

# **Bio 208: A Brief History of Genome Sequencing**

**Paul Magwene, Fall 2022**

# Landmarks in Genomics, I

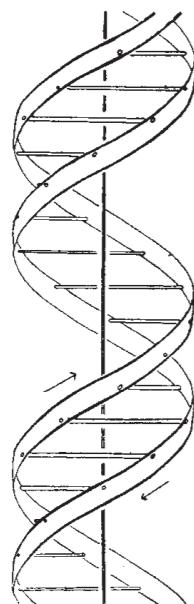
- 1944: DNA identified as hereditary material
- 1953: Structure of DNA
- 1965: First nucleic acid sequence
- 1966: Genetic code
- 1970: Efficient algorithms for pairwise global sequence alignment
- 1977: Efficient DNA sequencing methods
- 1977: First genome sequenced
- 1979: First algorithms for computational sequence assembly
- 1981: Efficient algorithms for local sequence alignment
- 1982: Genbank
- 1986: First automated DNA sequencers
- 1987: Sequencing using fluorescent ddNTPs

# Landmarks in Genomics, II

- 1990: Capillary electrophoresis
- 1990: BLAST software tools
- 1991: Large-scale automated DNA sequencers
- 1995: First genome of a cellular organism
- 1996: First eukaryotic genome
- 1998: First animal genome
- 2000: First plant genome
- 2001: Draft human genome
- 2002: Mouse genome
- 2005: First integrated systems for next generation sequencing
- 2009: First commercial single molecule sequencing

# Structure of DNA

## A Structure for Deoxyribose Nucleic Acid



This figure is purely diagrammatic. The two ribbons symbolize the two phosphate-sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

## Molecular Structure of Deoxypentose Nucleic Acids

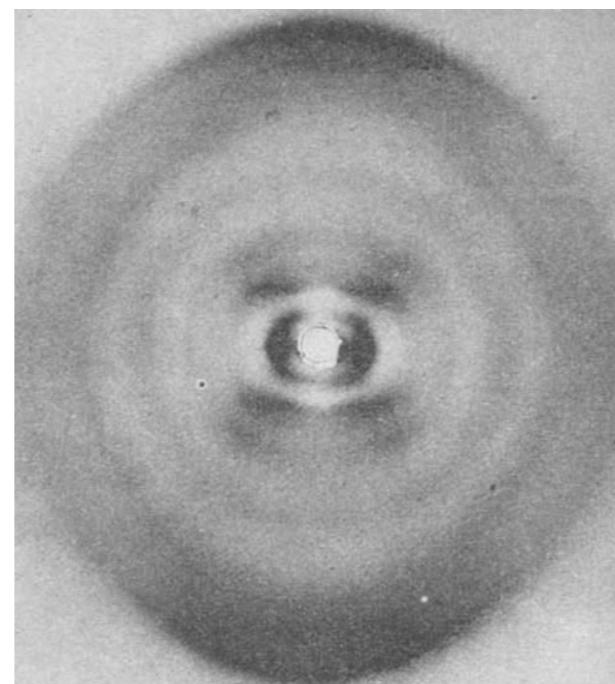
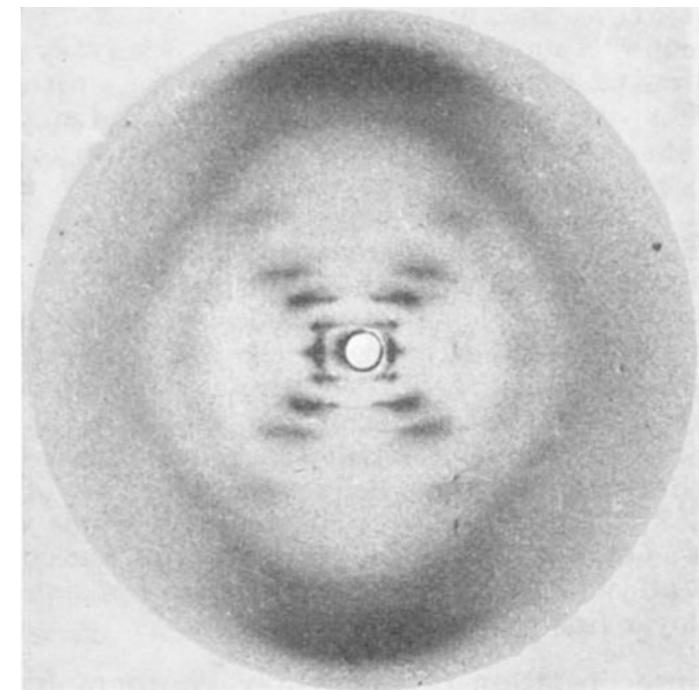


Fig. 1. Fibre diagram of deoxypentose nucleic acid from *B. coli*.  
Fibre axis vertical

## Molecular Configuration in Sodium Thymonucleate



Sodium deoxyribose nucleate from calf thymus. Structure B

J. D. WATSON  
F. H. C. CRICK

Medical Research Council Unit for the  
Study of the Molecular Structure of  
Biological Systems,  
Cavendish Laboratory, Cambridge.  
April 2.

M. H. F. WILKINS  
Medical Research Council Biophysics  
Research Unit,

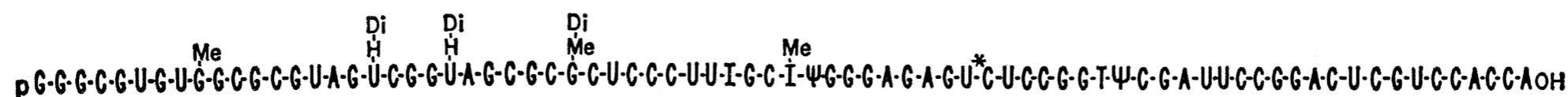
A. R. STOKES  
H. R. WILSON  
Wheatstone Physics Laboratory,  
King's College, London.  
April 2.

ROSALIND E. FRANKLIN\*  
R. G. GOSLING

Wheatstone Physics Laboratory,  
King's College, London.  
April 2.

# First complete nucleotide sequence: alanine tRNA of *Saccharomyces cerevisiae*

## STRUCTURE OF AN ALANINE RNA



## Structure of a Ribonucleic Acid

Author(s): Robert W. Holley, Jean Apgar, George A. Everett, James T. Madison, Mark Marquisee, Susan H. Merrill, John Robert Penswick and Ada Zamir

Source: *Science*, Mar. 19, 1965, New Series, Vol. 147, No. 3664 (Mar. 19, 1965), pp. 1462-1465

Published by: American Association for the Advancement of Science

Stable URL: <https://www.jstor.org/stable/1715055>

# Maxam-Gilbert Sequencing

Proc. Natl. Acad. Sci. USA  
Vol. 74, No. 2, pp. 560-564, February 1977  
Biochemistry

## A new method for sequencing DNA

(DNA chemistry/dimethyl sulfate cleavage/hydrazine/piperidine)

ALLAN M. MAXAM AND WALTER GILBERT

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138

Contributed by Walter Gilbert, December 9, 1976

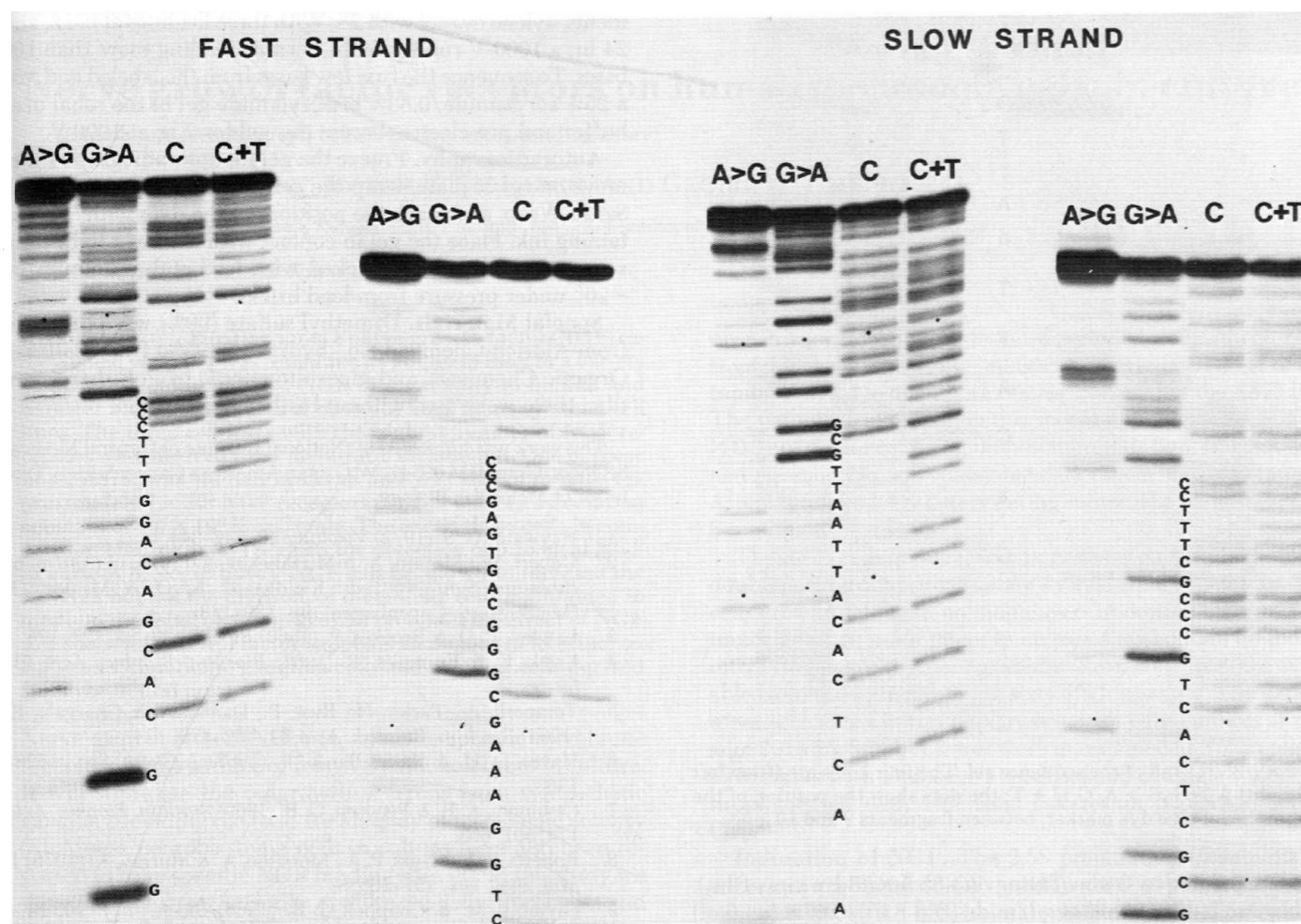


FIG. 2. Autoradiograph of a sequencing gel of the complementary strands of a 64-base-pair DNA fragment. Two panels, each with four reactions, are shown for each strand; cleavages proximal to the 5' end are at the bottom on the left. A strong band in the first column with a weaker band in the second arises from an A; a strong band in the second column with a weaker band in the first is a G; a band appearing in both the third and fourth columns is a C; and a band only in the fourth column is a T. To derive the sequence of each strand, begin at the bottom of the left panel and read upward until the bands are not resolved; then, pick up the pattern at the bottom of the right panel and continue upward. One-tenth

We have developed a new technique for sequencing DNA molecules. The procedure determines the nucleotide sequence of a terminally labeled DNA molecule by breaking it at adenine, guanine, cytosine, or thymine with chemical agents. Partial cleavage at each base produces a nested set of radioactive fragments extending from the labeled end to each of the positions of that base. Polyacrylamide gel electrophoresis resolves these single-stranded fragments; their sizes reveal *in order* the points of breakage. The autoradiograph of a gel produced from four different chemical cleavages, each specific for a base in a sense we will describe, then shows a pattern of bands from which the sequence can be read directly. The method is limited only by the resolving power of the polyacrylamide gel; in the current state of development we can sequence inward about 100 bases from the end of any terminally labeled DNA fragment.

