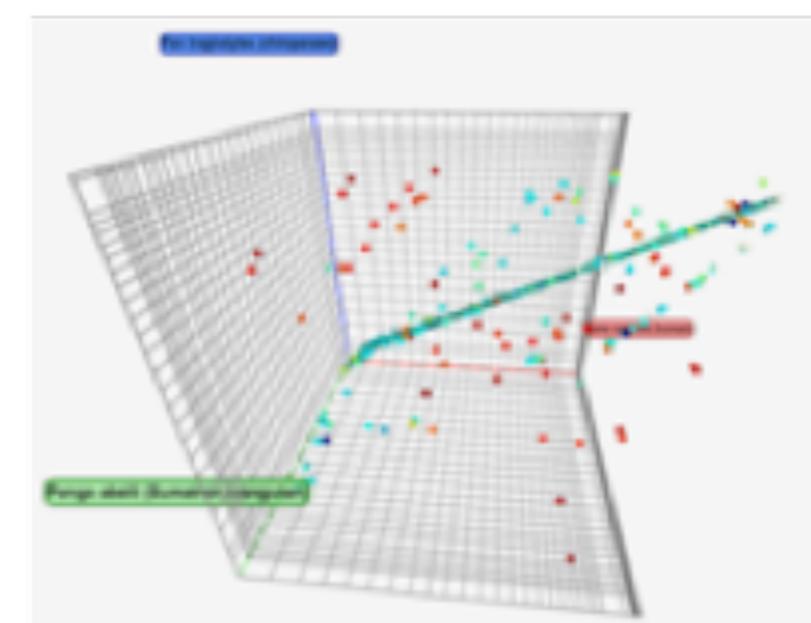


# Computational Genomics

## History of Genomics





---

# The Genomic Landscape: *circa 2010*

---

**Eric Green, M.D., Ph.D.  
Director, NHGRI**

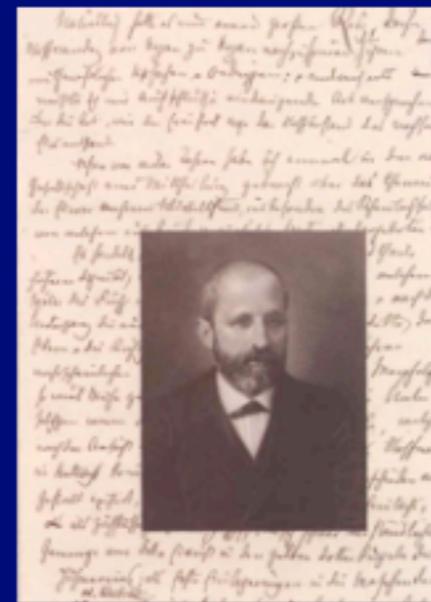


# Foundational Milestones in Genetics & Genomics



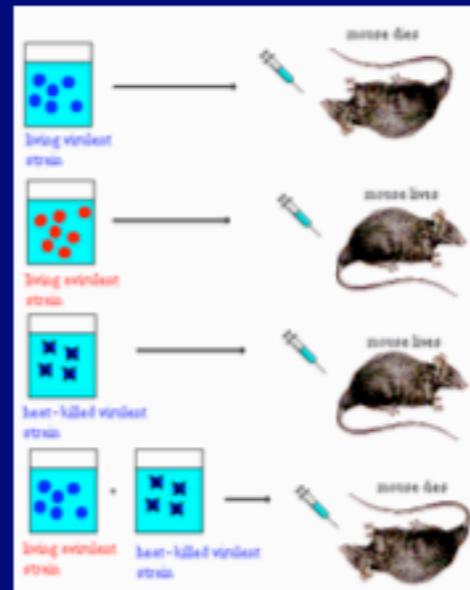
Mendel

1865



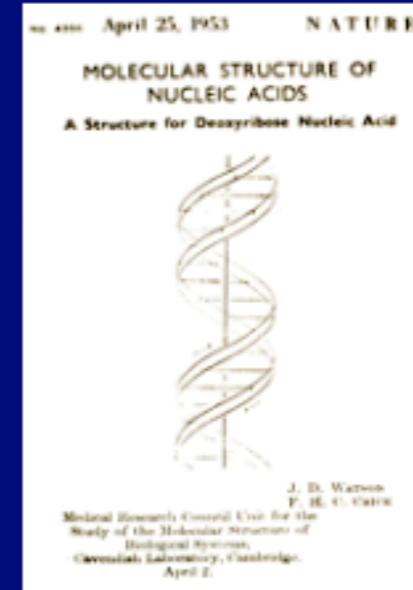
Miescher

1871



Avery

1944



Watson  
& Crick

1953



# Human Genome Project

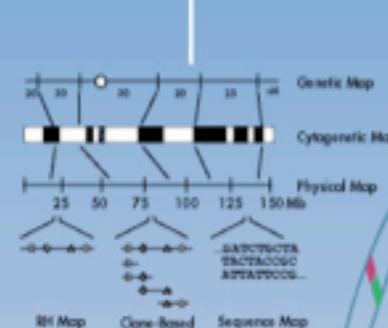
1990

Human Genome Project (HGP) launched in the U.S.



1991

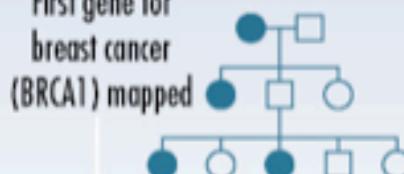
First U.S. Genome Centers established



Ethical, Legal, and Social Implications (ELSI) programs founded at NIH and DOE

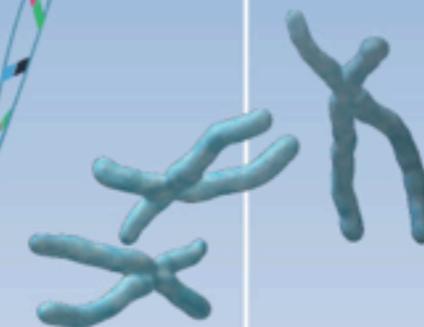


First gene for breast cancer (BRCA1) mapped



1992

Second-generation human genetic map developed



Rapid data release guidelines established by NIH and DOE

1993

New five-year plan for the HGP in the U.S. published



Sanger Centre founded (later renamed Wellcome Trust Sanger Institute)



The Wellcome Trust

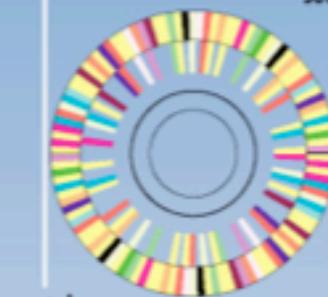
1994

HGP's human genetic mapping goal achieved



1995

HGP's human physical mapping goal achieved



First bacterial genome (*H. influenzae*) sequenced

First human gene map established

Pilot projects for human genome sequencing begin in U.S.

