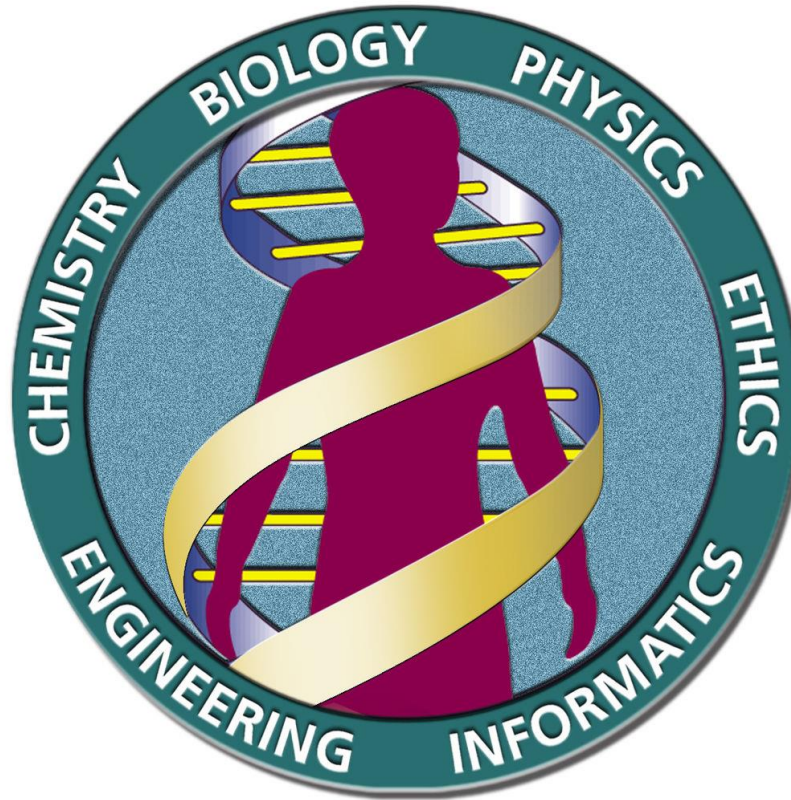




Build a human reference genome



Yunhao Wang Ph.D.

September 25, 2020

Department of Biomedical Informatics, Ohio State University

Initial sequencing and analysis of the human
Human Genome Project
Eric S. Lander



The Sequence of the Human Genome
Celera Genomics
J. Craig Venter



Release of human reference genome

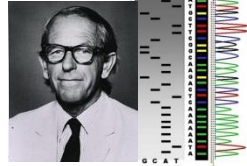
Release date	GRC version	NCBI version	UCSC version
Dec 2013	GRCh38		hg38
Feb 2009	GRCh37		hg19
Mar 2006		NCBI 36	hg18
May 2004		NCBI 35	hg17
Jul 2003		NCBI 34	
Apr 2003		NCBI 33	

Human gene annotation library

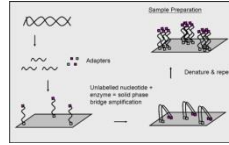
	RefSeq NCBI UCSC	Ensembl / GENCODE
ID	Multiple loci	Unique
Criteria	Stringent	Basic: stringent Comprehensive: loose

