

DNA 测序技术的发展

三代测序的优势与应用

王强

December 27, 2016

南京大学生命科学学院

Outline

一代: Sanger 法电泳测序

二代: (短读长的) 高通量测序

三代: (长读长的) 单分子测序

一代: Sanger 法电泳测序

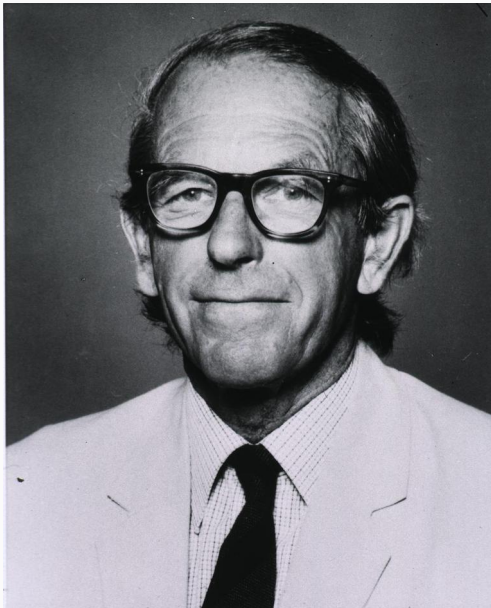


Figure 1. Frederick Sanger

- 1955 年, 第一个蛋白质测序, 胰岛素
- 1958 年, 诺贝尔化学奖
- 1975 年, 双脱氧法, ddNTP
- 1977 年, 第一个基因组, ϕ -X174 噬菌体
- 1980 年, 再度获得诺贝尔化学奖

人类基因组计划

- 1985 年, 美国能源部正式提出人类基因组测序.
- 1990 年, 正式启动人类基因组测序.



