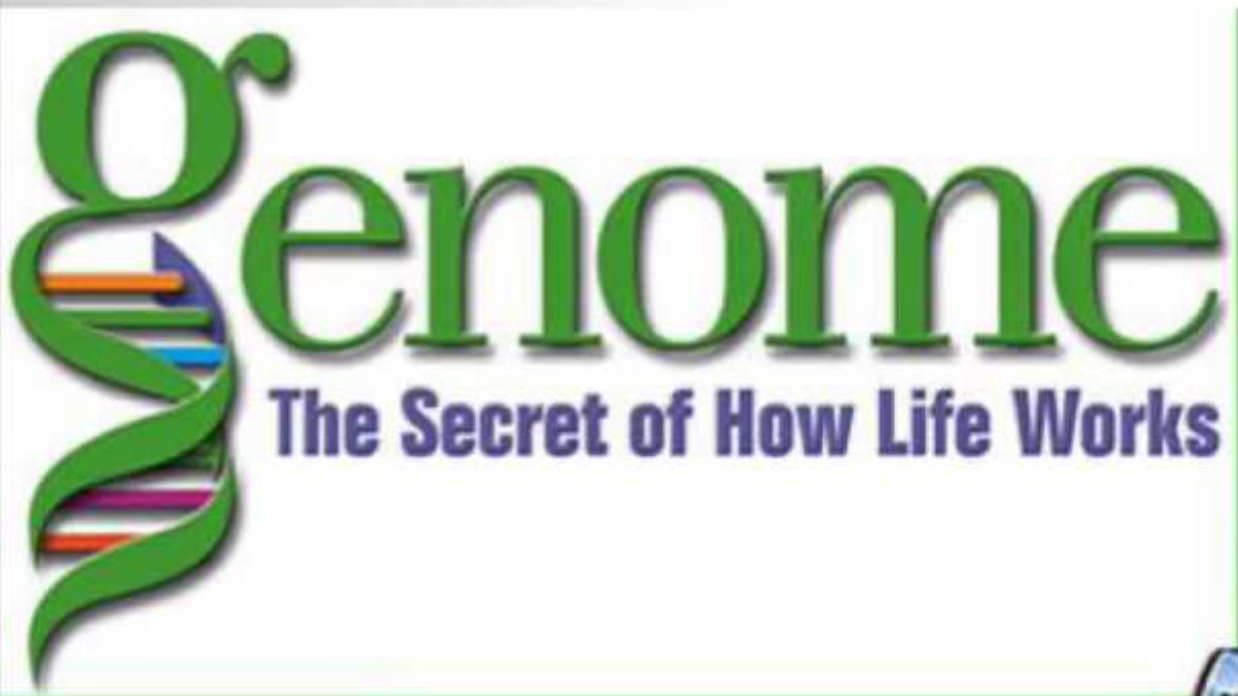


The Human Genome Project



Steven Salzberg

- Proposed in 1987 by the U.S. Department of Energy (not NIH!)
- Biology's “Manhattan project”