We attack DNA with reagents that first damage and then remove a base from its sugar. The exposed sugar is then a weak point in the backbone and easily breaks; an alkali- or amine-catalyzed series of  $\beta$ -elimination reactions will cleave the sugar completely from its 3' and 5' phosphates. The reaction with the bases is a limited one, damaging only 1 residue for every 50 to 100 bases along the DNA. The second reaction to cleave the DNA strand must go to completion, so that the molecules finally analyzed do not have hidden damages. The purine-specific reagent is dimethyl sulfate; the pyrimidine-specific reagent is hydrazine.

The sequencing requires DNA molecules, either double-stranded or single-stranded, that are labeled at one end of one strand with  $^{32}\text{P}$ . This can be a 5' or a 3' label. A restriction fragment of any length is labeled at both ends—for example, by being first treated with alkaline phosphatase to remove terminal phosphates and then labeled with  $^{32}\text{P}$  by transfer from  $\gamma$ -labeled ATP with polynucleotide kinase. There are then two strategies: either (i) the double-stranded molecule is cut by a second restriction enzyme and the two ends are resolved on a polyacrylamide gel and isolated for sequencing or (ii) the doubly labeled molecule is denatured and the strands are separated on a gel (1), extracted, and sequenced.

# Sanger Sequencing

Proc. Natl. Acad. Sci. USA  
Vol. 74, No. 12, pp. 5463-5467, December 1977  
Biochemistry

## DNA sequencing with chain-terminating inhibitors

(DNA polymerase/nucleotide sequences/bacteriophage  $\phi$ X174)

F. SANGER, S. NICKLEN, AND A. R. COULSON

Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, England

Contributed by F. Sanger, October 3, 1977

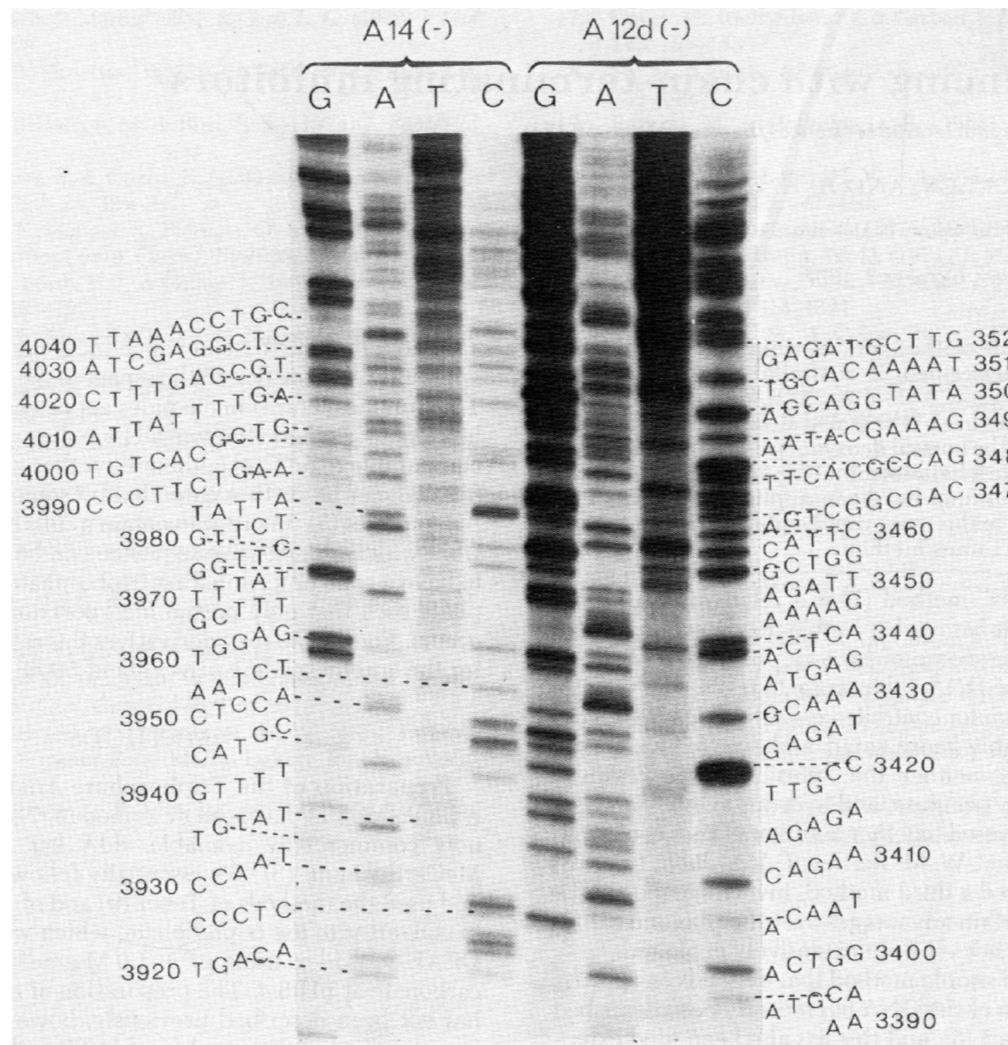


FIG. 1. Autoradiograph of the acrylamide gel from the sequence determination using restriction fragments A12d and A14 as primers on the complementary strand of  $\phi$ X174 DNA. The inhibitors used were (left to right) ddGTP, ddATP, ddTTP, and araCTP. Electrophoresis was on a 12% acrylamide gel at 40 mA for 14 hr. The top 10 cm of the gel is not shown. The DNA sequence is written from left to right and upwards beside the corresponding bands on the radioautograph. The numbering is as given in ref. 2.

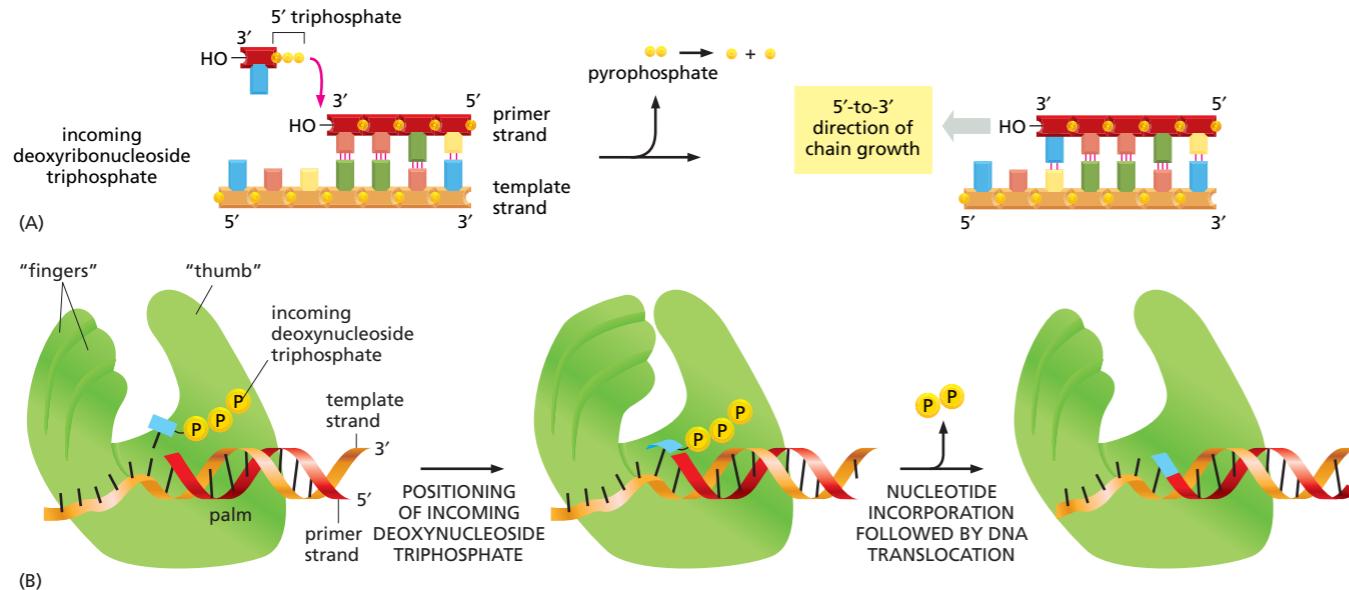


Image from Alberts et al. MBoC (5e)

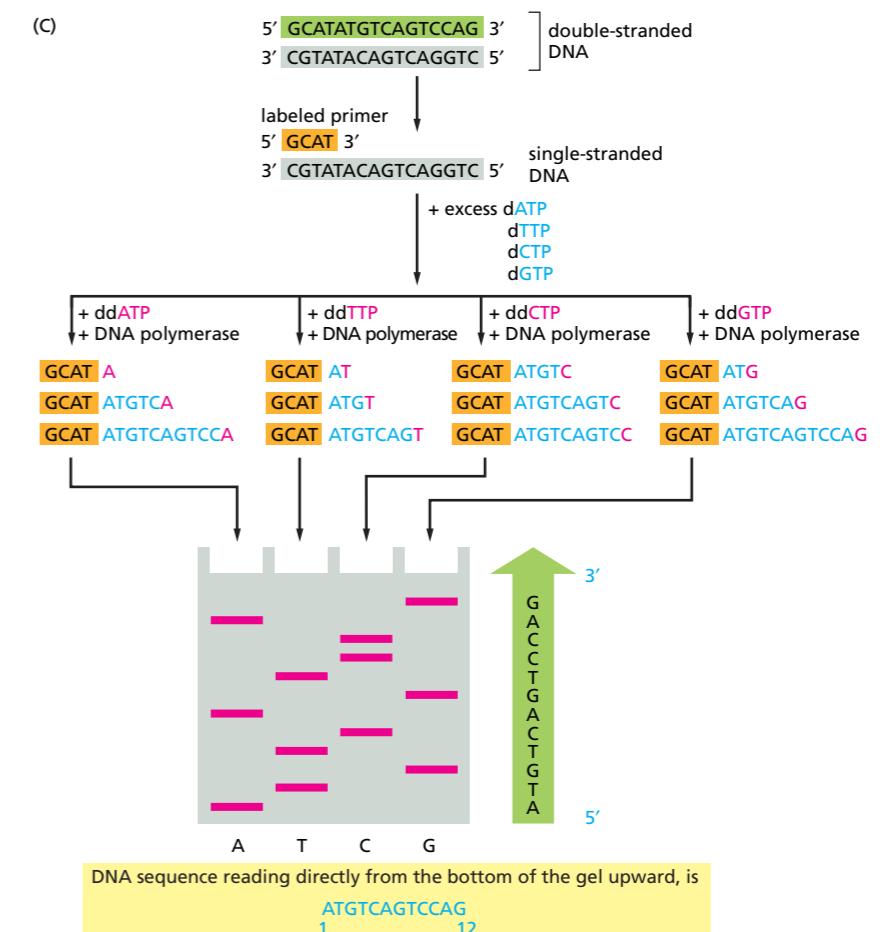
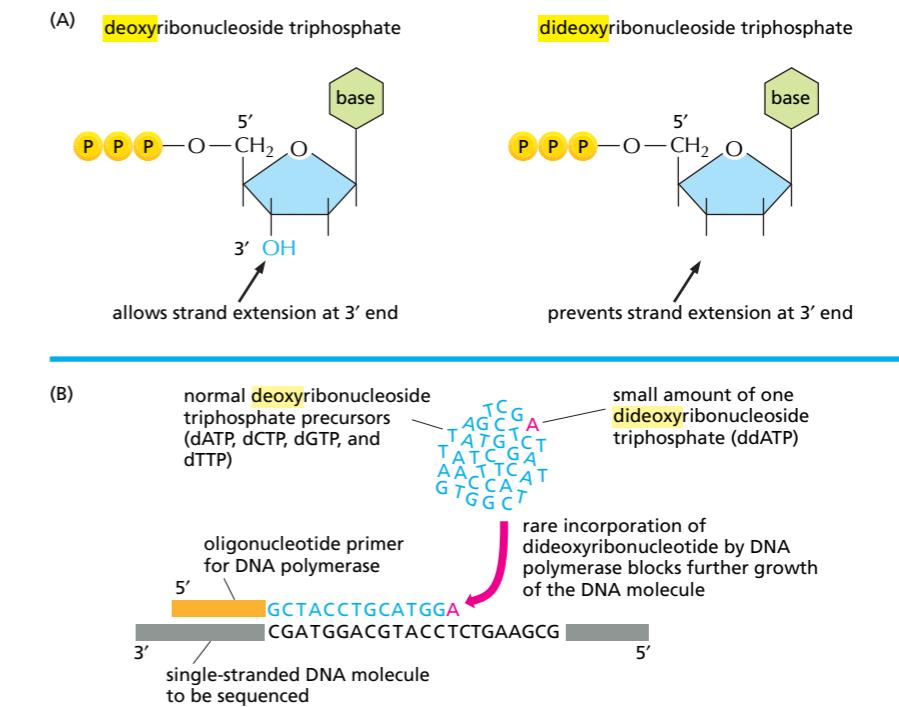