Figure 2. 聚丙烯酰胺凝胶电泳

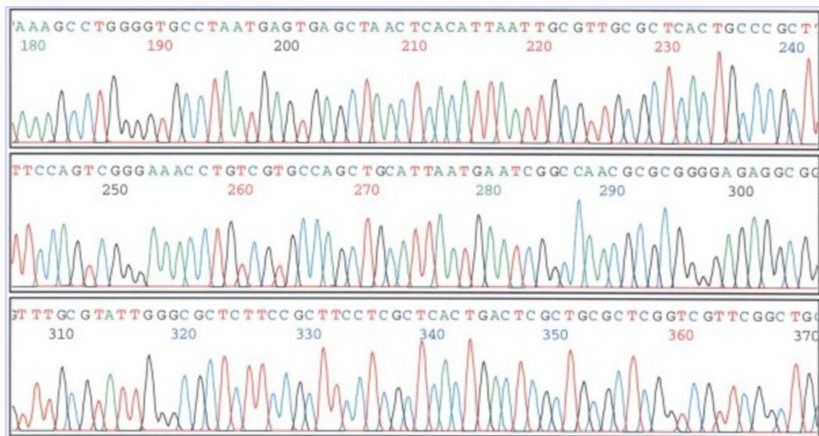


Figure 3. 毛细管电泳图谱



Figure 4. ABI 3730XL



Figure 5. 模板与测序室

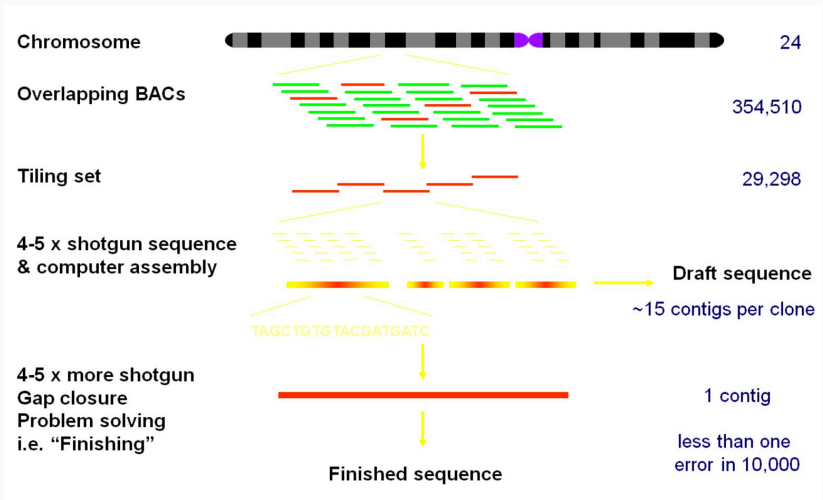


Figure 6. HGP sequencing strategy

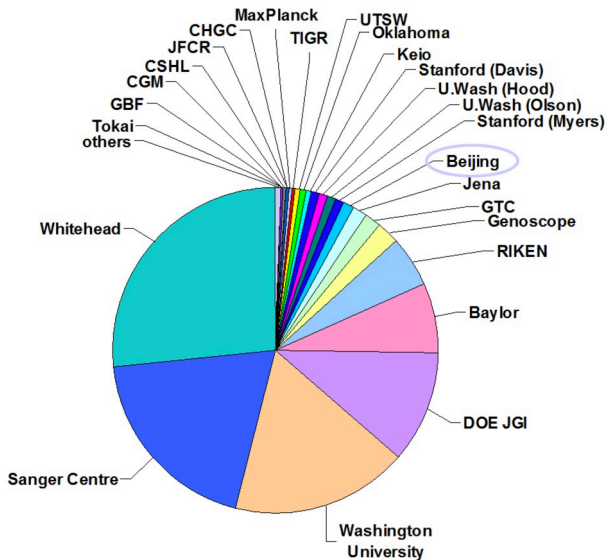
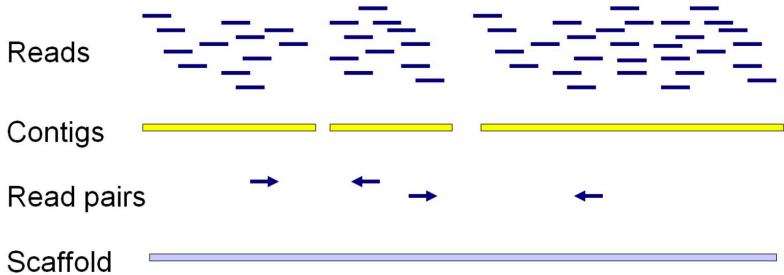


Figure 7. 贡献



Figure 8. J. Craig Venter (文特尔)



Then order scaffolds on the chromosomes
using the HGP clone map and other
publicly available maps

Figure 9. Celera assembly strategy

Enter Symbol

GET CHART

[COMPARE](#)

[EVENTS ▾](#)

[TECHNICAL INDICATORS ▾](#)

[CHART SETTINGS ▾](#)

[RESET](#)

Week of Mar 15, 2010 :

■ CRA 6.92



© [type Function] Yahoo! Inc.

2000

2002

2004

2006

2008

2010

[1D](#)

[5D](#)

[1M](#)

[3M](#)

[YTD](#)

[6M](#)

[1Y](#)

[2Y](#)

[5Y](#)

[Max](#)

FROM: Apr 28 1999

TO: Apr 19 2010



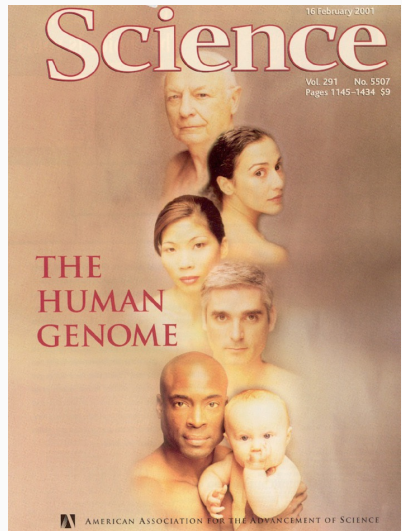
[Print](#) | [Share](#) | [Send Feedback](#)

Disclaimer. - Quotes delayed at least 15 minutes.