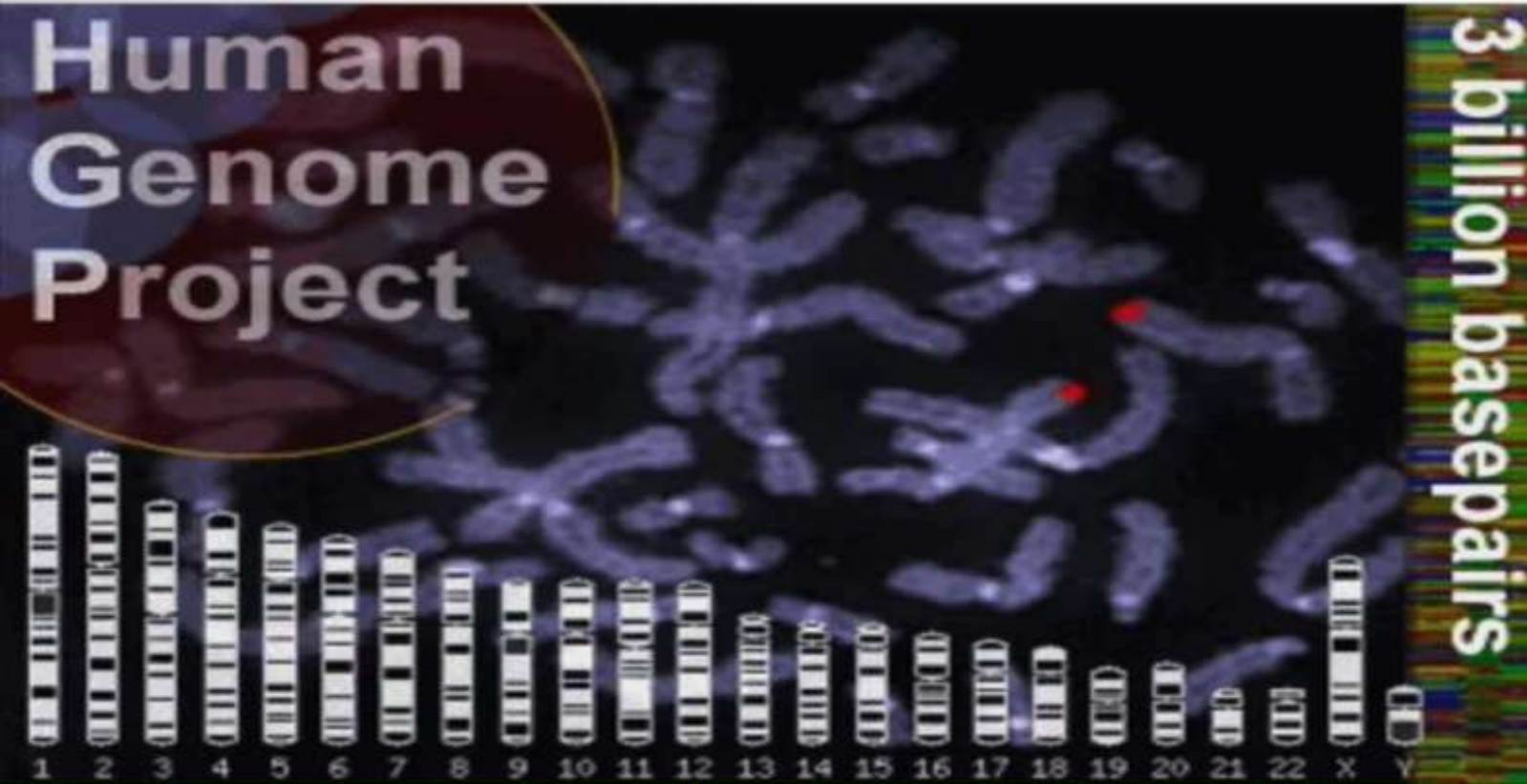
Human Genome Project: 1989 - present



- Officially started in 1989
- Joint effort of NIH and DOE in the United States, plus many other countries
- the Sanger Centre in England was the largest center outside the US