First archaeal genome sequenced

Yeast (*S. cerevisiae*) genome sequenced

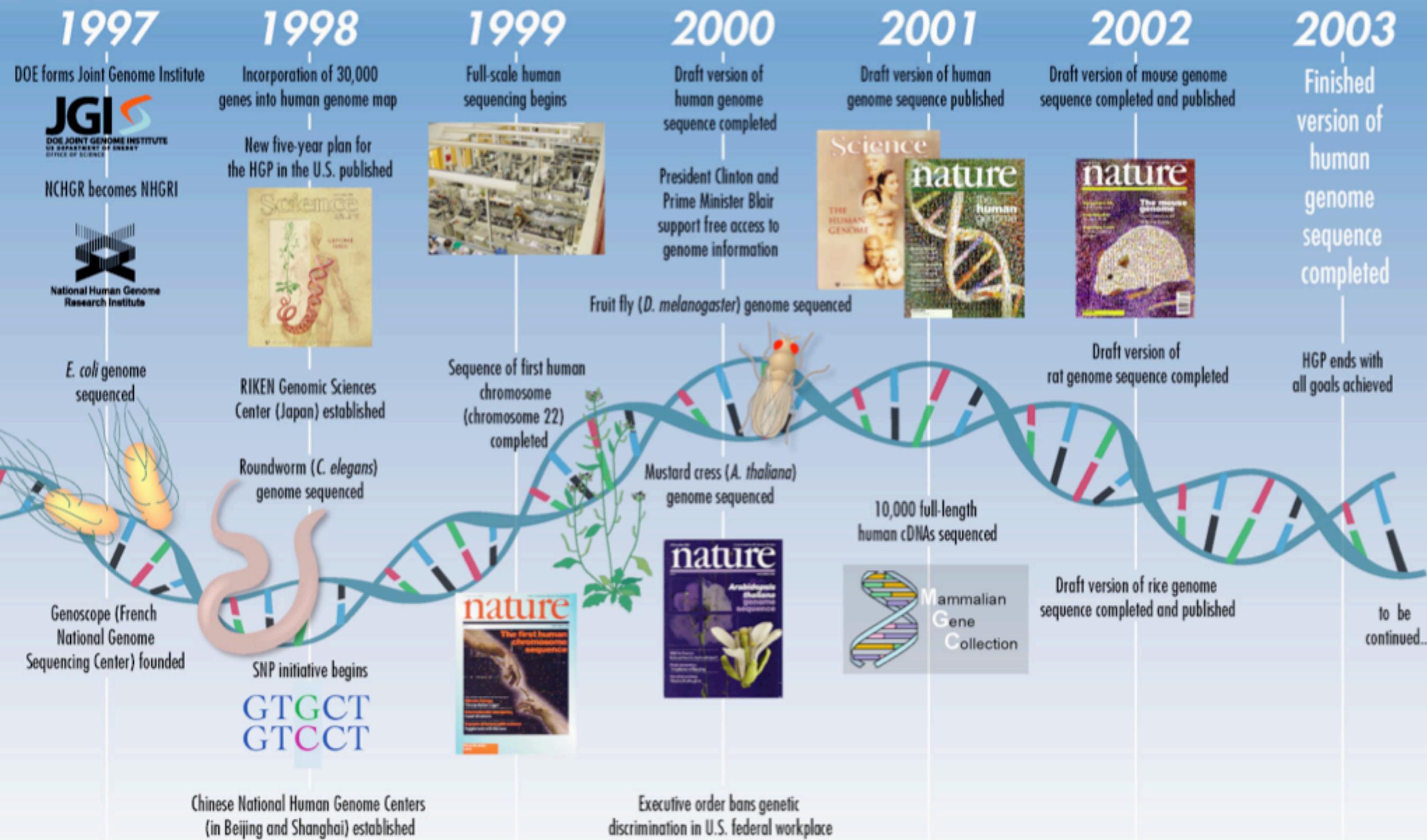


HGP's mouse genetic mapping goal achieved



Bermuda principles for rapid and open data release established

# Human Genome Project

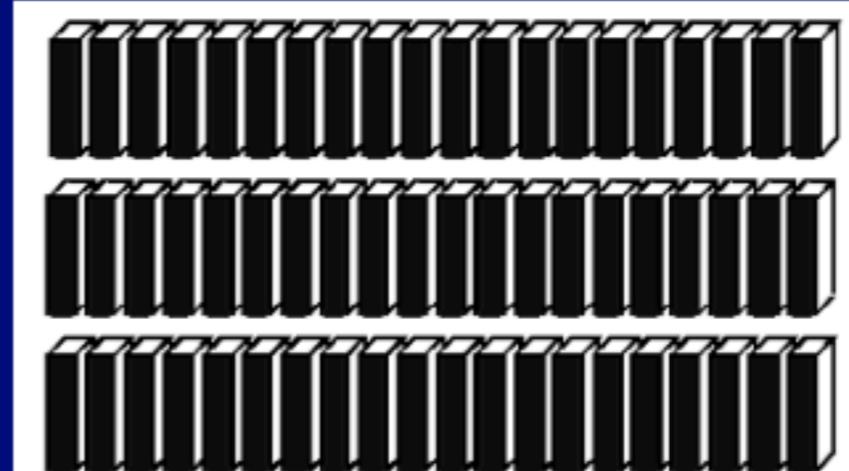


# **Outline**

- I. Fundamentals of Genome Mapping & Sequencing**
  
- II. Mapping & Sequencing in the Human Genome Project**
  
- III. Comparative Sequencing**
  
- IV. New Frontiers in Genomics**

# Genome Sizes

**Human Genome**  
**Mouse Genome**



**~3,000,000,000 bp**

**Fruit Fly Genome**



**~160,000,000 bp**

**Nematode Genome**



**~100,000,000 bp**

**Yeast Genome**



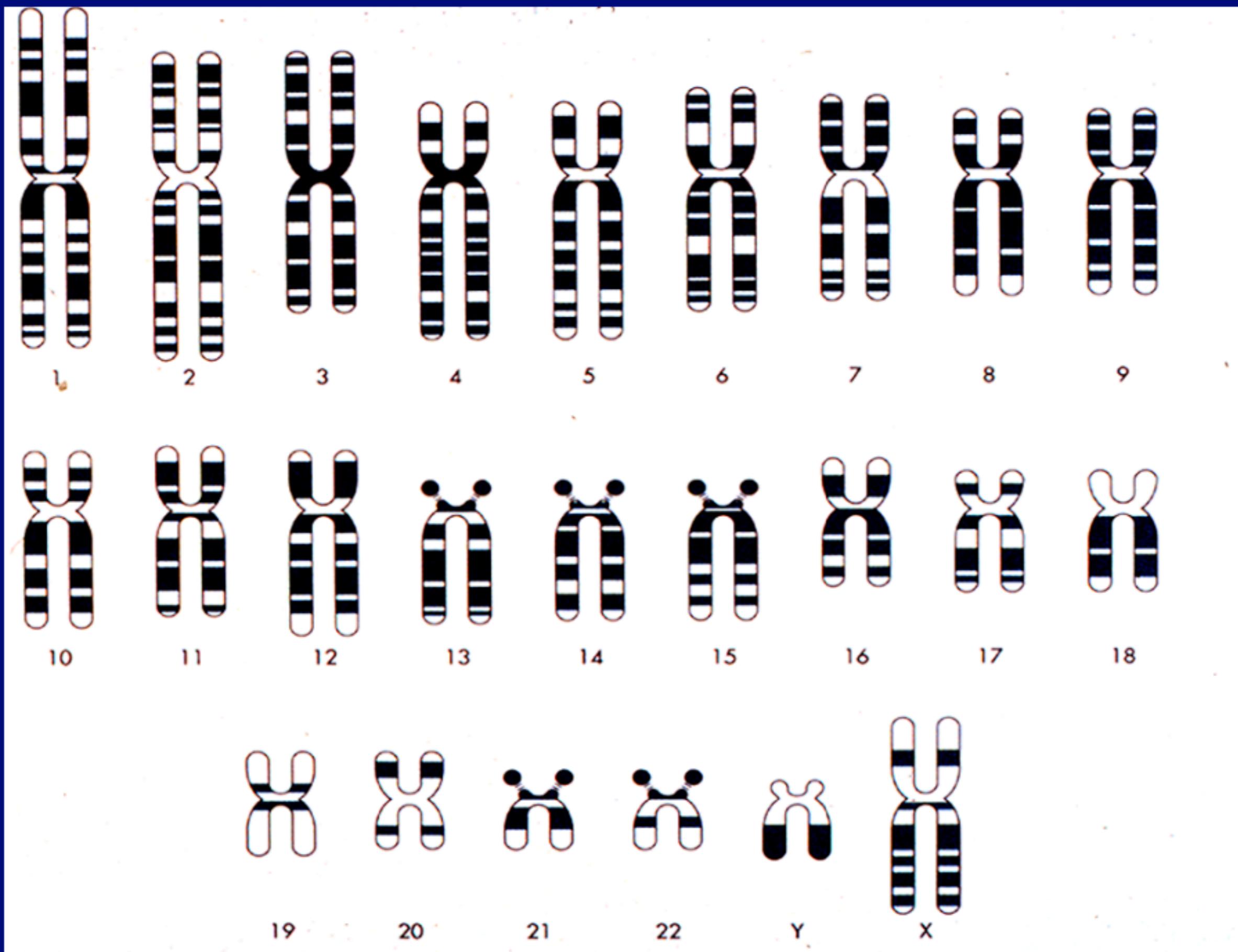
**~15,000,000 bp**

***E. coli* Genome**

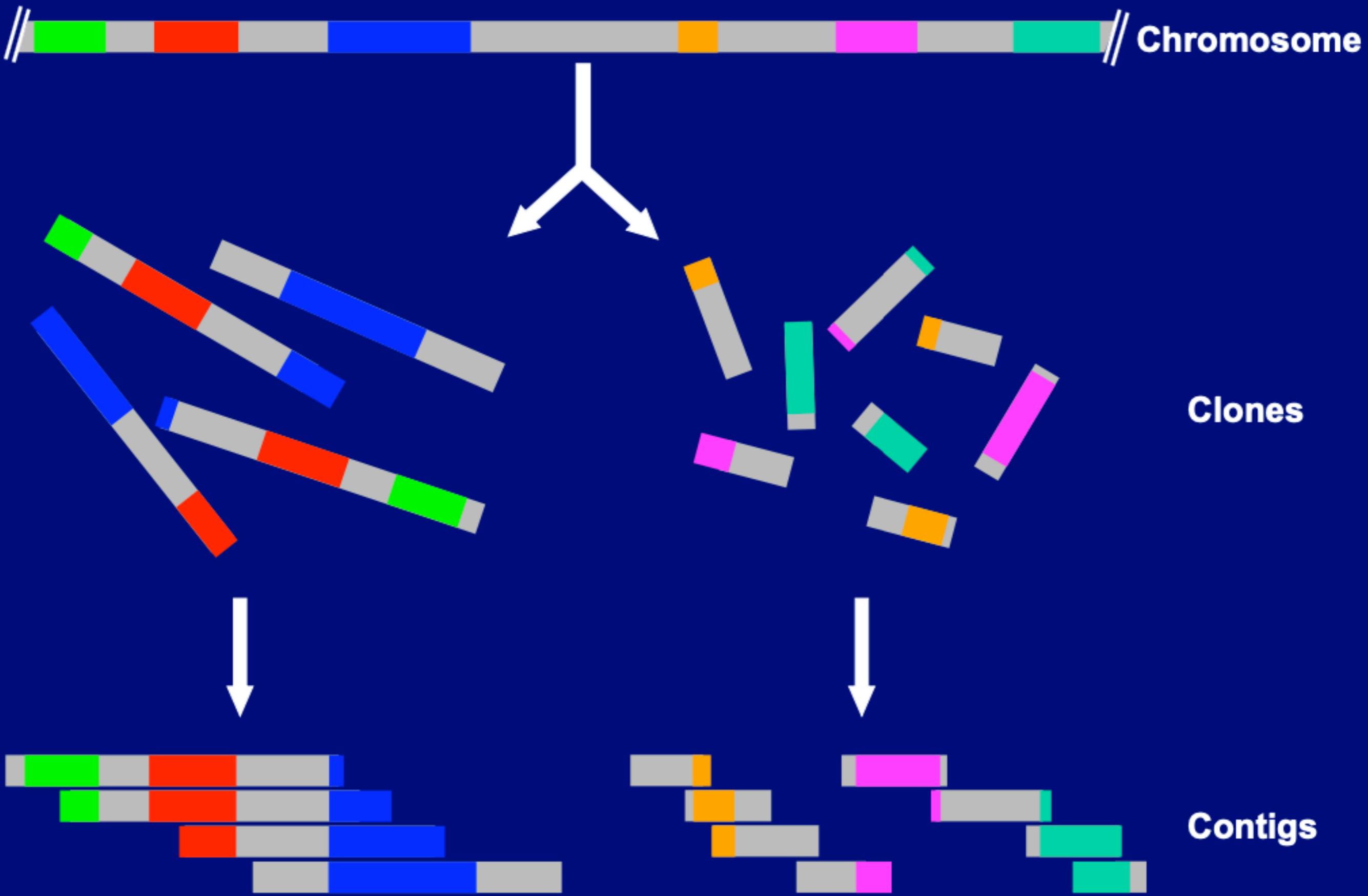


**~5,000,000 bp**

# The Human Cytogenetic Map

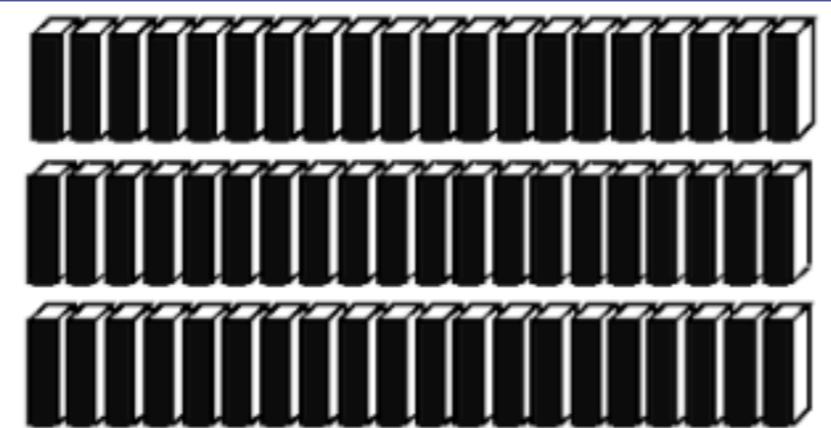


# Clone-Based Physical Mapping



## Genome Sizes

Human  
Mouse



~3,000,000,000 bp

Fruit Fly



~160,000,000 bp

Nematode



~100,000,000 bp

Yeast



~15,000,000 bp

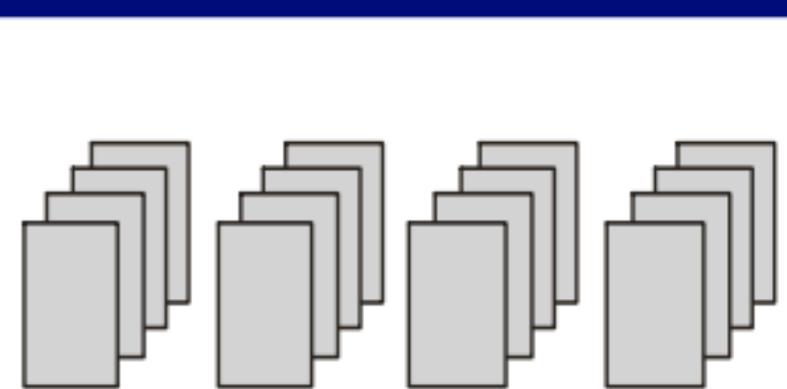
*E. coli*



~5,000,000 bp

## Cloning Capacity

YAC



~1,000,000 bp

BAC



~100,000 bp

Cosmid

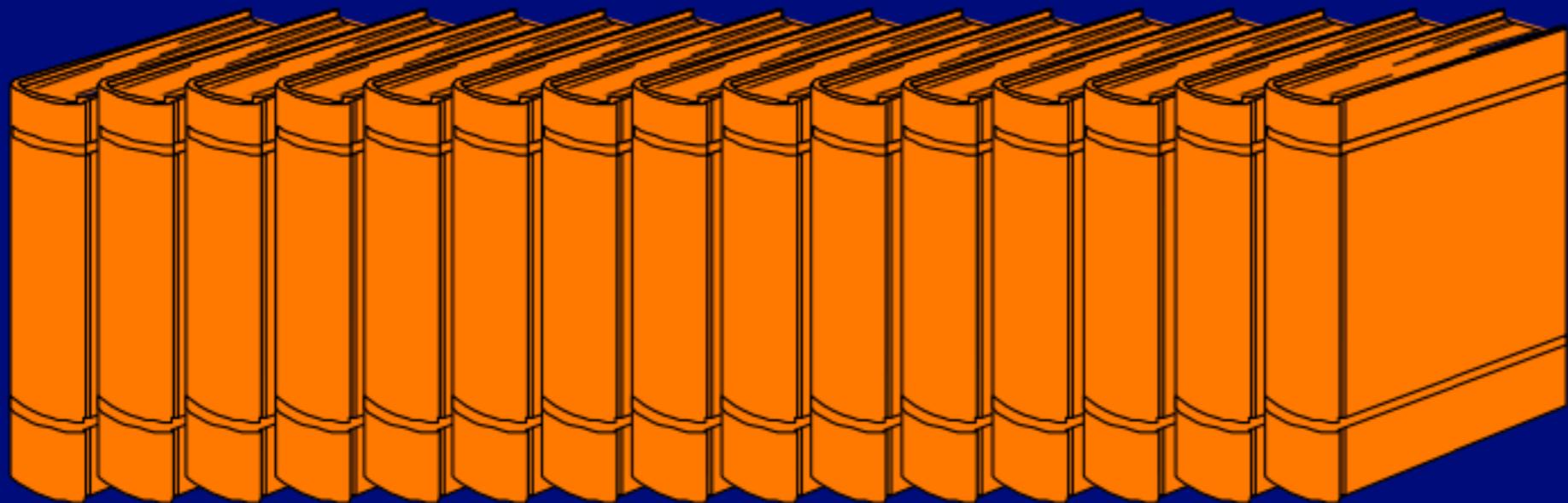


~45,000 bp

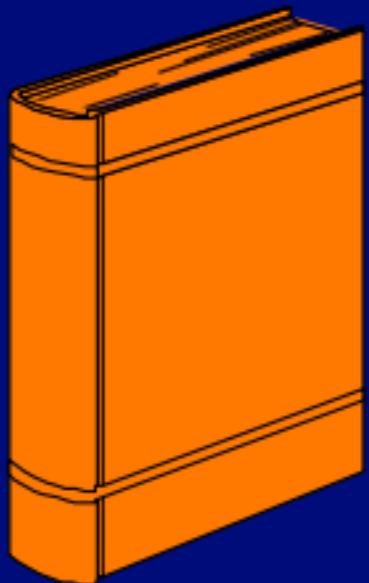
Bacteriophage



~25,000 bp