Figure 10. Celera stocks



(a) Nature



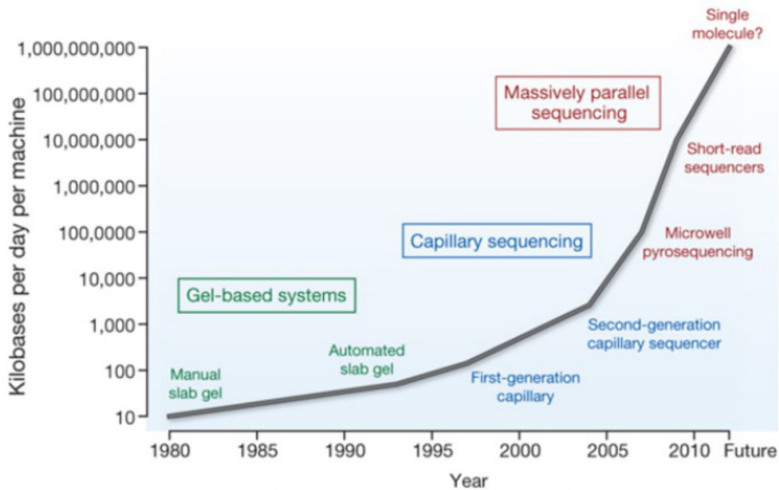
(b) Science

Figure 11. 基因组草图, 2001

二代: (短读长的) 高通量测序


Personal Genomes

- Craig Venter
- James Watson
- Stephen Quake
- George Church
- Marjolein Kriek
- Hermann Hauser
- Han Chinese
- Seong-Jin Kim
- Korean AK1
- Yoruban African NA18507
- 14 others sequenced by Complete Genomics
- Unknown number sequenced by Knome
- 6 genomes sequenced at high depth by the 1000 Genomes Project
- 180 genomes sequenced at low coverage by the 1000 Genomes Project
- Two acute myeloid leukemia patients



MR Stratton *et al. Nature* **458**, 719-724 (2009)

Figure 12. 测序能力的增长



NEW HiSeq 2500
Remarkable speed
and flexibility.

MiSeq
Simplicity, integration,
and ease-of-use.

[LEARN MORE](#)

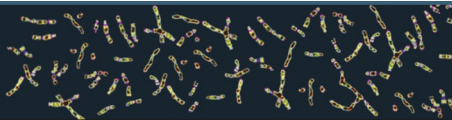
**Illumina announces
speed and
performance
enhancements.**

Introducing the HiSeq 2500 and Triple the
Output on MiSeq.

Figure 13. Illumina HiSeq

1000 Genomes

A Deep Catalog of Human Genetic Variation



[Home](#) [About](#) [Participants](#) [Data](#) [Contact](#) [Wiki](#)

1000 GENOMES PROJECT DATA RELEASE

SNP data downloads and genome browser representing four high coverage individuals

The first set of SNP calls representing the preliminary analysis of four genome sequences are now available to download through the [EBI FTP site](#) and the [NCBI FTP site](#). The README file dealing with the FTP structure will help you find the data you are looking for.

The data can also be viewed directly through the 1000 Genomes browser at <http://browser.1000genomes.org>. Launch the browser and [view a sample region here](#).

More information about the data release can be found in the [data section](#) of this web site.

Download the 1000 Genomes Browser Quick Start Guide

[Quick start \(pdf\)](#)

PRESS RELEASE

WEDNESDAY JUN. 11, 2008

[Three Sequencing Companies Join 1000 Genomes Project](#)

TUESDAY JAN. 22, 2008

[International Consortium Announces the 1000 Genomes Project](#)

LOG IN

Username:

Password:

([I forgot my password](#))

LINKS



[Download the meeting report](#)



[View the participants](#)

Figure 14. 千个基因组计划

What's in the NCBI FTP site?