- sequence 3 billion basepairs
- for \$1/base
- by 2005

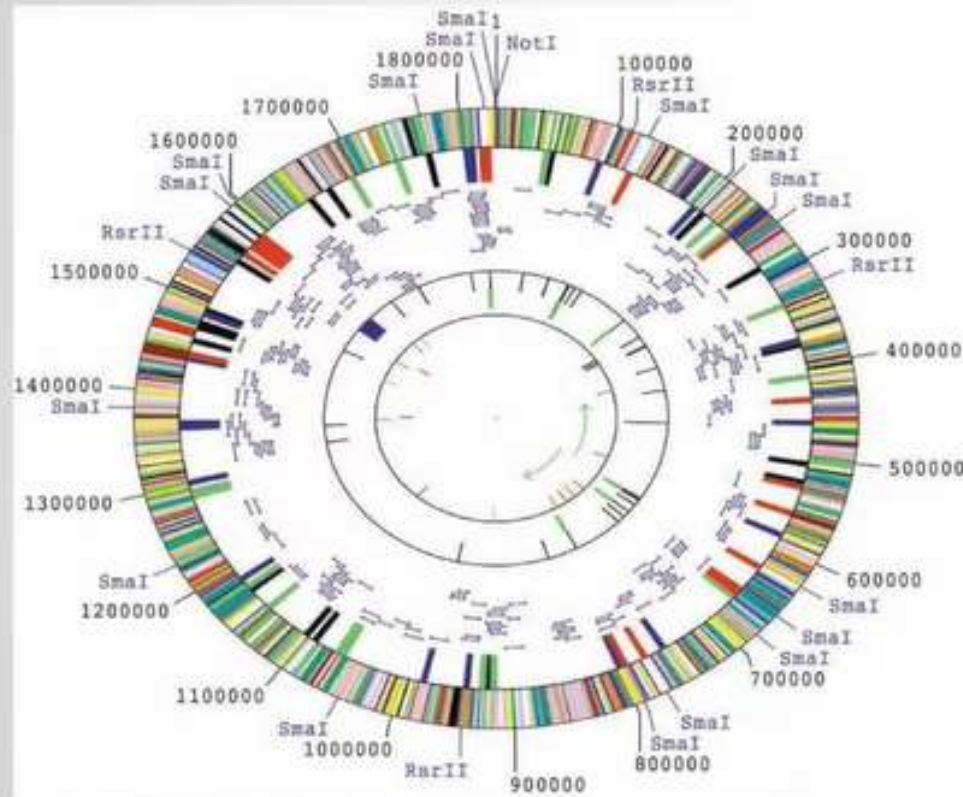
GOALS



The race to sequence the genome: early 1990s

- It wasn't a race – not yet!
- Scientists around the world were busy creating “maps”
 - Maps take small or large pieces of DNA and place them somewhere on the genome
 - Maps also take particular genes and identify their approximate location

1995: TIGR sequences first complete bacterial genome ever, *Haemophilus influenzae*

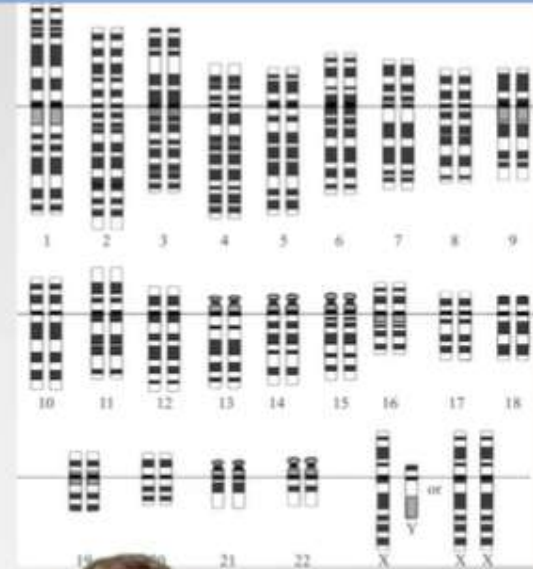


1.8 million bases
1742 genes

Project led by Craig Venter (TIGR) and Hamilton Smith (Johns Hopkins)

1998: the race begins

- new sequencing machine developed by Applied Biosystems
- Craig Venter, Ham Smith & others leave TIGR to form **Celera Genomics**, a for-profit company



The race nears the finish

- 1999: Craig Venter announces that Celera will finish by 2001
- 1999: NIH and the Sanger Centre announce that the public HGP will finish a “draft” genome by 2001
- 2000: NIH, Sanger Centre, and Celera talk about publishing jointly
- Late 2000: talks fall apart; 2 papers planned

June 2000: Bill Clinton, Tony Blair jointly
announce the completion of the human genome



Whose genome did we sequence?



What did the genome tell us?

letters to nature

Nature **201**, 847 (22 February 1964); doi:10.1038/201847a0

A Preliminary Estimate of the Number of Human Genes

F. VOGEL

Institut für Anthropologie und Humangenetik, University of Heidelberg, Germany.