**Genome**  
**(~3000 Mb)**



**Chromosome**  
**(~130 Mb)**

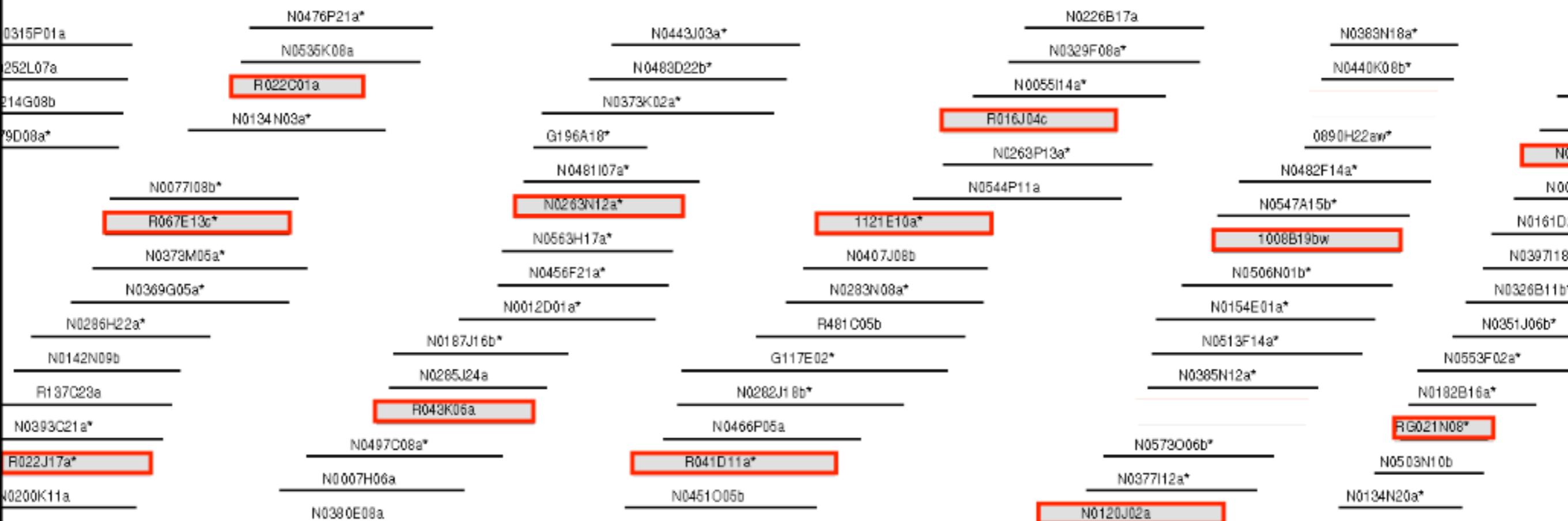
G	G	G	G	G	G	GATCGTCTAGAATCTC
G	G	G	G	G	G	GAGATCTCTGAGAGTC
G	G	G	G	G	G	GTGGGAAACTGTGTGA
T	T	T	T	T	T	TGTGACTAGCCACAGT
T	T	T	T	T	T	TGTGACTAGCCACAGT
T	T	T	T	T	T	TACGTGTGAGAGATGT
A	A	A	A	A	A	ATGATGCACCTGACCC
G	G	G	G	G	G	GGGTTTCACTCTCAAC
G	G	G	G	G	G	GACTCACTCCACCTCA
C	C	C	C	C	C	CCGGTTAGACATACAT
G	G	G	G	G	G	GAGGCCACCGCCGCT
G	G	G	G	G	G	GTGCACGTCCACCACC

**YAC**  
**(~0.5-1.0 Mb)**

GATCGTCTAGAATCTC  
GAGATCTCTGAGAGTC  
GTGGGAAACTGTGTGA  
TGTGACTAGCCACAGT  
TAGGTATTGGGGCATT  
TACGTGTGAGAGATGT  
ATGATGCACCTGACCC  
GGGTTTCACTCTCAAC  
GACTCACTCCACCTCA  
CCGGTTAGACATACAT  
GAGGCCACCGCCGCT  
GTGCACGTCCACCACC

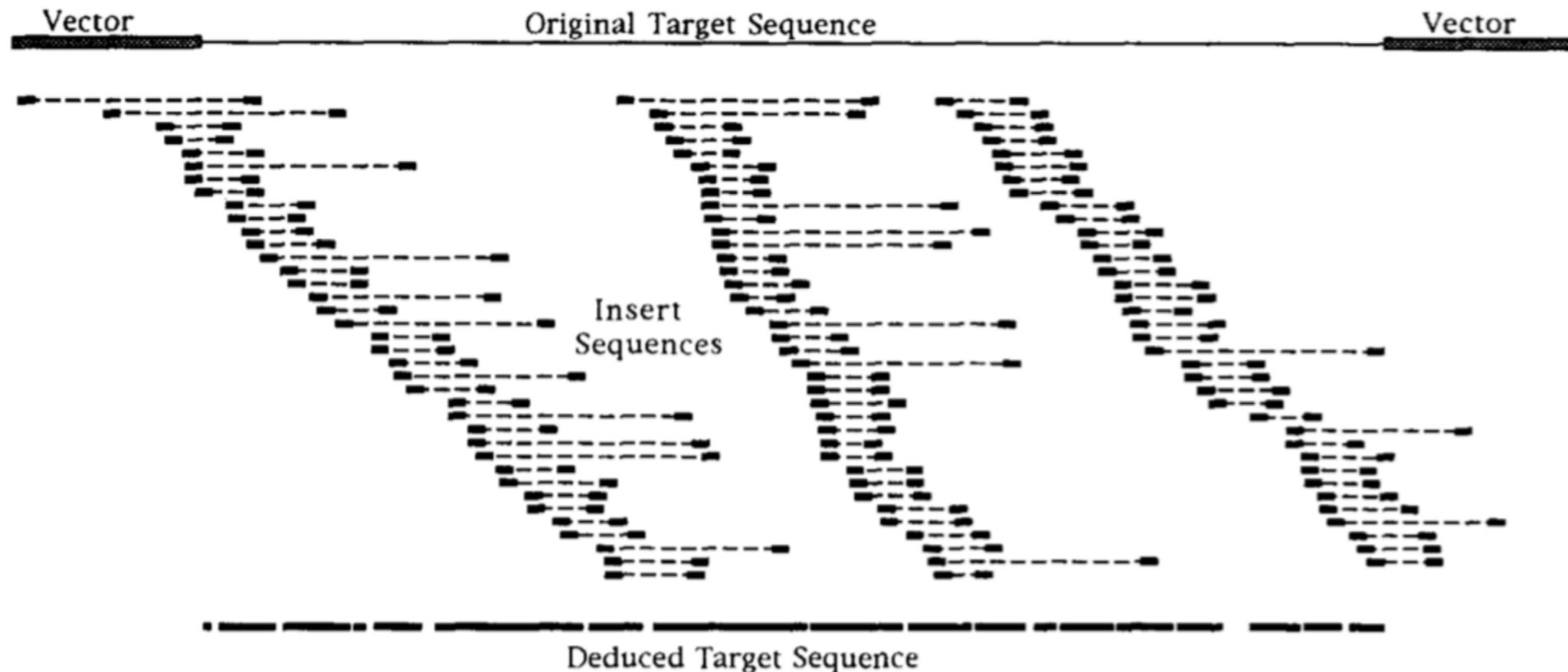
**BAC**  
**(~0.1-0.2 Mb)**

# Sequence-Ready BAC Contig Map



Pairwise End Sequencing: A Unified Approach  
to Genomic Mapping and Sequencing

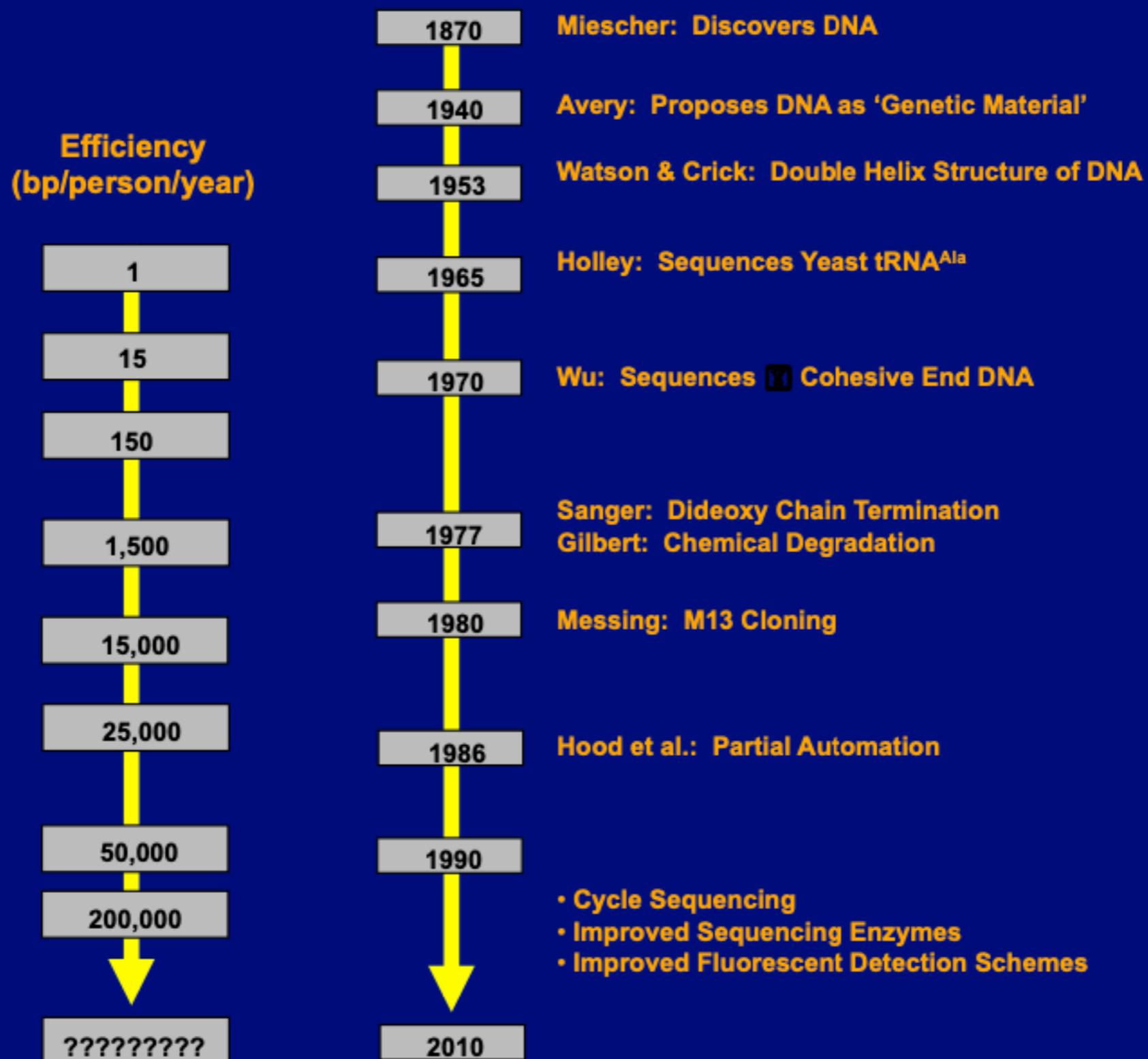
JARED C. ROACH,<sup>\*†</sup> CECILIE BOYSEN,<sup>†</sup> KAI WANG,<sup>\*</sup> AND LEROY HOOD\*



**FIG. 1.** A model “double-barrel shotgun” assembly. A 2.25 sequence redundancy produces 18 contigs that span 90% of an original target cosmid at 99.9% accuracy. Contig orientation and order are determined as shown. All but one gap are less than 400 bp; the remaining is 751 bp. More statistics are presented in Table 1.

