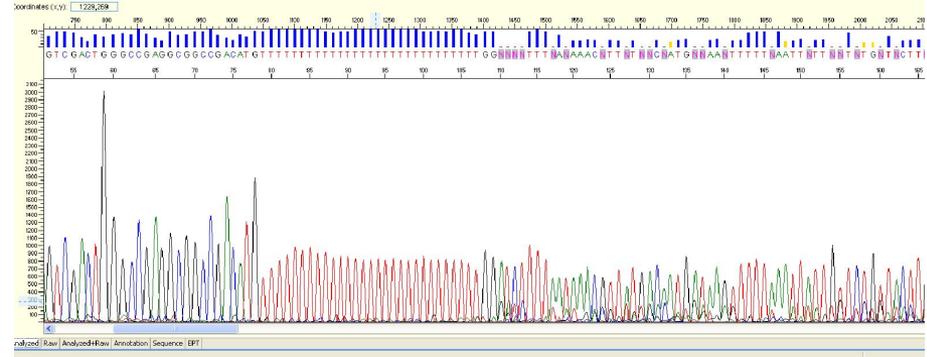
Phred 10 = 90% точность, Phred 20 = 99%, Phred 30 = 99,9%,



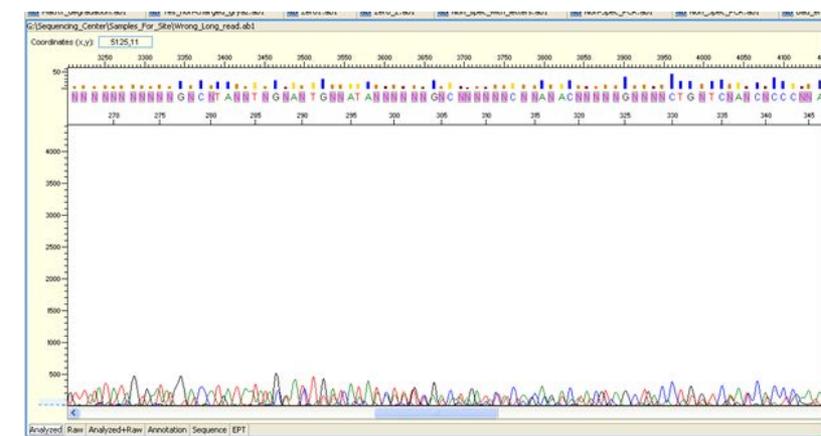
# Примеры ошибок и их соотнесение с Phred



