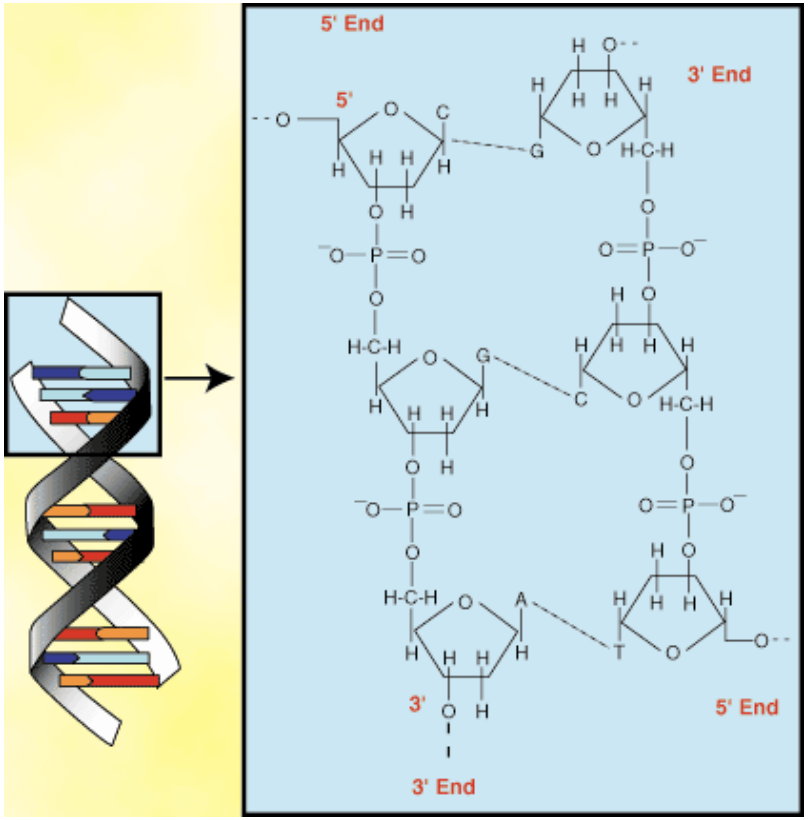
DNA Sequencing Technologies

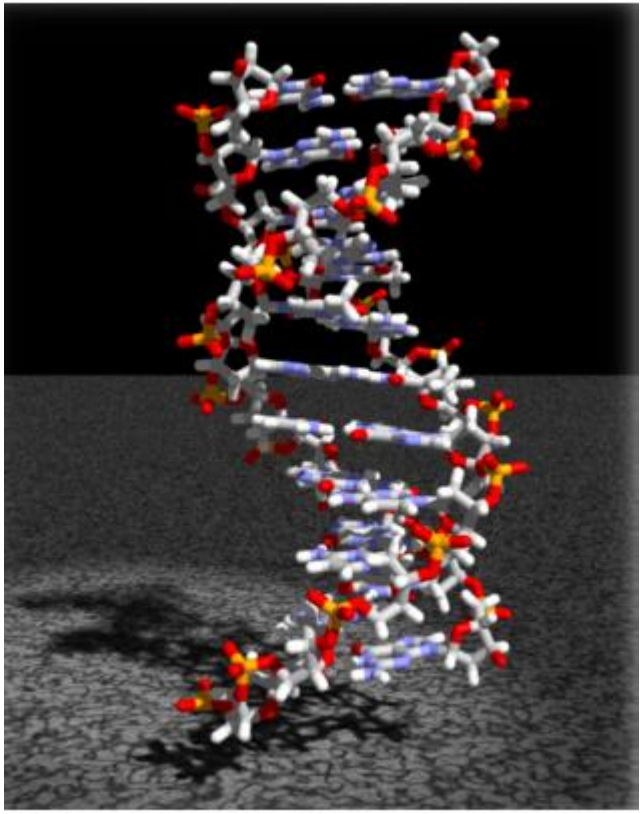
Aleksandra Radenovic
aleksandra.radenovic@epfl.ch

EPFL – Ecole Polytechnique Federale de Lausanne
Bioengineering Institute IBI

DNA –charged polymer

- DNA is a linear polymer molecule, i. e. a long chain of repeated subunits, called nucleotides. These come in four types, and thus DNA stores information as particular sequences of nucleotides



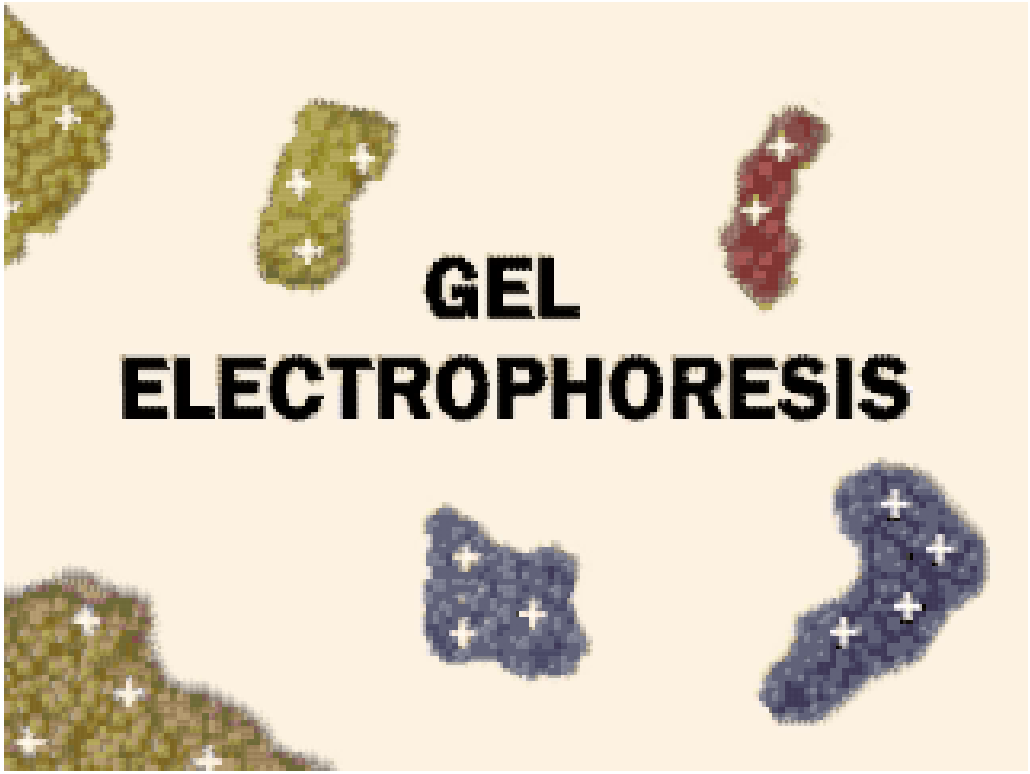


Phosphate carries negative charge

Molecular Biology of the Cell, Alberts, Bray, et al.