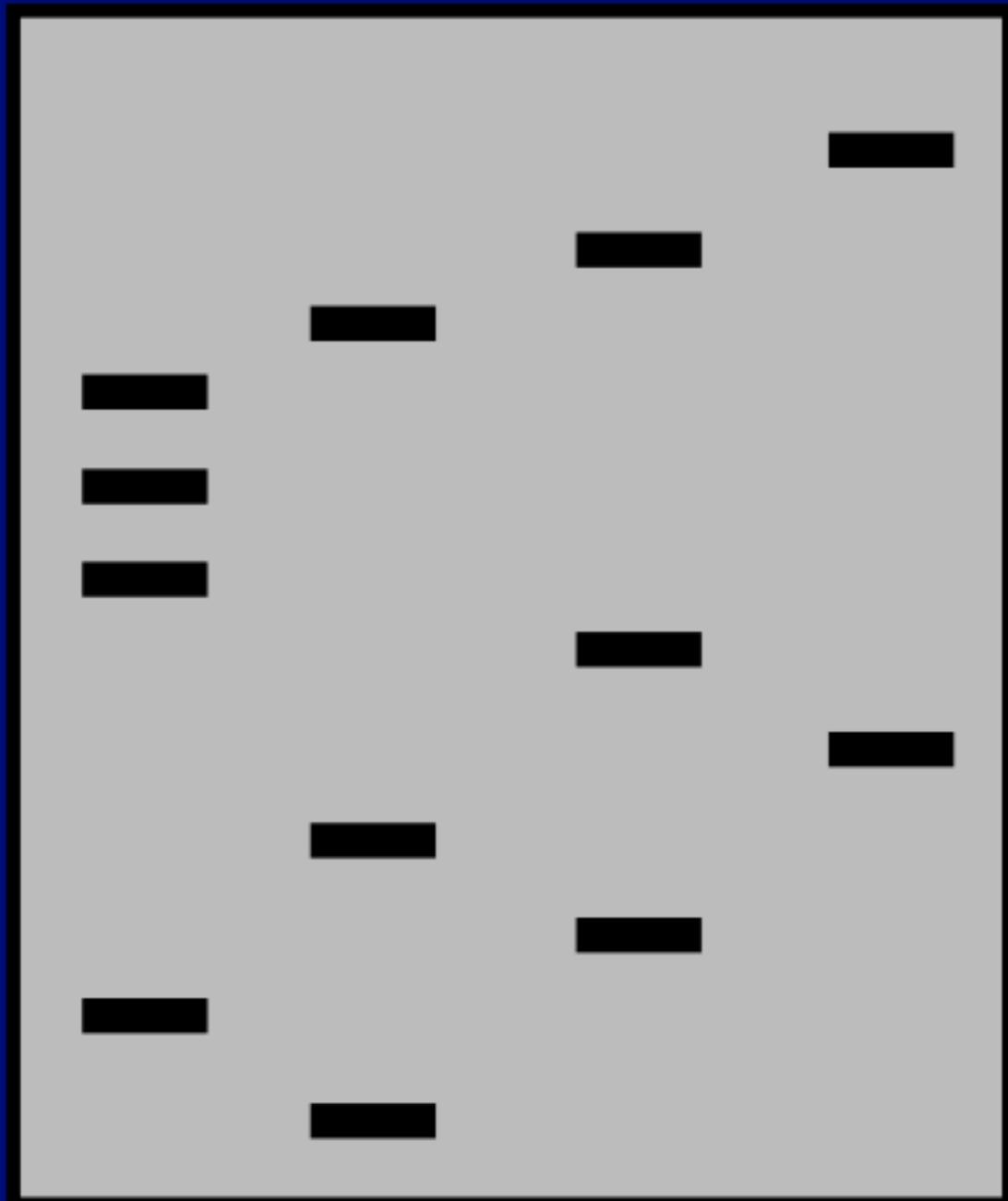
# DNA Sequencing

# History of DNA Sequencing



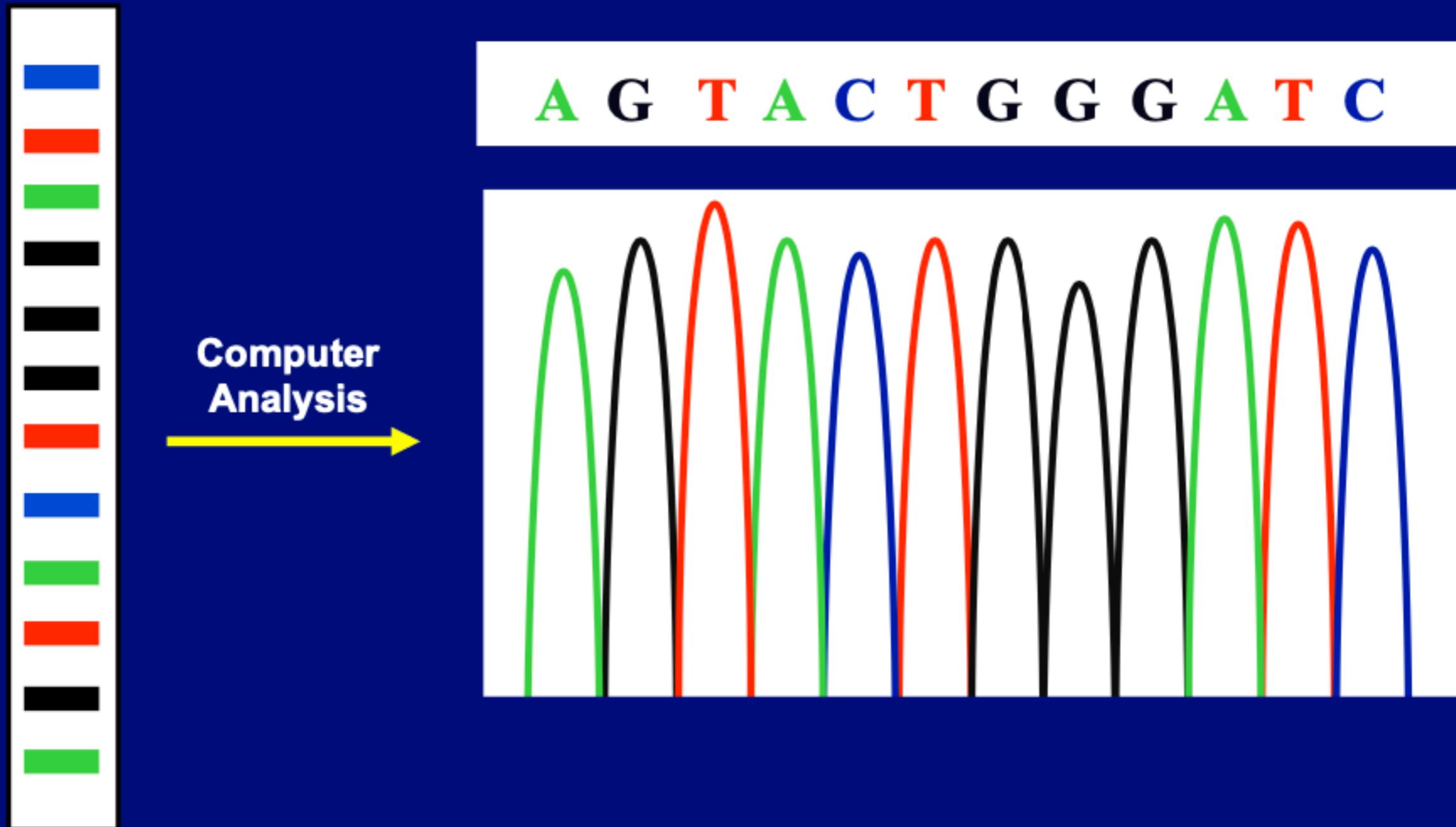
# DNA Tagged with Radioactivity

G A T C

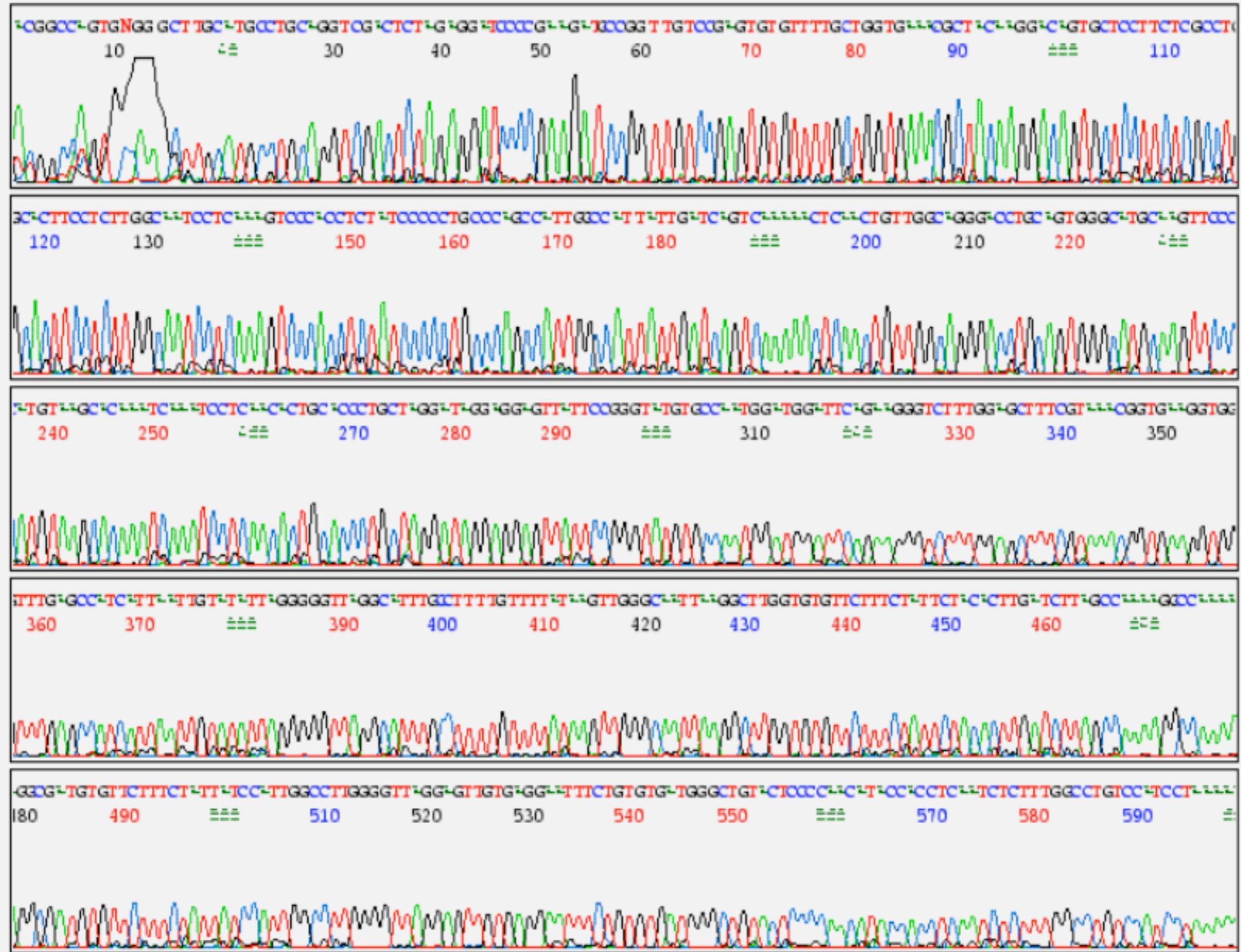


G: G Reaction  
A: A Reaction  
T: T Reaction  
C: C Reaction

# Analyzing Fluorescent DNA Sequencing Data

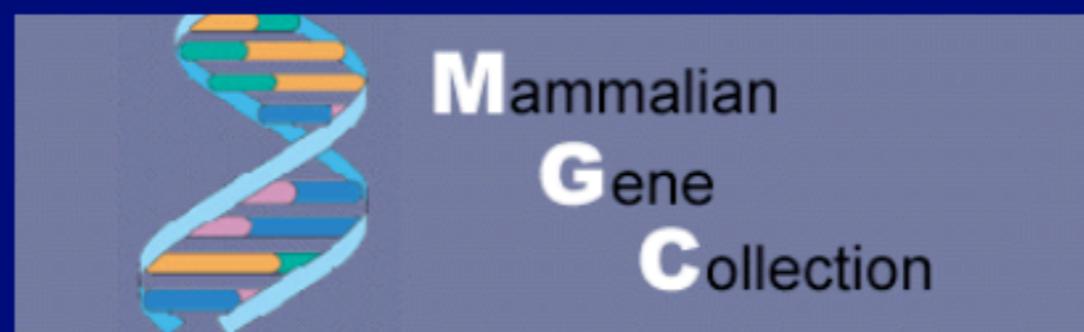


# Fluorescent DNA Sequencing Results



# **Analysis of Gene Expression**

- ESTs: **Expressed-Sequence Tags**
- SAGE: **Serial Analysis of Gene Expression**
- Full-Insert (Full-Length) cDNA Sequencing



[mgc.nci.nih.gov](http://mgc.nci.nih.gov)

# **Genome Sequencing**



# **Shotgun Sequencing**

# Subclone Construction

```
GATCGTCTAGAACATCTC  
GAGATCTCTGAGAGTC  
GTGGAAAACGTGTGAA  
TGTGACTAGCCACAGT  
TACGTGTGAGAGATGT  
ATGATGCACCTGACCC  
GGGTTCACTCTCAAC  
GACTCACTCCACCTCA  
GAGGCCACCGCCGCT  
GTGCACGTCCACCACC  
GATTATTACCATTTA  
ATCCTTAGGATTGACA
```

**BAC DNA**

GAT	GAT	GAT	GAT	GATCGTCTAGAACATCTC
GAA	GAA	GAA	GAA	GAGATCTCTGAGAGTC
GTC	GTC	GTC	GTC	GTGGAAAACGTGTGAA
TGT	TGT	TGT	TGT	TGTGACTAGCCACAGT
TAA	TAA	TAA	TAA	TACGTGTGAGAGATGT
ATG	ATG	ATG	ATG	ATGATGCACCTGACCC
GGG	GGG	GGG	GGG	GGGTTCACTCTCAAC
GAA	GAA	GAA	GAA	GACTCACTCCACCTCA
GTC	GTC	GTC	GTC	GAGGCCACCGCCGCT
GAT	GAT	GAT	GAT	GTGCACGTCCACCACC
ATC	ATC	ATC	ATC	GATTATTACCATTTA
				ATCCTTAGGATTGACA



**Prepare Multiple Copies**

The diagram shows a single row of DNA sequence segments, each identical to the original BAC DNA sequence, representing multiple copies of the BAC DNA.