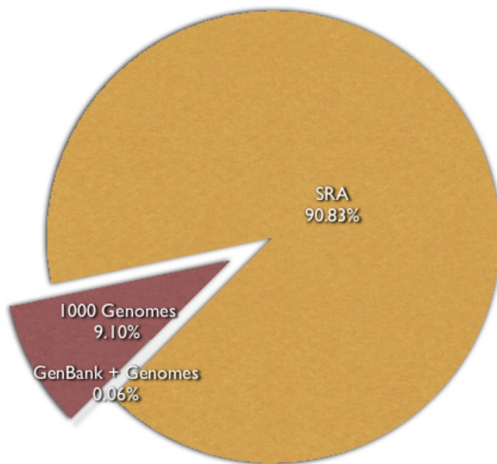


Figure 15. 爆发性增长的数据量

Cost per genome

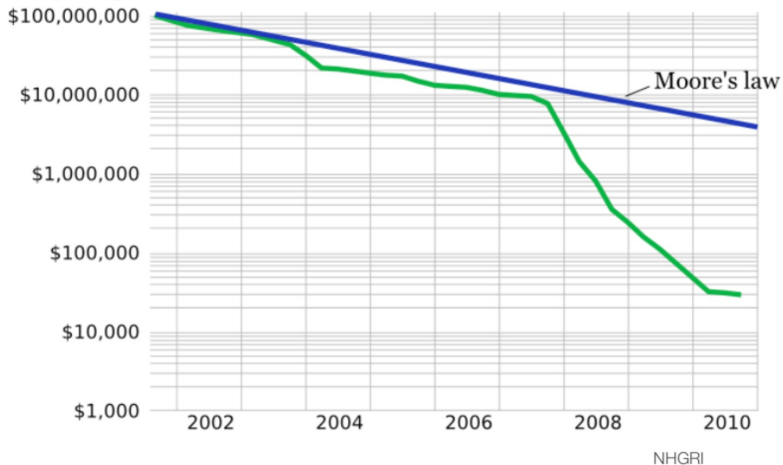


Figure 16. 比摩尔定律更快



Figure 17. HiSeq X

100 GB \approx 33X Human genome \approx 10,000 CNY

别再升级通量了, 测序仪卖不出去了.

CEO, 侬脑子哇特了?

其它二代方法

- Ion Torrent
- 454
- SOLiD

在最近与彭博社的一次采访中, 奥巴马透露了自己想要进入科技风投界的想法.

我与硅谷以及风投的交流, 极大地满足了我对科学和组织的兴趣... 你可以只花几千美元, 而不是十万美元, 就可以把个人的基因绘制出来. 你可以有能力辨识自己的(基因)倾向, 去生产对你这个个体而言最有效的药物. 这只是可以让我坐下来, 与别人谈几个小时的科技创业的例子之一.

三代: (长读长的) 单分子测序

二代的缺点

- 建库过程中扩增带来的偏性
- 高 GC 区域的覆盖度与准确性
- 读长短
 - ▶ 转座子 (transposons and retrotransposons)
 - ▶ 片段重复 (tandom or segment duplications)
 - ▶ 一般不超过 5 kbp

- <http://allseq.com/providers/>
- <http://www.pacb.com/products-and-services/service-providers/>



Providers

Platforms

- ▶ 454 (Roche)
- ▶ Illumina
- ▶ Ion Torrent
- ▶ Oxford Nanopore
- ▶ Pacific Biosciences
 - ☒ Sequel
 - ☒ RSII
- ▶ Sanger
- ▶ SOLiD (Thermo)