History of DNA sequencing technologies

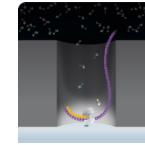
Sanger



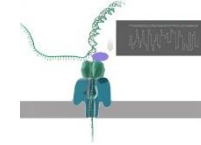
Illumina



PacBio



ONT



1977

First

2006

Second

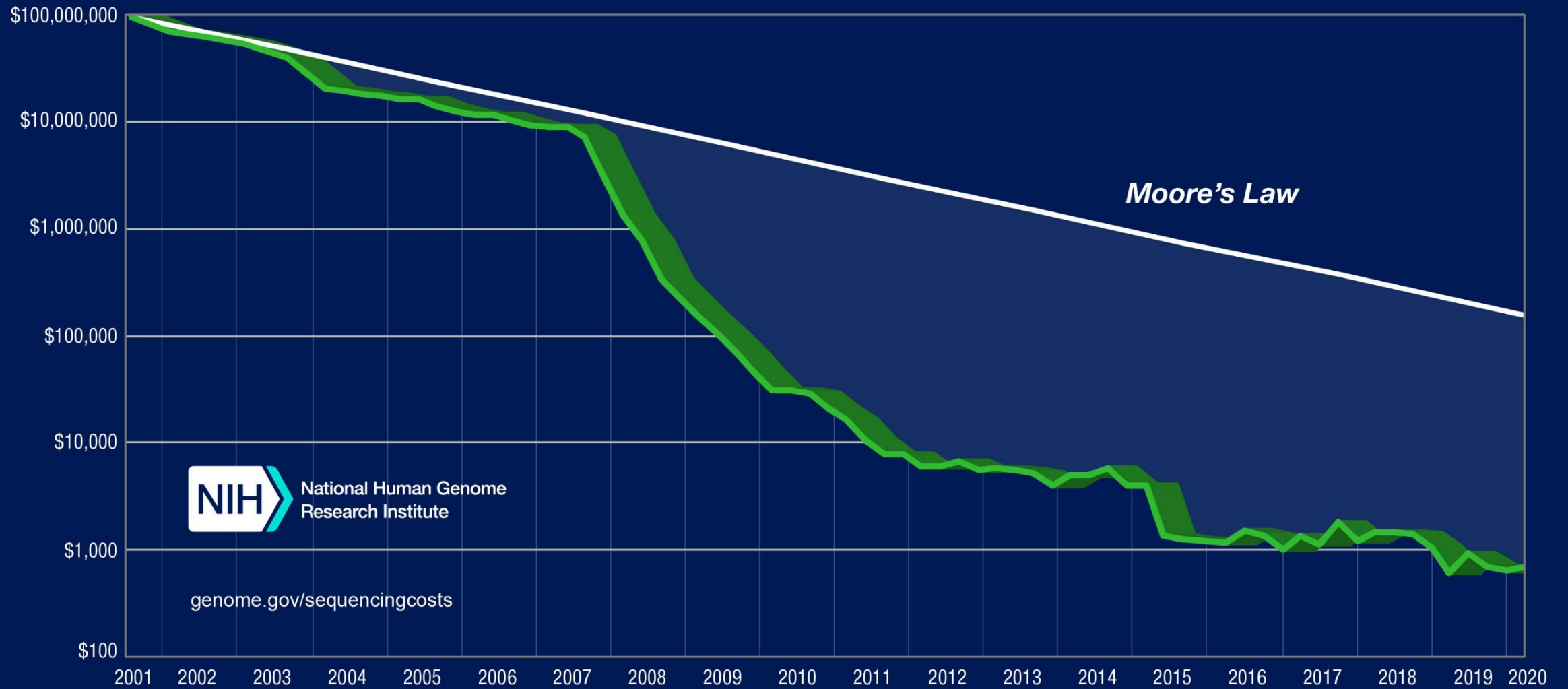
2009

2014

Third

➤ Length	<1 kb	2 x 150 bp	>10 kb (up to 2.3 Mb)
➤ Accuracy	>99%	~99%	>90%
➤ Throughput	low	high	high
➤ Cost	high	high ->low	?

Cost per Human Genome



Open Access Published: 05 October 2016

De novo assembly and phasing of a Korean human genome

Jeong-Sun Seo, Arang Rhie, Junsoo Kim, Sangjin Lee, Min-Hwan Sohn, Chang-Uk Kim, Alex Hastie, Han Cao, Ji-Young Yun, Jihye Kim, Junho Kuk, Gun Hwa Park, Juhyeok Kim, Hanna Ryu, Jongbum Kim, Mira Roh, Jeonghun Baek, Michael W. Hunkapiller, Jonas Korlach, Jong-Yeon Shin & Changhoon Kim