**Randomly Fragment**

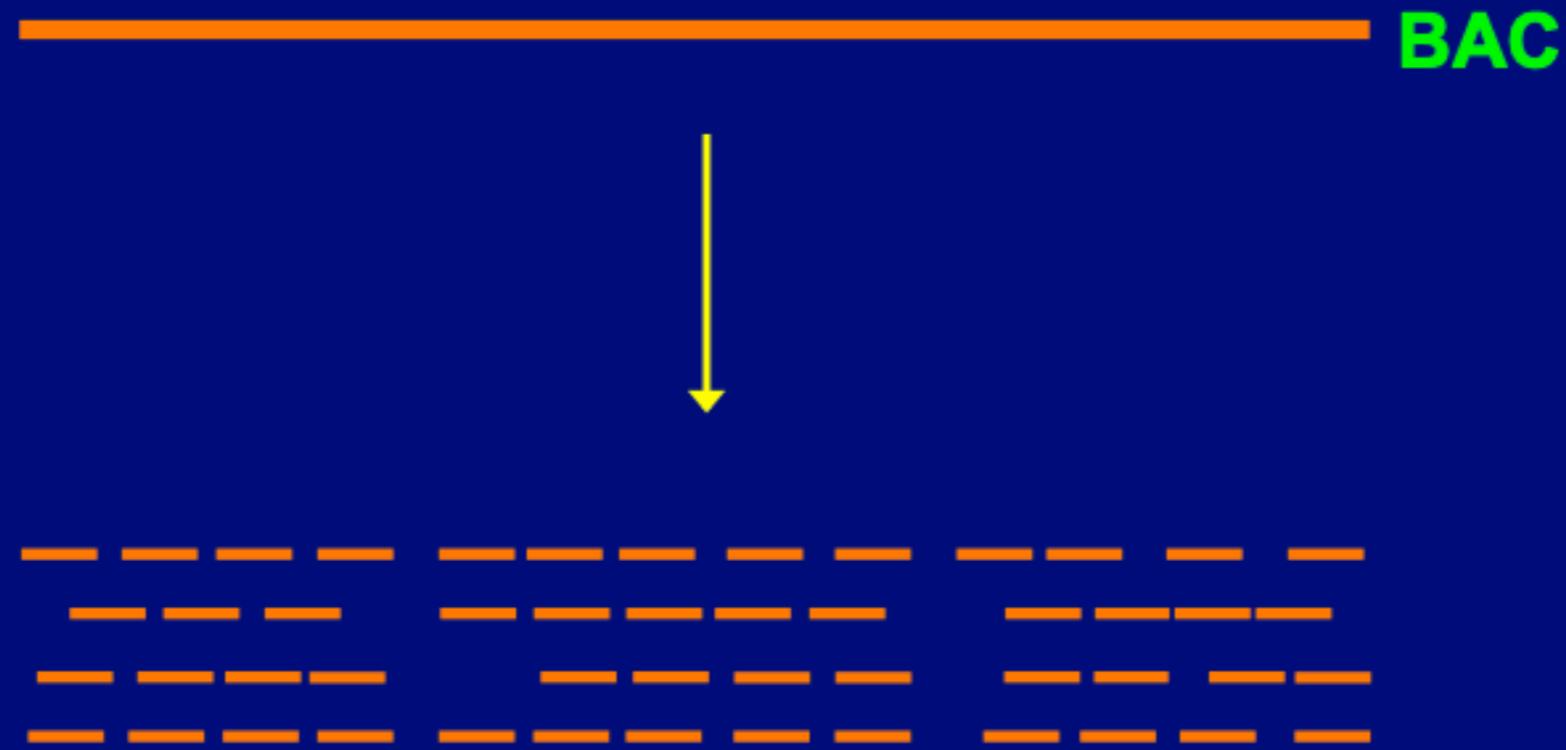
The diagram shows the BAC DNA sequence from the previous step being randomly cut into smaller, overlapping fragments represented by orange horizontal dashes.



**Subclone Fragments**

The diagram shows the randomly fragmented DNA pieces being cloned into plasmid vectors, represented by orange and pink rings.

# Shotgun Sequencing Strategy



# Poisson Calculations

The sequencing strategy for the shotgun approach follows the Lander and Waterman application of the Poisson distribution

The probability a base is not sequenced is given by:

$$P_0 = e^{-c}$$

Where:

- $c$  = fold sequence coverage ( $c=LN/G$ ),
- $LN$  = # bases sequenced, i.e.  $L$  = average sequencing read length and  $N$  = # reads
- $G$  = target sequence length
- $e = 2.718$  ( $e=2.718281828459$ )

Fold Coverage	$P_0 = e^{-c}$	% not sequenced	% sequenced
1	0.37	37%	63%
2	0.135	13.5%	87.5%
3	0.05	5%	95%
4	0.018	1.8%	98.2%
5	0.0067	0.6%	99.4%
6	0.0025	0.25%	99.75%
7	0.0009	0.09%	99.91%
8	0.0003	0.03%	99.97
9	0.0001	0.01%	99.99%
10	0.000045	0.005%	99.995%

**(a) Sequence reads**

Read 1 CACATACACATGG

Read 2 TCAATGGGGCTAA

Read 3 AGCACGGACTTGTCAACATACACATG

Read 4 ACACATGGAAATA

Read 5 GGGCTAATGATTGTCAC

Read 6 TGATTGTCACATA

Read 7 ATTCAATGAAGCACCGA

Read 8 GTCACATACACATGATCAATGGGG

↓ Use computer to assemble sequence reads

**(b)**

7 ATTCAATGAAGCACCGA

3 AGCACGGACTTGTCAACATACACATG

8 GTCACATACACATGATCAATGGGG

2 TCAATGGGGCTAA

5 GGGCTAATGATTGTCAC

6 TGATTGTCACATA

1 CACATACACATGG

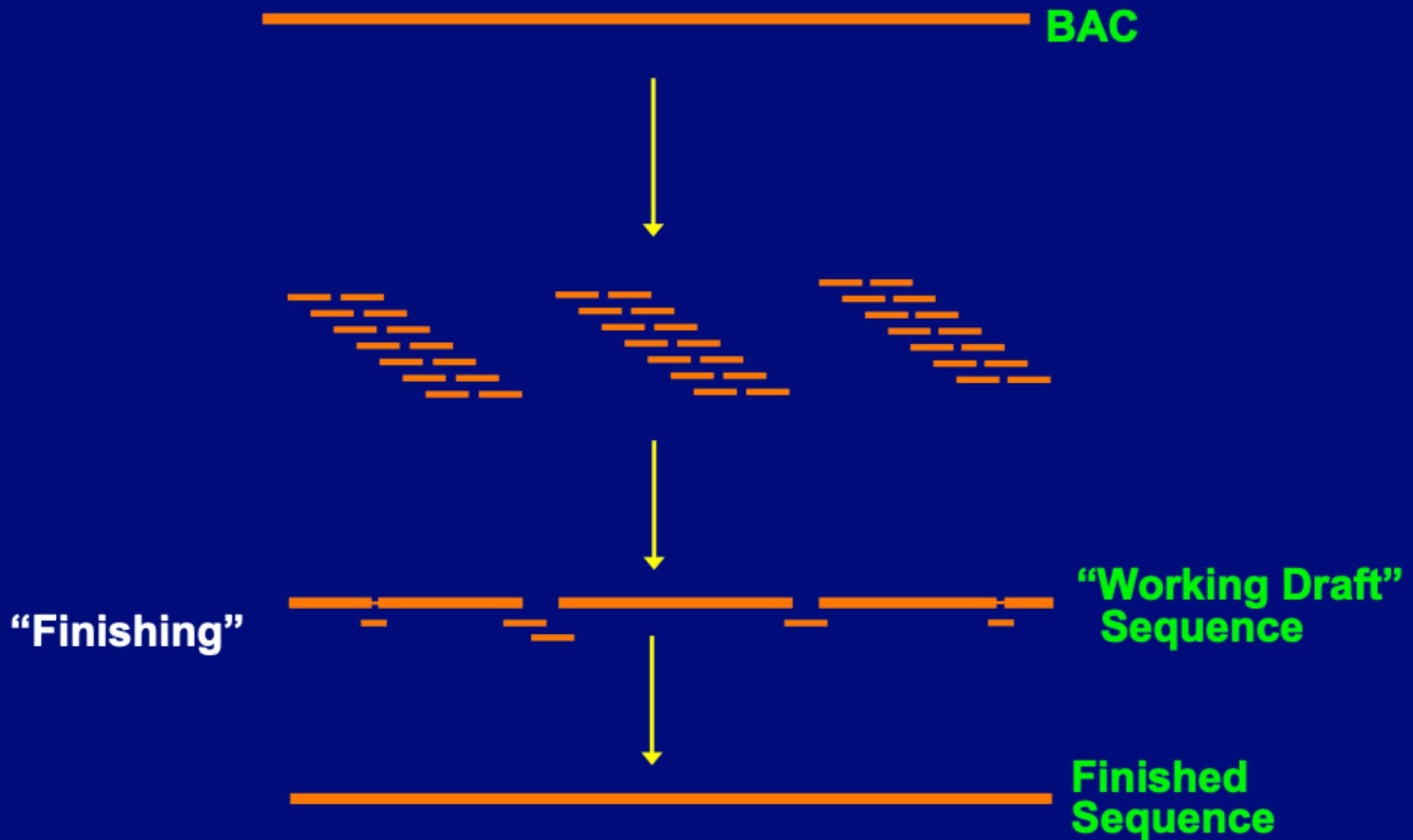
4 ACACATGGAAATA

↓ Assembled sequence

**(c)**

ATTCAATGAAGCACCGACTTGTCAACATACACATGATCAATGGGGCTAAATGATTGTCACATACACATGGAAATA

# Shotgun Sequencing Strategy



## Trace Window: Contig32

yg02f10.y1

con  
rd  
5 8030 8035 8040 8045 8050 8055  
365 360 355 350 345 340con  
edt  
V ABIScroll  
Together?  Yes  No

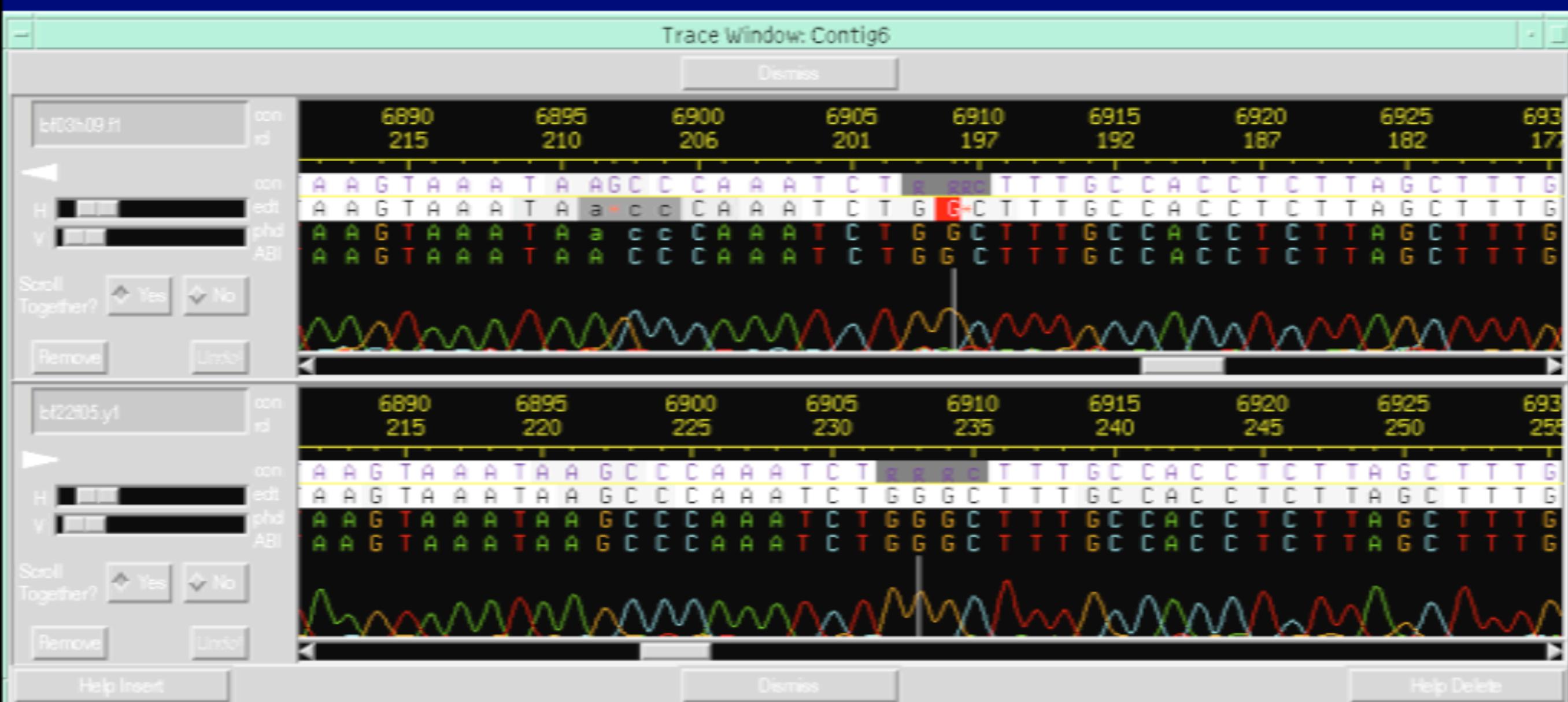
yg03g10.y1

con  
rd  
5 8030 8035 8040 8045 8050 8055  
454 449 444 439 434 429con  
edt  
V ABIScroll  
Together?  Yes  No

yg18a10.y1

con  
rd  
5 8030 8035 8040 8045 8050 8055  
173 168 163 158 153 148con  
edt  
V ABIScroll  
Together?  Yes  No