RECENT results of molecular genetics enable us to estimate the number of human genes, if certain assumptions are made. The following data are available: (1) The α -chain of human haemoglobin contains 141, the β -chain contains 146 amino-acids, corresponding to a molecular weight of about 17,000 each¹. Assuming a triplet code^{2,3} this means that the α - and β -chains are determined by 423 and 438 nucleotide pairs, respectively. According to 'Svedberg's law'⁴, many proteins consist of sub-units of the same order of magnitude (molecular weight of about 17,500). Hence, the assumption seems to be warranted that one average structural gene might have a length of about 450 nucleotide pairs. (2) The weight of one haploid human chromosome set in human spermatozoa is about 2.72×10^{-12} g. Granulocytes contain about 6.23×10^{-12} g; lymphocytes contain about 5.84×10^{-12} g (ref. 5). Extensive examinations have shown that the DNA content is constant in all resting cells of one species, which have the same number of chromosome sets, and depends on the degree of polyploidy^{5,6}. The assumption seems to be justified that most of the DNA works as genetic material, even if in some cells minor fractions with other functions might possibly be present⁷. In the following calculations the total amount of DNA in a haploid human chromosome set is estimated to be about 3×10^{-12} g. (3) Usually the genetic variants of human haemoglobins differ in one amino-acid substitution only^{1,8}. One structural gene can only produce one single type of genetically determined polypeptide chain. As much as we know, this applies for other genetically determined proteins as well. This means that the genetic information for these structural genes can only be present once. Any degree of polyteny for these loci in the germ cells is highly unlikely. As has been mentioned, however, the DNA content of diploid cells is about twice the content of (haploid) spermatozoa. We assume that the total genetic information is only present once.

6.7 million
genes?