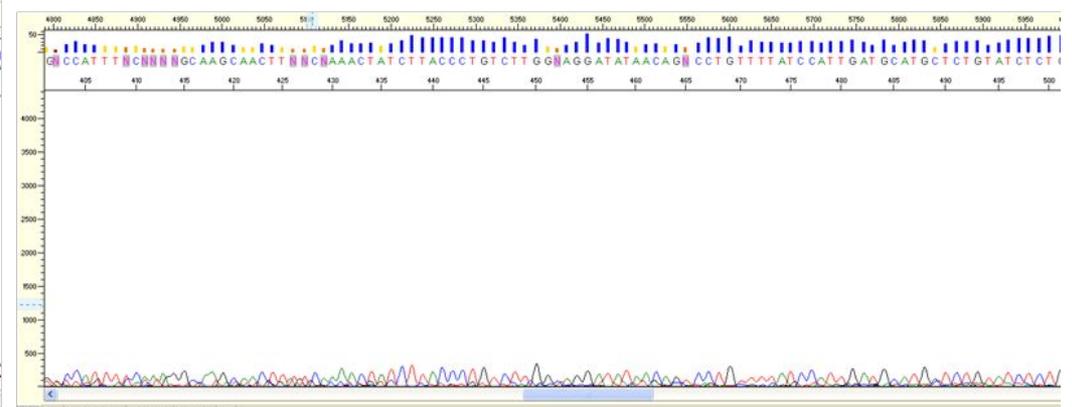
Компрессия



Проскальзывание полимеразы



Химерная последовательность - начало



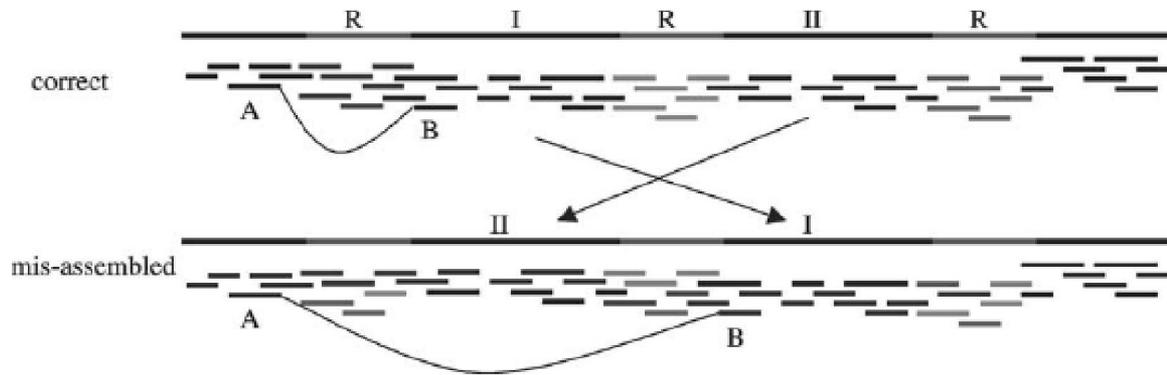
конец



# ДНК не является случайной последовательностью.



Основную часть ДНК в эукариотических геномах составляют повторы. Это делает крайне затруднительным правильное ассемблирование одиночных последовательностей в длинные геномные контиги.



Часть последовательностей ДНК выпадает при клонировании/секвенировании, образуя не заполняемые разрывы (gap).

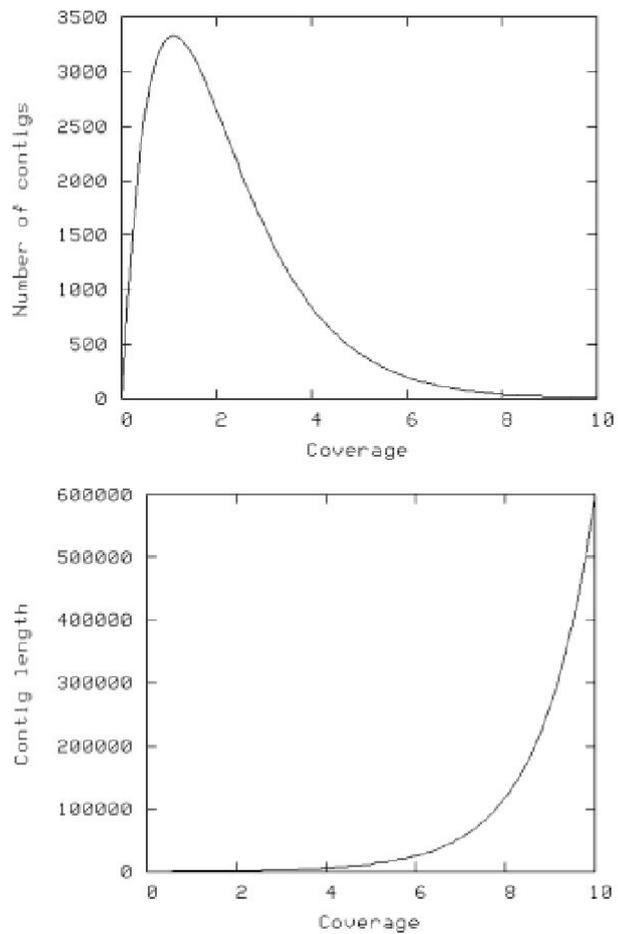




# Зависимость числа и длины контигов от покрытия



## SHOTGUN SEQUENCE ASSEMBLY



Разрывы в контигах образуются и по чисто статистическим причинам.