Image from Alberts et al. MBoC (5e)

# First Sequenced Genome: Bacteriophage ΦX174

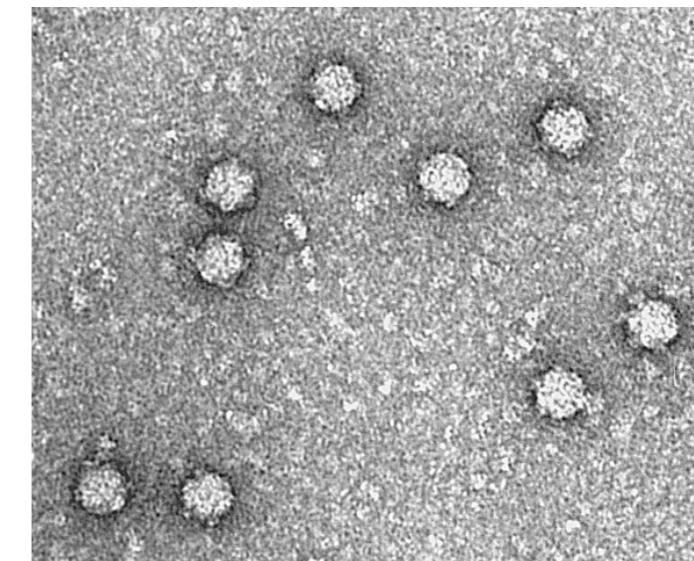
Nature Vol. 265 February 24 1977

## articles

### Nucleotide sequence of bacteriophage ΦX174 DNA

F. Sanger, G. M. Air\*, B. G. Barrell, N. L. Brown†, A. R. Coulson, J. C. Fiddes,  
C. A. Hutchison III‡, P. M. Slocombe§ & M. Smith\*

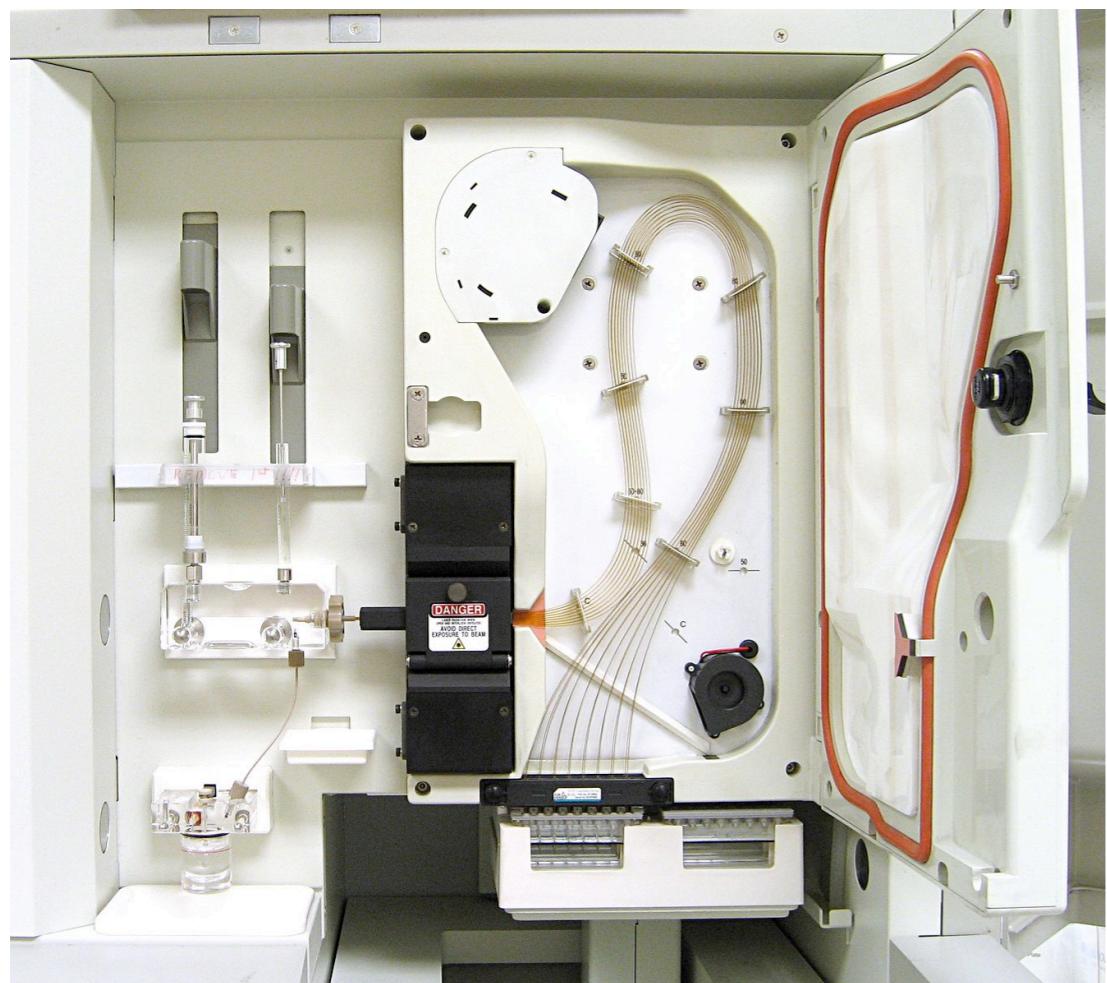
MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK



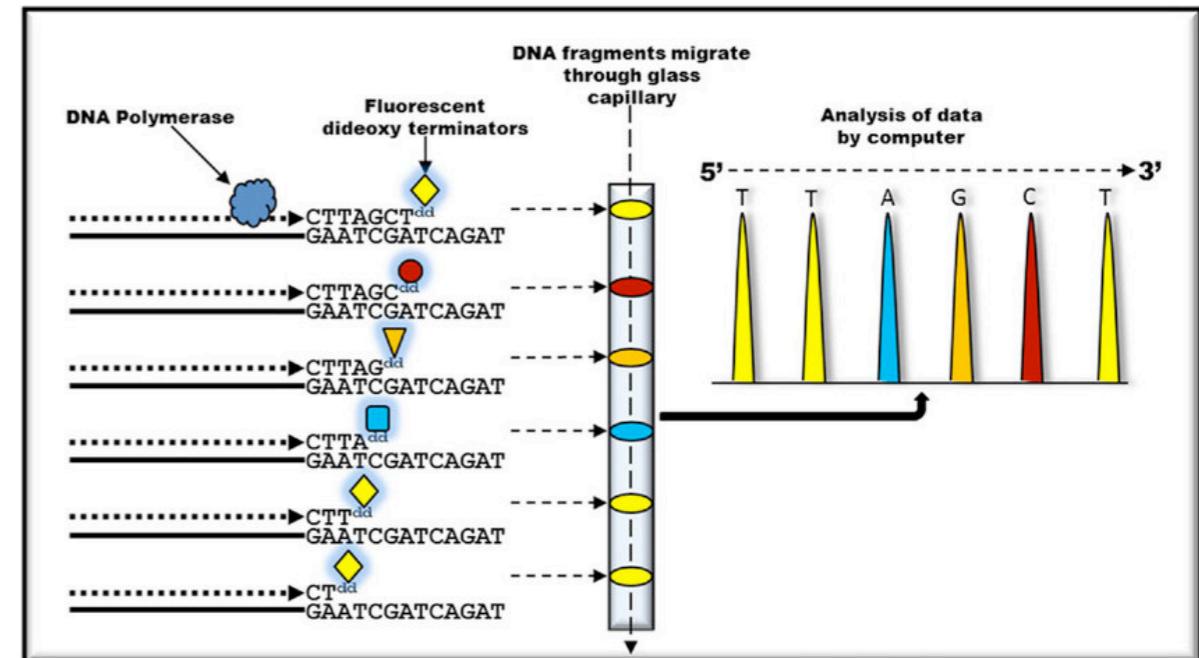
<http://studia.scienceontheweb.net/visualization.php>

