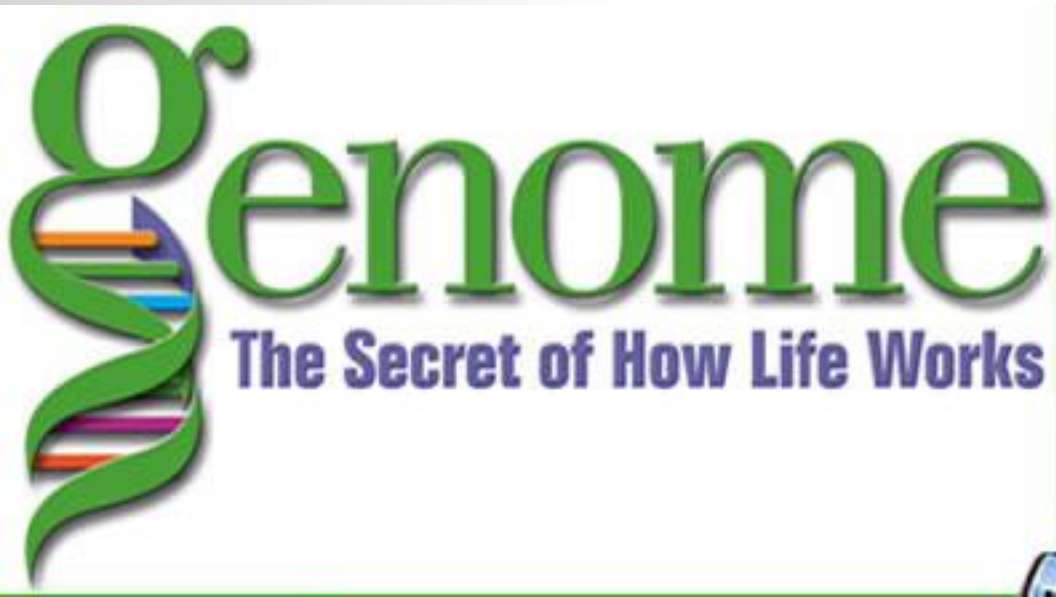


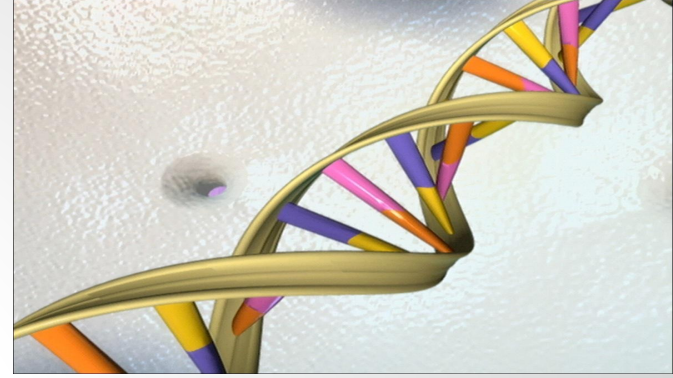
The Human Genome Project

Steven Salzberg

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



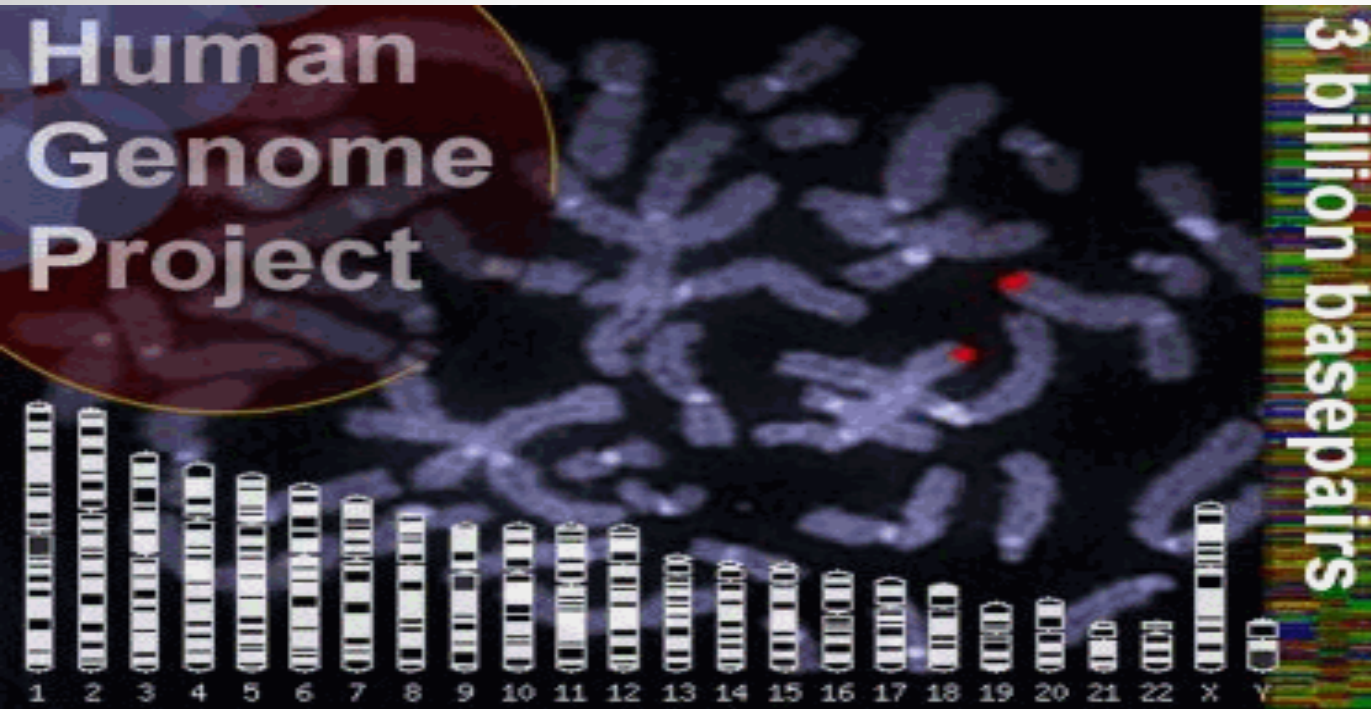
Human Genome Project: 1989 - present



- 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

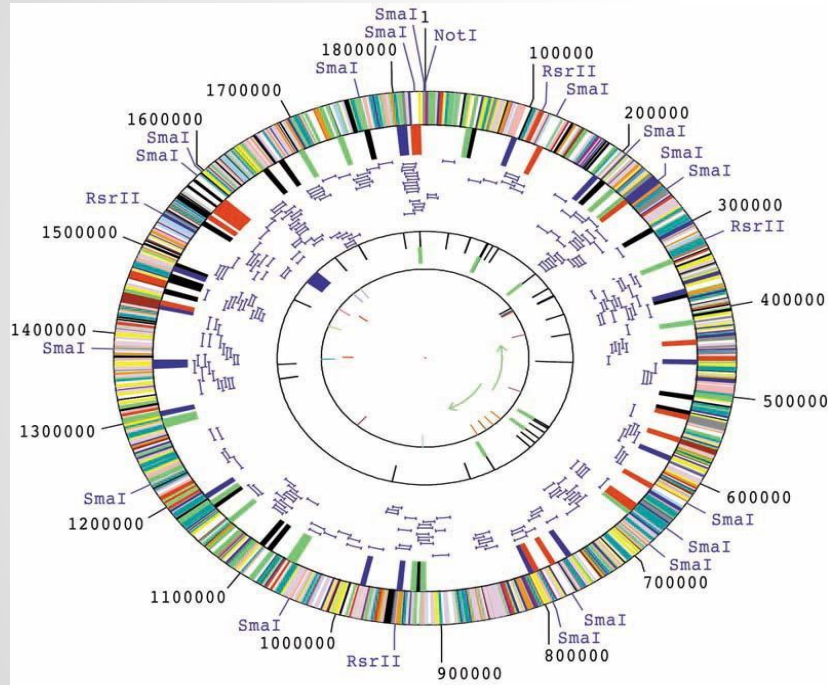
GOALS



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*

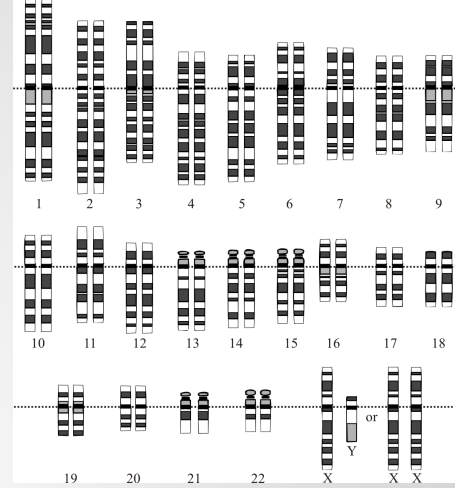


1.8 million bases
1742 genes

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 to sequence the genome: 1999-2000

- 1999: NIH merges its efforts into 3 larger centers
- 2000: Celera sequences and publishes the complete genome of the fruit fly, *Drosophila melanogaster*
 - Proves the whole-genome shotgun technique works on a 20X larger scale than previously