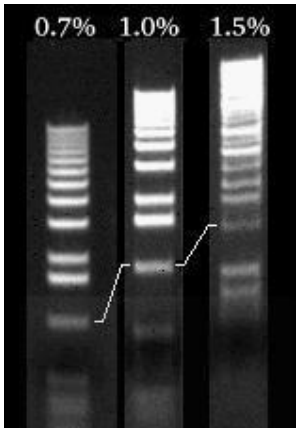
Gel electrophoresis

- Determining the length of a DNA molecule
- Gel electrophoresis involves driving charged molecules, such as DNA, through a porous gel matrix by means of an applied electric field. The gel matrix exerts a frictional force which increases with molecule size, and thus molecules of different
- sizes move at different speeds, separating as they move



$$f \cdot \vec{v} = q \cdot \vec{E}$$

$$\mu = \frac{\vec{v}}{\vec{E}} = \frac{q}{f} = \frac{z \cdot e}{f}$$

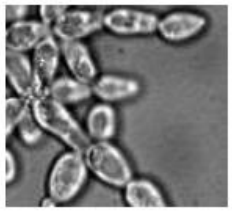


DNA Sequencing

- Genome Projects


M. Tuberculosis


H. pylori


S. cervisiae


C. elegans


Grape wine


Barley


Arabidopsis


Wheat


Medicago


Potato


Wasp


Maccaca


Neanderthal


James Watson


Cancer

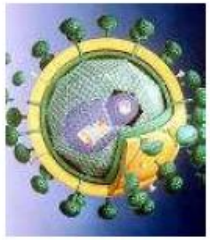

Melon


Salamander


Bonobo


Honeybee


Mammut


HIV


H5N1


Toma

Why is Genome Sequencing Important?

- To understand how the genome as a whole works – how genes work together to direct growth, development and maintenance of an entire organism
- Understand how gene expression is regulated in a particular environment
- To study gene expression in a specific tissue, organ or tumor
- To study human variation
- To study how humans relate to other organisms
- To find correlations how genome information relates to development of cancer, susceptibility to certain diseases and drug metabolism (pharmacogenomics)

GOALS find sequence variability from cell to cell

DNA Sequencing

Israeli city to build dog poop DNA database

September 19, 2008 by Mark O'Neill |

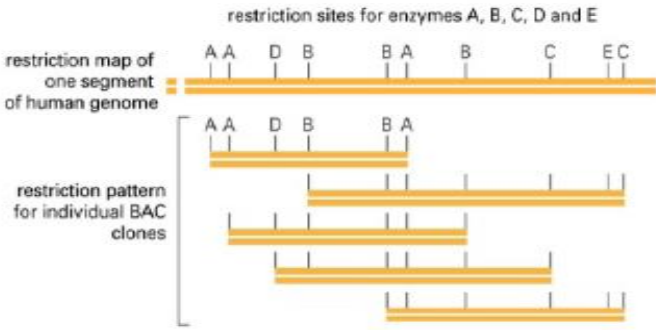
By **Mark O'Neill**
Contributing Writer, [GAS]

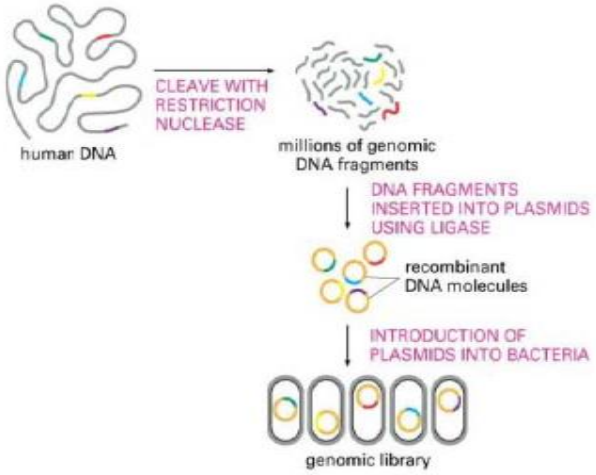
OK, now I *really* have heard everything!

The Israeli city of Petah Tikva is seemingly fed up with all the doggie poop lying around on their streets. So their solution is to set up a dog poop DNA database. To build this, all the dog owners of the city are being asked to take their canine pals into the local vet to have their dog's mouth swabbed. Rover or Tiddles can then have their DNA information entered into the computer.

DNA Sequencing: How is Genome Sequencing Done?

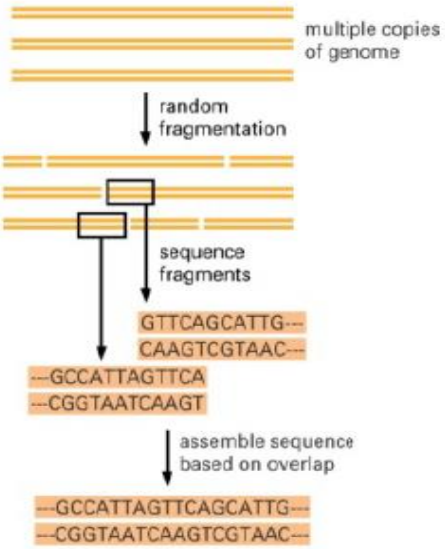
- **How is Genome Sequencing Done?**
- ***Clone by clone***





Create a crude physical map of the whole genome before sequencing with restriction enzymes Break the genome into overlapping fragments and insert them into BACs and transfect into E.coli

- ***Shotgun sequencing***



Break genome into random fragments, sequence each of the fragments and assemble fragments based on sequence overlaps

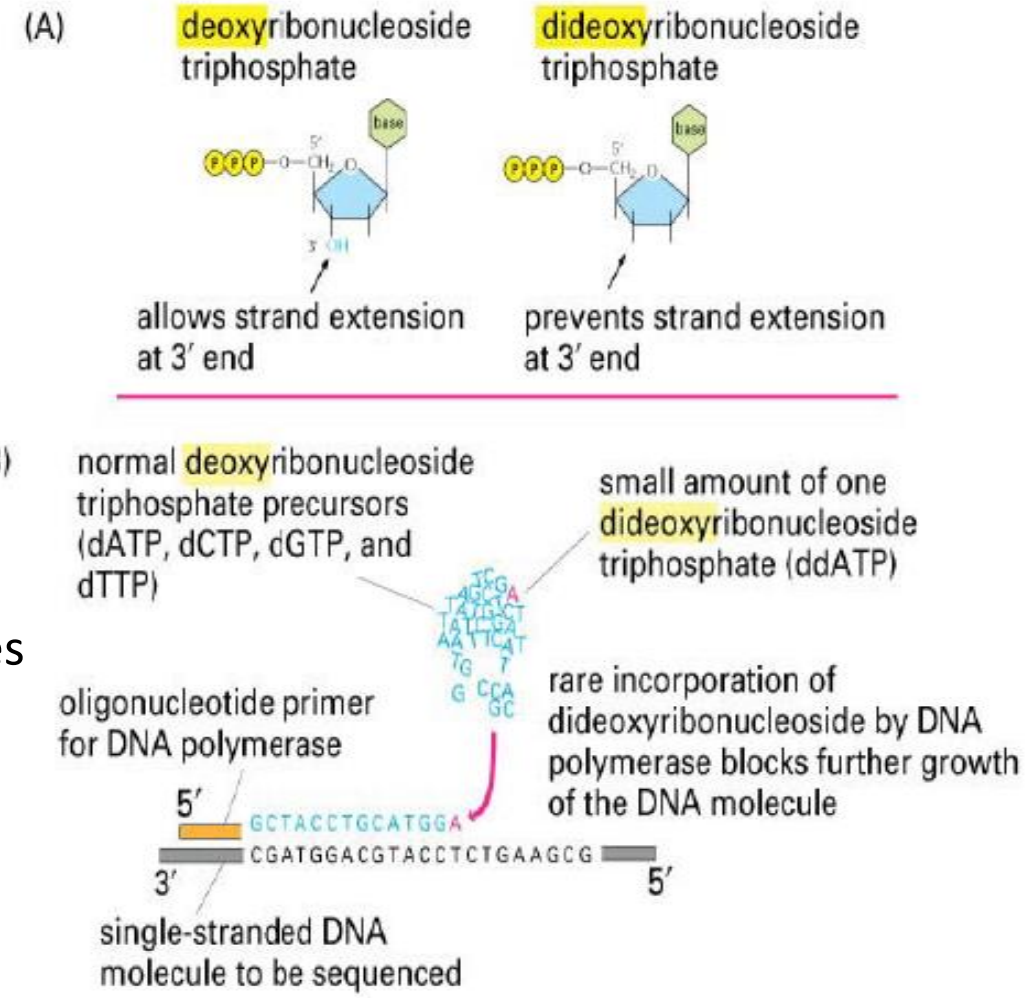
Figure 10-9 Essential Cell Biology, 2/e. (© 2004 Garland Science)

DNA Sequencing: How is Genome Sequencing Done?