FIG. 3. Dependence of number and length of contigs on genome coverage.



# Нелинейность в выборе оптимальных параметров Shotgun



Для экономии средств в ходе проекта необходим корректный выбор среднего размера встоек. Кроме того, стремление к максимальной длине прочтения в ущерб количеству прочтений может оказать строго обратный эффект.

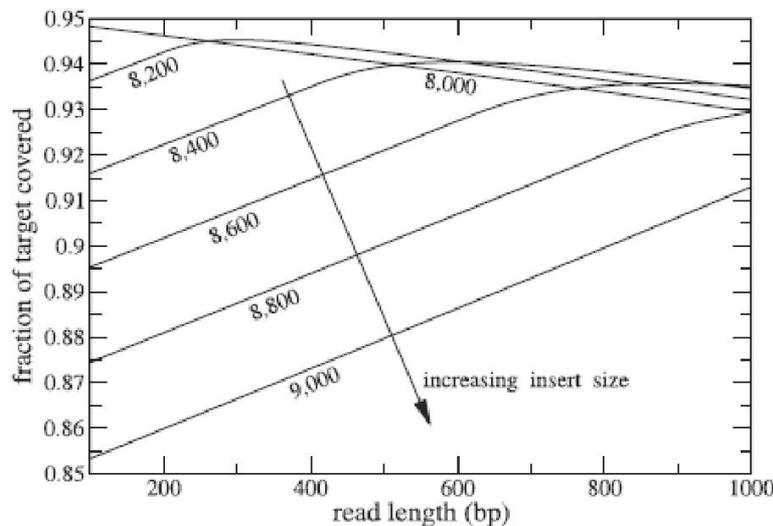


FIG. 6. Coverage response in the parameter neighborhood of the optimum for a 16-kb target at threefold sequence redundancy for a series of insert lengths.

Для проекта секвенирования генома лейшмании (35 Mb) [Laurentino et al., 2004] библиотекой на основе pUC18 со встройкой 2 kb покрыто 15% генома

Изменения в степени покрытия (%) для 16-kb района при 3-х кратном покрытии (48 kb суммарных данных) для различных величин клонированных встоек