GAGTTTATCGCTTCATGACGCAGAAGTTAACACTTCGGATATTCTGATGAGTCGAAAATTATCTTGATAAAGCAGGAATTACTACTGCTGTTACGAATTAAATCGAAGTGGACTGCTGGCGAAAATGAGAAAATTGACCTATC  
CTTGCAGCTGAGAAGCTTACTTGCACCTCAACTAACGATTGTCAAAACTGACCGTGGATGAGGAGAAGTGGCTTAATATGCTGGCACGTTGCAAGGACTGGTTAGATATGAGTCACATTGTCATG  
TAGAGATTCTCTTGTGACATTAAAGAGCGTGGATTACTATCTGAGTCGATGCTCAACCCTAATAGGAAGAAATCATGAGTCAGTTACTGAACAATCCGACGTTCCAGACCGCTTGGCCTCTATTAGCTCATTGAG  
CTTGCCTTGGATTAAACGAAGATGATTGAGTCTTGTGAGTAACAAAGTTGGATTGCTACTGACCGCTCTCGTCTGAGGCTTGGAGGCTTGGCTTATGGTACGCTGGACTTGTGGATAACCTCGCTTCTGCT  
CCTGTTGAGTTATTGCTGCCGCTATTGCTTATTATGTTCATCCGTCACATTCAAACGGCCTGTCATCATGGAAGGCCGCTGAATTACGGAAAACATTATTAAATGGCGTCAAGCGCTCCGGTAAAGCCGCTGAATTGTCGCGTTAC  
CTTGCCTGACGCCAGAACACTGACGTTCTACTGACGCCAGAAGAAAACGTGCGTCAAAATTACGTGCGGAAGGAGTGATGTAATGCTAAAGTAAAAACGTTCTGGCGCTGCCCTGGTGTCCGAGCCGTTGCGAGGACTAA  
AGGCAAGCGTAAGGCCTGCTTGGTATGAGGTGGTCAACAATTAAATTGAGGGCTTGGCCCTACTTGAGGATAAAATTATGCTAATATTCAAACCTGGCGCCAGCGTATGCCGATGACCTTCCATTTGGCTTCTG  
CTGGTCAGATTGGCTTATTACCAATTCAACTACTCCGGTATCGTGGCACTCCTCGAGATGGACGCCGTTGGCGCTCTCGTCTCCATTGCGTGTGGCCTGCTATTGACTCTACTGTAGACATTAAATTGTC  
CCTCATCGTCACGTTATGGTGAACAGTGGATTAAGTTCATGAAGGATGGTAAATGCCACTCCTCTCCGACTGTTAACACTACTGGTTATATTGACCATGCCGTTTCTGGCACGATTAACCCGATACCAATAAAATCCCTAAGCA  
TTGTTTCAAGGTTATTGAATATCTATAACAATTTAAAGGCCGTTGACCGTACCGAGGCTAACCTAATGAGCTTAATCAAGATGATGCTCGTTATGGTTCCGTTGCTGCCATCTCAAAACATTGGACTGCTCCG  
TTCCCTCTGAGACTGAGCTTCTGCCAAATGACGACTTCTACCACATCTATTGACATTATGGCTGCAAGCTGCTTATGCTAATTGCTACTGACCAAGAAGCTGATTACTCATGAGCGTACCATGATGTTATTCTCATTGGA  
GGTAAAACCTTTATGACGCTGACAACCGCTTACTTGTCATGCCCTTAATCTGGCATCTGGCTATGATGTTGAGGACTGACCAAAACGTGTTAGGCCAGTTCTGGCTGTTCAACAGACCTATAAACATTGTCGCCG  
TTTCTTGTCTGAGCATGCCACTATGTTACTCTTGCCTGTTCCGCTACTGCGACTAAAGAGATTGACTACCTAACGCTAAAGGTGCTTGTACTATACCGATATTGCTGGCGACCCCTGTTGATGGCAACTGCCG  
CGCGTAAATTCTATGAGGATGTTTCCGTTGGTATTGCTTAAGAAGTTAAGATTGCTGAGGGTCAGTGGTATGCTTATGCCCTCGTATGTTCTCTGCTTACCTTCTGAAGGCTTCCATTCAAGGACCGCCT  
TCTGGTATTGCAAGAACCGTACTTATTGCCACCATGATTGACCACTGTTCCAGTCCGTCAGTTGCTGAGTGGAAATAGTCAGGTTAAATTAAATGTGACCGTTATCGCAATCGCAGCACTCGCATTCAATGACTTC  
GTGATAAAAGATTGAGTGTGAGGTTATAACGCCGAAGCGTAAAAATTAAATTGGCGCTGAGGGTTGACCAAGCGAGCGGGTAGGTTCTGCTTAGGAGTTAACATGTTCAGACTTTATTCTGCCATAATTCAAACCTT  
TTTTCTGATAAGCTGGTCTCACTTCTGTTACTCCAGCTTCTGCCACCTGTTTACAGACACCTAACGCTAACGTTATATTGATAGTTGACGTTAACGTTGCTGGTAATGGTGGTTCTCATTGCAATTGAGATA  
CATCTGCAACGCCGCTAACGGTTGTTCTGTTGCTGATATTGCTTGTGCTGTTGCTGCTTGGCTGCTTGGCTGACTACCCCTCCGACTGCCATGATGTTATCCTTGAATGGT  
GCCATGATGGGGTATTACCGTCAAGGACTGTTGACTATTGACGCTCTCCCCGACGCCGGCAATAACGTTATGTTGGTTCATGGTTGGCTAACCTTACCGCTAACATGCCGGATTGGTTGCTGAATCAGGTTAT  
TAAAGAGATTATTGCTCCAGCCACTTAAGTGAGGTGATTGTTGGCTATTGCTGGCGTATTGCTCTGCTGGTGGCCATGCTAACATTGTTGGAGGCGTCAAAAGCCGCTCCGGTGGCATTCAAGGTGATG  
GCTTGCCTACCGATAACAATGCTAGGCATGGGTGATGCTGGTATTAAATGCCATTCAAGGCTCTAACCTGATGAGGCCCCCTAGTTGTTCTGGCTATGGCTAACAGTGGTAAAGGACTTCTGAAGGTA  
TGCAGGCTGGCACTCTGCCGTTCTGATAAGTGCTTATTGGTGGACTTGGCAAGTCTGCCGCTGATAAAGGAAAGGACTCTGTGATTATCTGCTGCTGATTCTGAGCTTAATGCTGGAGCGTGGTGTGATGCT  
TCCTCTGCTGGTATTGGTGGCGATTGAGAATCAAAGAGCTTAAAGTCAACTGGACAATCAGAAAGAGATTGCCGAGATGCAAATGAGACTCAAAAGAGATTGCTGGCATTCACTCGCGACTTCAGGCCAGAACATCGAA  
AGACCAGGTATATGCAACAAATGAGATGCTTACTCAACAGAAGGAGTCACTGCTCGCGTGCCTATTATGAAACACCAATCTTCAAGCAACAGCAGGTTCCGAGATTGCTGCAACATGCTCAACCGCTG  
GTCAGTATTTCACCAATGACCAAATCAAAGAAATGACTCGCAAGGTTAGTGTGAGGTTGACTTAGTCATGCAACACGCAAGATCAGCGGTATGGCTCTCTCATATTGGCGTACTGCAAAGATATTCTAATGCGTACTGATGCT  
GCTTCTGGTGGTTGATATTTCATGGTATTGATAAAGCTGTTGCCGACTTGGAAACAATTCTGGAAAGACGGTAAAGCTGATGGTATTGGCTTAATTGCTCTAGGAAATAACCGTCAGGATTGACACCCCTCCAATTGATGTTT  
CATGCCCTAACGCTTGGAGGCTTATTGTTCTTATTACCTCTGATAATGTCACGCTGATTATTGACTTGGCTGAGGCTATGAGGCTTAAACCTGCTATTGAGGCTTGGCATTGACGGTCTGCTGATGAGTTGATCTGTTACTGAG  
GCTTCCATAAGCAGATGGATAACCGCATCAAGCTCTGGAAAGAGATTCTGCTTGGTGGAGTATGCTAACGGCTTGGCTGATAATGGTATATGTTGACGCCATAAGGCTGCTTCTGACGGTCTGATGAGTTGATCTGTTACTGAG  
AAGTTAATGGTGAATTGGCACAATGCTACAATGTGCTCCCCAACCTGATATTAAACACTATAGACCAACGCCCGAAGGGGACGAAAAATTGTTAGAGAAGCAGGTTACGCAAGCTGGCTGCTGATGAGC  
CCCTCTTAAAGGATATTGCGATGAGTATAATTACCCAAAAAGAAAGGTTAACGGATGAGTGTCAAGATTGCTGGAGGCCCTCAACTATGAAATCGCTGAGGCTTGTATTGAGGCTTGTGATGCAATGCGACAGGCTCATGCTG  
ATGGTTGTTATGTTGACTCTCACGTTGGCTGACGACCGATTAGAGGCGTTTATGATAATCCAATGCTTGTGACTATTGCTGATATTGGCTGATGGTCTGCGAGGGCTGCAAGGCTAATGATTGACACGCC  
GACTGCTATCAGTATTGCTGAGTATGGTACAGCTAACGGCTCTGCTGAGTACAGCTAACGGCTCTGCTGAGTACAGCTAACGGCTCTGCTGAGTACAGCTAACGGCTCTGCTGAGTACAGCTAACGGCT  
GCAAACATGGCTTATGGTACAGTATGCCATGCGACTCGCTAACGGCTCTGCTGAGTACAGCTAACGGCTCTGCTGAGTACAGCTAACGGCTCTGCTGAGTACAGCTAACGGCTCTGCTGAGTACAGCTAACGGCT  
AATACGTTAACAAAAGTCAGATATGGACCTTGCTGCTAAAGGTCTAGGAGCTAAAGAACAACACTAACAAACCAAGCTGTCGACTTCCCAAGAAGCTGTTGAGAATCAGAATGAGGCCACTTCGGGATGAAATGCTACA  
ATGACAATCTGTCACGGAGTGCTTAATCCAACCTACCAAGCTGGGTTACGACGCCGCTCAACCAGATATTGAGAAGCAGAAGGAGATGAGGCTGGAAAAGTTACTGTAGCCGACGTTGGCGCGCAACC  
TGTGACGACAAATCTGCTCAAATTATGCGCCTGATAAAATGATTGGCGTATCCAACCTGCA

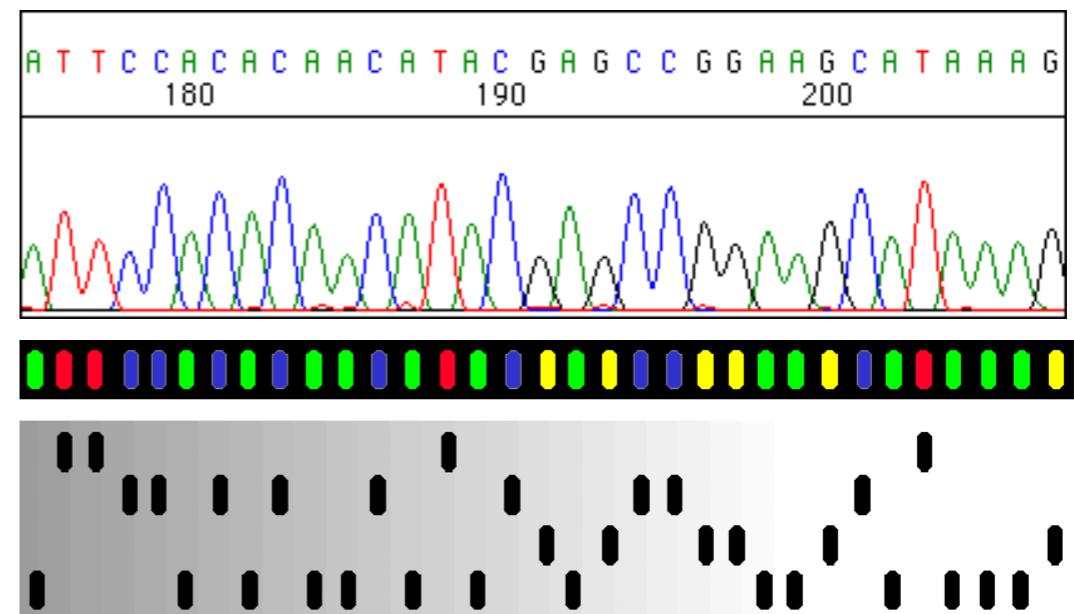
# Automated Capillary Sequencing



Michael Pereckas,  
CC BY-SA 2.0 <<https://creativecommons.org/licenses/by-sa/2.0/>>, via Wikimedia Commons



from Hawkins, G. 2017. [10.1016/B978-0-12-803077-6.00005-9](https://doi.org/10.1016/B978-0-12-803077-6.00005-9)



**Figure 1-2** Four-color/one-lane fluorescent sequencing vs. one-color/four-lane method such as radioactive sequencing

Image from ABI. 200. Automated Sequencing: Chemistry Guide.