# Sequence Finishing: Resolving Ambiguities



\*\*\* Sequence Finishing: Remains Relatively Expensive \*\*\*

# **Historically Significant Genome Sequencing Projects**

# First Eukaryotic Genome Sequence

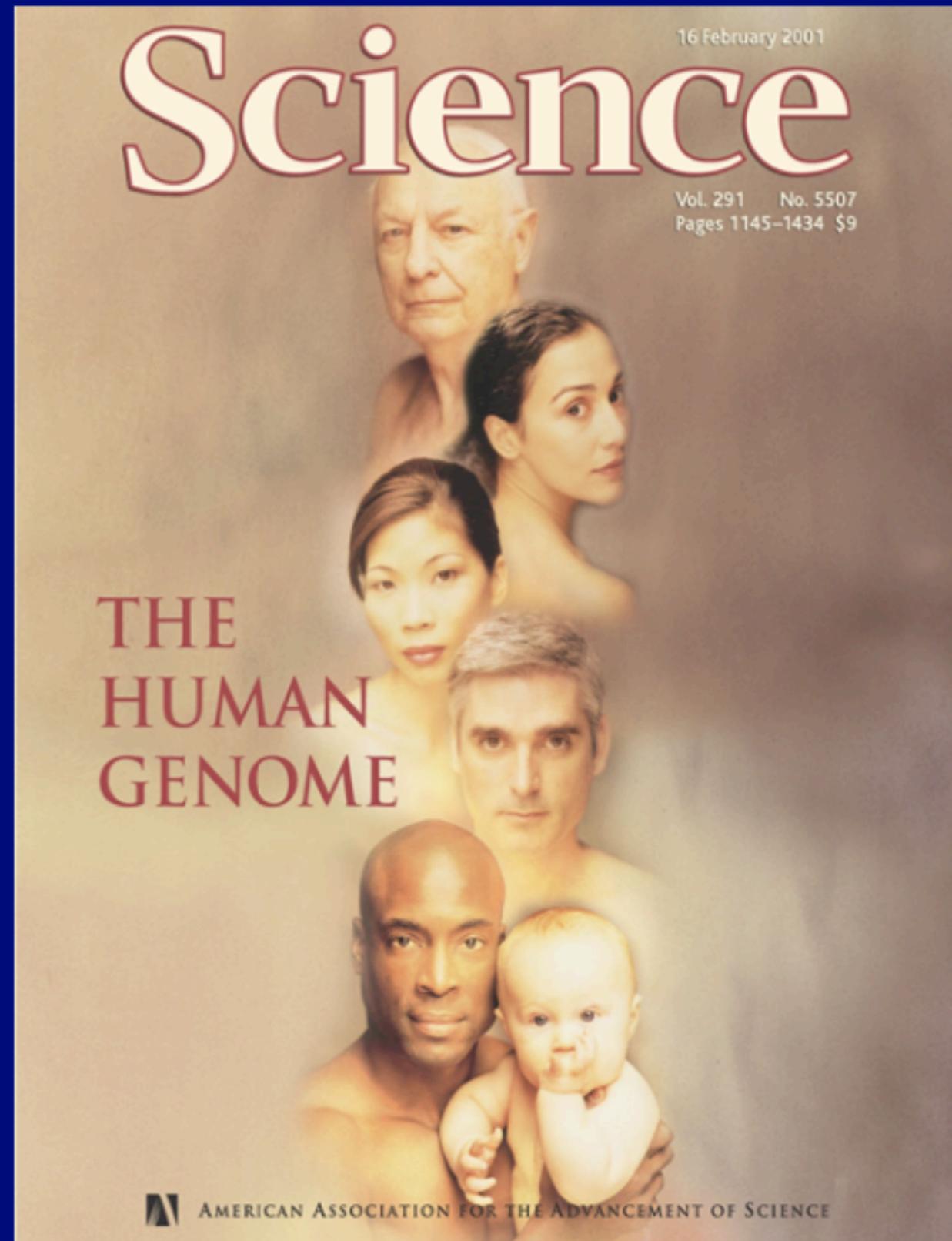


Goffeau et al. (1997)

# February, 2001 Draft Sequence

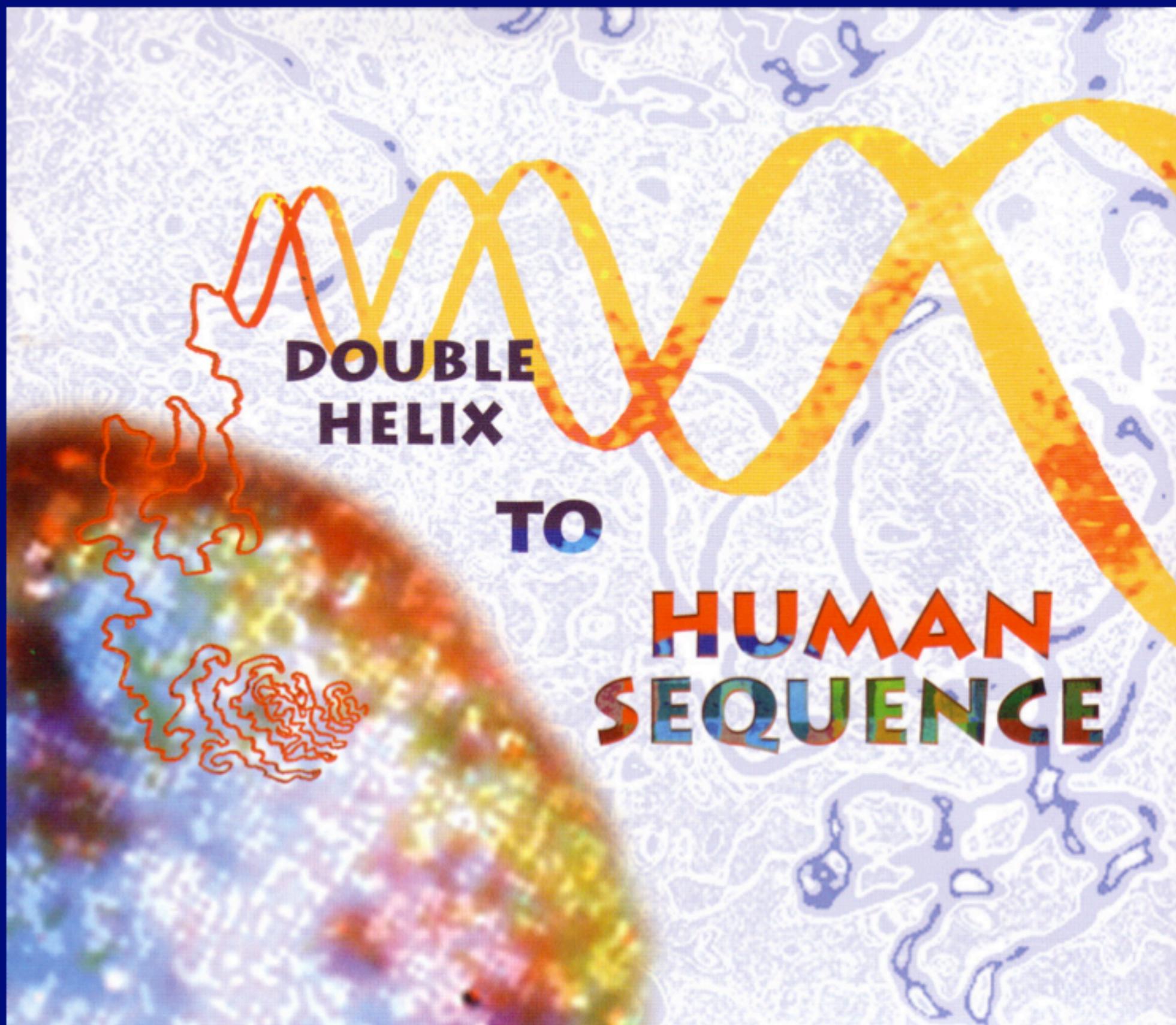


International Human Genome  
Sequencing Consortium (2001)



Venter et al. (2001)

# April, 2003 Completion



# October, 2004 Publication

21 October 2004

International weekly journal of science.

# nature

£19.00

**Tetraodon to human**  
Evolutionary history in genome sequences

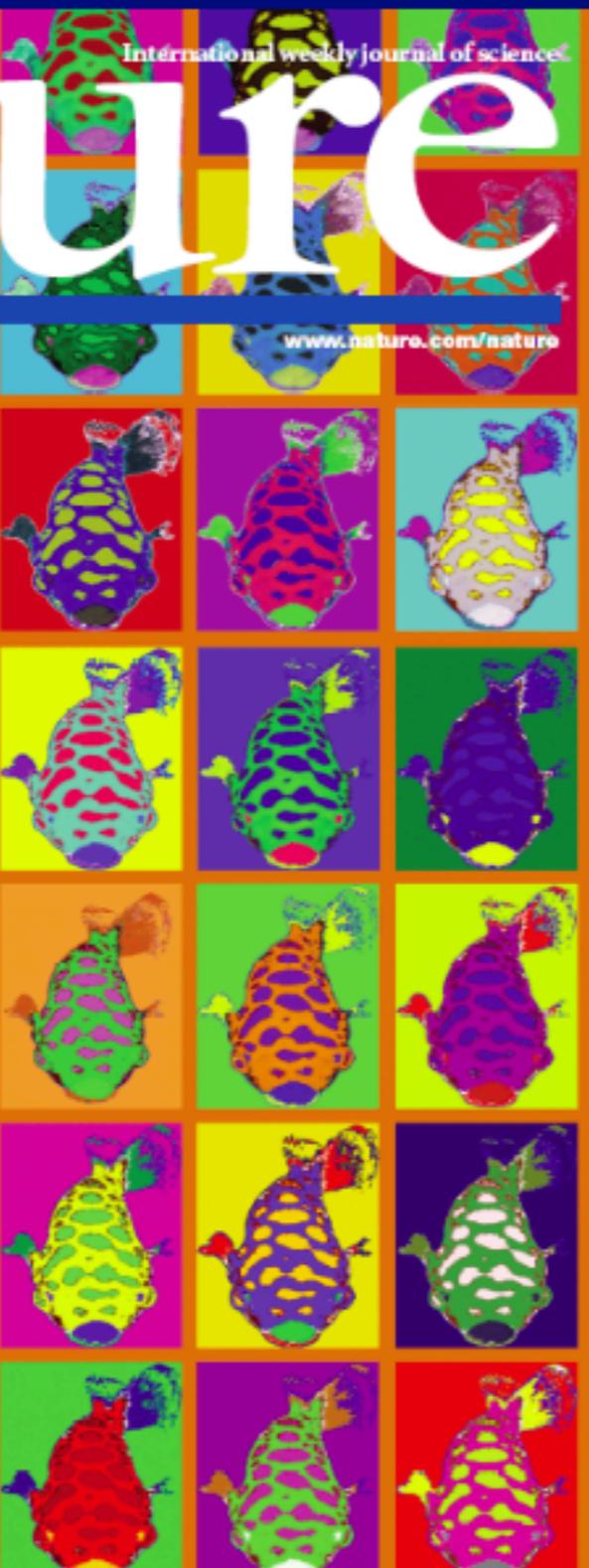
**General relativity**  
Did the orbit move for you?

**The human genome**  
Going the last mile

**Antibiotics crisis**  
Market forces fail to deliver

**Medical ethics**  
Choosing deafness

**naturejobs think Finland**



articles

## Finishing the euchromatic sequence of the human genome

International Human Genome Sequencing Consortium\*

\*A list of authors and their affiliation appears in the Supplementary Information

The sequence of the human genome encodes the genetic instructions for human physiology, as well as rich information about human evolution. In 2001, the International Human Genome Sequencing Consortium reported a draft sequence of the euchromatic portion of the human genome. Since then, the international collaboration has worked to convert this draft into a genome sequence with high accuracy and nearly complete coverage. Here, we report the result of this finishing process. The current genome sequence (Build 35) contains 2.85 billion nucleotides interrupted by only 341 gaps. It covers ~99% of the euchromatic genome and is accurate to an error rate of ~1 event per 100,000 bases. Many of the remaining euchromatic gaps are associated with segmental duplications and will require focused work with new methods. The near-complete sequence, the first for a vertebrate, greatly improves the precision of biological analyses of the human genome including studies of gene number, birth and death. Notably, the human genome seems to encode only 20,000–25,000 protein-coding genes. The genome sequence reported here should serve as a firm foundation for biomedical research in the decades ahead.

The Human Genome Project (HGP) was launched in 1990 with the goal of obtaining a highly accurate sequence of the vast majority of the euchromatic portion of the human genome. The initial work followed a two-pronged approach: (1) the mapping of the human and mouse genomes<sup>1–3</sup> to allow the study of inherited disease and provide a crucial scaffold for genome assembly; and (2) the sequencing of organisms with smaller, simpler genomes<sup>4–6</sup> to serve as a testbed for method development and assist in interpreting the human genome. With success along both paths, the sequencing of the human genome itself eventually became feasible. The International Human Genome Sequencing Consortium (IHGSC), an open collaboration involving twenty centres in six countries, was formed to carry out this component of the HGP.

In February 2001, the IHGSC<sup>7</sup> and Celera Genomics<sup>8</sup> each reported draft sequences providing a first overall view of the human genome. These sequences allowed systematic study of the human genome itself, including identification of genes, combinatorial architecture of proteins, regional differences in genome composition, distribution and history of transposable elements, distribution of polymorphism and relationship between genetic recombination and physical distance. Moreover, systematic knowledge of the human genome has enabled new tools and approaches that have markedly accelerated biomedical research.

Both draft sequences, however, had important shortcomings. The IHGSC sequence, for example, omitted ~10% of the euchromatic genome; it was interrupted by ~150,000 gaps; and the order and orientation of many segments within local regions had not been established. The IHGSC thus turned to the challenge of completing the sequence of the euchromatic genome. Operationally, a finished sequence was defined as having an error rate of, at most, one event per  $10^4$  bases, and the goal for completion was coverage in finished sequence of at least 95% of the euchromatic genome, with the only gaps being those refractory to all available techniques<sup>9</sup> (see <http://www.genome.gov/10000923>). The goal was challenging because the human genome is replete with such features as dispersed repeats and large segmental duplications, which greatly complicate the determination of genome structure and sequence. In fact, near-complete sequences have been obtained so far only for three multicellular organisms: the nematode<sup>10</sup>, mustard weed<sup>11</sup> and the fruitfly<sup>12</sup>. These genomes are all roughly 30-fold smaller than the human genome and have much simpler structure.