- **How is Genome Sequencing Done?**
- The key principle of the Sanger method was the use of dideoxynucleotide triphosphates (ddNTPs) as DNA chain terminators.

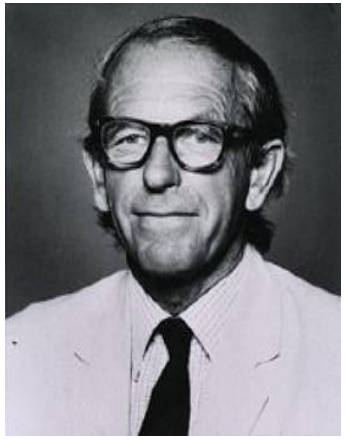
- "Normal" DNA synthesis:
  - DNA strand as template
  - Primer
  - Deoxynucleotides
  - Polymerase enzyme
  - Use several cycles to amplify

DNA synthesis is carried out in the presence of limiting amounts of dideoxy-ribonucleoside triphosphates that results in chain termination trough chain termination fragments of distinct sizes are generated that can be separated by gel electrophoresis



DNA Sequencing: How is Genome Sequencing Done?

- Original method used radio-labeled primer or dideoxynucleotides This method required four separate DNA synthesis reactions to be separated by electrophoresis in four parallel lanes. The gel needs to be dried, exposed to film, developed and manually read. Approx. 150 bases read length



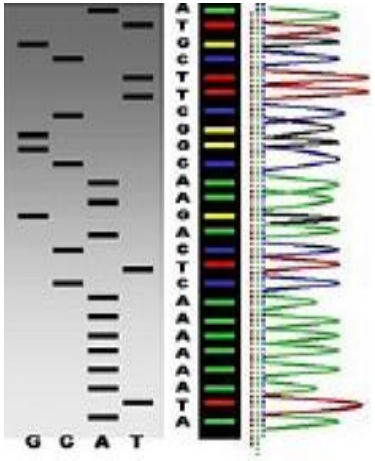
Frederick Sanger
Nobel Prize (1980)

Summary – Sequencing Method Established

- Need of four reactions in parallel
- Heat labile polymerase
- Use of radioactivity
- Low resolution on gels
-Approx. 150 nucleotides read length
- time consuming

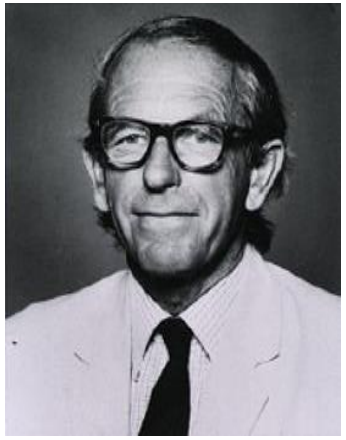
Improvements

-Use of fluorescently labeled dideoxynucleotides /one-lane electrophoresis
-Introduction of capillary electrophoresis to increase resolution (up to **1,000** ntes)
-Use of heat stable polymerase (Taq) Automation



DNA Sequencing: How is Genome Sequencing Done?

- Original method used radio-labeled primer or dideoxynucleotides This method required four separate DNA synthesis reactions to be separated by electrophoresis in four parallel lanes. The gel needs to be dried, exposed to film, developed and manually read. Approx. 150 bases read length



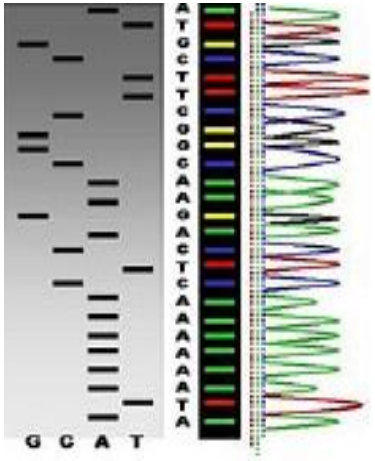
Frederick Sanger
Nobel Prize (1980)

Summary – Sequencing Method Established

- Need of four reactions in parallel
- Heat labile polymerase
- Use of radioactivity
- Low resolution on gels
-Approx. 150 nucleotides read length
- time consuming

Improvements

-Use of fluorescently labeled dideoxynucleotides /one-lane electrophoresis
-Introduction of capillary electrophoresis to increase resolution (up to **1,000** ntes)
-Use of heat stable polymerase (Taq) Automation