Human genome paper: *Nature* 409 (Feb 2001), 860-921

articles

Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium*

** A partial list of authors appears on the opposite page. Affiliations are listed at the end of the paper.*

- 30,000 - 40,000 genes

Science

Vol. 291, No. 5507
February 9, 2001

THE
HUMAN
GENOME

Human genome paper: Science 291 (Feb 2001), 1304-51

Sequence of the Human Genome

- 26,588 genes

Mark D. Adams,¹ Eugene W. Myers,¹ Peter W. Li,¹ Richard J. Mural,¹
 Hamilton O. Smith,¹ Mark Yandell,¹ Cheryl A. Evans,¹ Robert A. Holt,¹
 Gocayne,¹ Peter Amanatides,¹ Richard M. Balow,¹ Daniel H. Huson,¹
 Portman,¹ Qing Zhang,¹ Chinnappa D. Kodira,¹ Xiangqun H. Zheng,¹ Lin Chen,¹
 Gopalakrishnan Subramanian,¹ Paul D. Thomas,¹ Jinghui Zhang,¹
 Miklos,² Catherine Nelson,³ Samuel Broder,⁴ Andrew G. Clark,⁵ Joe Nadeau,⁶
 Jaisick,⁷ Norton Zinder,⁷ Arnold J. Levine,⁷ Richard J. Roberts,⁸ Mel Simon,⁹
 Michael Hunkapiller,¹⁰ Randall Bolanos,¹¹ Arthur Delcher,¹² Ian Dew,¹³ Daniel Fasulo,¹⁴
 Jana Flores,¹⁵ Aaron Halpern,¹⁶ Sridhar Hannenhalli,¹⁷ Saul Kravitz,¹⁸ Samuel Levy,¹⁹
 Reinert,²⁰ Karin Remington,²¹ Jane Abu-Threideh,²² Ellen Beasley,²³ Kendra Biddick,²⁴
 Linda Brandon,²⁵ Michele Cargill,²⁶ Ishwar Chandramouliswaran,²⁷ Rosane Charlab,²⁸
 Li,²⁹ Zuoming Deng,³⁰ Valentina Di Francesco,³¹ Patrick Dunn,³² Karen Ellbeck,³³
 Andrei E. Gabrielian,³⁴ Weinu Gan,³⁵ Wangmao Ge,³⁶ Fangcheng Gong,³⁷ Zhiping Gu,³⁸
 J. Heiman,³⁹ Maureen E. Higgins,⁴⁰ Rui-Ru Ji,⁴¹ Zhaoxi Ke,⁴² Karen A. Ketchum,⁴³
 Zhongwu Lai,⁴⁴ Yiding Lei,⁴⁵ Zhenya Li,⁴⁶ Jiajin Li,⁴⁷ Yong Liang,⁴⁸ Xiaoying Lin,⁴⁹ Fu Lu,⁵⁰
 Gennady V. Merkulov,⁵¹ Natalia Milshina,⁵² Helen M. Moore,⁵³ Ashwinkumar K. Malik,⁵⁴
 Vaibhav A. Narayan,⁵⁵ Beena Neelam,⁵⁶ Deborah Nusskern,⁵⁷ Douglas B. Rusch,⁵⁸ Steven Salzberg,¹²
 Wei Shao,⁵⁹ Bixiong Shue,⁶⁰ Jingtao Sun,⁶¹ Zhen Yuan Wang,⁶² Aihui Wang,⁶³ Xin Wang,⁶⁴ Jian Wang,⁶⁵
 Ming-Hui Wei,⁶⁶ Ron Wides,⁶⁷ Chunlin Xiao,⁶⁸ Chunhua Yan,⁶⁹ Allison Yao,⁷⁰ Jane Ye,⁷¹ Ming Zhan,⁷²
 Weiqing Zhang,⁷³ Hongyu Zhang,⁷⁴ Qi Zhao,⁷⁵ Liansheng Zheng,⁷⁶ Fei Zhong,⁷⁷ Wenyang Zhong,⁷⁸
 Shiaoqing C. Zhu,⁷⁹ Shaying Zhao,⁸⁰ Dennis Gilbert,⁸¹ Suzanna Baumhueter,⁸² Gene Spier,⁸³
 Christine Carter,⁸⁴ Anibal Cravchik,⁸⁵ Trevor Woodage,⁸⁶ Feroze Ali,⁸⁷ Huijin An,⁸⁸ Aderonke Awe,⁸⁹
 Danita Baldwin,⁹⁰ Holly Baden,⁹¹ Mary Barnstead,⁹² Ian Barrow,⁹³ Karen Beeson,⁹⁴ Dana Busam,⁹⁵
 Amy Carver,⁹⁶ Angela Center,⁹⁷ Ming Lai Cheng,⁹⁸ Liz Curry,⁹⁹ Steve Danaher,¹⁰⁰ Lionel Davenport,¹⁰¹