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

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

Sequence of the Human Genome

• 26,588 genes

er,^{1*} Mark D. Adams,¹ Eugene W. Myers,¹ Peter W. Li,¹ Richard J. Mural,¹
on,¹ Hamilton O. Smith,¹ Mark Yandell,¹ Cheryl A. Evans,¹ Robert A. Holt,¹
Gocayne,¹ Peter Amanatides,¹ Richard M. Ballow,¹ Daniel H. Huson,¹
ortman,¹ Qing Zhang,¹ Chinnappa D. Kodira,¹ Xiangqun H. Zheng,¹ Lin Chen,¹
upski,² Gangadharan Subramanian,¹ Paul D. Thomas,¹ Jinghui Zhang,¹
Miklos,² Catherine Nelson,² Samuel Broder,¹ Andrew G. Clark,⁴ Joe Nadeau,⁵
asic,⁶ Norton Zinder,⁷ Arnold J. Levine,⁷ Richard J. Roberts,⁸ Mel Simon,⁹
ichael Hunkapiller,¹¹ Randall Bolanos,¹ Arthur Delcher,¹ Ian Dew,¹ Daniel Fasulo,¹
ana Florea,¹ Aaron Halpern,¹ Sridhar Hannenhalli,¹ Saul Kravitz,¹ Samuel Levy,¹
Reinert,¹ Karin Remington,¹ Jane Abu-Threideh,¹ Ellen Beasley,¹ Kendra Biddis,¹
nda Brandon,¹ Michele Cargill,¹ Ishwar Chandramouliswaran,¹ Rosane Charlab,¹
li,¹ Zuoming Deng,¹ Valentina Di Francesco,¹ Patrick Dunn,¹ Karen Eilbeck,¹
ndrei E. Gabrielian,¹ Weiniu Gan,¹ Wangmao Ge,¹ Fangcheng Gong,¹ Zhiping Gu,¹
s J. Heiman,¹ Maureen E. Higgins,¹ Rui-Ru Ji,¹ Zhaoxi Ke,¹ Karen A. Ketchum,¹
Zhongwu Lai,¹ Yiding Lei,¹ Zhenya Li,¹ Jiayin Li,¹ Yong Liang,¹ Xiaoying Lin,¹ Fu Lu,¹
Gennady V. Merkulov,¹ Natalia Milshina,¹ Helen M. Moore,¹ Ashwinikumar K. Nair,¹
Vaibhav A. Narayan,¹ Beena Neelam,¹ Deborah Nusskern,¹ Douglas B. Rusch,¹ Steven Salzberg,^{1,2}
Wei Shao,¹ Bixiong Shue,¹ Jingtao Sun,¹ Zhen Yuan Wang,¹ Aihui Wang,¹ Xin Wang,¹ Jian Wang,¹
Ming-Hui Wei,¹ Ron Wides,¹ Chunlin Xiao,¹ Chunhua Yan,¹ Alison Yao,¹ Jane Ye,¹ Ming Zhan,¹
Weiqing Zhang,¹ Hongyu Zhang,¹ Liansheng Zheng,¹ Fei Zhong,¹ Wenyuan Zhong,¹
Shiaoping C. Zhu,¹ Shaying Zhu,¹² Dennis Gilbert,¹ Suzanna Baumhueter,¹ Gene Spier,¹
Christine Carter,¹ Anibal Cravchik,¹ Trevor Woodage,¹ Feroze Ali,¹ Huijin An,¹ Aderonke Awe,¹
Danita Baldwin,¹ Holly Baden,¹ Mary Barnstead,¹ Ian Barrow,¹ Karen Beeson,¹ Dana Busam,¹
Amy Carver,¹ Angela Center,¹ Ming Lai Cheng,¹ Liz Curry,¹ Steve Danaher,¹ Lionel Davenport,¹
Raymond Desilets,¹ Susanne Dietz,¹ Kristina Dodson,¹ Lisa Doup,¹ Steven Ferriera,¹ Neha Garg,¹
Andres Gluecksmann,¹ Brit Hart,¹ Jason Haynes,¹ Charles Haynes,¹ Cheryl Heiner,¹ Suzanne Hladun,¹
Damon Hostin,¹ Jarrett Houck,¹ Timothy Howland,¹ Chinyere Ibegwam,¹ Jeffery Johnson,¹
Francis Kalush,¹ Lesley Kline,¹ Shashi Koduru,¹ Amy Love,¹ Felecia Mann,¹ David May,¹
Steven McCawley,¹ Tina McIntosh,¹ Ivy McMullen,¹ Mee Moy,¹ Linda Moy,¹ Brian Murphy,¹
Keith Nelson,¹ Cynthia Pfannkoch,¹ Eric Pratt,¹ Vinita Puri,¹ Hina Qureshi,¹ Matthew Reardon,¹
Robert Rodriguez,¹ Yu-Hui Rogers,¹ Deanna Romblad,¹ Bob Ruhfel,¹ Richard Scott,¹ Cynthia Sitter,¹
Michelle Smallwood,¹ Erin Stewart,¹ Renee Strong,¹ Ellen Suh,¹ Reginald Thomas,¹ Ni Ni Tint,¹
Sukye Tse,¹ Claire Vech,¹ Gary Wang,¹ Jeremy Wetter,¹ Sherita Williams,¹ Monica Williams,¹
Sandra Windsor,¹ Emily Winn-Deen,¹ Keriellen Wolfe,¹ Jayshree Zaveri,¹ Karena Zaveri,¹
Josep F. Abril,¹⁴ Roderic Guigó,¹⁴ Michael J. Campbell,¹ Kimmen V. Sjolander,¹ Brian Karlak,¹
Anish Kejarwal,¹ Huaiyu Mi,¹ Betty Lazareva,¹ Thomas Hatton,¹ Apurva Narechania,¹ Karen Diemer,¹
Anushya Muruganujan,¹ Nan Guo,¹ Shinji Sato,¹ Vineet Bafna,¹ Sorin Istrail,¹ Ross Lippert,¹
Russell Schwartz,¹ Brian Walenz,¹ Shibu Yooseph,¹ David Allen,¹ Anand Basu,¹ James Baxendale,¹
Louis Blick,¹ Marcelo Caminha,¹ John Carnes-Stine,¹ Parris Caulk,¹ Yen-Hui Chiang,¹ My Coyne,¹
Carl Dahlke,¹ Anne Deslattes Mays,¹ Maria Dombroski,¹ Michael Donnelly,¹ Dale Ely,¹ Shiva Esparham,¹
Carl Foster,¹ Harold Gire,¹ Stephen Glanowski,¹ Kenneth Glasser,¹ Anna Glodek,¹ Mark Gorokhov,¹
Ken Graham,¹ Barry Gropman,¹ Michael Harris,¹ Jeremy Hell,¹ Scott Henerson,¹ Jeffrey Hoover,¹
Donald Jennings,¹ Catherine Jordan,¹ James Jordan,¹ John Kasha,¹ Leonid Kagan,¹ Cheryl Kraft,¹
Alexander Levitsky,¹ Mark Lewis,¹ Xiangjun Liu,¹ John Lopez,¹ Daniel Ma,¹ William Majoros,¹
Joe McDaniel,¹ Sean Murphy,¹ Matthew Newman,¹ Trung Nguyen,¹ Ngoc Nguyen,¹ Marc Nodell,¹
Sue Pan,¹ Jim Peck,¹ Marshall Peterson,¹ William Rowe,¹ Robert Sanders,¹ John Scott,¹
Michael Simpson,¹ Thomas Smith,¹ Arlan Sprague,¹ Timothy Stockwell,¹ Russell Turner,¹ Eli Venter,¹
Mei Wang,¹ Meiyuan Wen,¹ David Wu,¹ Mitchell Wu,¹ Ashley Xia,¹ Ali Zandieh,¹ Xiaohong Zhu¹

Steven Salzberg,¹²

the evolving human gene count

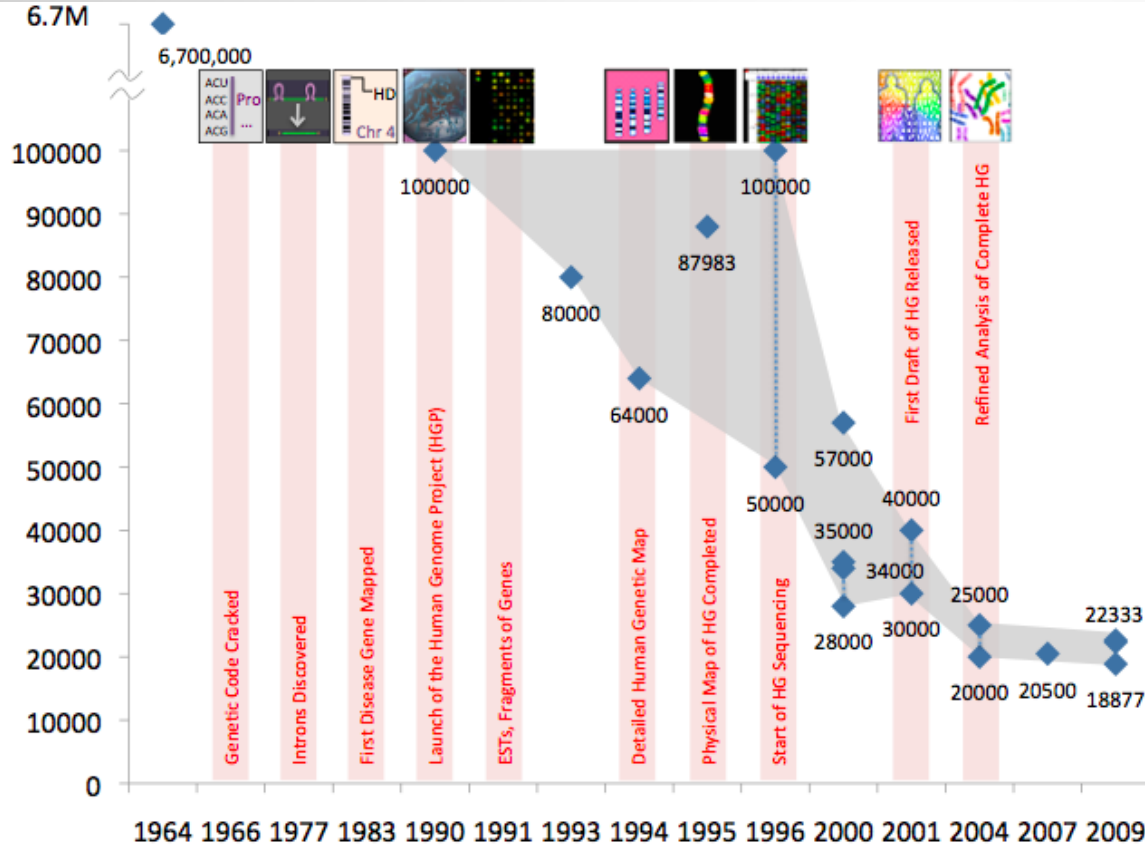


Figure from M. Pertea and S.L. Salzberg, *Genome Biology* 2010, 11:206

- 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!