Human Genome Project Sequencing Centers

Whole Genome Shotgun Sequencing

<u>Personalized genome Sequencing</u>

<u>Goals:</u>

Link genome with phenotype

Provide personalized diet and medicine

(???) designer babies, big-brother insurance companies

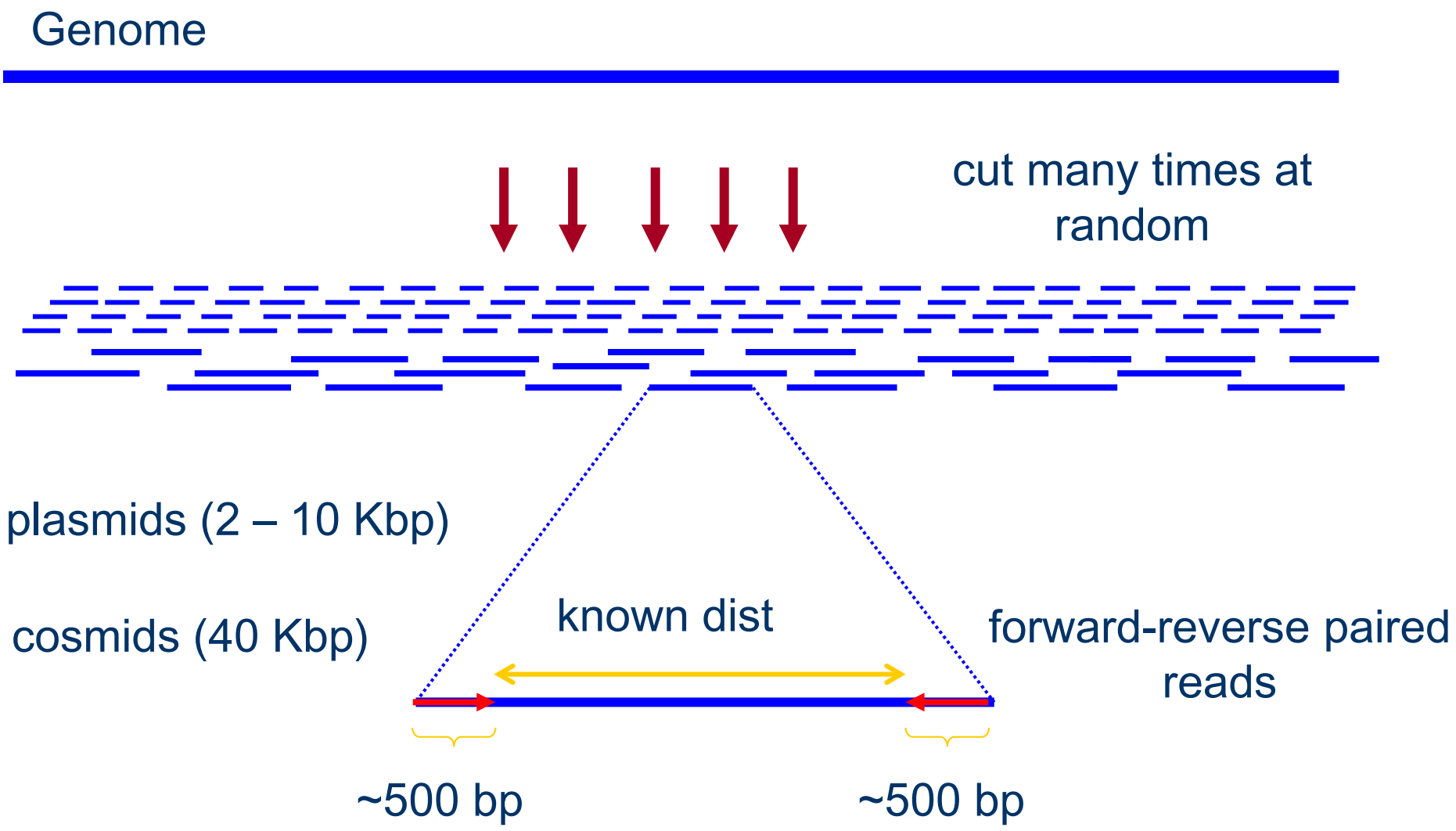
<u>Timeline:</u>

Inexpensive sequencing: 2010-2017

Genotype–phenotype

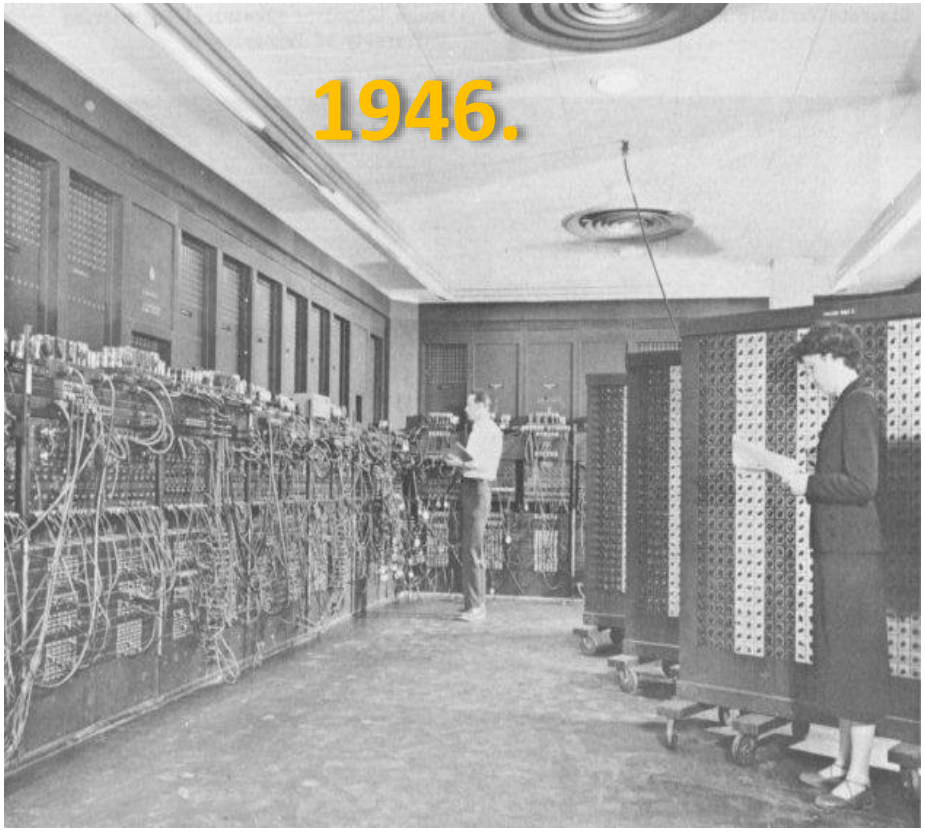
Personalized drugs: 2017-???

Whole Genome Shotgun Sequencing



Successful technology

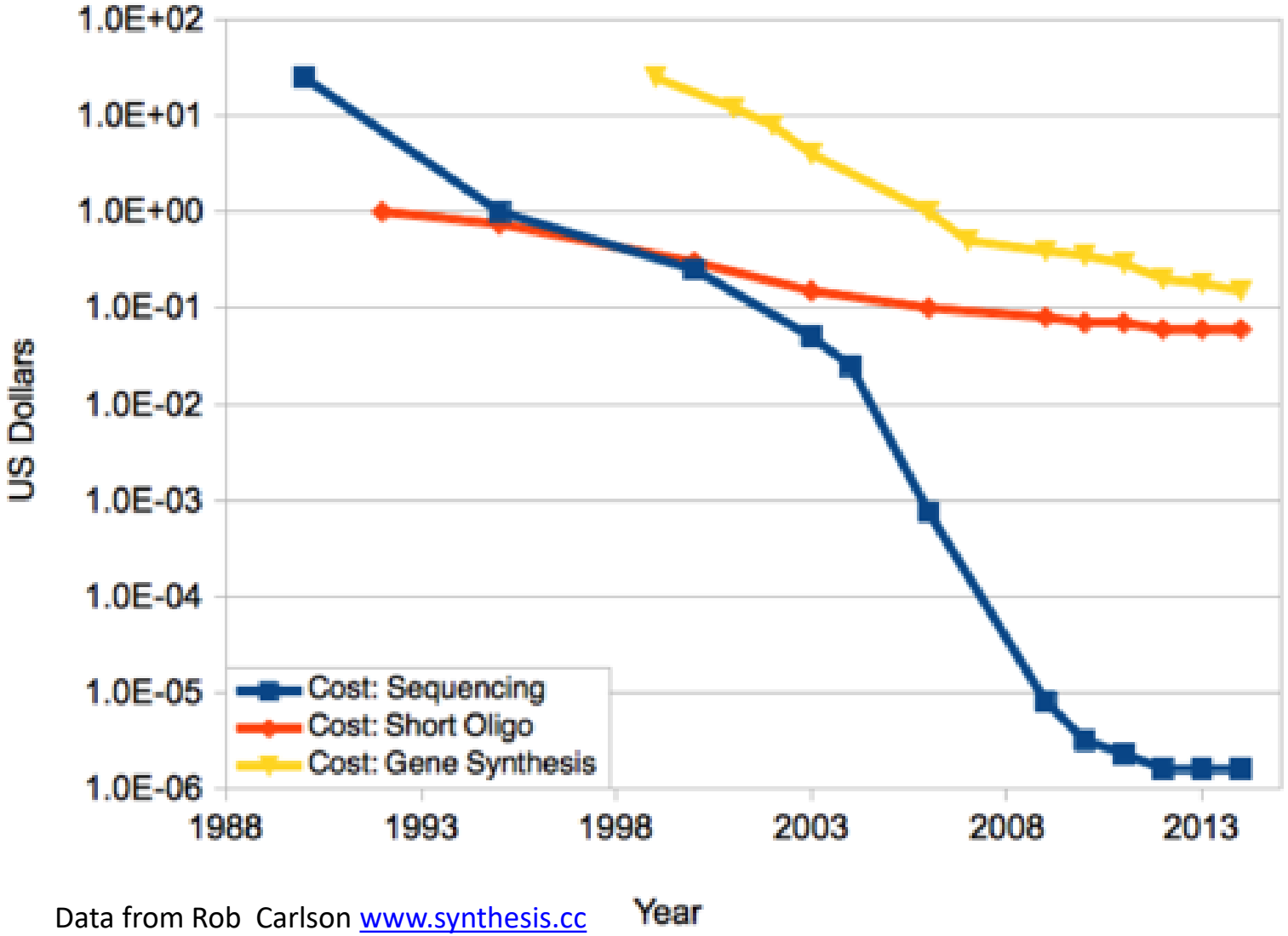
Affordable, reliable, straightforward to use, and easy to adapt to new applications