 Raymond Desilets,¹⁰² Susanne Dietz,¹⁰³ Kristina Dodson,¹⁰⁴ Lisa Doup,¹⁰⁵ Steven Ferreira,¹⁰⁶ Neha Garg,¹⁰⁷
 Andres Gluecksmann,¹⁰⁸ Brit Hart,¹⁰⁹ Jason Haynes,¹¹⁰ Charles Haynes,¹¹¹ Cheryl Heiner,¹¹² Suzanne Hladun,¹¹³
 Damon Hostin,¹¹⁴ Jarrett Houck,¹¹⁵ Timothy Howland,¹¹⁶ Chinyere Ibegwam,¹¹⁷ Jeffery Johnson,¹¹⁸
 Francis Kalush,¹¹⁹ Lesley Kline,¹²⁰ Shashi Koduru,¹²¹ Amy Love,¹²² Felecia Mann,¹²³ David May,¹²⁴
 Steven McCawley,¹²⁵ Tina McIntosh,¹²⁶ Ivy McFullen,¹²⁷ Mee Moy,¹²⁸ Linda Moy,¹²⁹ Brian Murphy,¹³⁰
 Keith Nelson,¹³¹ Cynthia Pfannkuch,¹³² Eric Pratts,¹³³ Vinita Puri,¹³⁴ Hina Qureshi,¹³⁵ Matthew Reardon,¹³⁶
 Robert Rodriguez,¹³⁷ Yu-Hui Rogers,¹³⁸ Deanna Romblad,¹³⁹ Bob Ruhfel,¹⁴⁰ Richard Scott,¹⁴¹ Cynthia Sitter,¹⁴²
 Michelle Smallwood,¹⁴³ Erin Stewart,¹⁴⁴ Renee Strong,¹⁴⁵ Ellen Suh,¹⁴⁶ Reginald Thomas,¹⁴⁷ Ni Ni Tint,¹⁴⁸
 Sukyee Tse,¹⁴⁹ Claire Vech,¹⁵⁰ Gary Wang,¹⁵¹ Jeremy Wetter,¹⁵² Sherita Williams,¹⁵³ Monica Williams,¹⁵⁴
 Sandra Windsor,¹⁵⁵ Emily Winn-Deen,¹⁵⁶ Keriellen Wolfe,¹⁵⁷ Jayshree Zaveri,¹⁵⁸ Karena Zaveri,¹⁵⁹
 Josep F. Abril,¹⁶⁰ Roderic Guigó,¹⁶¹ Michael J. Campbell,¹⁶² Kimmen V. Sjolander,¹⁶³ Brian Karlak,¹⁶⁴
 Anish Kejariwal,¹⁶⁵ Huaiyu Mi,¹⁶⁶ Betty Lazareva,¹⁶⁷ Thomas Hatton,¹⁶⁸ Apurva Narechania,¹⁶⁹ Karen Diemer,¹⁷⁰
 Anushya Muruganujan,¹⁷¹ Nan Guo,¹⁷² Shinji Sato,¹⁷³ Vineet Bafna,¹⁷⁴ Sorin Istrail,¹⁷⁵ Ross Lippert,¹⁷⁶
 Russell Schwartz,¹⁷⁷ Brian Walenz,¹⁷⁸ Shibu Yooseph,¹⁷⁹ David Allen,¹⁸⁰ Anand Basu,¹⁸¹ James Baxendale,¹⁸²
 Louis Blick,¹⁸³ Marcelo Caminha,¹⁸⁴ John Carnes-Stine,¹⁸⁵ Parris Caulk,¹⁸⁶ Yen-Hui Chiang,¹⁸⁷ My Coyne,¹⁸⁸
 Carl Dahlke,¹⁸⁹ Anne Deslattes Mays,¹⁹⁰ Maria Dombroski,¹⁹¹ Michael Donnelly,¹⁹² Dale Ely,¹⁹³ Shiva Esparham,¹⁹⁴
 Carl Foster,¹⁹⁵ Harold Gire,¹⁹⁶ Stephen Glanowski,¹⁹⁷ Kenneth Glasser,¹⁹⁸ Anna Glodek,¹⁹⁹ Mark Gorokhov,²⁰⁰
 Ken Graham,²⁰¹ Barry Gropman,²⁰² Michael Harris,²⁰³ Jeremy Heil,²⁰⁴ Scott Henderson,²⁰⁵ Jeffrey Hoover,²⁰⁶
 Donald Jennings,²⁰⁷ Catherine Jordan,²⁰⁸ James Jordan,²⁰⁹ John Kasha,²¹⁰ Leonid Kagan,²¹¹ Cheryl Kraft,²¹²
 Alexander Levitsky,²¹³ Mark Lewis,²¹⁴ Xiangjun Liu,²¹⁵ John Lopez,²¹⁶ Daniel Ma,²¹⁷ William Majoros,²¹⁸
 Joe McDaniel,²¹⁹ Sean Murphy,²²⁰ Matthew Newman,²²¹ Trung Nguyen,²²² Ngoc Nguyen,²²³ Marc Nodell,²²⁴
 Sue Pan,²²⁵ Jim Peck,²²⁶ Marshall Peterson,²²⁷ William Rowe,²²⁸ Robert Sanders,²²⁹ John Scott,²³⁰
 Michael Simpson,²³¹ Thomas Smith,²³² Arlan Sprague,²³³ Timothy Stockwell,²³⁴ Russell Turner,²³⁵ Eli Venter,²³⁶
 Mei Wang,²³⁷ Melyuan Wen,²³⁸ David Wu,²³⁹ Mitchell Wu,²⁴⁰ Ashley Xia,²⁴¹ Ali Zandieh,²⁴² Xiaohong Zhu²⁴³