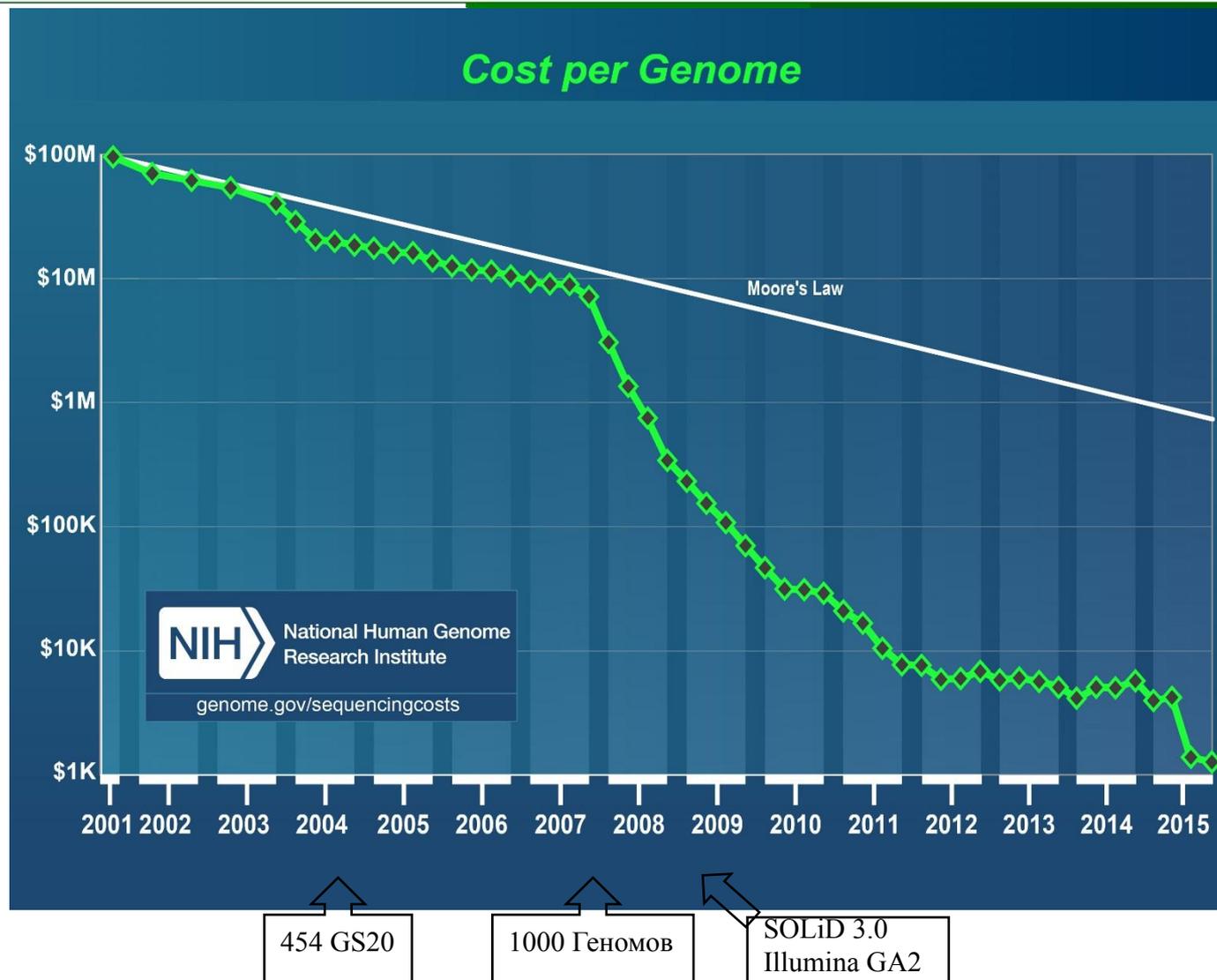
# Этапы биоинформатического анализа



- 1) Собственно сборка секвенированных последовательностей в геном.
- 2) Структурная аннотация, включающая :
  - идентификацию различных элементов генома
  - выявление ORF и их локализация в определённых районах
  - определение экзон-интронной структуры генов, их кодирующих районов и вариантов транскрипции/сплайсинга
  - локализация регуляторных районов гена
- 3) Функциональная аннотация: связь различных районов генома с их биологической функцией, в том числе с биохимическими процессами, заболеваниями, регуляторными процессами в организме в целом и т.п.
- 4) Подтверждение сходства физической карты хромосом и взаимного расположения контигов\скаффолдов



# Революция цены в геномных исследованиях





# Второе поколение секвенаторов – NGS – создатели геномики

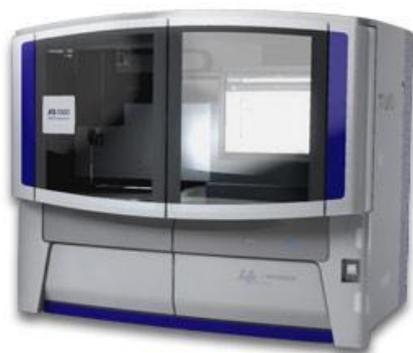


**Roche FLX Titanium**



©2010, Illumina Inc. All rights reserved.

**Illumina HiSeq**



**SOLiD 5500**



**Ion Proton**