http://en.wikipedia.org/wiki/ENIAC

http://en.wikipedia.org/wiki/DNA_sequencer

Price per base for DNA sequencing and synthesis



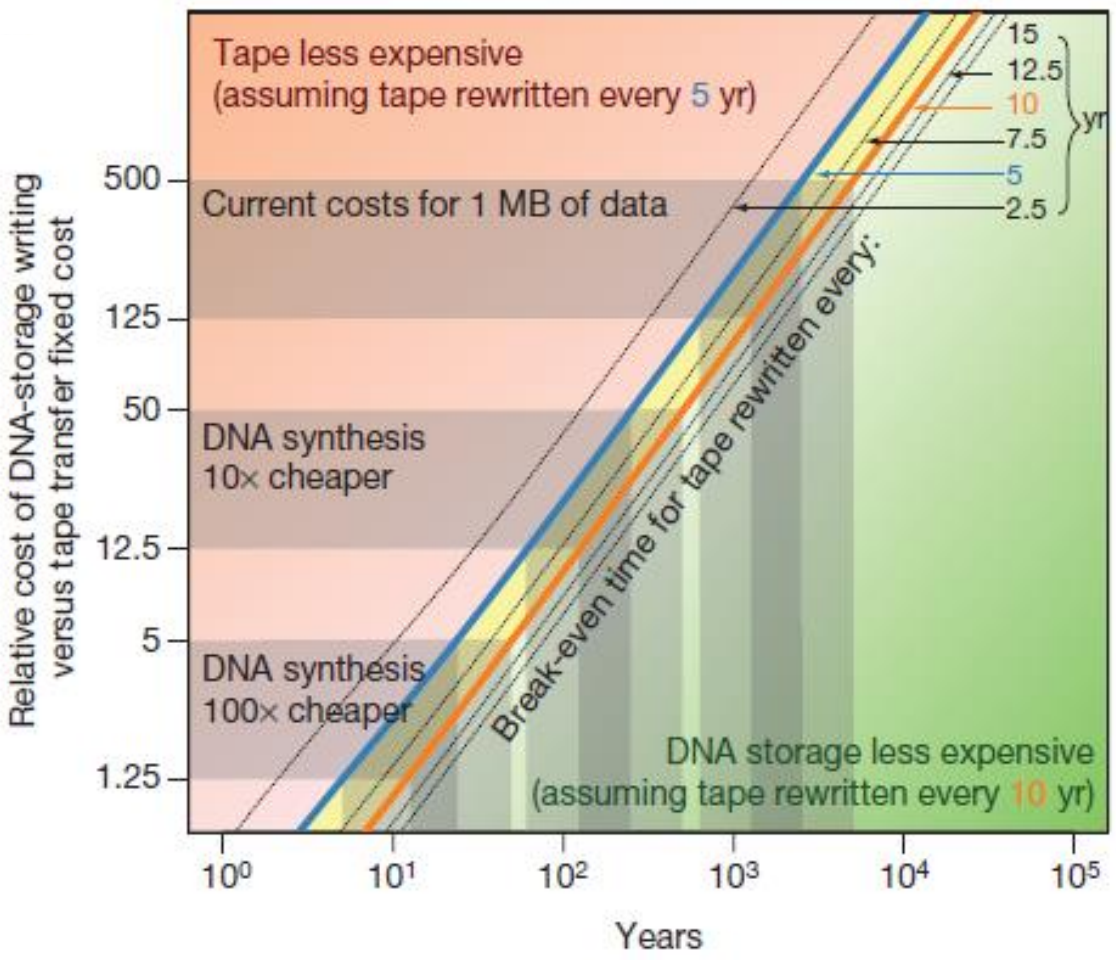
Data from Rob Carlson www.synthesis.cc

Synthesized DNA- for information storage

effective solution for rarely accessed archives

high-capacity and low-maintenance

The speed of DNA-storage writing and reading are not competitive with current technology

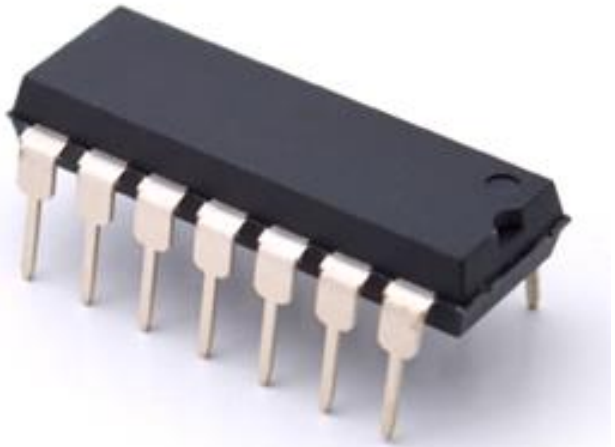


Goldman et at. Nature **2013 .**

Could we advance DNA reading and writing?

Vacuum tubes: slow, expensive, fragile

Transistors: much simpler, much smaller, much cheaper, more reliable, no warm up, much faster.



Integrated circuits: miniaturization added to all the existing benefits, enabled unthought-of possibilities