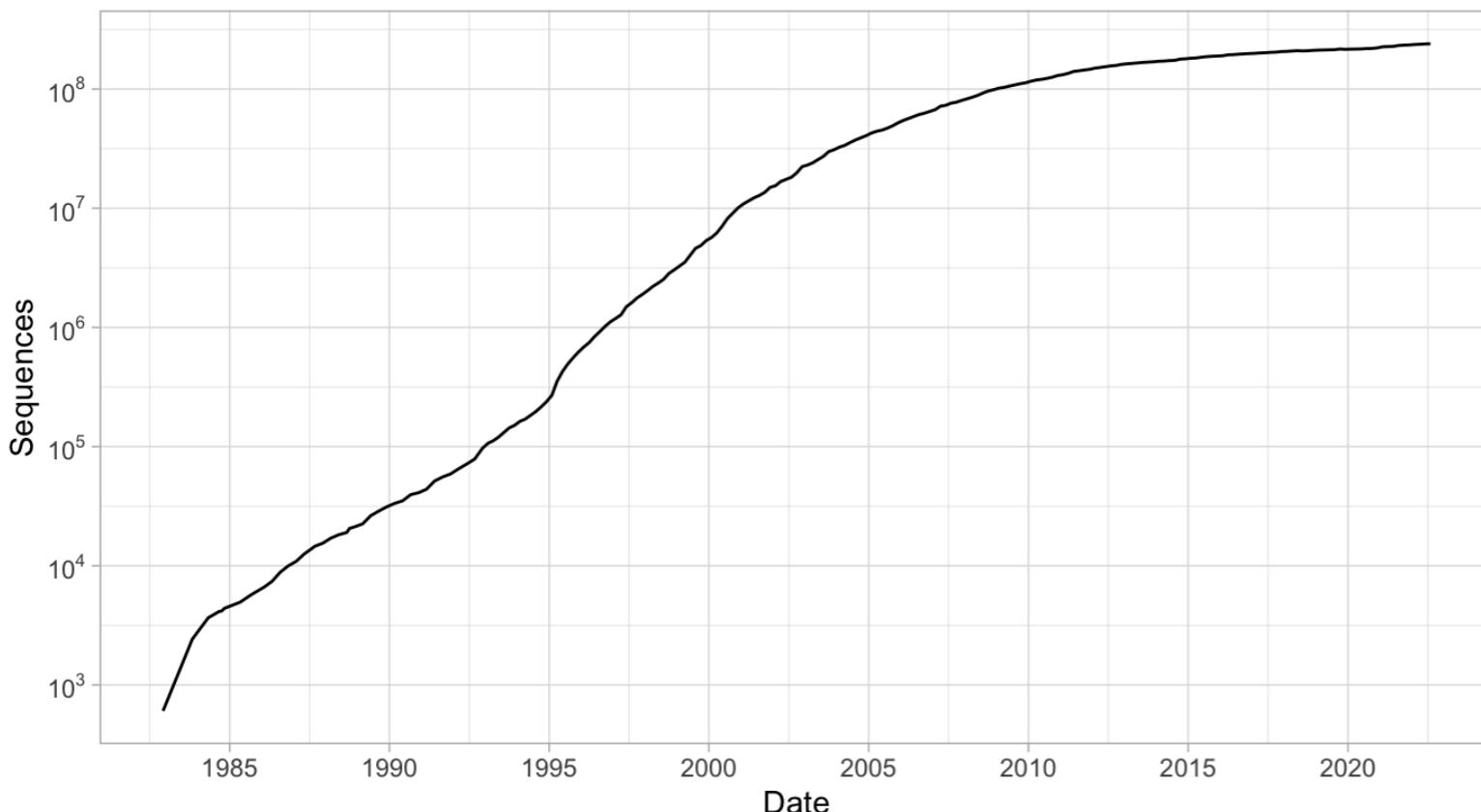
# GenBank

<https://www.ncbi.nlm.nih.gov/genbank/>

“GenBank® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (*Nucleic Acids Research*, 2013 Jan;41(D1):D36-42). GenBank is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI. These three organizations exchange data on a daily basis.”

## Growth of GenBank

Source: [www.ncbi.nlm.nih.gov/genbank/statistics/](http://www.ncbi.nlm.nih.gov/genbank/statistics/)



**Table 1.** Growth of GenBank Divisions

Division	Description	Base pairs <sup>a</sup>	Annual increase <sup>b</sup>
VRL	Viruses	39 351 597 469	575.68%
UNA	Unannotated	4 421 782	550.93%
INV	Invertebrates	108 680 334 593	450.00%
ROD	Rodents	23 336 550 435	93.02%
PRI	Primates	15 165 437 356	72.97%
WGS	Whole genome shotgun data	13 888 187 863 722	57.08%
TLS	Targeted Loci Studies	39 930 167 315	43.50%
MAM	Other mammals	28 568 850 588	37.06%
VRT	Other vertebrates	85 320 979 451	34.22%
BCT	Bacteria	130 518 385 589	32.07%
PLN	Plants	350 590 744 188	30.12%
TSA	Transcriptome shotgun data	454 757 992 932	19.31%
PHG	Phages	935 884 237	19.59%
PAT	Patent sequences	29 588 418 021	11.85%
ENV	Environmental samples	7 394 414 660	9.46%
SYN	Synthetic	7 994 601 379	0.78%
HTC	High-throughput cDNA	737 423 641	0.57%
HTG	High-throughput genomic	27 800 219 072	0.07%
EST	Expressed sequence tags	43 324 455 796	0.05%
GSS	Genome survey sequences	26 380 049 011	0.01%
STS	Sequence tagged sites	640 923 137	0.00%
<b>TOTAL</b>	All GenBank sequences	<b>15 309 209 714 374</b>	<b>54.79%</b>

<sup>a</sup>Release 245 (8/2021).

<sup>b</sup>Relative to release 239 (8/2020).

# First Genomes of Cellular Organisms

## Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd

Robert D. Fleischmann, Mark D. Adams, Owen White, Rebecca A. Clayton, Ewen F. Kirkness, Anthony R. Kerlavage, Carol J. Bult, Jean-Francois Tomb, Brian A. Dougherty, Joseph M. Merrick, Keith McKenney, Granger Sutton, Will FitzHugh, Chris Fields,\* Jeannine D. Gocayne, John Scott, Robert Shirley, Li-Ing Liu, Anna Glodek, Jenny M. Kelley, Janice F. Weidman, Cheryl A. Phillips, Tracy Spriggs, Eva Hedblom, Matthew D. Cotton, Teresa R. Utterback, Michael C. Hanna, David T. Nguyen, Deborah M. Sudek, Rhonda C. Brandon, Leah D. Fine, Janice L. Fritchman, Joyce L. Fuhrmann, N. S. M. Geoghegan, Cheryl L. Gnehm, Lisa A. McDonald, Keith V. Small, Claire M. Fraser, Hamilton O. Smith, J. Craig Venter†

1995

## Complete Genome Sequence of the Methanogenic Archaeon, *Methanococcus jannaschii*

Carol J. Bult, Owen White, Gary J. Olsen, Lixin Zhou, Robert D. Fleischmann, Granger G. Sutton, Judith A. Blake, Lisa M. Fitzgerald, Rebecca A. Clayton, Jeannine D. Gocayne, Anthony R. Kerlavage, Brian A. Dougherty, Jean-Francois Tomb, Mark D. Adams, Claudia I. Reich, Ross Overbeek, Ewen F. Kirkness, Keith G. Weinstock, Joseph M. Merrick, Anna Glodek, John L. Scott, Neil S. M. Geoghegan, Janice F. Weidman, Joyce L. Fuhrmann, Dave Nguyen, Teresa R. Utterback, Jenny M. Kelley, Jeremy D. Peterson, Paul W. Sadow, Michael C. Hanna, Matthew D. Cotton, Kevin M. Roberts, Margaret A. Hurst, Brian P. Kaine, Mark Borodovsky, Hans-Peter Klenk, Claire M. Fraser, Hamilton O. Smith, Carl R. Woese, J. Craig Venter\*

1996

## Life with 6000 Genes

A. Goffeau,\* B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, S. G. Oliver

The genome of the yeast *Saccharomyces cerevisiae* has been completely sequenced through a worldwide collaboration. The sequence of 12,068 kilobases defines 5885 potential protein-encoding genes, approximately 140 genes specifying ribosomal RNA, 40 genes for small nuclear RNA molecules, and 275 transfer RNA genes. In addition, the