Certifications

- ☐ CLIA
- ☐ CAP
- ☐ CSPro
- ☐ HIPAA

Sequencing

APPLY FILTERS

CLEAR FILTERS

Search Providers



Provider Name ^	Location ^	Commercial/Non-profit	Sequencing ^	Bioinformatics ^
Admera Health	N.J, USA	Commercial	✓	✓
DNA Link	Korea	Commercial	✓	✓
Earlham Institute	UK	Non-profit	✓	✓
Macrogen Clinical Laboratory	USA	Commercial	✓	✓
Mount Sinai School of Medicine Genomics Core Facility	USA	Non-profit	✓	✓
Theragen Etxex Bio	South Korea	Commercial	✓	✓
University of Maryland – Genomics Resource Center	USA	Non-profit	✓	✓

Page 1 of 1

Providers Per Page: 10 20 50

Figure 18. 商业化服务列表

PACBIO SYSTEMS INSTALL BASE ~160 UNITS WORLDWIDE



Figure 19. 2015 年全球装机约 160 台

高通量PacBio Sequel测序平台

浏览次数: 157

日期: 2016年4月29日 16:00

武汉菲沙基因信息有限公司联合美国 Pacific Biosciences 公司在武汉东湖综合保税区共建高通量测序中心, 引进了不少于 6 台 Sequel 三代测序系统, 将建成大规模的三代测序中心.

RS II 与 Sequel 对比

	RS II (P6-C4)	Sequel
运行时间	240 min	240 min
输出量	0.5-1 Gb	5-10 Gb
每日输出量	2 Gb	20 Gb
平均读长	10-15 kb	10-15 kb
单程准确率	~86%	~86%
30X 准确率	>99.999%	>99.999%
Reads 数	50k	500k
平台价格	\$700k	\$350k
运行成本	\$400	\$850

适用范围

	Sequel	原因
人类全基因组	Ok/ Good	贵; 利于鉴定结构变异及组装
小基因组	Good	长读长, 只需要较低的通量
靶向测序	Good	长读长, 只需要较低的通量
转录组	Poor/ Good	贵; 可得到全长的转录本
宏基因组	Poor/Ok	贵; 利于 de novo 组装
外显子组	Poor	贵; 长读长对外显子没有用处
表达谱	Poor	贵
ChIP-Seq	Poor	贵

信息学的问题

SMRT Analysis Software 包括了一些第三方程序:

■ 编程语言

- ▶ Java 7
- ▶ Mono 3
- ▶ Perl 5.8
- ▶ Python 2.7
- ▶ Scala 2.9

■ 平台

- ▶ Tomcat 7.0.23
- ▶ MySQL 5.1.73
- ▶ Docutils

■ 生物信息学工具

- ▶ Celera Assembler 8.1
- ▶ GMAP

其它三代方法

- Helicos, 读长过短, 已经破产
- Oxford Nanopore

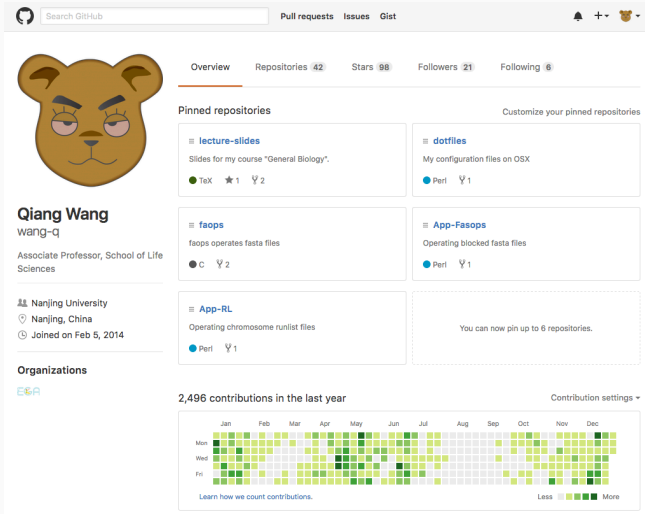


Figure 20. <https://github.com/wang-q>