We describe here the results of a multiyear effort by the IHGSC

towards the goal of a complete human sequence. The number of gaps has been reduced 400-fold to only 341, most of which are associated with segmental duplications and will require new methods for resolution. The assembled near-complete genome sequence has an error rate of only ~1 event per 100,000 bases; it contains 2.85 billion nucleotides and covers ~99% of the euchromatic genome. This paper describes the current genome sequence and the process used to produce it; examines the accuracy and completeness of the sequence; and illustrates biological analyses made possible by the sequence. We do not attempt here a comprehensive analysis of the contents of the human genome. An initial analysis was previously reported<sup>13</sup> and a series of papers is being written describing the individual chromosomes<sup>14–18</sup>, including annotation of genes and other features.

### Current genome sequence

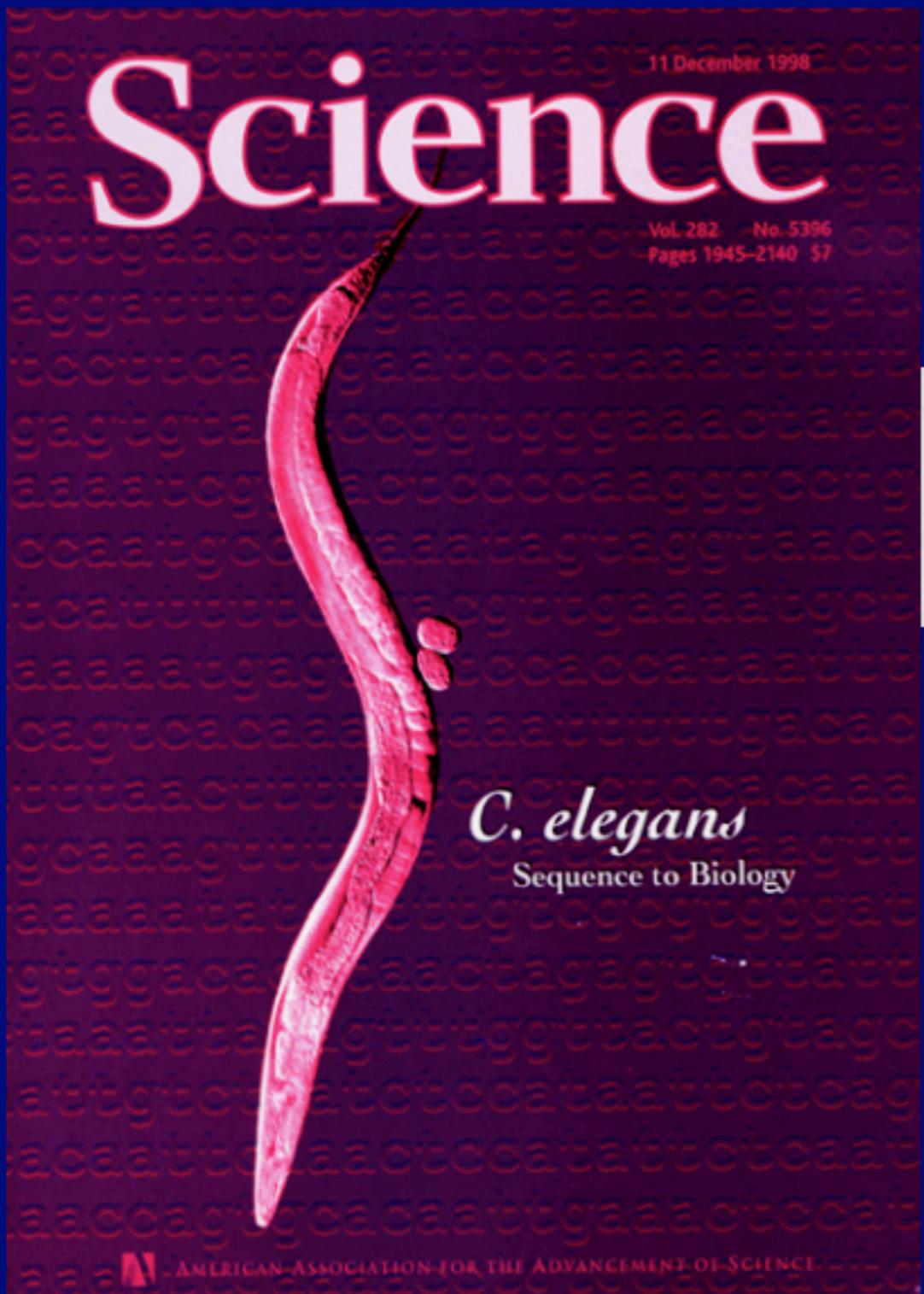
#### Finishing process

The process of converting the initial draft sequence into a near-complete sequence is referred to as 'finishing'. It is a complex iterative process that proceeds simultaneously at multiple scales, ranging from single nucleotides to the integrity of whole chromosomes. The fundamental challenge is that genomic regions that are not well represented or readily resolved through random shotgun sequencing tend to be highly enriched in problematic sequences. Resolving such regions required the development of special approaches, which evolved substantially over time and varied among centres.

Broadly, the finishing process involved two distinct components: (1) producing finished maps, consisting of continuous and accurate paths of overlapping large-insert clones spanning the euchromatic region of each chromosome arm; and (2) producing finished clones, consisting of continuous and accurate nucleotide sequence across each large-insert clone. In practice, these two components were tightly intertwined in that progress in each often depended on results from the other. The components are described in Boxes 1 and 2. Further information about the finishing process and finishing standards can be found in the Supplementary Information (Note 1) and at <http://www.genome.gov/10000923>.

In total, we generated a shotgun sequence from 59,208 large-insert clones (total length ~5.84 gigabases (Gb)) and finished the sequence from 45,742 of these clones (total length ~3.67 Gb). The clones consisted primarily of bacterial artificial chromosomes