1996

# First genomes of multicellular organisms

1998

## Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology

The *C. elegans* Sequencing Consortium\*

2000

## The Genome Sequence of *Drosophila melanogaster*

Mark D. Adams,<sup>1,\*</sup> Susan E. Celinker,<sup>2</sup> Robert A. Holt,<sup>1</sup> Cheryl A. Evans,<sup>1</sup> Jeannine D. Gocayne,<sup>1</sup> Peter G. Amanatides,<sup>1</sup> Steven E. Scherer,<sup>3</sup> Peter W. Li,<sup>1</sup> Roger A. Hoskins,<sup>2</sup> Richard F. Galle,<sup>2</sup> Reed A. George,<sup>2</sup> Suzanna E. Lewis,<sup>4</sup> Stephen Richards,<sup>2</sup> Michael Ashburner,<sup>5</sup> Scott N. Henderson,<sup>1</sup> Granger G. Sutton,<sup>1</sup> Jennifer R. Wortman,<sup>1</sup> Mark D. Yandell,<sup>1</sup> Qing Zhang,<sup>1</sup> Lin X. Chen,<sup>1</sup> Rhonda C. Brandon,<sup>1</sup> Yu-Hui C. Rogers,<sup>1</sup> Robert G. Blazej,<sup>2</sup> Mark Champe,<sup>2</sup> Barret D. Pfeiffer,<sup>2</sup> Kenneth H. Wan,<sup>2</sup> Clare Doyle,<sup>2</sup> Evan G. Baxter,<sup>2</sup> Gregg Heit,<sup>6</sup> Catherine R. Nelson,<sup>4</sup> George L. Gabor Miklos,<sup>7</sup> Josep F. Abril,<sup>8</sup> Anna Agbayani,<sup>2</sup> Hui-Jin An,<sup>1</sup> Cynthia Andrews-Pfannkoch,<sup>1</sup> Danita Baldwin,<sup>1</sup> Richard M. Ballew,<sup>1</sup> Anand Basu,<sup>1</sup> James Baxendale,<sup>1</sup> Leyla Bayraktaroglu,<sup>9</sup> Ellen M. Beasley,<sup>1</sup> Karen Y. Beeson,<sup>1</sup> P. V. Benos,<sup>10</sup> Benjamin P. Berman,<sup>2</sup> Deepali Bhandari,<sup>1</sup> Slava Bolshakov,<sup>11</sup> Dana Borkova,<sup>12</sup> Michael R. Botchan,<sup>13</sup> John Bouck,<sup>3</sup> Peter Brokstein,<sup>4</sup> Phillippe Brottier,<sup>14</sup> Kenneth C. Burtis,<sup>15</sup> Dana A. Busam,<sup>1</sup> Heather Butler,<sup>16</sup> Edouard Cadieu,<sup>17</sup> Angela Center,<sup>1</sup> Ishwar Chandra,<sup>1</sup> J. Michael Cherry,<sup>18</sup> Simon Cawley,<sup>19</sup> Carl Dahlke,<sup>1</sup> Lionel B. Davenport,<sup>1</sup> Peter Davies,<sup>1</sup> Beatriz de Pablos,<sup>20</sup> Arthur Delcher,<sup>1</sup> Zuoming Deng,<sup>1</sup> Anne Deslattes Mays,<sup>1</sup> Ian Dew,<sup>1</sup> Suzanne M. Dietz,<sup>1</sup> Kristina Dodson,<sup>1</sup> Lisa E. Doup,<sup>1</sup> Michael Downes,<sup>21</sup> Shannon Dugan-Rocha,<sup>3</sup> Boris C. Dunkov,<sup>22</sup> Patrick Dunn,<sup>1</sup> Kenneth J. Durbin,<sup>3</sup> Carlos C. Evangelista,<sup>1</sup> Concepcion Ferraz,<sup>23</sup> Steven Ferriera,<sup>1</sup> Wolfgang Fleischmann,<sup>5</sup> Carl Fosler,<sup>1</sup> Andrei E. Gabrielian,<sup>1</sup> Neha S. Garg,<sup>1</sup> William M. Gelbart,<sup>9</sup> Ken Glasser,<sup>1</sup> Anna Glodek,<sup>1</sup> Fangcheng Gong,<sup>1</sup> J. Harley Gorrell,<sup>3</sup> Zhiping Gu,<sup>1</sup> Ping Guan,<sup>1</sup> Michael Harris,<sup>1</sup> Nomi L. Harris,<sup>2</sup> Damon Harvey,<sup>4</sup> Thomas J. Heiman,<sup>1</sup> Judith R. Hernandez,<sup>3</sup> Jarrett Houck,<sup>1</sup> Damon Hostin,<sup>1</sup> Kathryn A. Houston,<sup>2</sup> Timothy J. Howland,<sup>1</sup> Ming-Hui Wei,<sup>1</sup> Chinyere Ibegwam,<sup>1</sup> Mena Jalali,<sup>1</sup> Francis Kalush,<sup>1</sup> Gary H. Karpen,<sup>21</sup> Zhaoxi Ke,<sup>1</sup> James A. Kennison,<sup>24</sup> Karen A. Ketchum,<sup>1</sup> Bruce E. Kimmel,<sup>2</sup> Chinnappa D. Kodira,<sup>1</sup> Cheryl Kraft,<sup>1</sup> Saul Kravitz,<sup>1</sup> David Kulp,<sup>6</sup> Zhongwu Lai,<sup>1</sup> Paul Lasko,<sup>25</sup> Yiding Lei,<sup>1</sup> Alexander A. Levitsky,<sup>1</sup> Jiayin Li,<sup>1</sup> Zhenya Li,<sup>1</sup> Yong Liang,<sup>1</sup> Xiaoying Lin,<sup>26</sup> Xiangjun Liu,<sup>1</sup> Bettina Mattei,<sup>1</sup> Tina C. McIntosh,<sup>1</sup> Michael P. McLeod,<sup>3</sup> Duncan McPherson,<sup>1</sup> Gennady Merkulov,<sup>1</sup> Natalia V. Milshina,<sup>1</sup> Clark Mobarry,<sup>1</sup> Joe Morris,<sup>6</sup> Ali Moshrefi,<sup>2</sup> Stephen M. Mount,<sup>27</sup> Mee Moy,<sup>1</sup> Brian Murphy,<sup>1</sup> Lee Murphy,<sup>28</sup> Donna M. Muzny,<sup>3</sup> David L. Nelson,<sup>3</sup> David R. Nelson,<sup>29</sup> Keith A. Nelson,<sup>1</sup> Katherine Nixon,<sup>2</sup> Deborah R. Nusskern,<sup>1</sup> Joanne M. Pacleb,<sup>2</sup> Michael Palazzolo,<sup>2</sup> Gjange S. Pittman,<sup>1</sup> Sue Pan,<sup>1</sup> John Pollard,<sup>1</sup> Vinita Puri,<sup>1</sup> Martin G. Reese,<sup>4</sup> Knut Reinert,<sup>1</sup> Karin Remington,<sup>1</sup> Robert D. C. Saunders,<sup>30</sup> Frederick Scheeler,<sup>1</sup> Hua Shen,<sup>3</sup> Bixiang Christopher Shue,<sup>1</sup> Inga Sidén-Kiamos,<sup>11</sup> Michael Simpson,<sup>1</sup> Marian P. Skupski,<sup>1</sup> Tom Smith,<sup>1</sup> Eugene Spier,<sup>1</sup> Allan C. Spradling,<sup>31</sup> Mark Stapleton,<sup>2</sup> Renee Strong,<sup>1</sup> Eric Sun,<sup>1</sup> Robert Svirskas,<sup>32</sup> Cyndee Tector,<sup>1</sup> Russell Turner,<sup>1</sup> Eli Venter,<sup>1</sup> Aihui H. Wang,<sup>1</sup> Xin Wang,<sup>1</sup> Zhen-Yuan Wang,<sup>1</sup> David A. Wassarman,<sup>33</sup> George M. Weinstock,<sup>3</sup> Jean Weissbach,<sup>14</sup> Sherita M. Williams,<sup>1</sup> Trevor Woodage,<sup>1</sup> Kim C. Worley,<sup>3</sup> David Wu,<sup>1</sup> Song Yang,<sup>2</sup> Q. Alison Yao,<sup>1</sup> Jane Ye,<sup>1</sup> Ru-Fang Yeh,<sup>19</sup> Jayshree S. Zaveri,<sup>1</sup> Ming Zhan,<sup>1</sup> Guangren Zhang,<sup>1</sup> Qi Zhao,<sup>1</sup> Liansheng Zheng,<sup>1</sup> Xiangqun H. Zheng,<sup>1</sup> Fei N. Zhong,<sup>1</sup> Wenyan Zhong,<sup>1</sup> Xiaojun Zhou,<sup>3</sup> ShaoPing Zhu,<sup>1</sup> Xiaohong Zhu,<sup>1</sup> Hamilton O. Smith,<sup>1</sup> Richard A. Gibbs,<sup>3</sup> Eugene W. Myers,<sup>1</sup> Gerald M. Rubin,<sup>34</sup> J. Craig Venter<sup>1</sup>