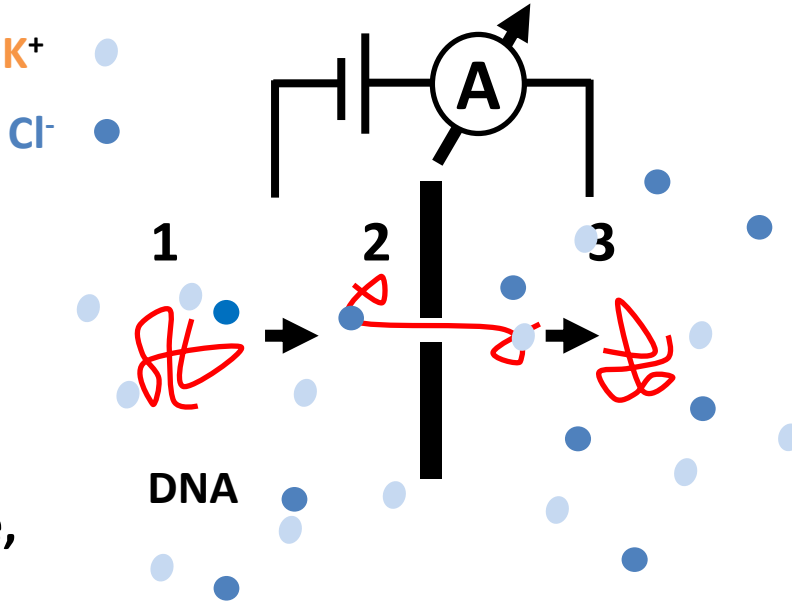
Nanopores

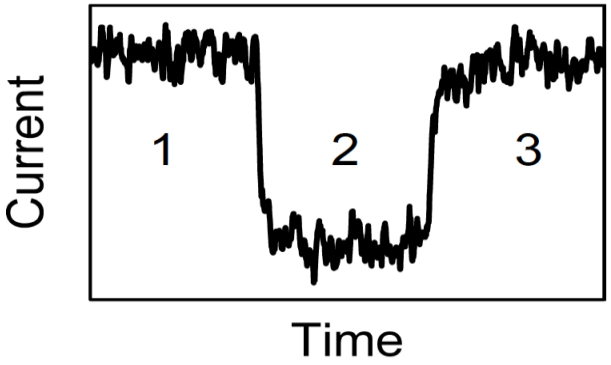
Single molecules –sensors

Sensitive to atomic composition

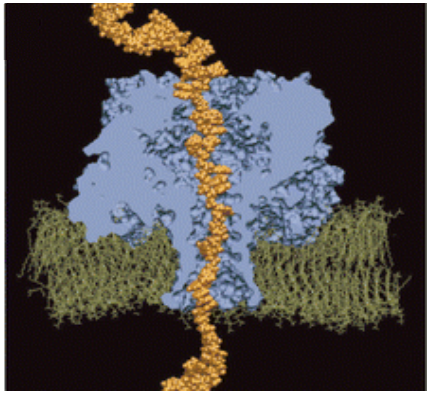
Sensing is intrinsic, no bleaching, nondestructive, repeatable

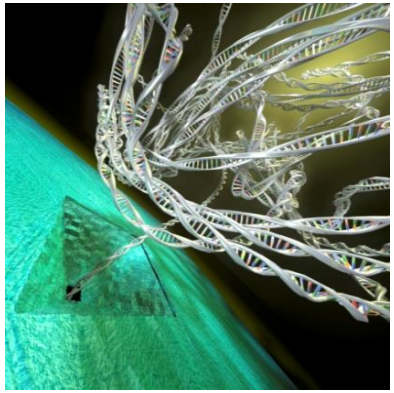
Sensing occurs in solution





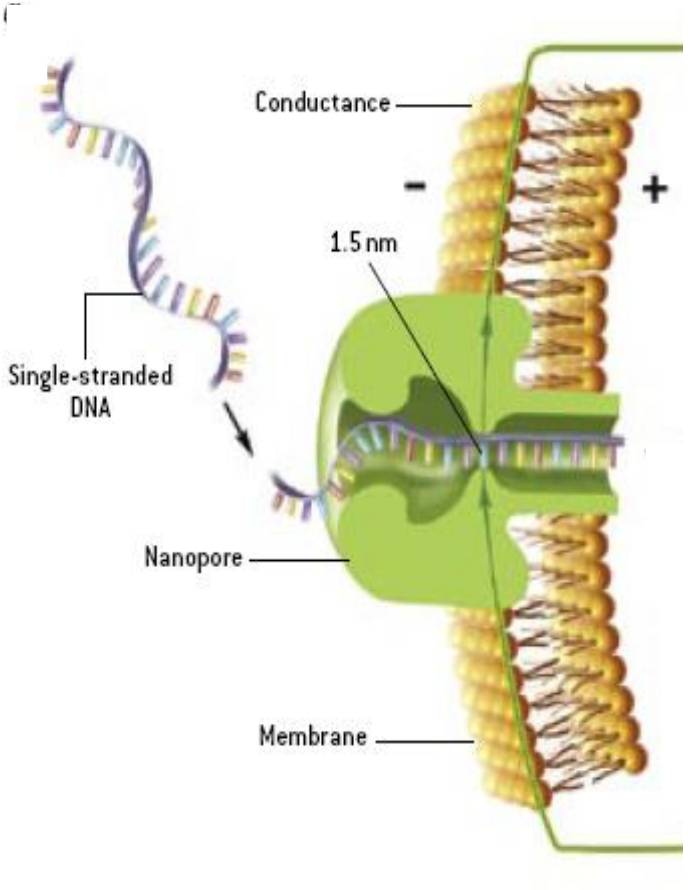
$$I = \frac{U}{R} = UG$$

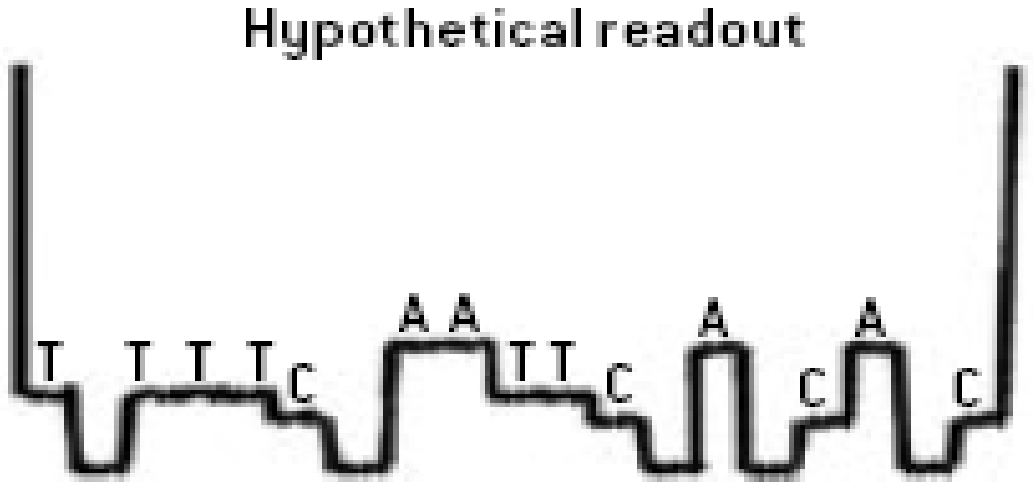




Akeson M. et al. **1999** Meller A. et al. **2000** Howorka S. et al. **2001**
Li J, et al. **2001.** Dekker C. et. al. **2007**

The Goal: Automated Rapid DNA Sequencing with Nanopores





Advantages of nanopore sequencing:

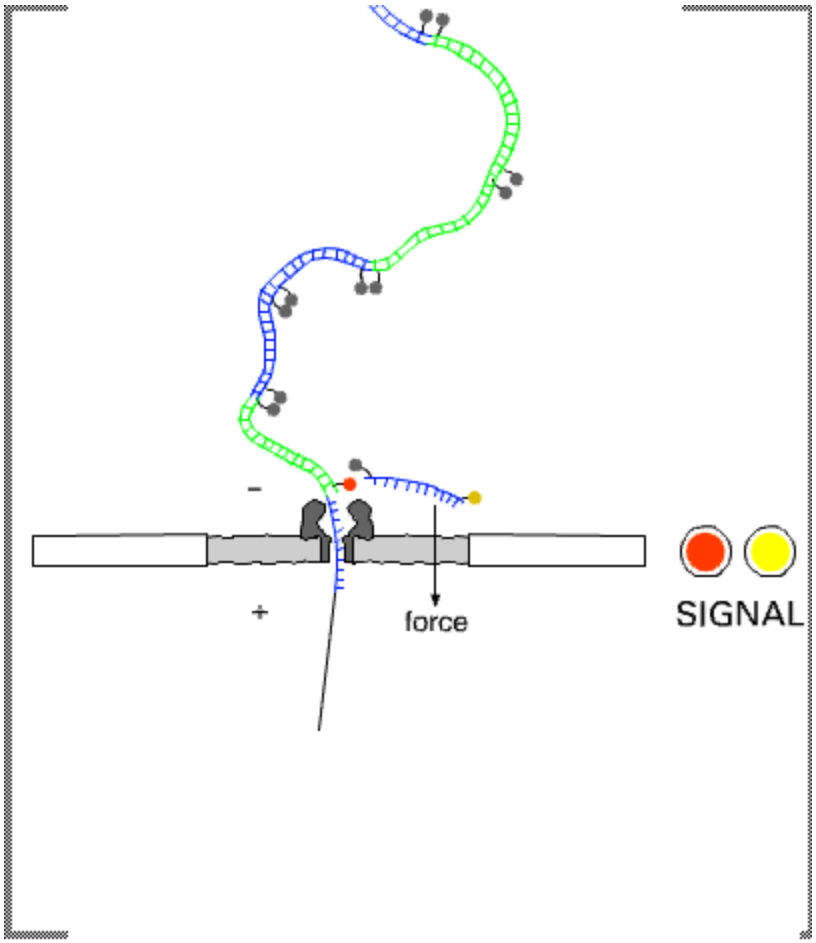
- long-read lengths
- single molecule
- no amplification
- label-free
- electrical detection

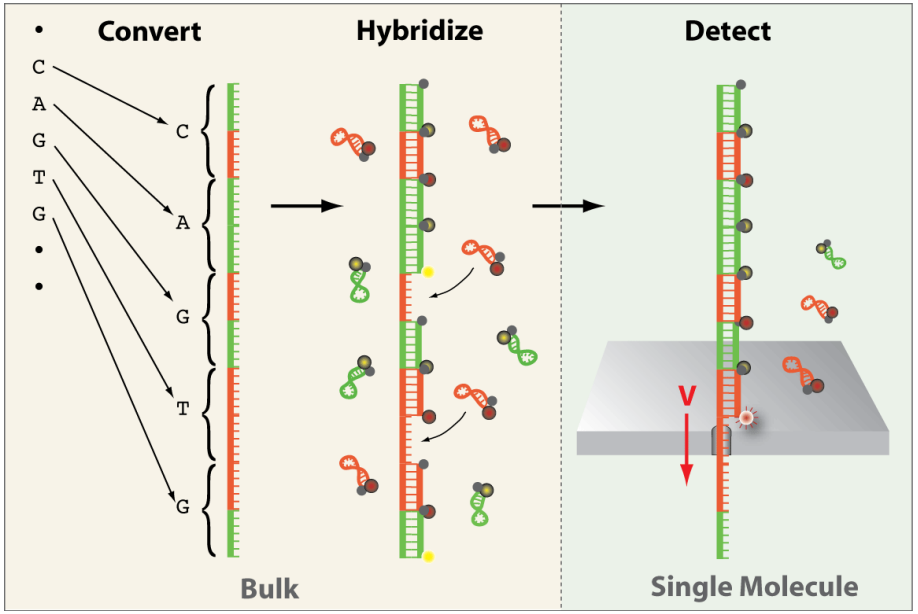
DNA can be sequenced using nanopores only if its dynamics through the pore can be controlled…

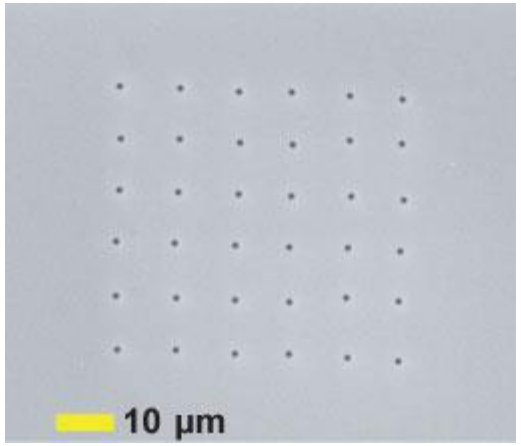
Kasianowicz J. **1996.** Branton D, et al. **2008**

The Goal: Automated Rapid DNA Sequencing with Nanopores

- **Ultra Fast DNA Sequencing Using Nanopores and Optical Probes**



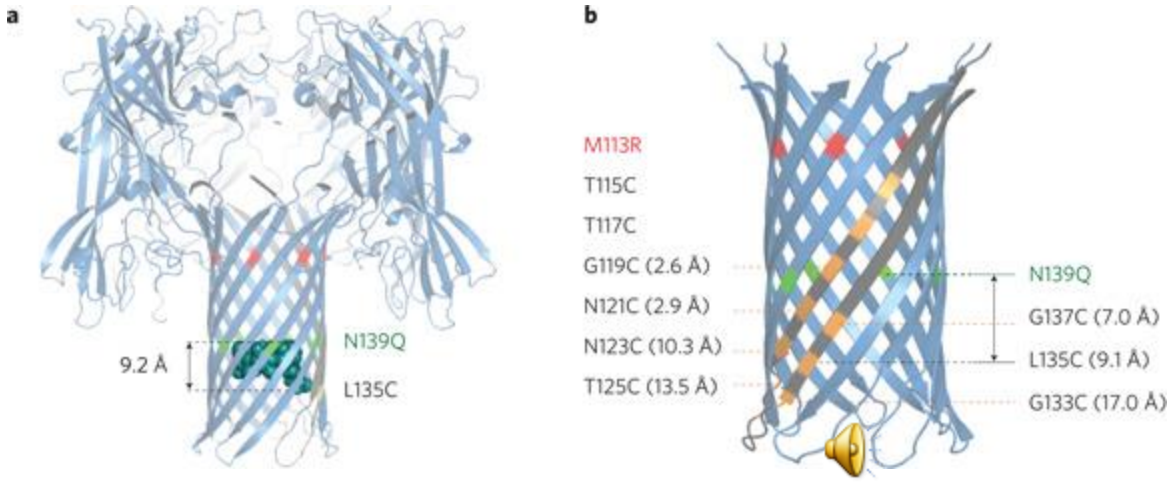


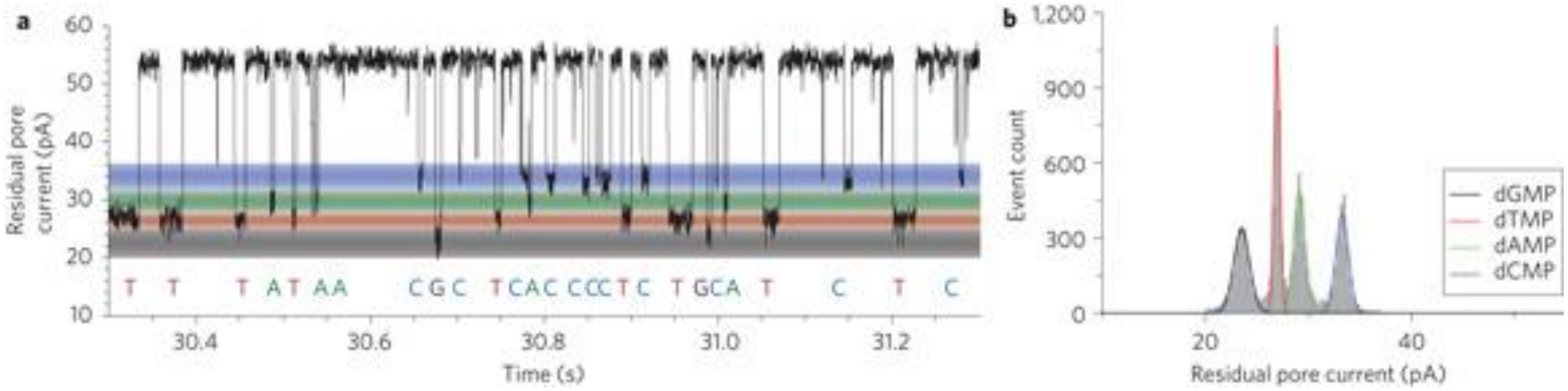


. D. Branton *et. al. Nature Biotech.* **26**, 1146-53. (2008)