# First Animal Genome Sequence

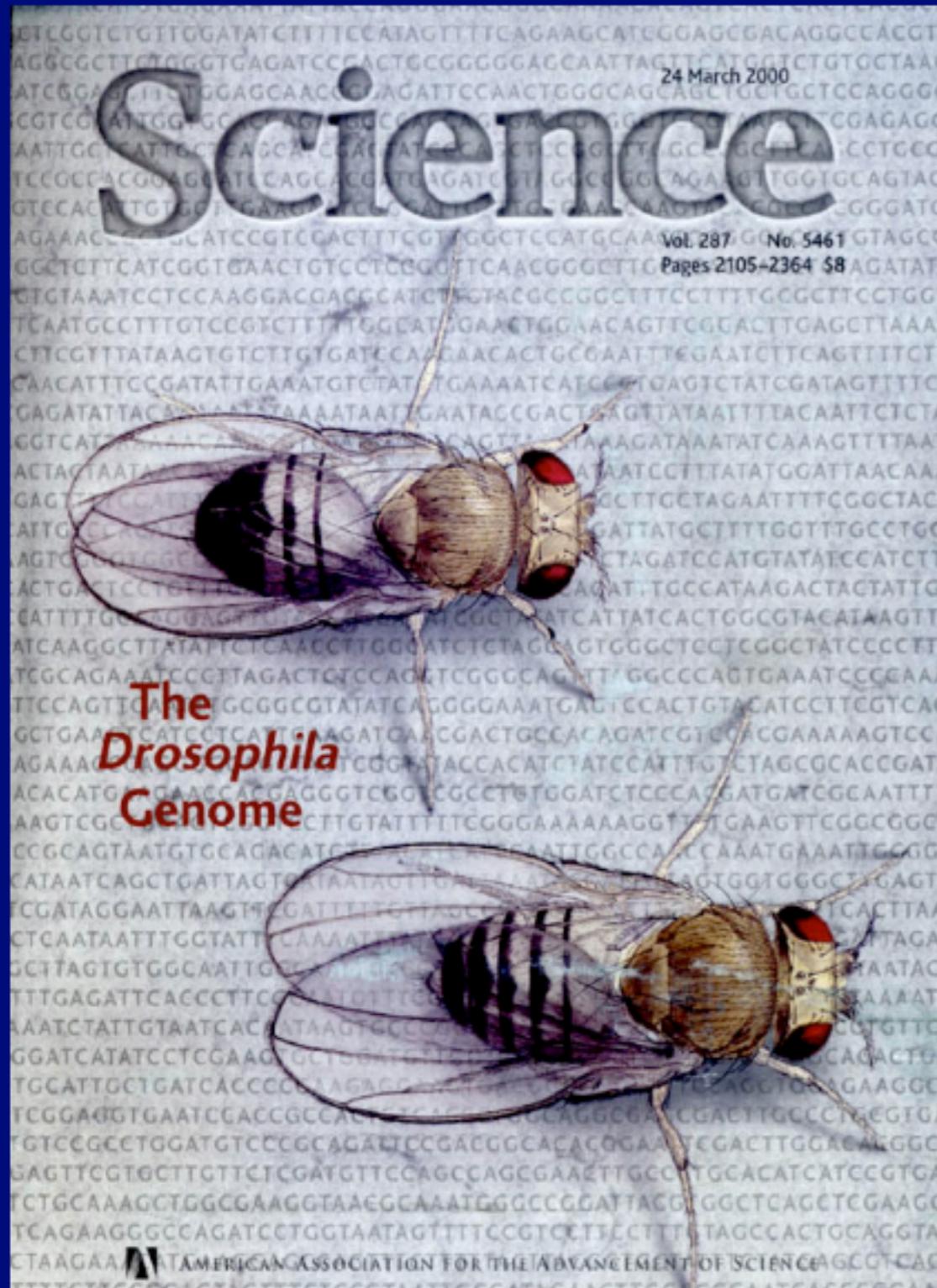


## Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology

The *C. elegans* Sequencing Consortium\*

***C. elegans* Sequencing Consortium (1998)**

# Second Animal Genome Sequence



# Science

24 March 2000

Vol. 287, No. 5461  
Pages 2105–2364 \$8 AGATAT



The  
*Drosophila*  
Genome



AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

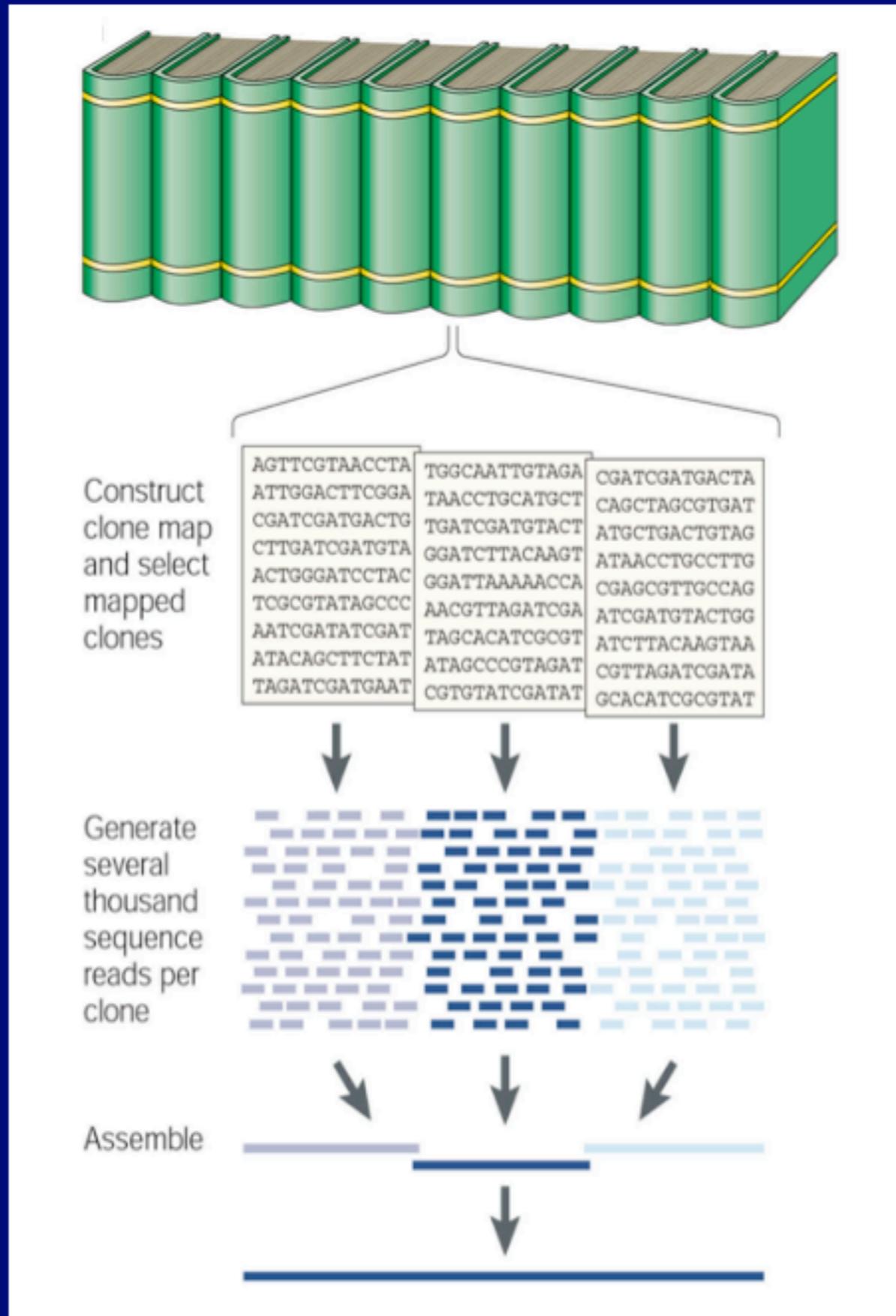
THE DROSOPHILA GENOME  
REVIEW

## The Genome Sequence of *Drosophila melanogaster*

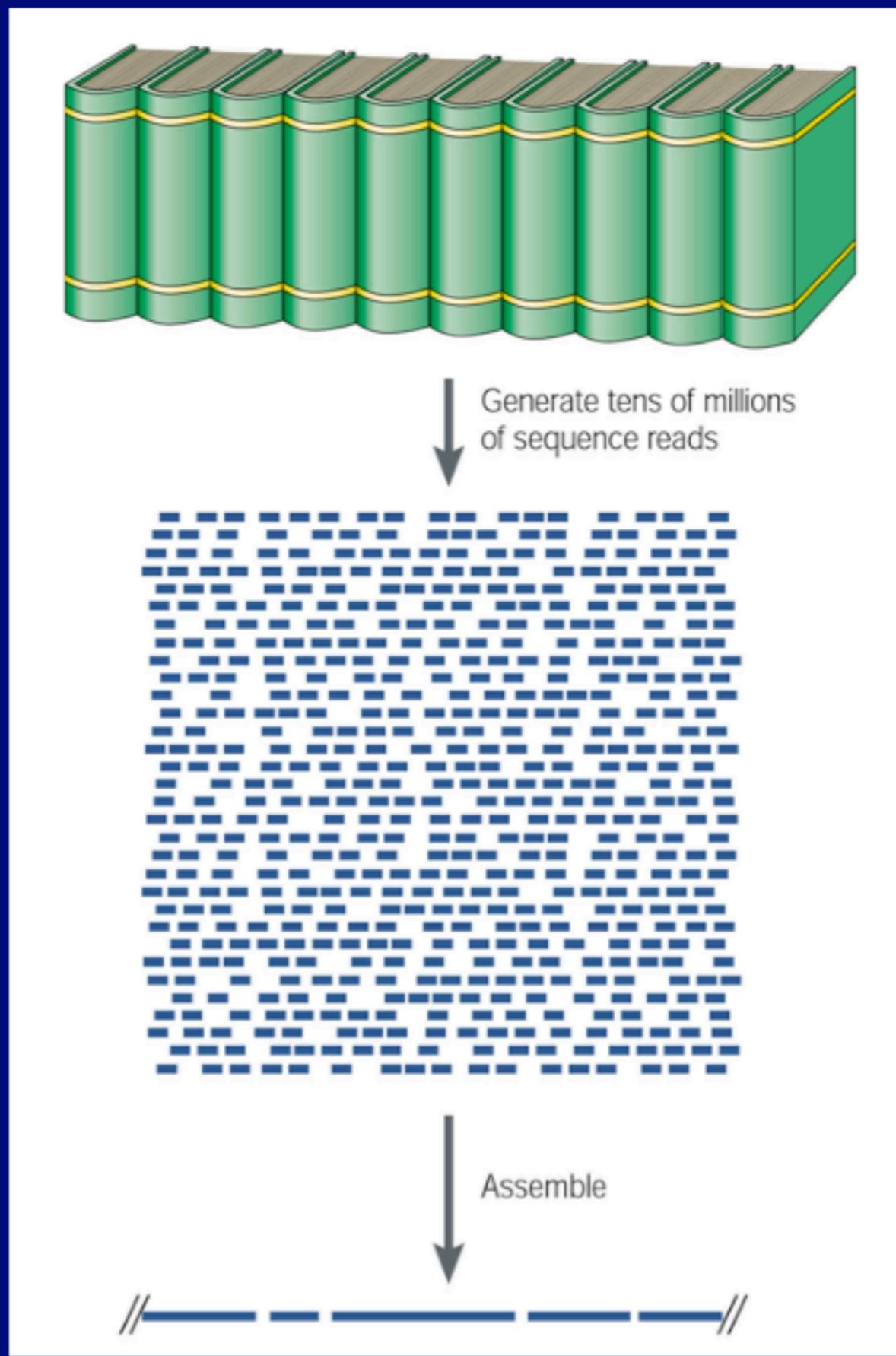
Mark D. Adams,<sup>1\*</sup> Susan E. Celinker,<sup>2</sup> Robert A. Holt,<sup>1</sup> Cheryl A. Evans,<sup>1</sup> Jeannine D. Gocayne,<sup>1</sup> Peter G. Amanatides,<sup>1</sup> Steven E. Scherer,<sup>3</sup> Peter W. Li,<sup>1</sup> Roger A. Hoskins,<sup>2</sup> Richard F. Gallo,<sup>2</sup> Reed 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 Holt,<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. Balliew,<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> Phillip 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 Dahike,<sup>1</sup> Lionel B. Davenport,<sup>1</sup> Peter Davies,<sup>1</sup> Beatriz de Pablo,<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 Ferreira,<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>2</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 Moberly,<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. Paclob,<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 Reiner,<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. Wasserman,<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> Shaojing 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>

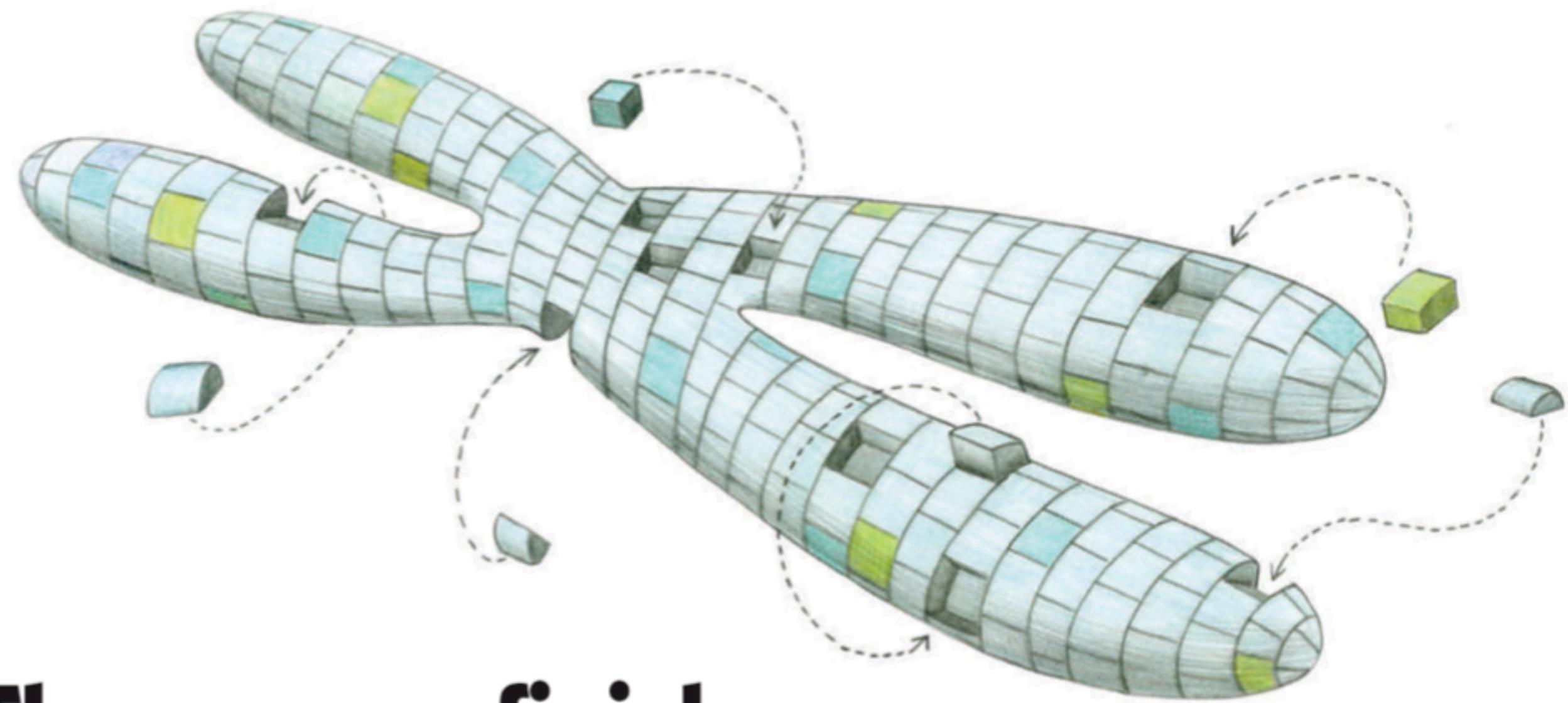
Adams et al. (2000)

# Clone-Based Shotgun Sequencing



# Whole-Genome Shotgun Sequencing





# The genome finishers

Dedicated scientists are working hard to close the gaps, fix the errors and finally complete the human genome sequence. **Elie Dolgin** looks at how close they are.

*Nature* (2009)



# Human Genome Project

D.T. Max: What Darwin Can Teach Us About Jane Austen; Alex Witcher: An Erma Bombeck for Military Wives

# The New York Times Magazine

NOVEMBER 8, 2003 • SECTION 8

Chem

og &  
B

## The Dawn of Genomic Medicine

How a pediatrician working with the Amish is changing what it means to diagnose and treat disease.

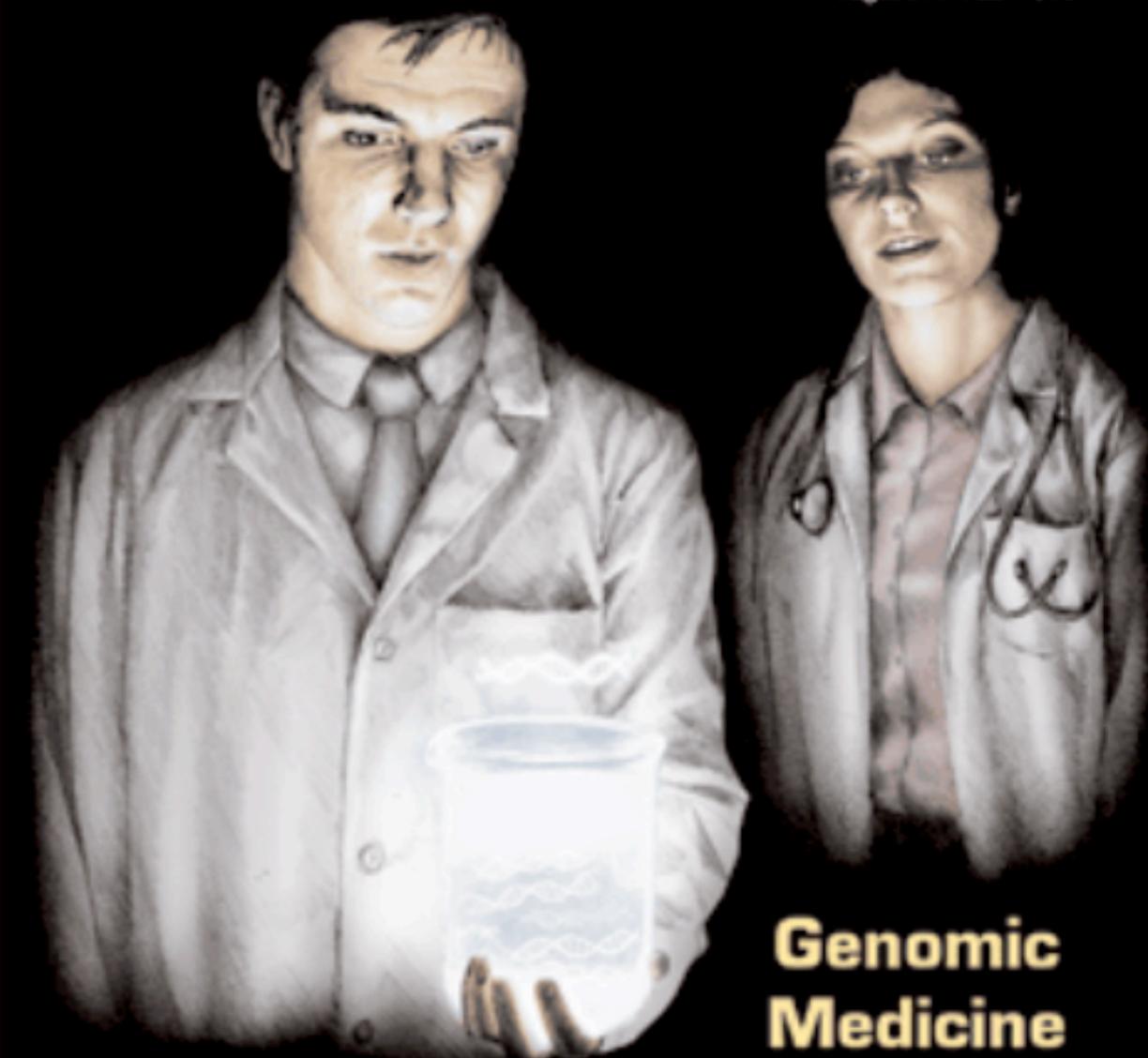
By Lisa Belkin

Why Waste Your Time Voting? (See Freakonomics, Page 30)

24 October 2003

# Science

Vol. 302 No. 5645  
Pages 517–728 \$10



Genomic  
Medicine



AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE