

Guest lecturer: Dr. Richard Allen White III
Post-doc: Friesen lab

Learning Objectives

1. Brief introduction of the rhizosphere and carbon cycling
2. Learn about ‘Omics,’ technology terms
3. Introduction to multiomics
4. History of genome sequencing
5. Learning about sequencing technologies, how do you get a bacterial genome sequence?
6. Review command line

CRISPR-Cas accelerates phage evolution

Trudy Morrison

Despite the growing use of the CRISPR-Cas system for targeted mutagenesis, little attention has been paid to understanding the basic biology of this defense system in bacteria-phage relationships. Rao *et al.* suggest that the system is a driver of phage, and perhaps bacterial, evolution. The authors characterized phage escape mutants restricted by high- and low-restriction spacers. Although the frequency of phage mutants that escaped high-restriction sequences was predictably low, the mutation frequency of phages that escaped low-restriction sequences was extremely high, about six orders of magnitude greater than spontaneous mutation frequencies.

Link to article: <http://science.sciencemag.org/content/359/6377/756.full>

Link to research article: <http://advances.sciencemag.org/content/4/2/eaar4134>

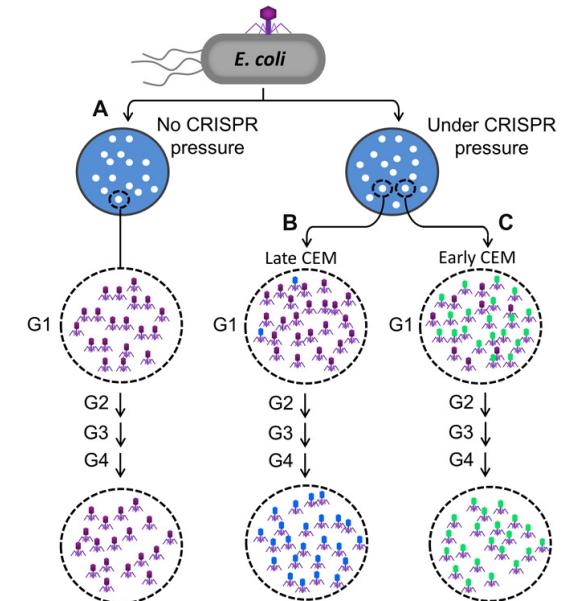


Figure 2 (from article): A model for CRISPR-Cas9–drive evolution of phage T4 genome.

Original article/source of figure: Tao P, Wu X, Rao V. 2018. Unexpected evolutionary benefit to phages imparted by bacterial CRISPR-Cas9. *Sci Adv.* 4(2):eaar4134. doi: 10.1126/sciadv.aar4134.

Announcements

- Phytobacteriology in the News [Week 3](#)
- Good work on the News & Views draft titles! First draft due **Thurs Sept 13th** (midnight), please submit to folder [here](#)
- Paper for next week: Pan-genomes! Form [here](#)

De Maayer et al. BMC Genomics 2014, 15:404
<http://www.biomedcentral.com/1471-2164/15/404>



RESEARCH ARTICLE

Open Access

Analysis of the *Pantoea ananatis* pan-genome reveals factors underlying its ability to colonize and interact with plant, insect and vertebrate hosts

Pieter De Maayer^{1,2*}, Wai Yin Chan², Enrico Rubagotti³, Stephanus N Venter², Ian K Toth^{2,4}, Paul R J Birch^{2,4,5} and Teresa A Coutinho²

Introduction - Dr. White III

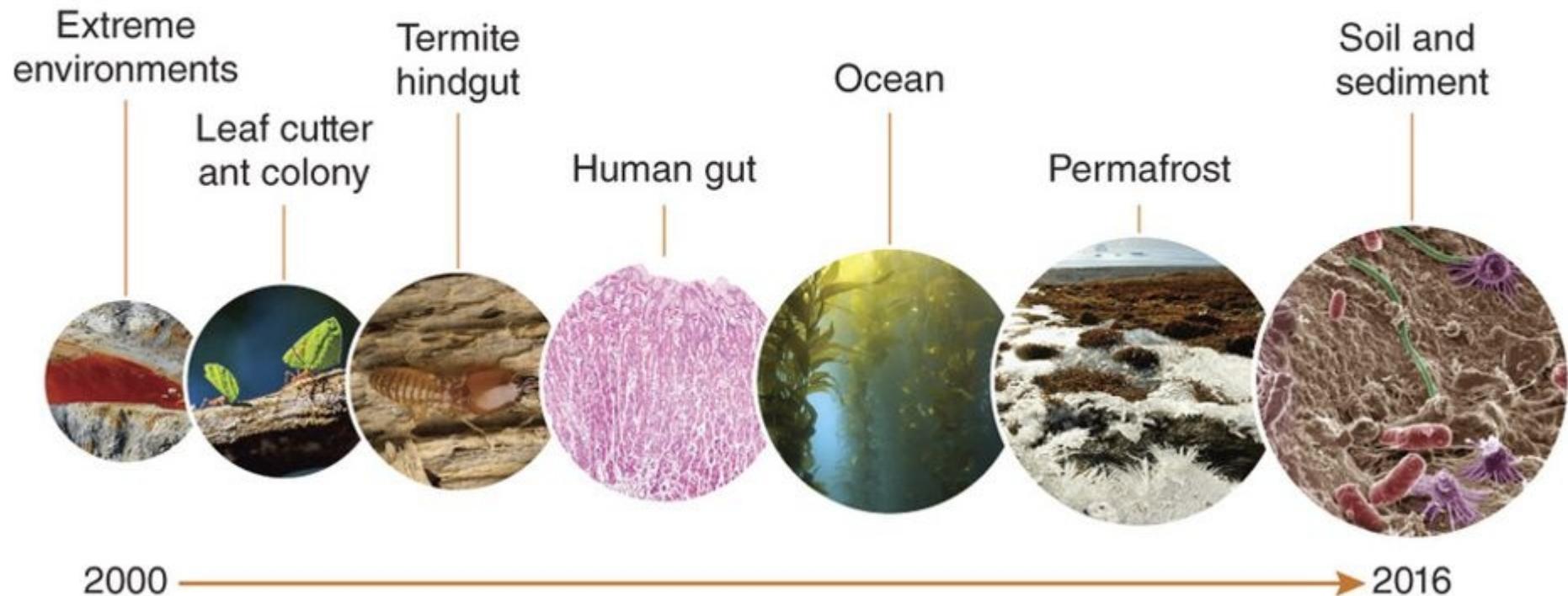
5



Research - Dr. White III

6

Microbiome complexity and multi-omics analysis timeline

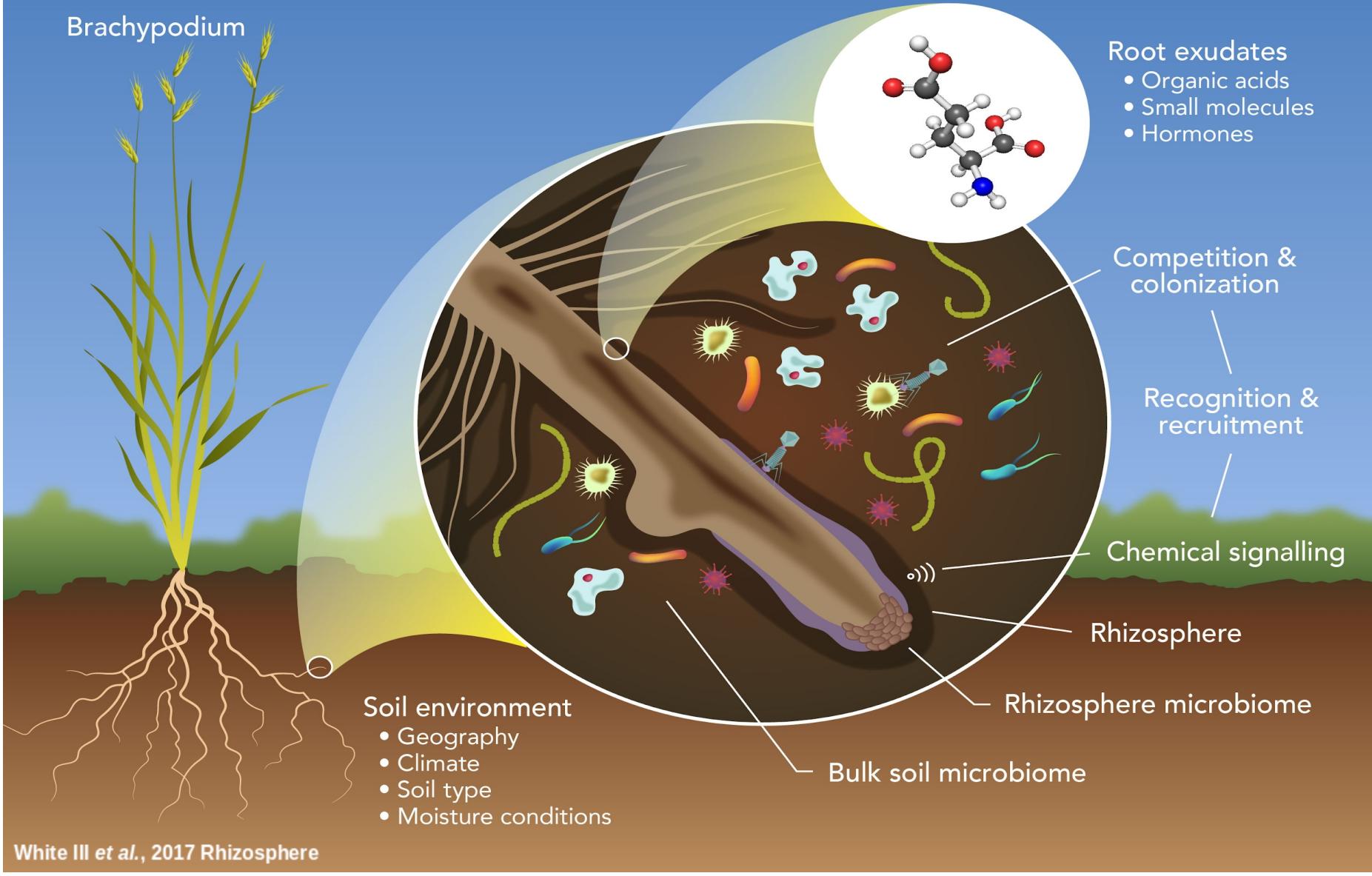


White III et al., 2016. Nature Protocols

Link to article: <https://www.nature.com/articles/nprot.2016.148>

Introduction of Rhizosphere

7

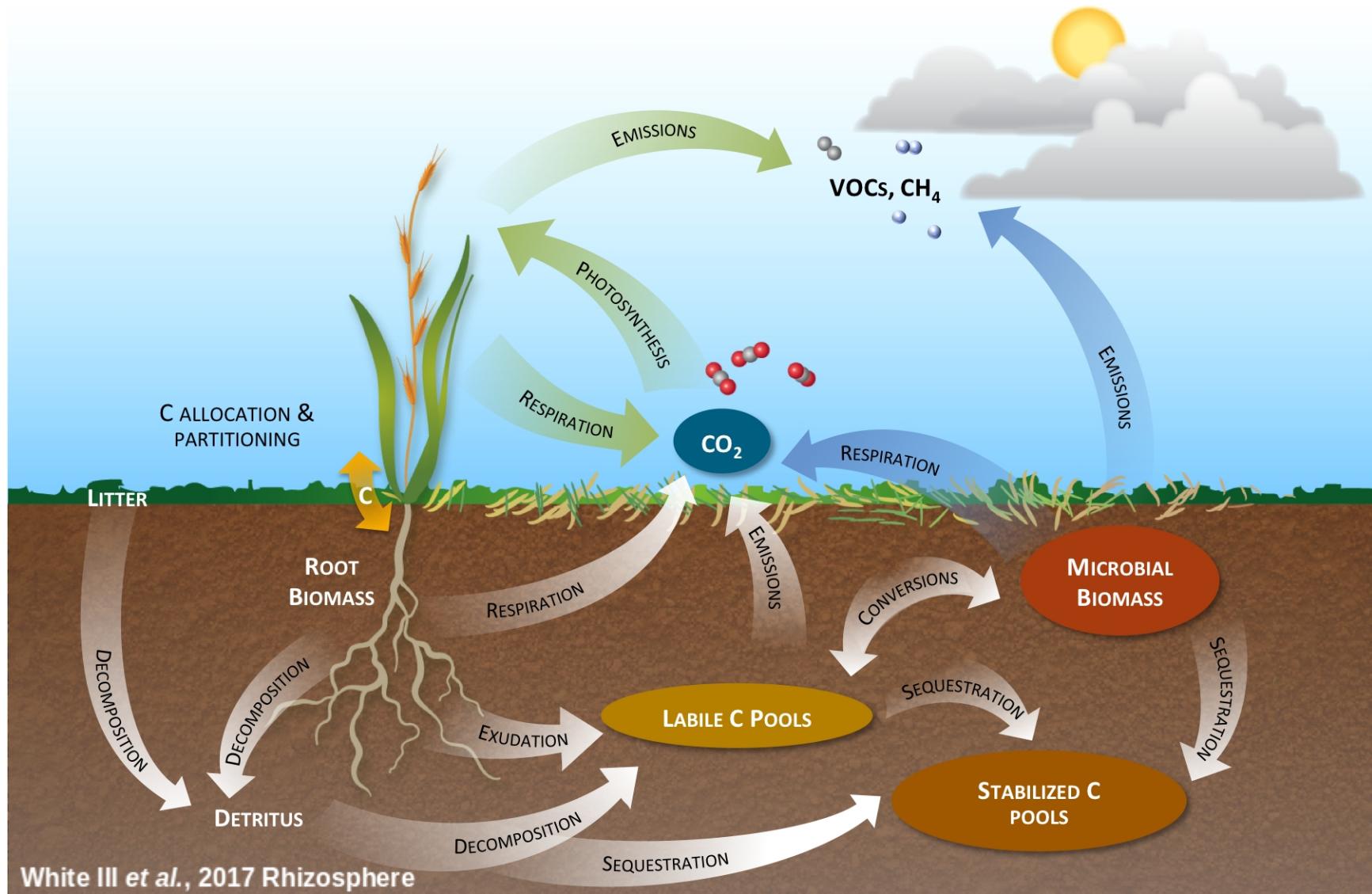


Link to article: <https://www.sciencedirect.com/science/article/pii/S2452219817300666>

Dr. Friesen - PLP514 - 2018-Sep-06 - L06

Rhizosphere impacting carbon cycling

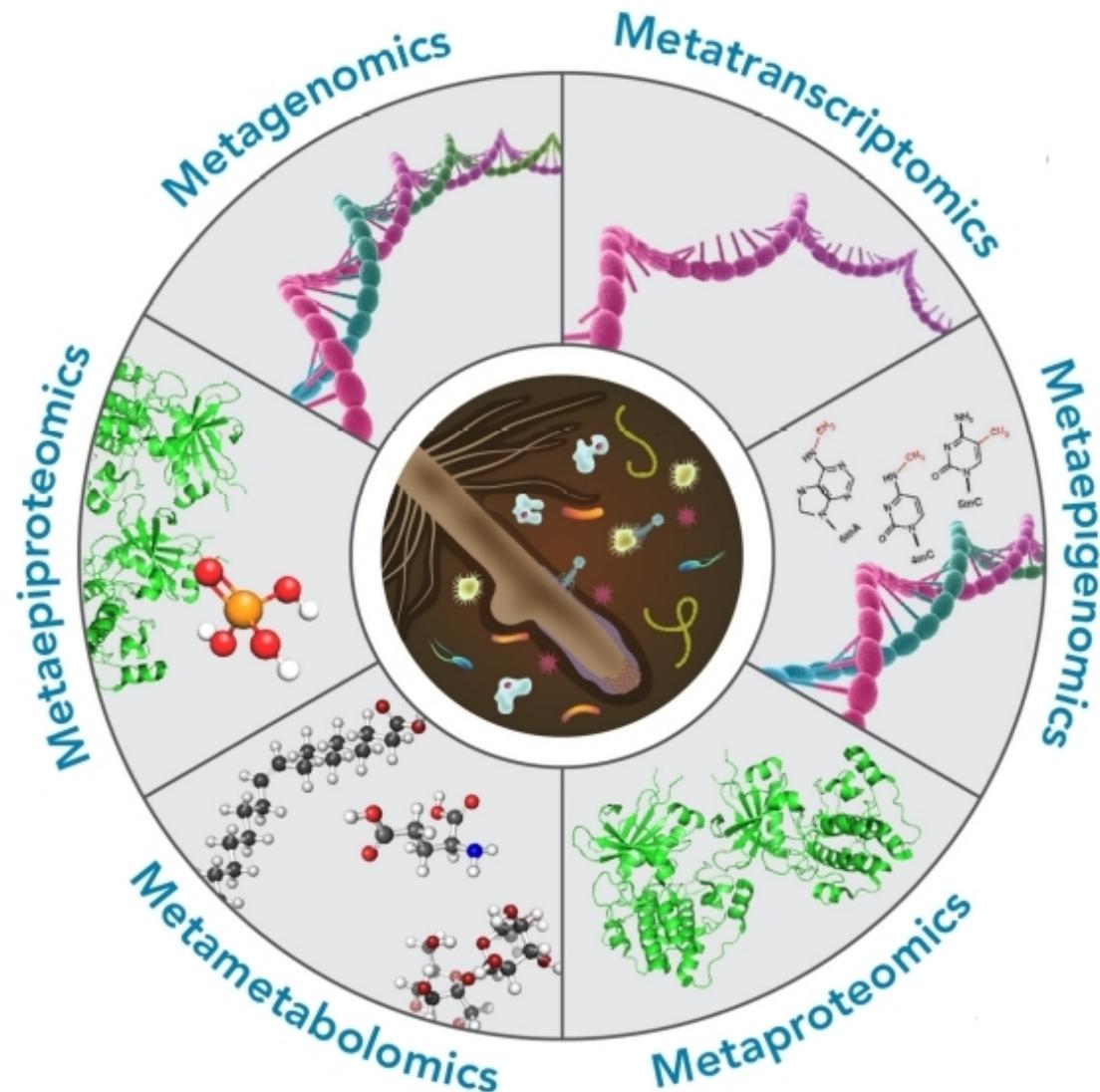
8



Link to article: <https://www.sciencedirect.com/science/article/pii/S24522198173004596>

Dr. Friesen - PLP514 - 2018-Sep-06 - L06

Omic terms



A typical multiomic study

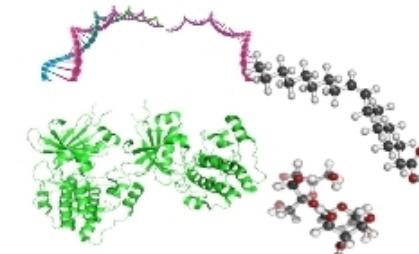
10



Environmental sample



Single step -
multiomic extraction



Library prep (NGS)
Fractionation (MS)



Sequencing



Mass Spectrometry

Data



Analysis

Direct Read or
Assembly (NGS)

Spectral features
analysis

White III et al., 2016. Nature Protocols

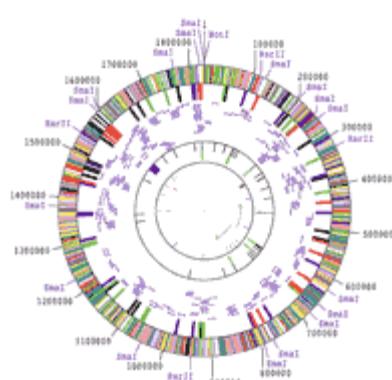
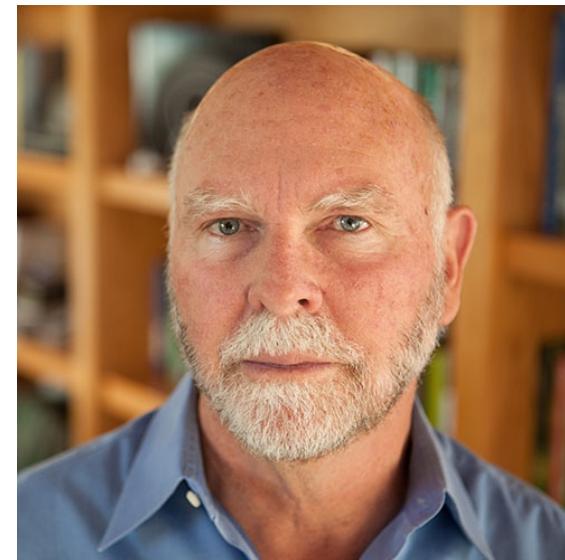
Link to article: <https://www.nature.com/articles/nprot.2016.148>

History of genome sequencing

- 1972 - First gene to be sequenced was MS2 coat protein (510 bp)
- 1976 - First RNA genome sequenced was Bacteriophage MS2 (3.56 kb)
- 1977 - First DNA genome sequenced was Bacteriophage phiX174 (5.38 kb)
- 1979 - First plasmid to be sequenced was pBR322 (4.3 kb)
- 1984 - First mammalian DNA virus sequenced was Epstein-Barr virus (172 kb)
- 1992 - First chromosome sequenced was Yeast chromosome III (315 kb)
- 1995 - First cellular genome sequenced was *Hemophilus influenzae* (1.8 Mb)
- 1995 - First archaeal genome sequenced was *Methanococcus jannaschii* (1.7 Mb)
- 1996 - First eukaryotic genome sequenced was *Saccharomyces cerevisiae* (12 MB)
- 1998 - First multicellular organism sequenced was *Ceanorhabditis elegans* (97 Mb)
- 2000 - First plant genome sequenced was *Arabidopsis thaliana* (125 Mb)
- 2001 - First mammalian draft genome *Homo sapiens* (3.3 Gb)

Who sequenced the first cellular genome?

J. Craig Venter



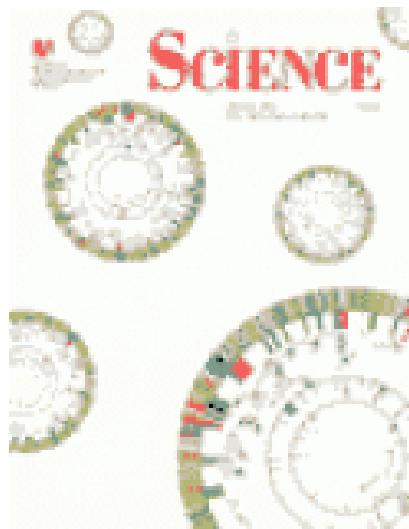
RESEARCH ARTICLES

Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd

RD Fleischmann, MD Adams, O White, RA Clayton, EF Kirkness, AR Kerlavage, CJ Bult, JF Tomb, BA Dougherty, JM Merrick, et al.

+ See all authors and affiliations

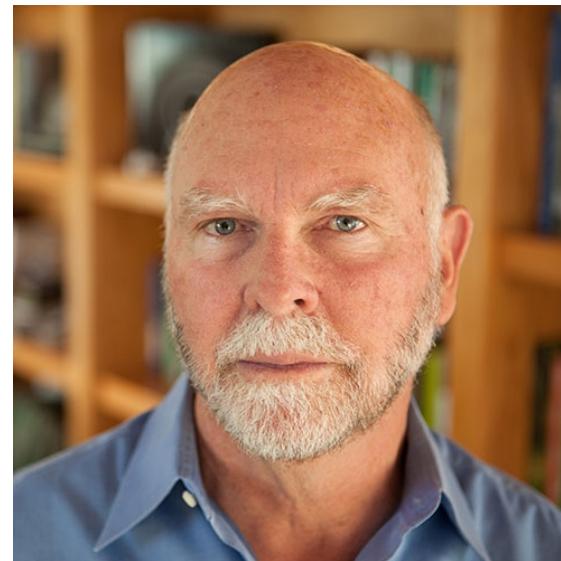
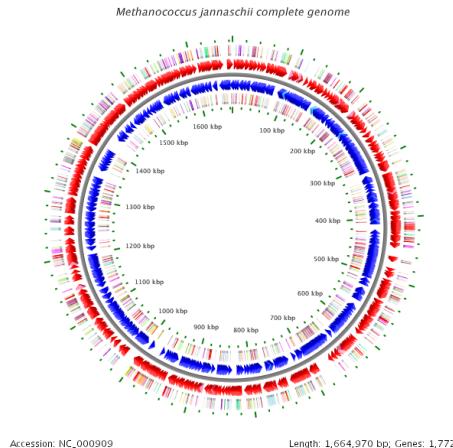
Science 28 Jul 1995:
Vol. 269, Issue 5223, pp. 496-512
DOI: 10.1126/science.7542800



http://www.genomenewsnetwork.org/resources/timeline/1995_Haemophilus.php

Who did the first archaeal genome sequence?

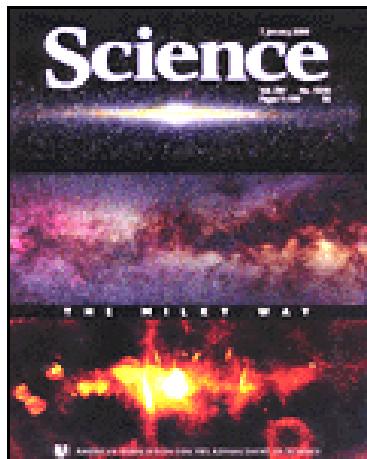
J. Craig Venter



■ ARTICLE

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*

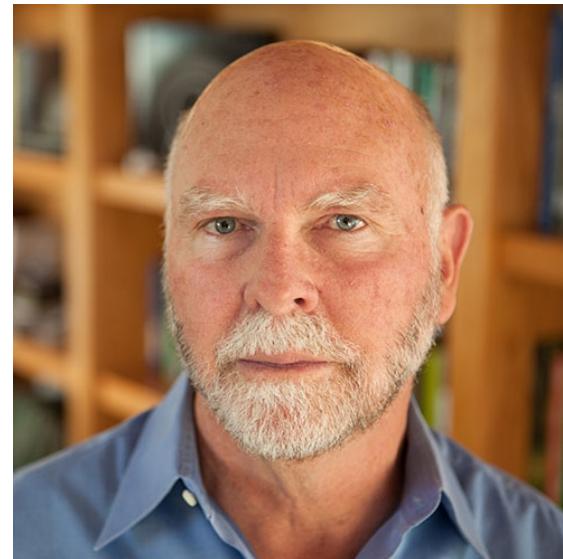


<https://www.yourgenome.org/facts/timeline-organisms-that-have-had-their-genomes-sequenced>

Who did the first human genome sequence?

Dr. Friesen - PLP514 - 2018-Sep-06 - L06

J. Craig Venter



THE HUMAN GENOME

The Sequence of the Human Genome

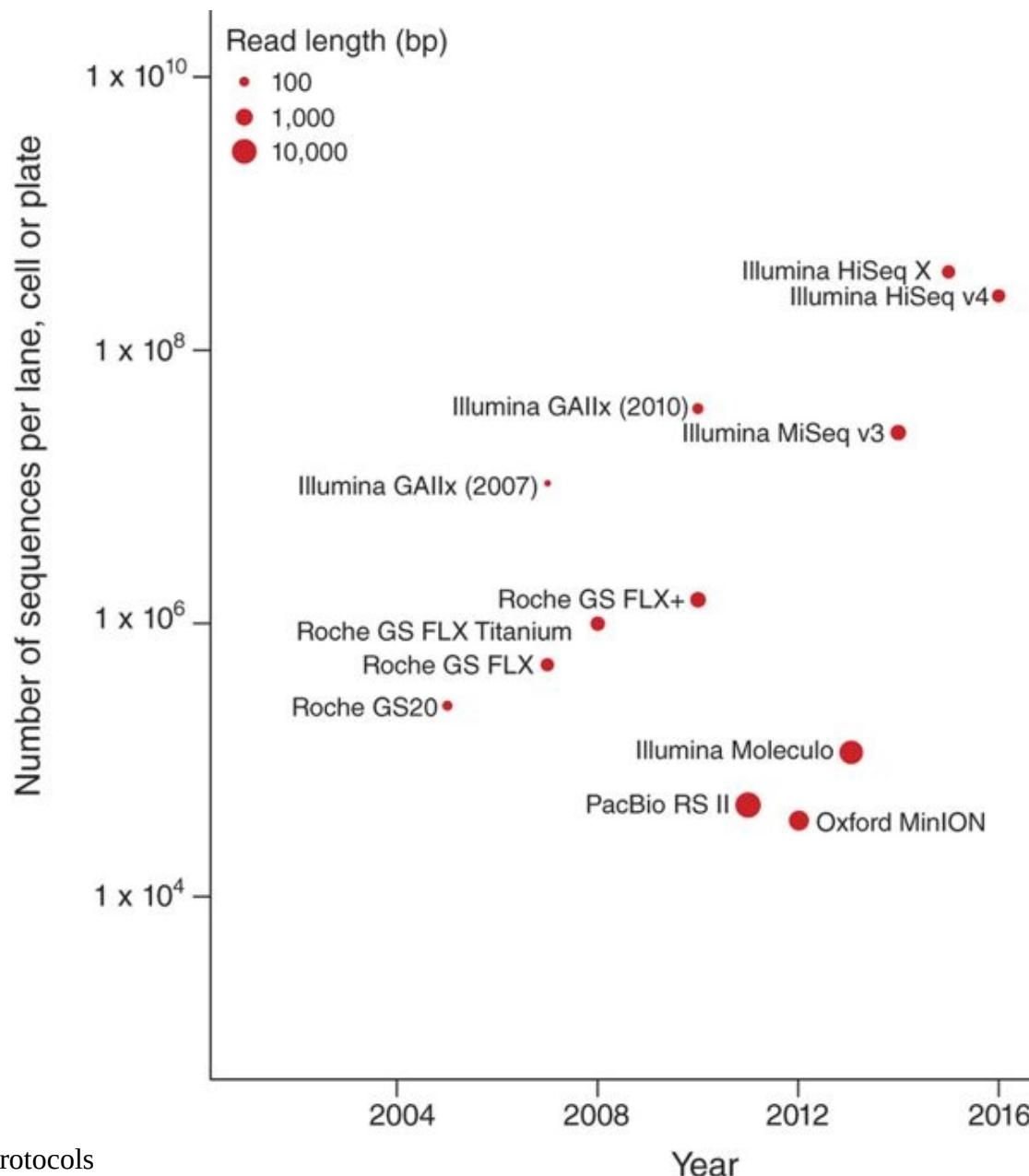
J. Craig Venter,^{1*} Mark D. Adams,¹ Eugene W. Myers,¹ Peter W. Li,¹ Richard J. Mural,¹ Granger G. Sutton,¹ Hamilton O. Smith,¹ Mark Yandell,¹ Cheryl A. Evans,¹ Robert A. Holt,¹ Jeannine D. Gocayne,¹ Peter Amanatides,¹ Richard M. Ballew,¹ Daniel H. Huson,¹ Jennifer Russo Wortman,¹ Qing Zhang,¹ Chinnappa D. Kodira,¹ Xiangqun H. Zheng,¹ Lin Chen,¹ Marian Skupski,¹ Gangadharan Subramanian,¹ Paul D. Thomas,¹ Jinghui Zhang,¹ George L. Gabor Miklos,² Catherine Nelson,³ Samuel Broder,¹ Andrew G. Clark,⁴ Jon Nadeau,⁵ Victor A. McKusick,⁶ Norton Zinder,⁷ Arnold J. Levine,⁷ Richard J. Roberts,⁸ Mel Simon,⁹

1304

16 FEBRUARY 2001 VOL 291 SCIENCE www.sciencemag.org

<https://www.yourgenome.org/facts/timeline-organisms-that-have-had-their-genomes-sequenced>

Sequencing technologies



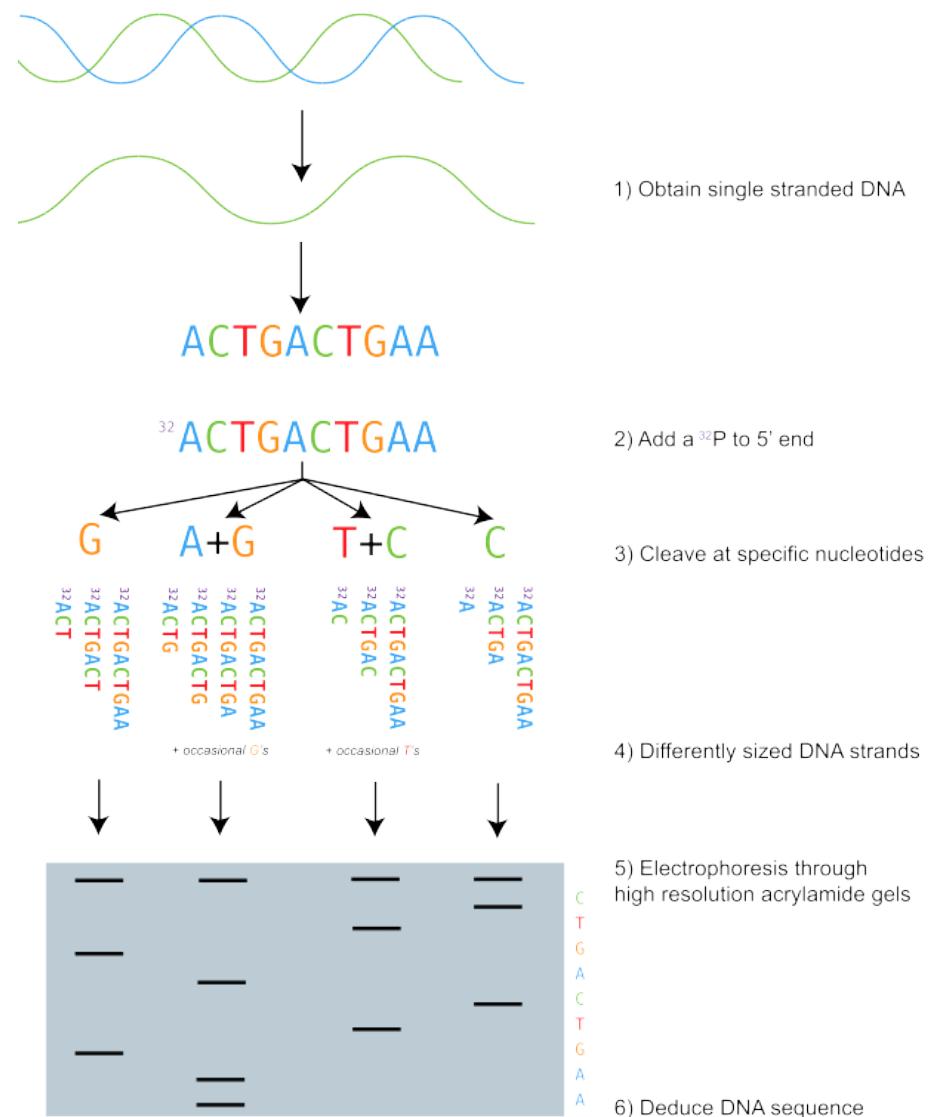
Sequencing platform - maxam gilbert

16

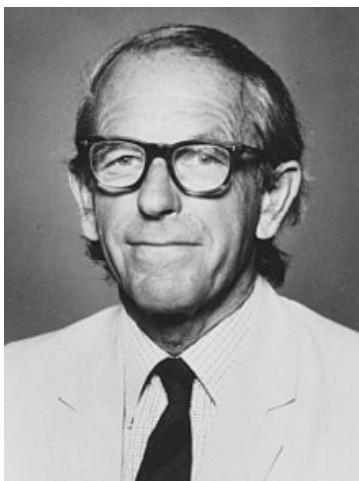


Walter Gilbert
1932 - Current

1976-1977

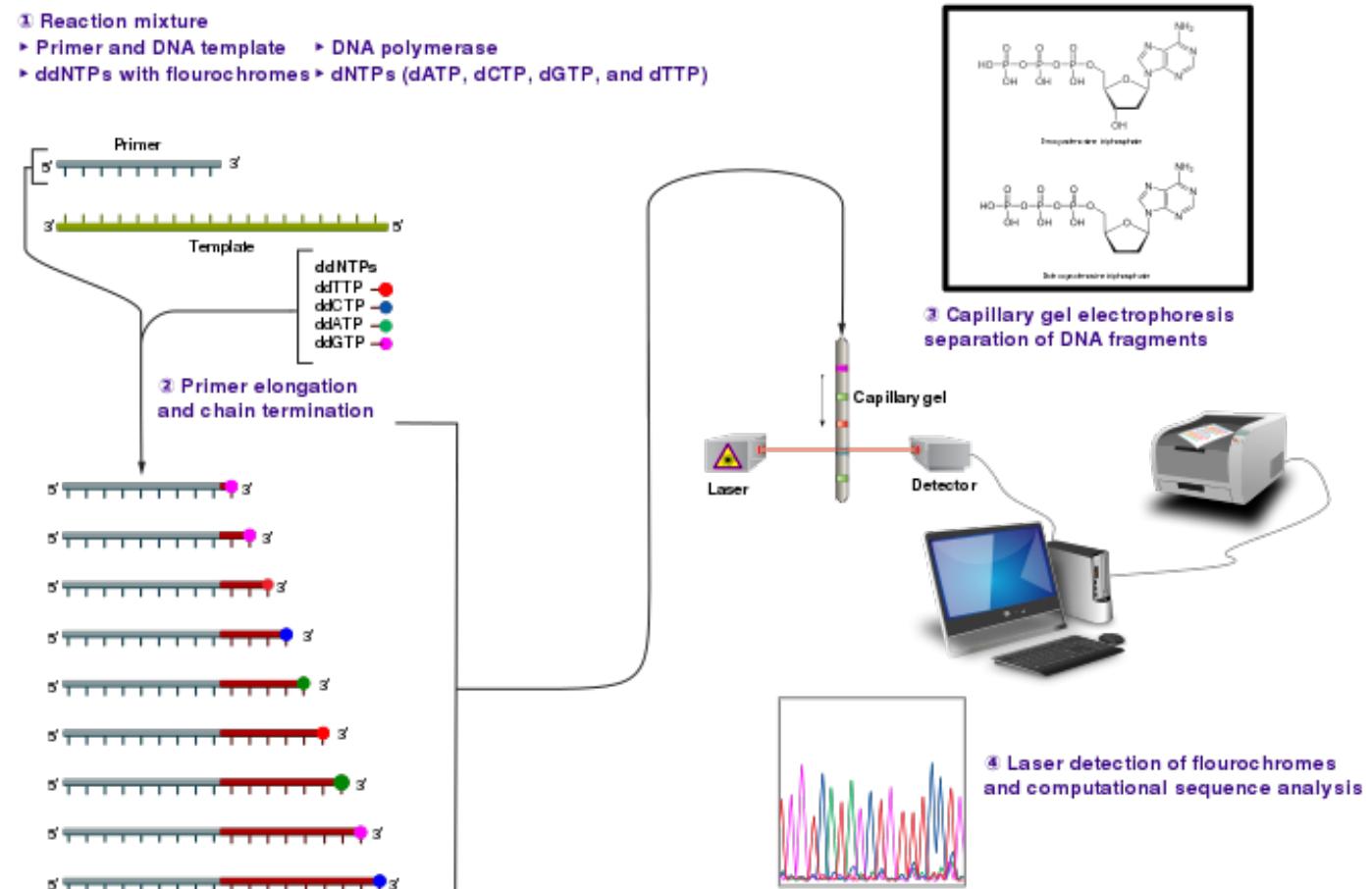
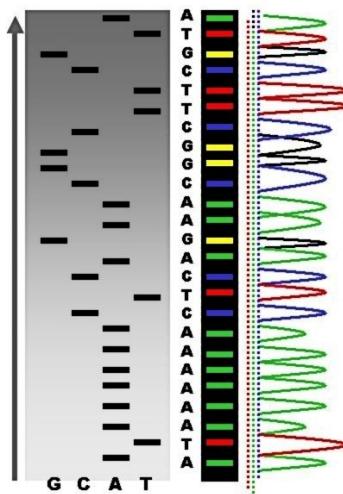


Sequencing platform - Sanger method dideoxy-chain termination

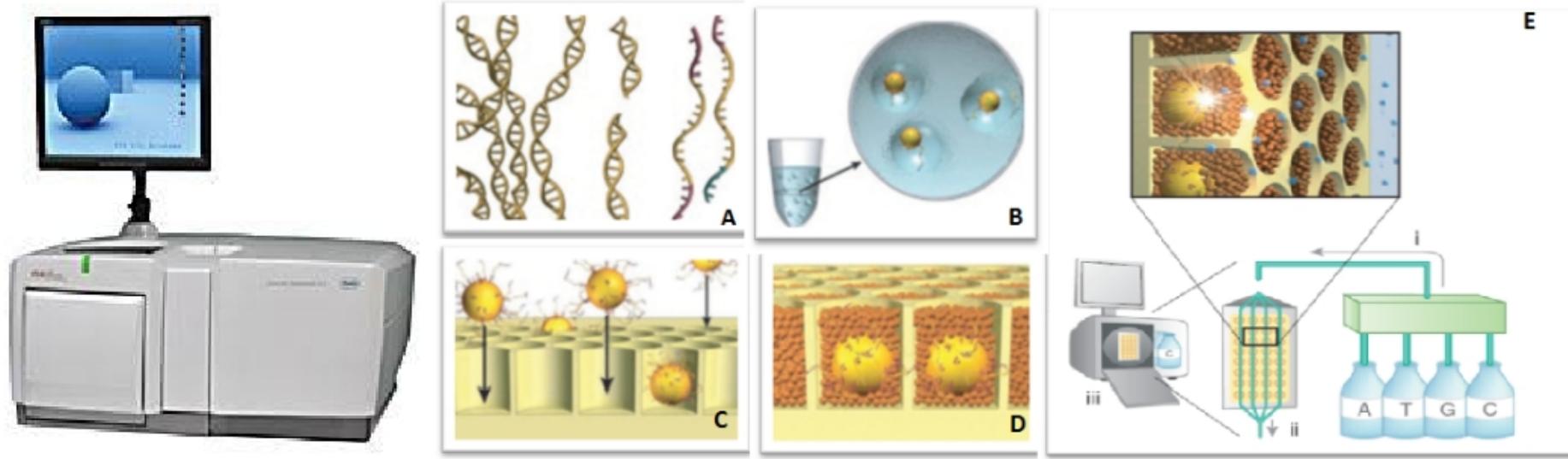


Fred Sanger
1932 - Current

1976-1977



Sequencing platform - pyrosequencing Roche 454



PRO

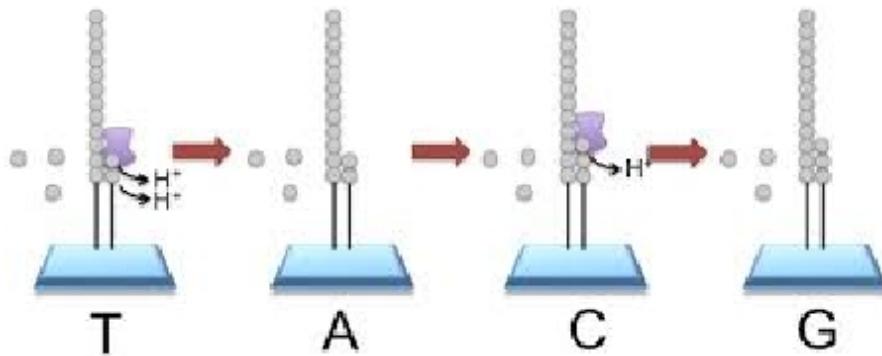
- First NGS sequencer
- Shorter than sanger
- Up to 1 M reads

CON

- Few reads ~ 1 M
- High Insertion/Deletion (INDEL) errors (homopolymers)
- NOW EXTINCT 2016

Sequencing platform - pH detection based

Life Tech Ion Torrent/Ion Proton



Ion PGM™ Sequencer



Ion Proton™ Sequencer

**PRO**

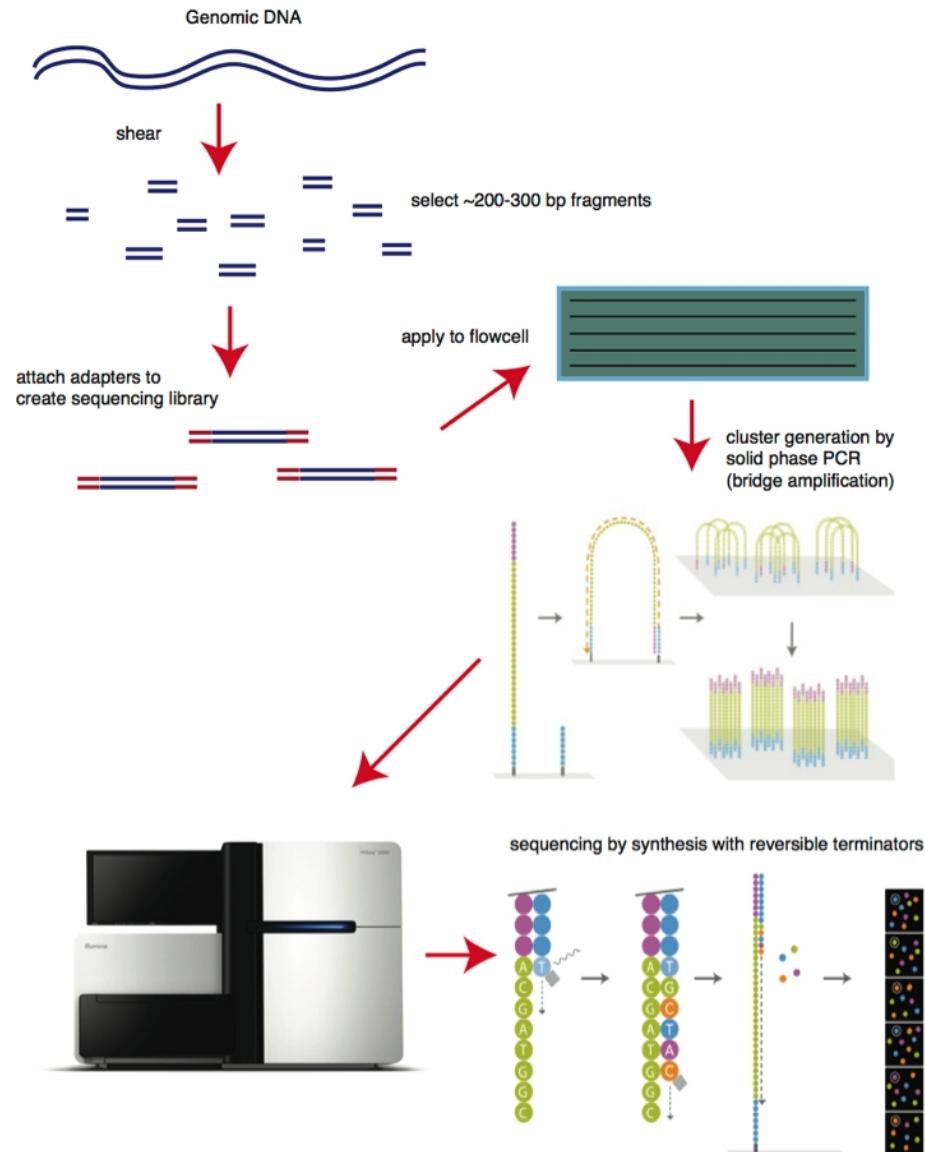
- NGS sequencer
- Shorter than sanger
- Up to 1 M reads

CON

- Few reads ~ 1 M
- High Insertion/Deletion (INDEL) errors (homopolymers)
- NOW EXTINCT ?

Sequencing platform - Reversible dideoxy-chain termination

Illumina



PRO

- NGS sequencer
- Very short (~100 bp, 250 bp)
- Up to 100 B reads

CON

- Short reads
- PCR based
- EXTINCT WHEN ?

Sequencing platform - Reversible dideoxy-chain termination

Illumina hybrid subassembly - Moleculo



RESEARCH ARTICLE



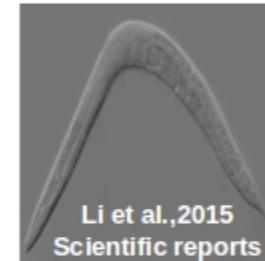
The genome sequence of the colonial chordate, *Botryllus schlosseri*

Ayelet Voskoboinik^{1,2*}, Norma F Neff^{3†}, Debashis Sahoo^{4†}, Aaron M Newman^{1†}, Dmitry Pushkarev^{1†}, Winston Koh^{3†}, Benedetto Passarelli³, H Christina Fan³, Gary L Mantas⁵, Karla J Palmeri^{1,2}, Katherine J Ishizuka^{1,3}, Carmela Gissi⁶, Francesca Griggio⁶, Rachel Ben-Shlomo⁷, Daniel M Corey⁸, Lolita Penland⁹, Richard A White III⁹, Irving L Weissman^{1,2,4,8}, Stephen R Quake^{1,2}

¹Department of Pathology, Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, Stanford, United States; ²Hopkins Marine Station, Stanford University, Pacific Grove, United States; ³Departments of Applied Physics and Bioengineering, Howard Hughes Medical Institute, Stanford University, Stanford, United States; ⁴Dipartimento di Bioscienze, Università degli Studi di Milano, Milano, Italy; ⁵Department of Biology, University of Haifa-Oranim, Tivon, Israel; ⁶Ludwig Center for Cancer Stem Cell Research and Medicine, Stanford University School of Medicine, Stanford, United States



Li et al., 2014
Plos One

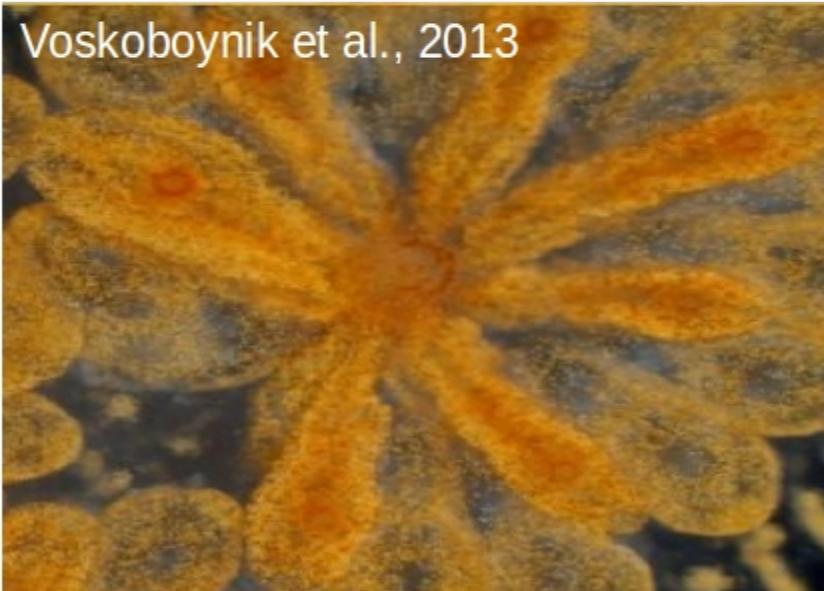


Li et al., 2015
Scientific reports



Barrio et al., 2016
eLife

Voskoboinik et al., 2013



Sharon et al., 2015
Genome Research



Kuleshov et al., 2016
Nature Methods

Why not soil?

Sequencing platform - Reversible dideoxy-chain termination

Illumina hybrid subassembly - Molecular

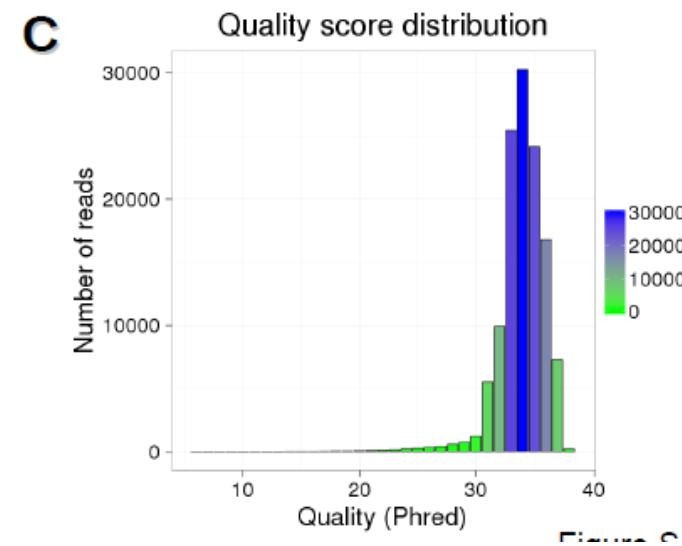
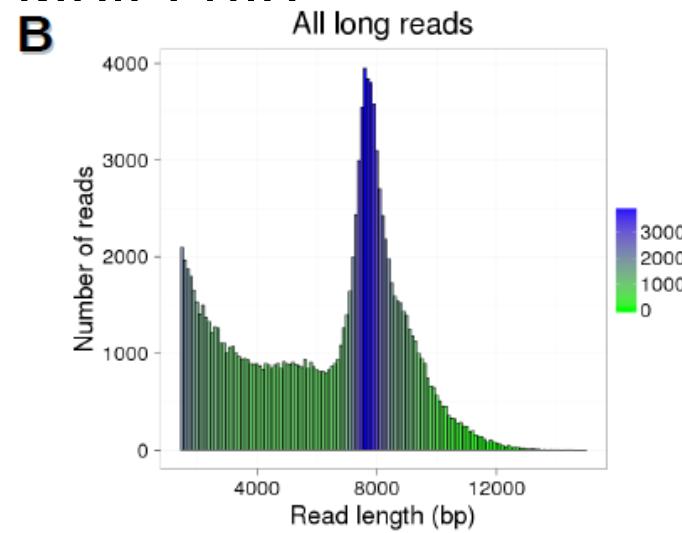
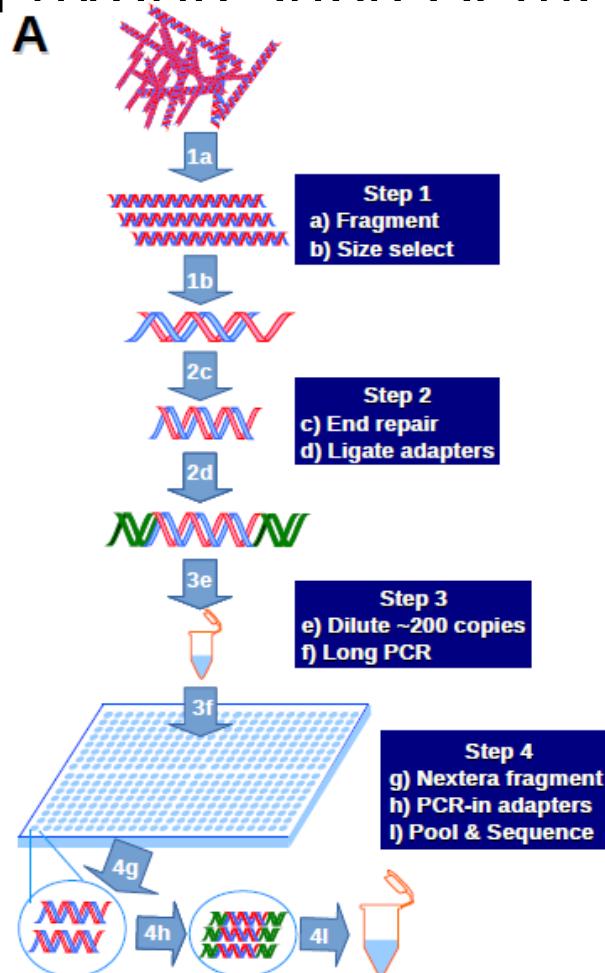


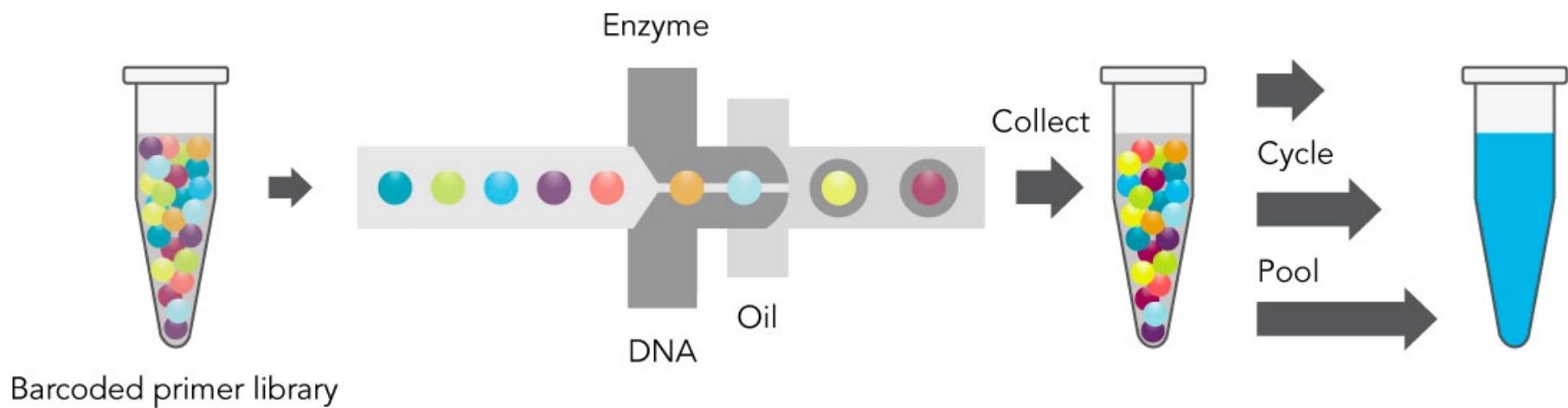
Figure S4

White III et al., 2016. *Msystems*

Link to article: <https://mSystems.asm.org/content/1/3/e00045-16>

Sequencing platform - Reversible dideoxy-chain termination

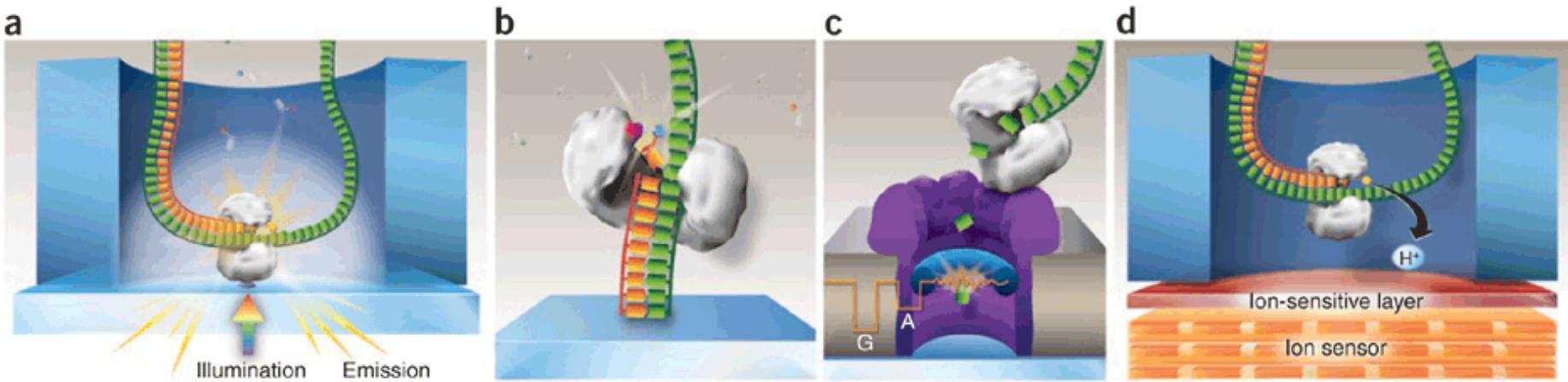
10X technology - Molecule in a droplet



- PRO**
- NGS sequencing library tech (~1 ng)
 - Long >100k partition
 - Phasing genomes

- CON**
- Expensive (\$5 run/\$75k instrument)
 - PCR based

Sequencing platform - Single molecule Pacific Biosciences (PacBio) zero-mode waveguides



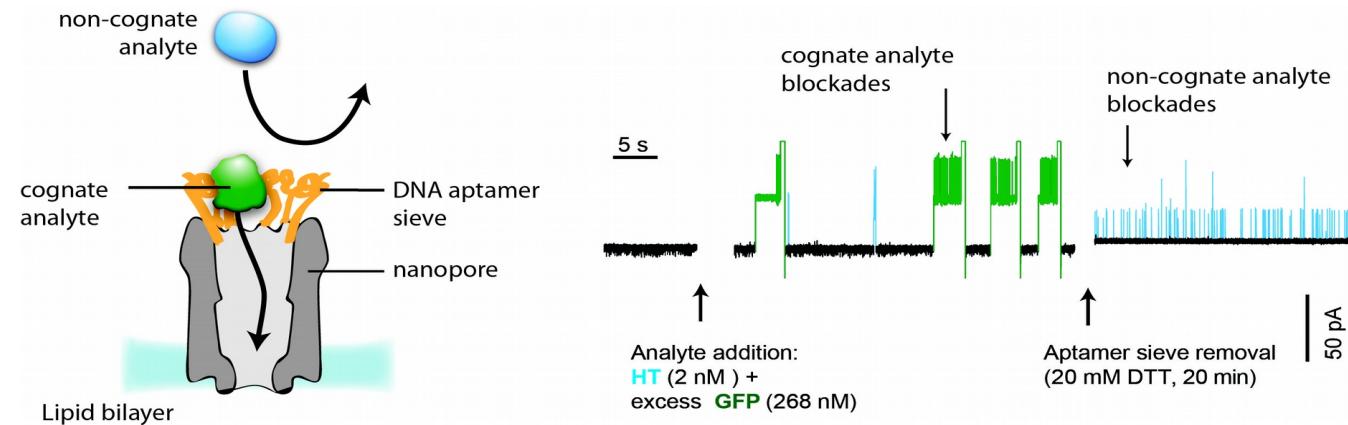
PRO

- Long Reads (>20 kb)
- Improved assembly/scaffold size
- Detects novel and native DNA modifications

CON

- Prior to 2013 long reads (<Q20) very low quality (<Q25)
- High Insertion/Deletion (INDEL) errors
- Higher cost than Illumina

Sequencing platform - Single molecule Oxford nanopore sequencing



PRO

- Long Reads (>20 kb)
- Improved assembly/scaffold size
- Detects novel and native DNA modifications

CON

- Prior to 2013 long reads (<Q20) very low quality (<Q25)
- High Insertion/Deletion (INDEL) errors
- Higher cost than Illumina

Sequencing platform - Single molecule Oxford nanopore sequencing



MinION
10 to 20 Gb ~6 h
Weights 100 g
PC/Laptop portable
~\$1k



GridION
100 GB per run
5 minIONs
\$50k



PromethION
48 MinIONs
Each cell 3,000 nanopores
Total 144,000 nanopores
\$150k



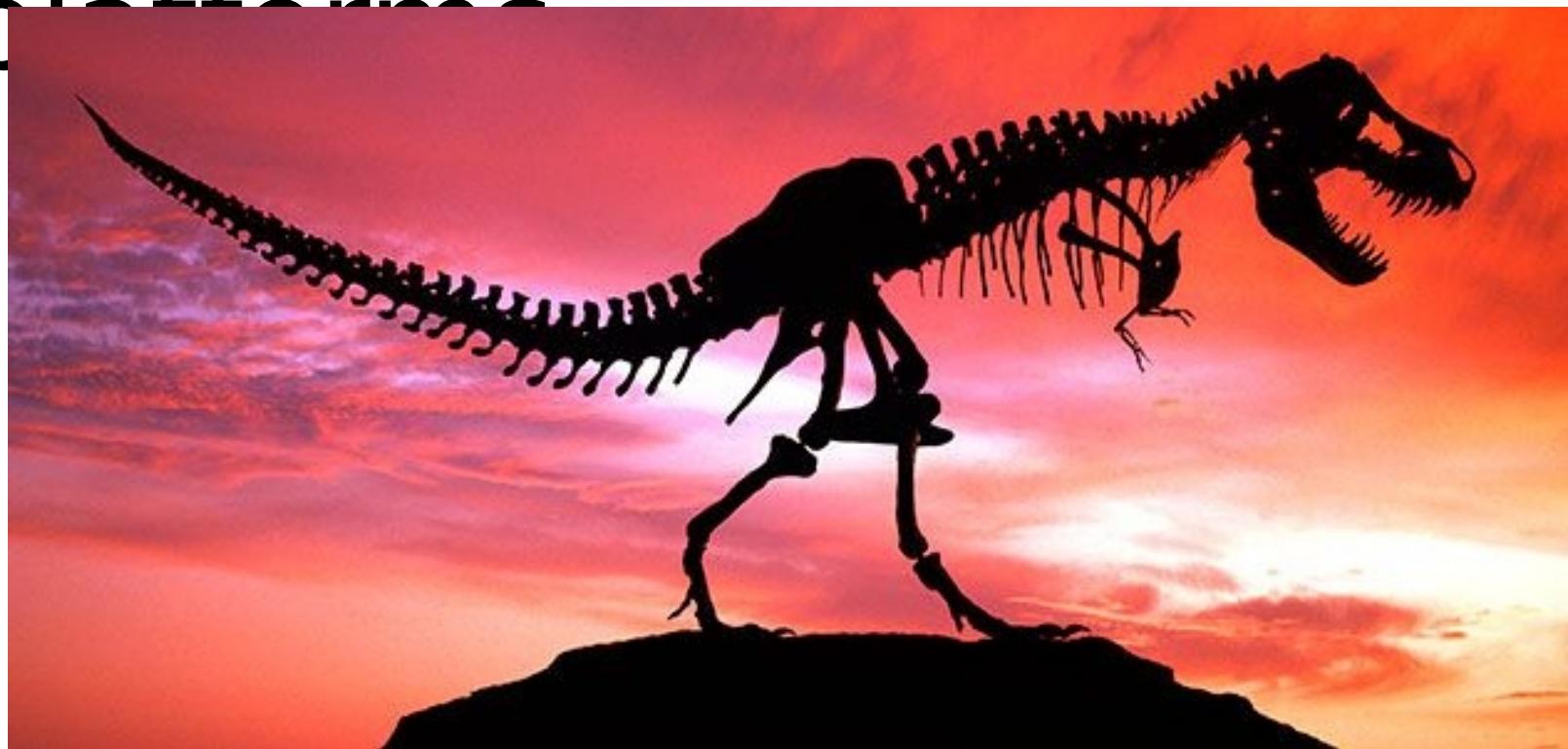
SmidgION

Sequencing cost by

	Date	Time	Read length	Quality	Amount	MB cost
454 FLX+	2005 (2011)	24 h	700 bp	> Q20	1 Gb	\$10.00
Ion torrent	2011 (2014)	2 h	400 bp	> Q20	1 Gb	\$1.00
Illumina MiSeq	2011 (2014)	27 h	2x300 bp	> Q30	15 Gb	\$0.15
Illumina HiSeq 2500	2010 (2014)	1 - 10 days	2x250 bp	> Q30	3000 GB	\$0.05
Illumina Moleculo	2012 (2014)	2 days	5 - 15 kb	> Q30	2 Gb	\$0.35
Illumina Xten	2014	14 days	2x100 bp	> Q30	1.8 Tb	\$0.001
PacBio	2011 (2013)	30 m - 4 h	10 - 40 kb	> Q20	500 Mb	\$0.13

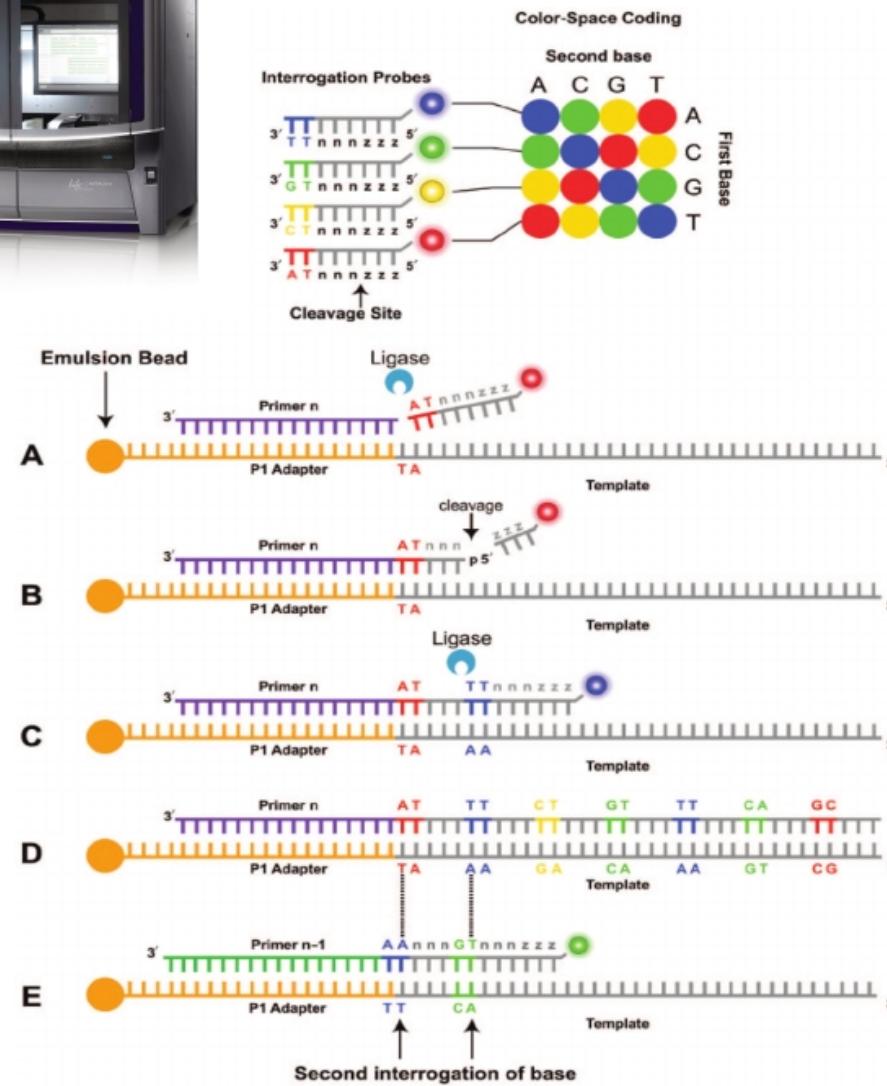
Extinct Sequencing

platforms



Sequencing platform - Sequencing by ligation

Life Tech SOiLD platform



PRO

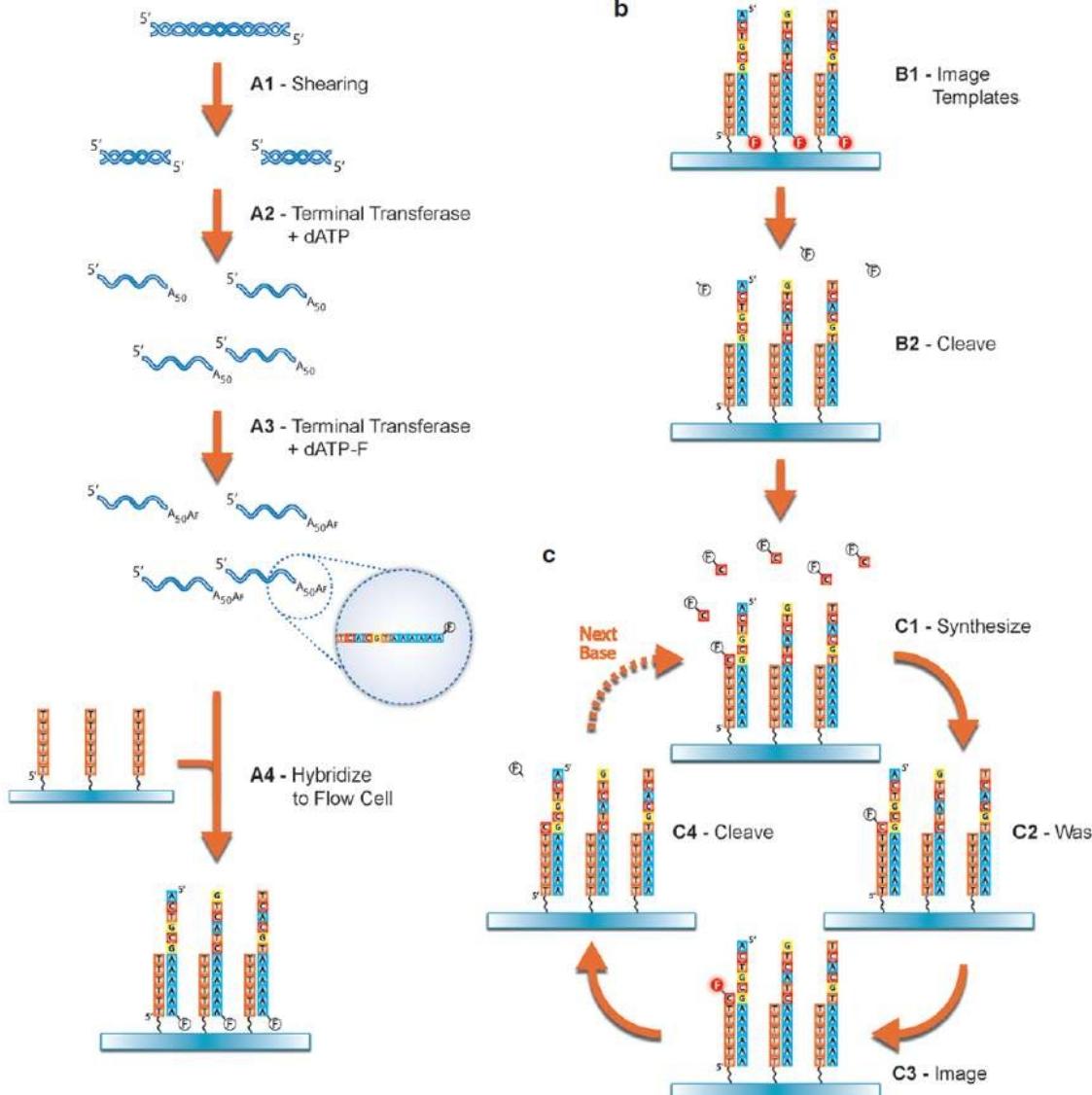
- NGS sequencer
- Very short (<100 bp)
- Up to 10 B reads

CON

- Short reads
- emPCR based
- NOW EXTINCT 2017

Sequencing platform - single molecule fluorescent sequencing

Helicos Biosciences heliscope



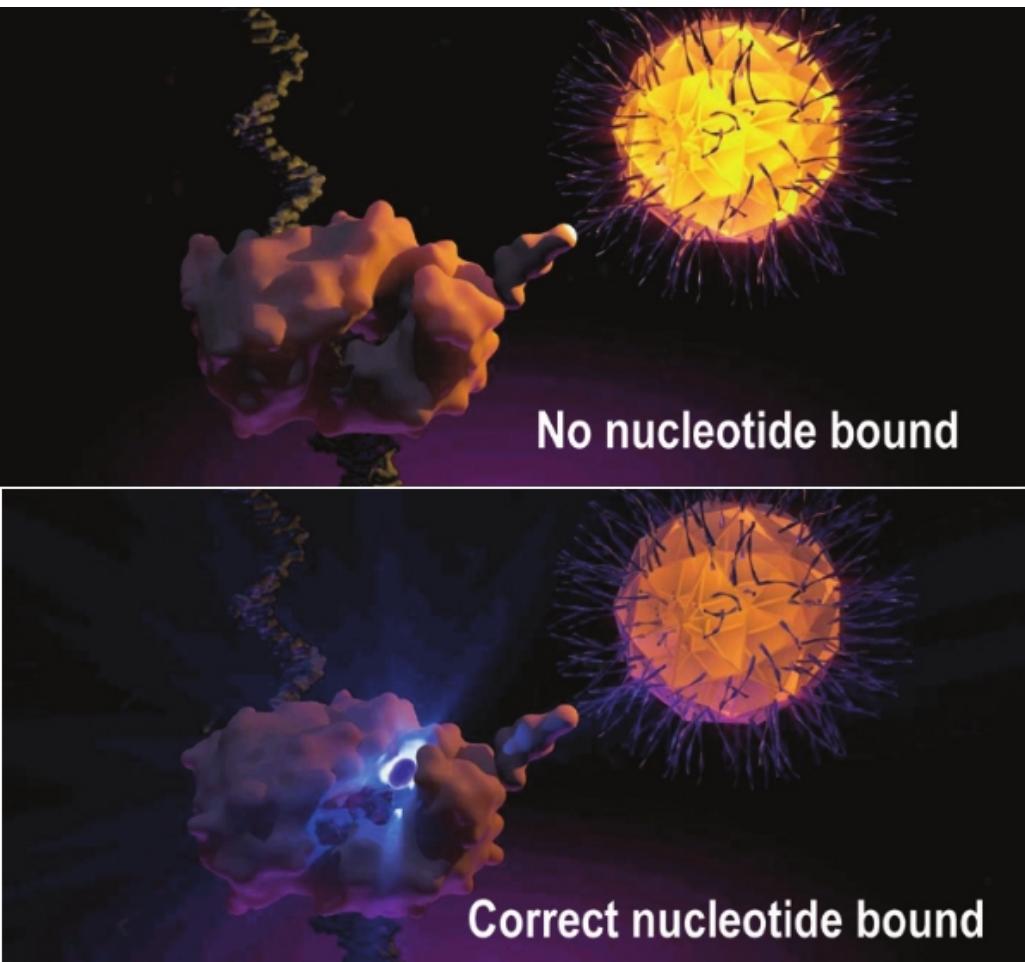
PRO

- NGS sequencer
- Very short (<50 bp)
- Up to 1 B reads
- Direct RNA (polyA oligos)

CON

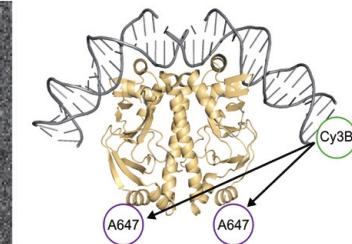
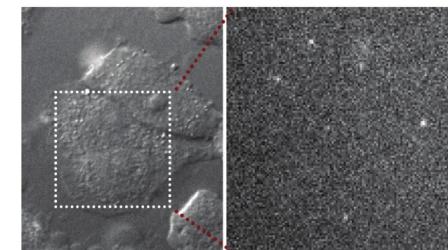
- Short reads
- expensive
- NOW EXTINCT 2012

Sequencing platform - single molecule FRET sequencing Life Tech “Project Starlight”



- PRO
 - NGS sequencer
 - Uses FRET
 - Truly single molecule

- CON
 - Can it sequence reads?
 - NOW EXTINCT 2010



Data review

*.fna, .fasta, .fa ?

*.faa ?

*.gff ?

*.gbk ?

*.fq or fastq ?

*.gtf Gene transfer format (GTF)

Read vs. sequence?

Contig?

kmer?

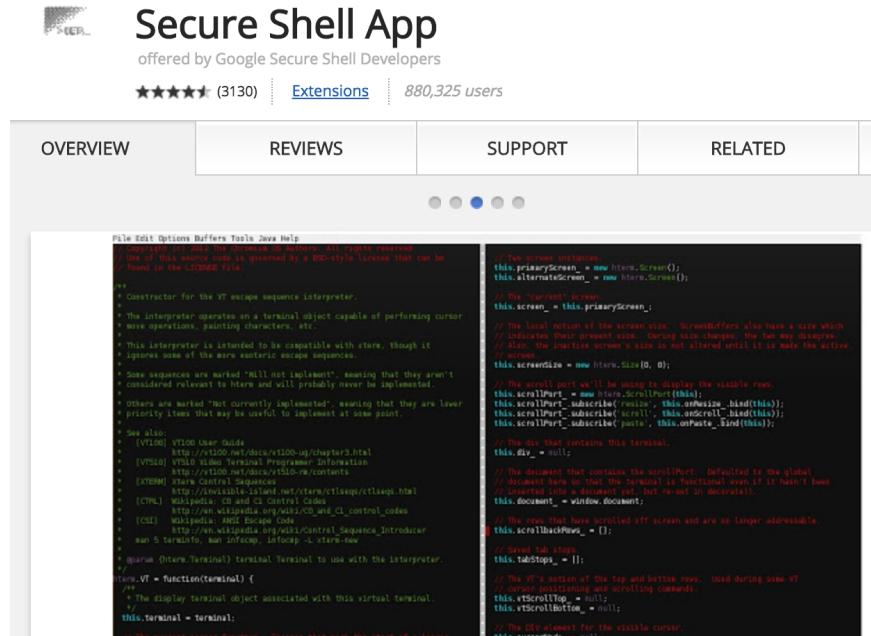
Command Line - Terminal

every have one?

Mac: Terminal

Windows: PUTTY

Chrome: Secure Shell (extension; runs command line in your browser! How cool is that?)



Command line - Review

cd	?	grep	?
pwd	?	sed	?
mkdir	?	rm	?
head	?	rm -r	?
tail	?	echo	?
more	?		?
less	?	>	?
mv	?	<	?
touch	?		
cp	?		
cat	?		

Command line - Review

- Make a folder with your name
- Make two files in as .txt (labeled 1 and 2)
- Write your name in file 1
- Write your name in file 2 with _a at the end
- Combine those two files to a new file
- Copy the whole folder of your name with a new name
- Then show me that you can remove it