Open Access Published: 19 March 2018

Linear assembly of a human centromere on the Y chromosome

Miten Jain, Hugh E Olsen, Daniel J Turner, David Stoddart, Kira V Bulazel, Benedict Paten, David Haussler, Huntington F Willard, Mark Akeson & Karen H Miga

Open Access Published: 29 January 2018

Nanopore sequencing and assembly of a human genome with ultra-long reads

Miten Jain, Sergey Koren, Karen H Miga, Josh Quick, Arthur C Rand, Thomas A Sasani, John R Tyson, Andrew D Beggs, Alexander T Dilthey, Ian T Fiddes, Sunir Malla, Hannah Marriott, Tom Nieto, Justin O'Grady, Hugh E Olsen, Brent S Pedersen, Arang Rhie, Hollian Richardson, Aaron R Quinlan, Terrance P Snutch, Louise Tee, Benedict Paten, Adam M Phillippy, Jared T Simpson, Nicholas J Loman & Matthew Loose -Show fewer authors

Article Open Access Published: 14 July 2020

Telomere-to-telomere assembly of a complete human X chromosome

Karen H. Miga, Sergey Koren, Arang Rhie, Mitchell R. Vollger, Ariel Gershman, Andrey Bzikadze, Shelise Brooks, Edmund Howe, David Porubsky, Glennis A. Logsdon, Valerie A. Schneider, Tamara Potapova, Jonathan Wood, William Chow, Joel Armstrong, Jeanne Fredrickson, Evgenia Pak, Kristof Tigyi, Milinn Kremitzki, Christopher Markovic, Valerie Maduro, Amalia Dutra, Gerard G. Bouffard, Alexander M. Chang, Nancy F. Hansen, Amy B. Wilfert, Françoise Thibaud-Nissen, Anthony D. Schmitt, Jon-Matthew Belton, Siddarth Selvaraj, Megan Y. Dennis, Daniela C. Soto, Ruta Sahasrabudhe, Gulhan Kaya, Josh Quick, Nicholas J. Loman, Nadine Holmes, Matthew Loose, Urvashi Surti, Rosa ana Risques, Tina A. Graves Lindsay, Robert Fulton, Ira Hall, Benedict Paten, Kerstin Howe, Winston Timp, Alice Young, James C. Mullikin, Pavel A. Pevzner, Jennifer L. Gerton, Beth A. Sullivan, Evan E. Eichler & Adam M. Phillippy -

Article Open Access Published: 04 May 2020

Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes

Kishwar Shafin, Trevor Pesout, Ryan Lorig-Roach, Marina Haukness, Hugh E. Olsen, Colleen Bosworth, Joel Armstrong, Kristof Tigyi, Nicholas Maurer, Sergey Koren, Fritz J. Sedlazeck, Tobias Marschall, Simon Mayes, Vania Costa, Justin M. Zook, Kelvin J. Liu, Duncan Kilburn, Melanie Sorensen, Katy M. Munson, Mitchell R. Vollger, Jean Monlong, Erik Garrison, Evan E. Eichler, Sofie Salama, David Haussler, Richard E. Green, Mark Akeson, Adam Phillippy, Karen H. Miga, Paolo Carnevali, Miten Jain & Benedict Paten -Show

bioRxiv THE PREPRINT SERVER FOR BIOLOGY

HOME | ALERTS

Search

bioRxiv is receiving many new papers on coronavirus SARS-CoV-2. A reminder: these are preliminary reports that have not undergone peer review. They should not be regarded as conclusive, used to guide clinical practice/health-related behavior, or be reported in news media.

New Results Comments (1)

The structure, function, and evolution of a complete human chromosome 8

Glennis A. Logsdon, Mitchell R. Vollger, PingHsun Hsieh, Yafei Mao, Mikhail A. Liskovych, Sergey Koren, Sergey Nurk, Ludovica Mercuri, Philip C. Dishuck, Arang Rhie, Leonardo G. de Lima, David Porubsky, Andrey V. Bzikadze, Milinn Kremitzki, Tina A. Graves-Lindsay, Chirag Jain, Kendra Hoekzema, Shwetha C. Murali, Katherine M. Munson, Carl Baker, Melanie Sorensen, Alexandra M. Lewis, Urvashi Surti, Jennifer L. Gerton, Vladimir Larionov, Mario Ventura, Karen H. Miga, Adam M. Phillippy, Evan E. Eichler

Linear assembly of a human centromere on the Y chromosome

Miten Jain; Karen H Miga @ UCSC; 2018

• Topics

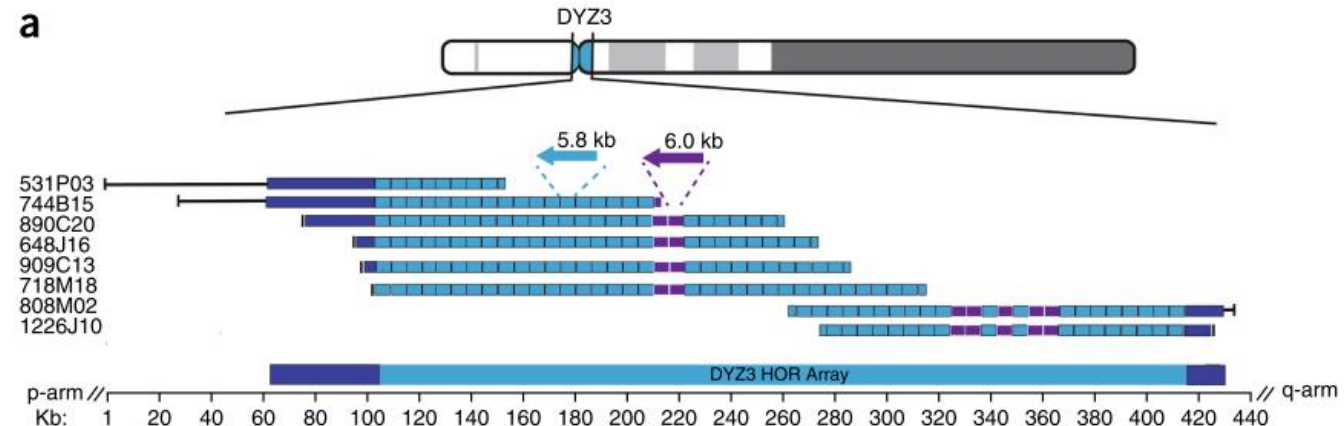
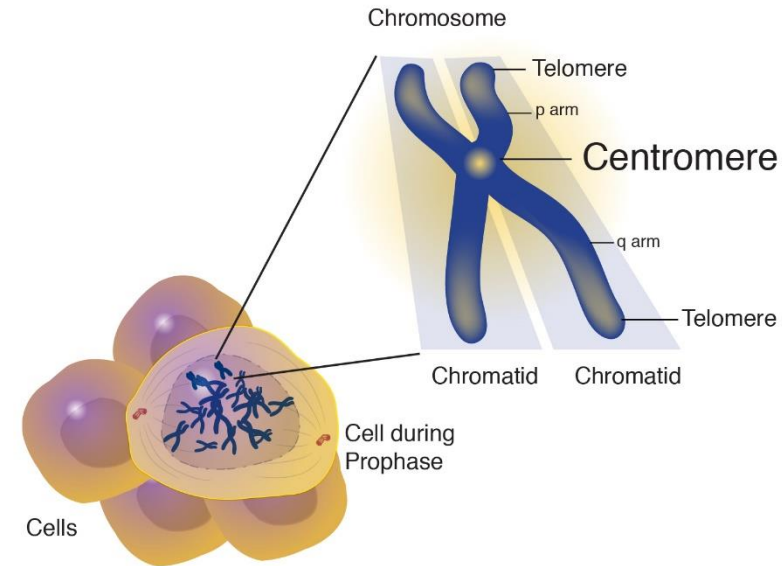
- Genome assembly

• Methods

- ONT + Illumina

• Results

- Complete assembly of tandemly repeated satellite DNAs (~350 kb)
- The first view for the array repeat organization and structure in a human centromere



Telomere-to-telomere assembly of a complete human X chromosome

Karen H Miga @ UCSC; Adam M. Phillippy @ NHGRI; 2020

• Topics

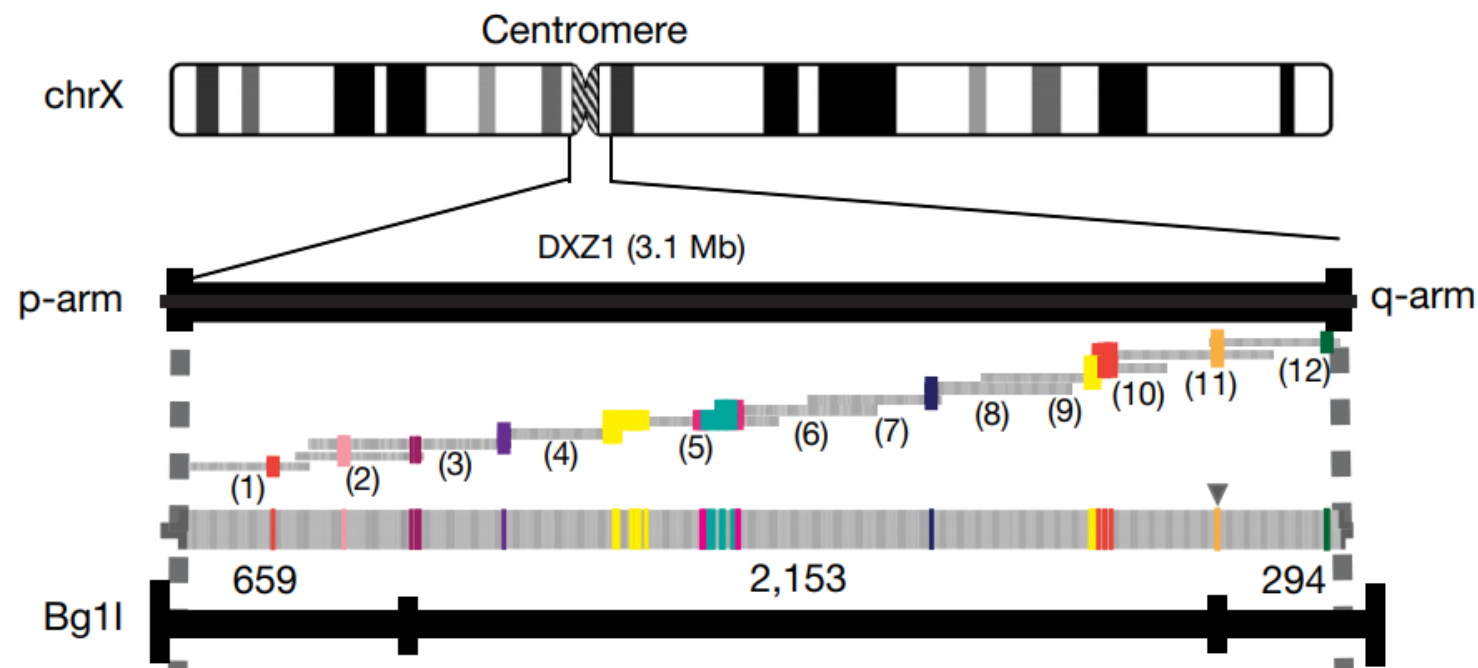
- Genome assembly

• Methods

- ONT + PacBio + 10X - Illumina

• Results

- The first finished human X chromosome
- 3.1 Mb centromeric satellite DNA array
- Close 29 gaps (1.1 Mb in total)



The structure, function, and evolution of a complete human chromosome 8

Evan E. Eichler @ University of Washington; 2020

• Topics

- Genome assembly

• Methods

- ONT + PacBio + Bionano

• Results

- The first linear assembly of a human autosome
- Close 5 long-standing gaps
 - 2.08 Mb centromeric satellite DNA array
 - 644 kb defensin copy number polymorphism
 - 863 kb variable number tandem repeat

Chromosome 8 gaps