2000

## Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*

The Arabidopsis Genome Initiative\*

# Mammalian Genomes

2001

## Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium\*

2002

## Initial sequencing and comparative analysis of the mouse genome

Mouse Genome Sequencing Consortium\*

2001

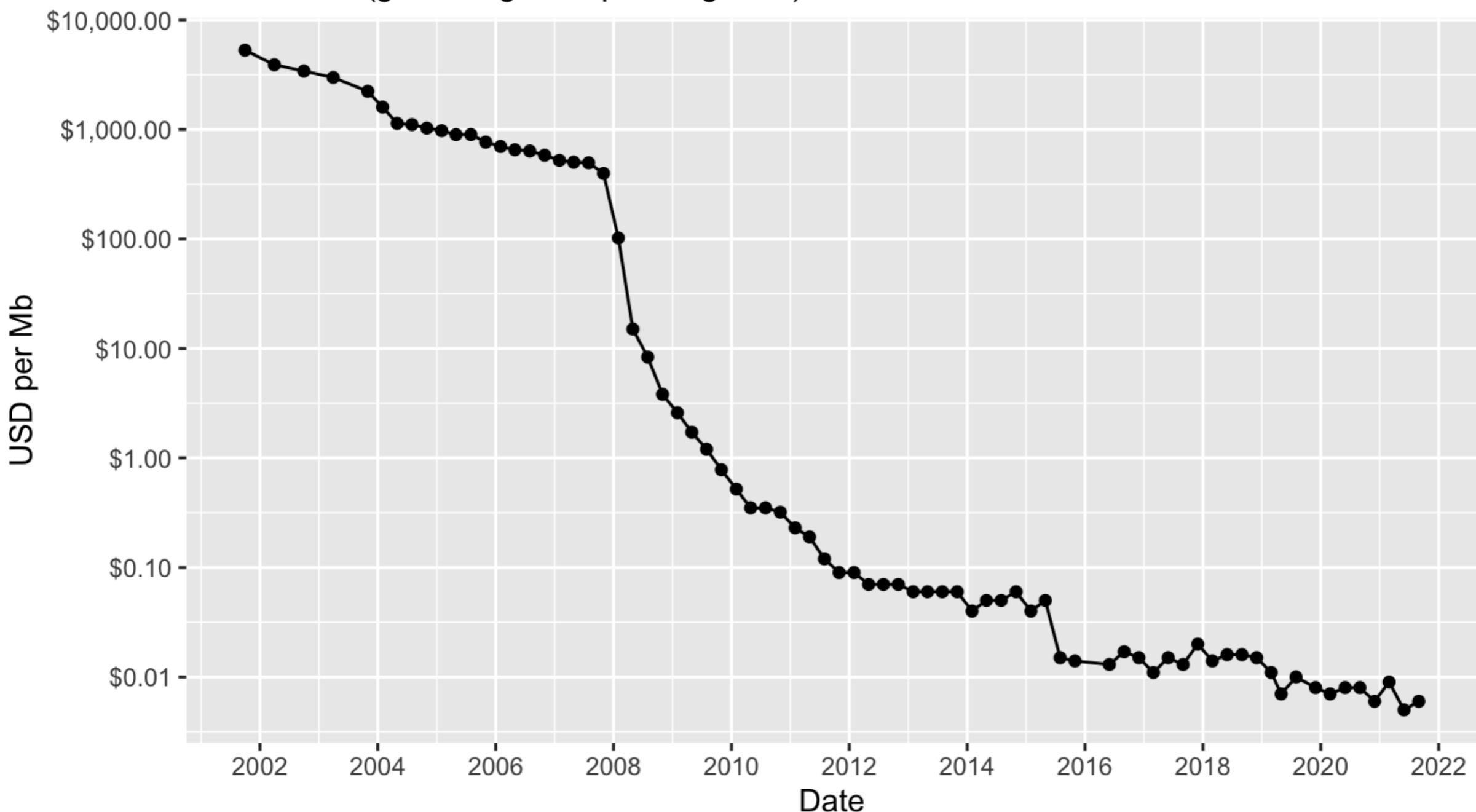
## The Sequence of the Human Genome

J. Craig Venter,<sup>1\*</sup> Mark D. Adams,<sup>1</sup> Eugene W. Myers,<sup>1</sup> Peter W. Li,<sup>1</sup> Richard J. Mural,<sup>1</sup> Granger G. Sutton,<sup>1</sup> Hamilton O. Smith,<sup>1</sup> Mark Yandell,<sup>1</sup> Cheryl A. Evans,<sup>1</sup> Robert A. Holt,<sup>1</sup> Jeannine D. Gocayne,<sup>1</sup> Peter Amanatides,<sup>1</sup> Richard M. Ballew,<sup>1</sup> Daniel H. Huson,<sup>1</sup> Jennifer Russo Wortman,<sup>1</sup> Qing Zhang,<sup>1</sup> Chinnappa D. Kodira,<sup>1</sup> Xiangqun H. Zheng,<sup>1</sup> Lin Chen,<sup>1</sup> Marian Skupski,<sup>1</sup> Gangadharan Subramanian,<sup>1</sup> Paul D. Thomas,<sup>1</sup> Jinghui Zhang,<sup>1</sup> George L. Gabor Miklos,<sup>2</sup> Catherine Nelson,<sup>3</sup> Samuel Broder,<sup>1</sup> Andrew G. Clark,<sup>4</sup> Joe Nadeau,<sup>5</sup> Victor A. McKusick,<sup>6</sup> Norton Zinder,<sup>7</sup> Arnold J. Levine,<sup>7</sup> Richard J. Roberts,<sup>8</sup> Mel Simon,<sup>9</sup> Carolyn Slayman,<sup>10</sup> Michael Hunkapiller,<sup>11</sup> Randall Bolanos,<sup>1</sup> Arthur Delcher,<sup>1</sup> Ian Dew,<sup>1</sup> Daniel Fasulo,<sup>1</sup> Michael Flanigan,<sup>1</sup> Liliana Florea,<sup>1</sup> Aaron Halpern,<sup>1</sup> Sridhar Hannenhalli,<sup>1</sup> Saul Kravitz,<sup>1</sup> Samuel Levy,<sup>1</sup> Clark Moberly,<sup>1</sup> Knut Reinert,<sup>1</sup> Karin Remington,<sup>1</sup> Jane Abu-Threideh,<sup>1</sup> Ellen Beasley,<sup>1</sup> Kendra Biddick,<sup>1</sup> Vivien Bonazzi,<sup>1</sup> Rhonda Brandon,<sup>1</sup> Michele Cargill,<sup>1</sup> Ishwar Chandramouliswaran,<sup>1</sup> Rosane Charlab,<sup>1</sup> Kabir Chaturvedi,<sup>1</sup> Zuoming Deng,<sup>1</sup> Valentina Di Francesco,<sup>1</sup> Patrick Dunn,<sup>1</sup> Karen Eilbeck,<sup>1</sup> Carlos Evangelista,<sup>1</sup> Andrei E. Gabrielian,<sup>1</sup> Weiniu Gan,<sup>1</sup> Wangmao Ge,<sup>1</sup> Fangcheng Gong,<sup>1</sup> Zhiping Gu,<sup>1</sup> Ping Guan,<sup>1</sup> Thomas J. Heiman,<sup>1</sup> Maureen E. Higgins,<sup>1</sup> Rui-Ru Ji,<sup>1</sup> Zhaoxi Ke,<sup>1</sup> Karen A. Ketchum,<sup>1</sup> Zhongwu Lai,<sup>1</sup> Yiding Lei,<sup>1</sup> Zhenya Li,<sup>1</sup> Jiayin Li,<sup>1</sup> Yong Liang,<sup>1</sup> Xiaoying Lin,<sup>1</sup> Fu Lu,<sup>1</sup> Gennady V. Merkulov,<sup>1</sup> Natalia Milshina,<sup>1</sup> Helen M. Moore,<sup>1</sup> Ashwinikumar K Naik,<sup>1</sup> Vaibhav A. Narayan,<sup>1</sup> Beena Neelam,<sup>1</sup> Deborah Nusskern,<sup>1</sup> Douglas B. Rusch,<sup>1</sup> Steven Salzberg,<sup>12</sup> Wei Shao,<sup>1</sup> Bixiong Shue,<sup>1</sup> Jingtao Sun,<sup>1</sup> Zhen Yuan Wang,<sup>1</sup> Aihui Wang,<sup>1</sup> Xin Wang,<sup>1</sup> Jian Wang,<sup>1</sup> Ming-Hui Wei,<sup>1</sup> Ron Wides,<sup>13</sup> Chunlin Xiao,<sup>1</sup> Chunhua Yan,<sup>1</sup> Alison Yao,<sup>1</sup> Jane Ye,<sup>1</sup> Ming Zhan,<sup>1</sup> Weiqing Zhang,<sup>1</sup> Hongyu Zhang,<sup>1</sup> Qi Zhao,<sup>1</sup> Liansheng Zheng,<sup>1</sup> Fei Zhong,<sup>1</sup> Wenyan Zhong,<sup>1</sup> Shaooping C. Zhu,<sup>1</sup> Shaying Zhao,<sup>12</sup> Dennis Gilbert,<sup>1</sup> Suzanna Baumhueter,<sup>1</sup> Gene Spier,<sup>1</sup> Christine Carter,<sup>1</sup> Anibal Cravchik,<sup>1</sup> Trevor Woodage,<sup>1</sup> Feroze Ali,<sup>1</sup> Huijin An,<sup>1</sup> Aderonke Awe,<sup>1</sup> Danita Baldwin,<sup>1</sup> Holly Baden,<sup>1</sup> Mary Barnstead,<sup>1</sup> Ian Barrow,<sup>1</sup> Karen Beeson,<sup>1</sup> Dana Busam,<sup>1</sup> Amy Carver,<sup>1</sup> Angela Center,<sup>1</sup> Ming Lai Cheng,<sup>1</sup> Liz Curry,<sup>1</sup> Steve Danaher,<sup>1</sup> Lionel Davenport,<sup>1</sup> Raymond Desilets,<sup>1</sup> Susanne Dietz,<sup>1</sup> Kristina Dodson,<sup>1</sup> Lisa Douc,<sup>1</sup> Steven Ferriera,<sup>1</sup> Neha Garg,<sup>1</sup> Andres Gluecksmann,<sup>1</sup> Brit Hart,<sup>1</sup> Jason Haynes,<sup>1</sup> Charles Haynes,<sup>1</sup> Cheryl Heiner,<sup>1</sup> Suzanne Hladun,<sup>1</sup> Damon Hostin,<sup>1</sup> Jarrett Houck,<sup>1</sup> Timothy Howland,<sup>1</sup> Chinyere Ibegwam,<sup>1</sup> Jeffery Johnson,<sup>1</sup> Francis Kalush,<sup>1</sup> Lesley Kline,<sup>1</sup> Shashi Koduru,<sup>1</sup> Amy Love,<sup>1</sup> Felecia Mann,<sup>1</sup> David May,<sup>1</sup> Steven McCawley,<sup>1</sup> Tina McIntosh,<sup>1</sup> Ivy McMullen,<sup>1</sup> Mee Moy,<sup>1</sup> Linda Moy,<sup>1</sup> Brian Murphy,<sup>1</sup> Keith Nelson,<sup>1</sup> Cynthia Pfankoch,<sup>1</sup> Eric Pratts,<sup>1</sup> Vinita Puri,<sup>1</sup> Hina Qureshi,<sup>1</sup> Matthew Reardon,<sup>1</sup> Robert Rodriguez,<sup>1</sup> Yu-Hui Rogers,<sup>1</sup> Deanna Romblad,<sup>1</sup> Bob Ruhfel,<sup>1</sup> Richard Scott,<sup>1</sup> Cynthia Sitter,<sup>1</sup> Michelle Smallwood,<sup>1</sup> Erin Stewart,<sup>1</sup> Renee Strong,<sup>1</sup> Ellen Suh,<sup>1</sup> Reginald Thomas,<sup>1</sup> Ni Ni Tint,<sup>1</sup> Sukyee Tse,<sup>1</sup> Claire Vech,<sup>1</sup> Gary Wang,<sup>1</sup> Jeremy Wetter,<sup>1</sup> Sherita Williams,<sup>1</sup> Monica Williams,<sup>1</sup> Sandra Windsor,<sup>1</sup> Emily Winn-Deen,<sup>1</sup> Keriellen Wolfe,<sup>1</sup> Jayshree Zaveri,<sup>1</sup> Karena Zaveri,<sup>1</sup> Josep F. Abril,<sup>14</sup> Roderic Guigó,<sup>14</sup> Michael J. Campbell,<sup>1</sup> Kimmen V. Sjolander,<sup>1</sup> Brian Karlak,<sup>1</sup> Anish Kejariwal,<sup>1</sup> Huaiyu Mi,<sup>1</sup> Betty Lazareva,<sup>1</sup> Thomas Hatton,<sup>1</sup> Apurva Narechania,<sup>1</sup> Karen Diemer,<sup>1</sup> Anushya Muruganujan,<sup>1</sup> Nan Guo,<sup>1</sup> Shinji Sato,<sup>1</sup> Vineet Bafna,<sup>1</sup> Sorin Istrail,<sup>1</sup> Ross Lippert,<sup>1</sup> Russell Schwartz,<sup>1</sup> Brian Walenz,<sup>1</sup> Shibu Yooseph,<sup>1</sup> David Allen,<sup>1</sup> Anand Basu,<sup>1</sup> James Baxendale,<sup>1</sup> Louis Blick,<sup>1</sup> Marcelo Caminha,<sup>1</sup> John Carnes-Stine,<sup>1</sup> Parris Caulk,<sup>1</sup> Yen-Hui Chiang,<sup>1</sup> My Coyne,<sup>1</sup> Carl Dahlke,<sup>1</sup> Anne Deslattes Mays,<sup>1</sup> Maria Dombroski,<sup>1</sup> Michael Donnelly,<sup>1</sup> Dale Ely,<sup>1</sup> Shiva Esparham,<sup>1</sup> Carl Fosler,<sup>1</sup> Harold Gire,<sup>1</sup> Stephen Glanowski,<sup>1</sup> Kenneth Glasser,<sup>1</sup> Anna Glodek,<sup>1</sup> Mark Gorokhov,<sup>1</sup> Ken Graham,<sup>1</sup> Barry Gropman,<sup>1</sup> Michael Harris,<sup>1</sup> Jeremy Heil,<sup>1</sup> Scott Henderson,<sup>1</sup> Jeffrey Hoover,<sup>1</sup> Donald Jennings,<sup>1</sup> Catherine Jordan,<sup>1</sup> James Jordan,<sup>1</sup> John Kasha,<sup>1</sup> Leonid Kagan,<sup>1</sup> Cheryl Kraft,<sup>1</sup> Alexander Levitsky,<sup>1</sup> Mark Lewis,<sup>1</sup> Xiangjun Liu,<sup>1</sup> John Lopez,<sup>1</sup> Daniel Ma,<sup>1</sup> William Majoros,<sup>1</sup> Joe McDaniel,<sup>1</sup> Sean Murphy,<sup>1</sup> Matthew Newman,<sup>1</sup> Trung Nguyen,<sup>1</sup> Ngoc Nguyen,<sup>1</sup> Marc Nodell,<sup>1</sup> Sue Pan,<sup>1</sup> Jim Peck,<sup>1</sup> Marshall Peterson,<sup>1</sup> William Rowe,<sup>1</sup> Robert Sanders,<sup>1</sup> John Scott,<sup>1</sup> Michael Simpson,<sup>1</sup> Thomas Smith,<sup>1</sup> Arlan Sprague,<sup>1</sup> Timothy Stockwell,<sup>1</sup> Russell Turner,<sup>1</sup> Eli Venter,<sup>1</sup> Mei Wang,<sup>1</sup> Meiyuan Wen,<sup>1</sup> David Wu,<sup>1</sup> Mitchell Wu,<sup>1</sup> Ashley Xia,<sup>1</sup> Ali Zandieh,<sup>1</sup> Xiaohong Zhu<sup>1</sup>

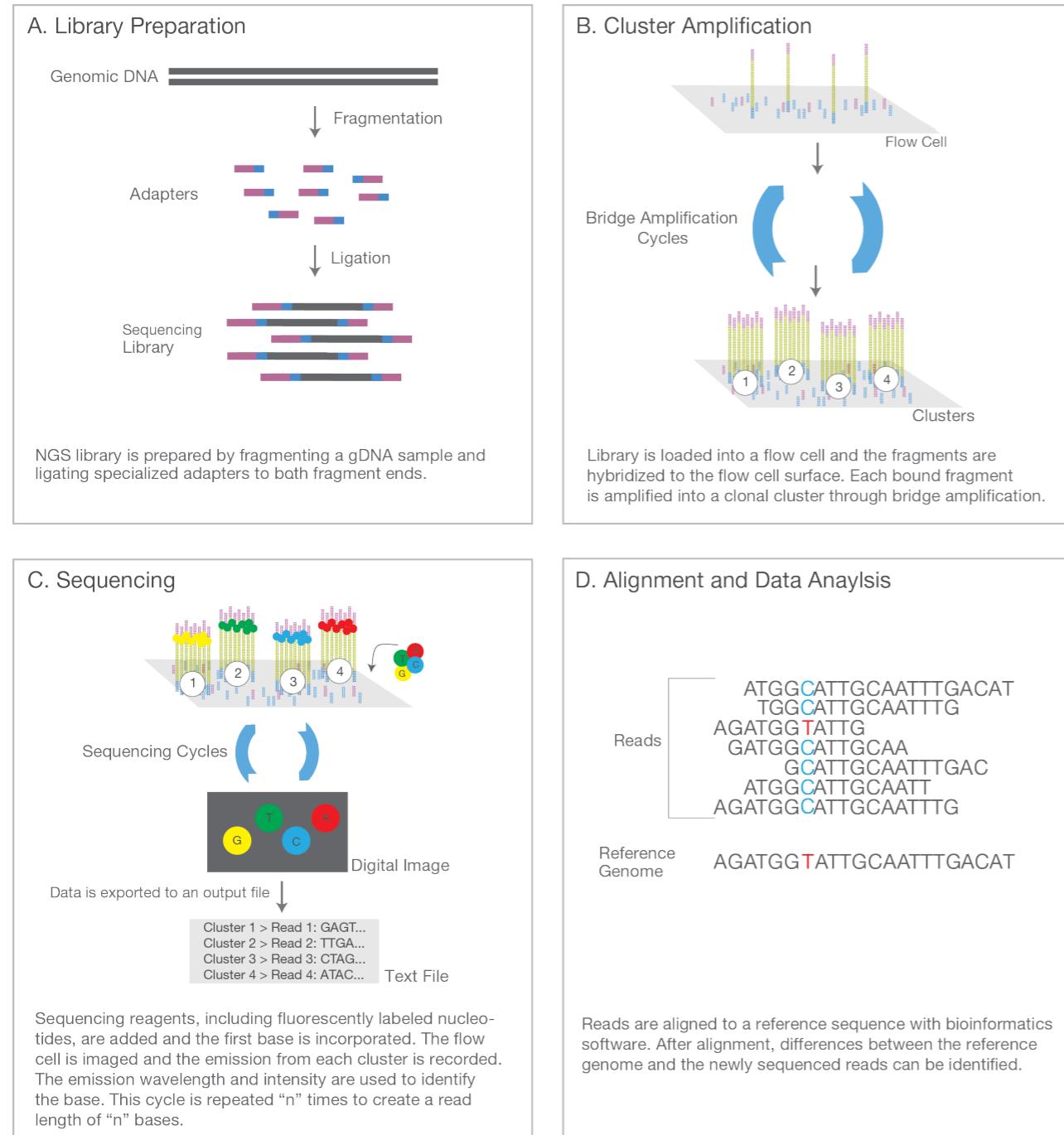
# Costs of Genome Sequencing Since 2001

Cost per Mb of Raw Sequence

Source: NIH ([genome.gov/sequencingcosts](http://genome.gov/sequencingcosts))