Steven Salzberg,¹²

the evolving human gene count

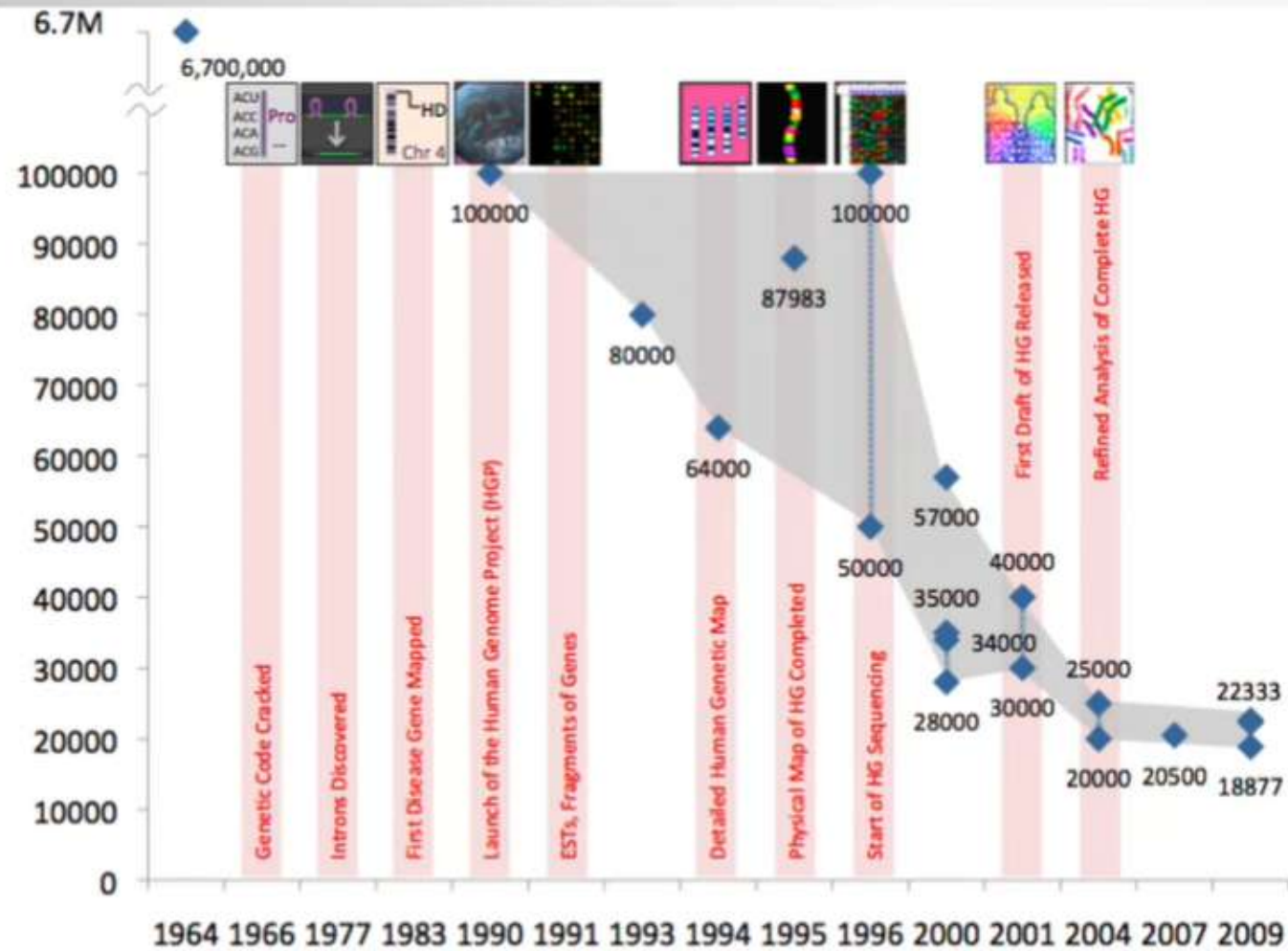


Figure from M. Pertea and S.L. Salzberg, *Genome Biology* 2010, 11:206



- sequence 3 billion basepairs - **yep!**
- for \$1/base - **\$1 per 700 bases!**
- by 2005 - **done in 2001!**
- Cost today: \$1 per 3,000,000 bases; 4000-fold cheaper!