# Illumina Sequencing



- Moderate length reads (75-300bp)
- Low-error rates
- High-throughput
- e.g. NovaSeq 6000 – up to 3000 Gb per run

Figure 3: Next-Generation Sequencing Chemistry Overview—Illumina NGS includes four steps: (A) library preparation, (B) cluster generation, (C) sequencing, and (D) alignment and data analysis.

from Illumina, An introduction to next generation sequencing technology

# PacBio Sequencing



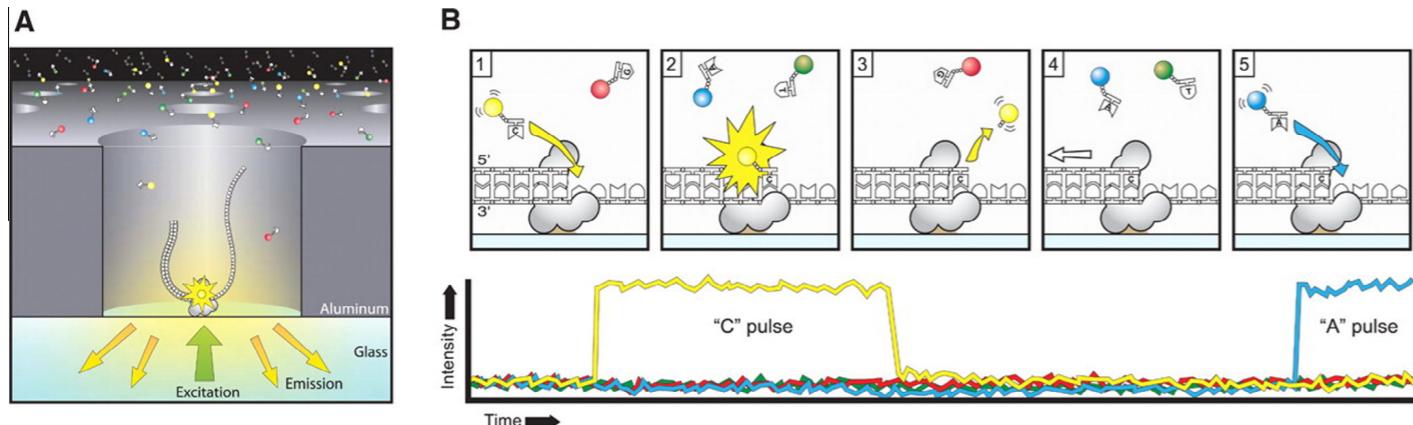
**Figure 1 SMRTbell template**

Hairpin adaptors (green) are ligated to the end of a double-stranded DNA molecule (yellow and purple), forming a closed circle. The polymerase (gray) is anchored to the bottom of a ZMW and incorporates bases into the read strand (orange). The image is adapted from [2] with permission from the Oxford University Press.



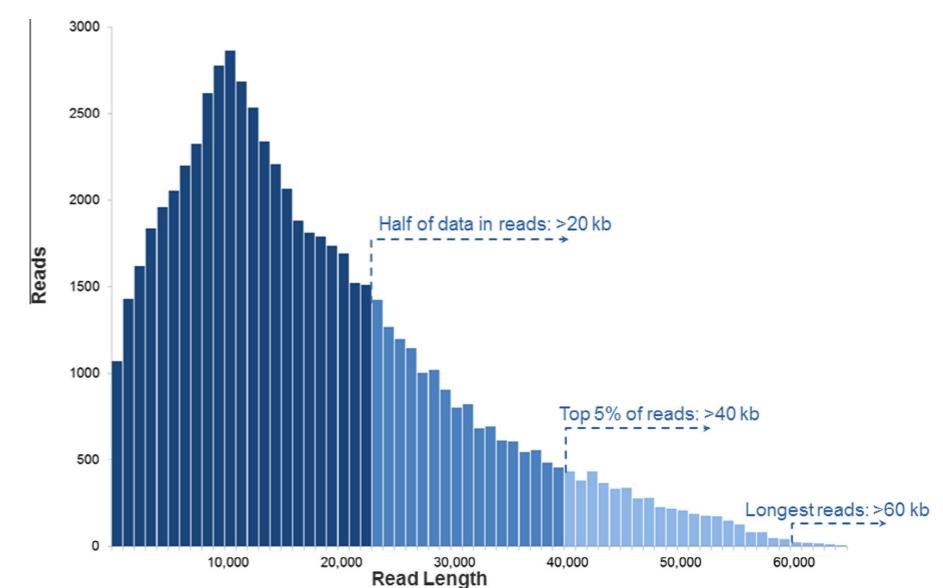
**Figure 2 A single SMRT cell**

Each SMRT cell contains 150,000 ZMWs. Approximately 35,000–75,000 of these wells produce a read in a run lasting 0.5–4 h, resulting in 0.5–1 Gb of sequence. The image is adapted with permission from Pacific Biosciences [3]. ZMW, zero-mode waveguide.



**Figure 3 Sequencing via light pulses**

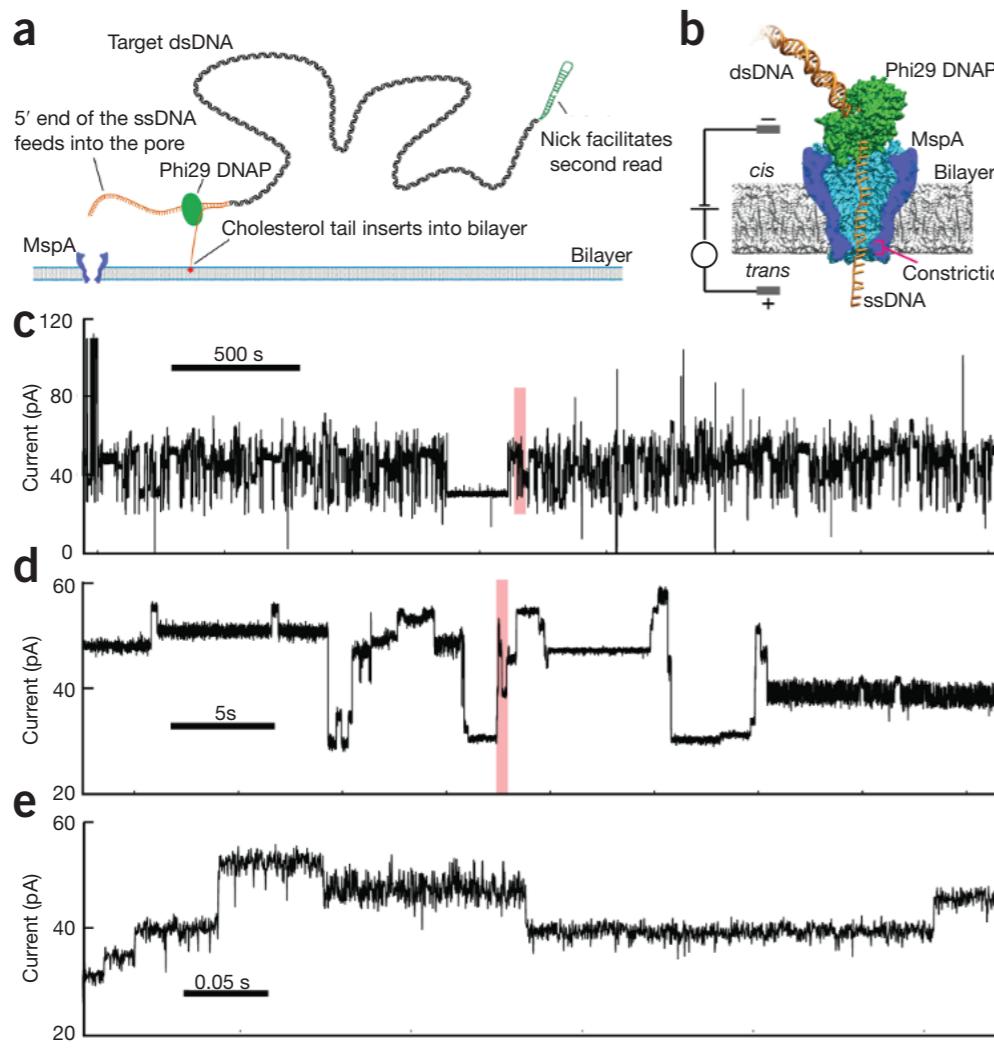
A. A SMRTbell (gray) diffuses into a ZMW, and the adaptor binds to a polymerase immobilized at the bottom. B. Each of the four nucleotides is labeled with a different fluorescent dye (indicated in red, yellow, green, and blue, respectively for G, C, T, and A) so that they have distinct emission spectra. As a nucleotide is held in the detection volume by the polymerase, a light pulse is produced that identifies the base. (1) A fluorescently-labeled nucleotide associates with the template in the active site of the polymerase. (2) The fluorescence output of the color corresponding to the incorporated base (yellow for base C as an example here) is elevated. (3) The dye-linker-pyrophosphate product is cleaved from the nucleotide and diffuses out of the ZMW, ending the fluorescence pulse. (4) The polymerase translocates to the next position. (5) The next nucleotide associates with the template in the active site of the polymerase, initiating the next fluorescence pulse, which corresponds to base A here. The figure is adapted from [4] with permission from The American Association for the Advancement of Science.



**Figure 4 PacBio RS II read length distribution using P6-C4 chemistry**

Data are based on a 20 kb size-selected *E. coli* library using a 4-h movie. Each SMRT cell produces 0.5–1 billion bases. The P6-C4 chemistry is currently the most advanced sequencing chemistry offered by PacBio. The figure is adapted with permission from Pacific Biosciences [8].

# Oxford Nanopore Sequencing



**Figure 1** Experimental schematic and raw data. (a) Method of adapting dsDNA for nanopore sequencing. The first adaptor (orange) includes a cholesterol tail that inserts into the membrane, increasing DNA capture rates<sup>32</sup>, whereas the long 5' single-stranded overhang facilitates insertion into the pore. A second adaptor (green) enables rereading of the DNA using the DNAP's synthesis mode<sup>11,17</sup>. (b) The protein nanopore MspA is shown in blue, phi29 DNAP in green and DNA in orange. An applied voltage across the membrane drives an ion current through the pore and an amplifier measures the current. DNA bases within the constriction determine the ion current. Phi29 DNAP steps DNA through the pore in single-nucleotide steps. (c–e) Raw data for a representative 3,000-s time window. The ion current changes as DNA is fed through the pore in single-nucleotide steps. Panels d and e each show the 1% section of the preceding panel's data shaded in red.



Figure from Shendure et al. 2017