The Goal: Automated Rapid DNA Sequencing with Nanopores

- cyclodextrin covalently attached inside alpha hemolysin pore





Single-channel recording showing dGMP, dTMP, dAMP and dCMP discrimination, with coloured bands

Clarke J. Nature Nanotechnology 4, 265 - 270 (2009)

Today: Oxford Nanopore

Oxford NANOPORE Technologies

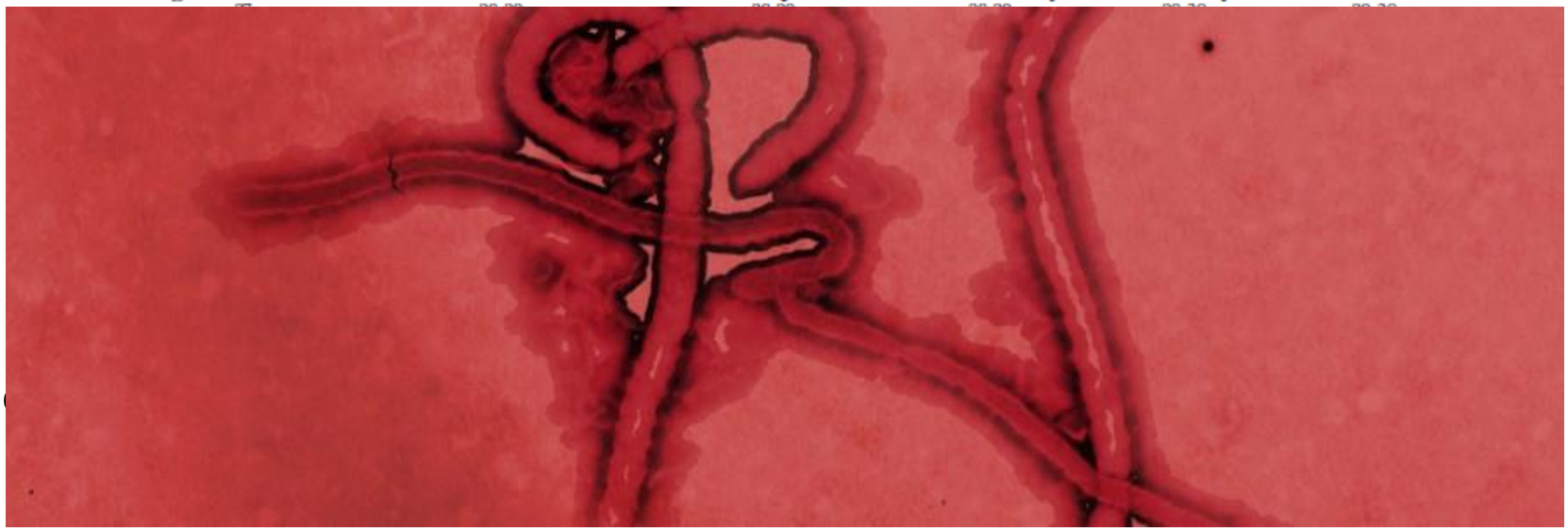
Real-time, portable genome sequencing for Ebola surveillance.

LETTER

doi:10.1038/nature16996

Real-time, portable genome sequencing for Ebola surveillance

Joshua Quick[1]*, Nicholas J. Loman[1]*, Sophie Duraffour[2,3]*, Jared T. Simpson[4,5]*, Ettore Severi[6]*, Lauren Cowley[7]*, Joseph Akoi Bore[2], Raymond Koundouno[2], Gytis Dudas[8], Amy Mikhail[7], Nobila Ouédraogo[9], Babak Afrough[2,10], Amadou Bah[2,11], Jonathan H. J. Baum[2,3], Beate Becker-Ziaja[2,3], Jan Peter Boettcher[2,12], Mar Cabeza-Cabrerizo[2,3], Álvaro Camino-Sánchez[2], Lisa L. Carter[2,13], Juliane Doerrbecker[2,3], Theresa Enkirch[2,14], Isabel García-Dorival[2,15], Nicole Hetzelt[2,12], Julia Hinzmann[2,12], Tobias Holm[2,3], Liana Eleni Kafetzopoulou[2,16], Michel Koropogui[2,17], Abigael Kosgey[2,18], Eeva Kuisma[2,10], Christopher H. Logue[2,10], Antonio Mazzarelli[2,19], Sarah Meisel[2,3], Marc Mertens[2,20], Janine Michel[2,12], Didier Ngabo[2,10], Katja Nitzsche[2,3], Elisa Pallasch[2,3], Livia Victoria Patrono[2,3], Jasmine Portmann[2,21], Johanna Gabriella Repits[2,22], Natasha Y. Rickett[2,15,23], Andreas Sachse[2,12], Katrin Singethan[2,24], Inês Vitoriano[2,10], Rahel L. Yemanaberhan[2,3], Elsa G. Zekeng[2,15,23], Trina Racine[25], Alexander Bello[25], Amadou Alpha Sall[26], Ousmane Faye[26], Oumar Faye[26],



Eb

Real-time, portable genome sequencing for Ebola surveillance.

LETTER

doi:10.1038/nature16996

Real-time, portable genome sequencing for Ebola surveillance

Joshua Quick[1]*, Nicholas J. Loman[1]*, Sophie Duraffour[2,3]*, Jared T. Simpson[4,5]*, Ettore Severi[6]*, Lauren Cowley[7]*, Joseph Akoi Bore[2], Raymond Koundouno[2], Gytis Dudas[8], Amy Mikhail[7], Nobila Ouédraogo[9], Babak Afrough[2,10], Amadou Bah[2,11], Jonathan H. J. Baum[2,3], Beate Becker-Ziaja[2,3], Jan Peter Boettcher[2,12], Mar Cabeza-Cabrerizo[2,3], Álvaro Camino-Sánchez[2], Lisa L. Carter[2,13], Juliane Doerrbecker[2,3], Theresa Enkirch[2,14], Isabel García-Dorival[2,15], Nicole Hetzelt[2,12], Julia Hinzmann[2,12], Tobias Holm[2,3], Liana Eleni Kafetzopoulou[2,16], Michel Koropogui[2,17], Abigael Kosgey[2,18], Eeva Kuisma[2,10], Christopher H. Logue[2,10], Antonio Mazzarelli[2,19], Sarah Meisel[2,3], Marc Mertens[2,20], Janine Michel[2,12], Didier Ngabo[2,10], Katja Nitzsche[2,3], Elisa Pallasch[2,3], Livia Victoria Patrono[2,3], Jasmine Portmann[2,21], Johanna Gabriella Repits[2,22], Natasha Y. Rickett[2,15,23], Andreas Sachse[2,12], Katrin Singethan[2,24], Inês Vitoriano[2,10], Rahel L. Yemanaberhan[2,3], Elsa G. Zekeng[2,15,23], Trina Racine[25], Alexander Bello[25], Amadou Alpha Sall[26], Ousmane Faye[26], Oumar Faye[26], N'Faly Magassouba[27], Cecelia V. Williams[28,29], Victoria Amburgey[28,29], Linda Winona[28,29], Emily Davis[29,30], Jon Gerlach[29,30]

Viruses > ssRNA viruses > ssRNA negative-strand viruses > Mononegavirales > Filoviridae >

Ebolavirus - 5 complete genomes

Retrieve sequences: -- Selec

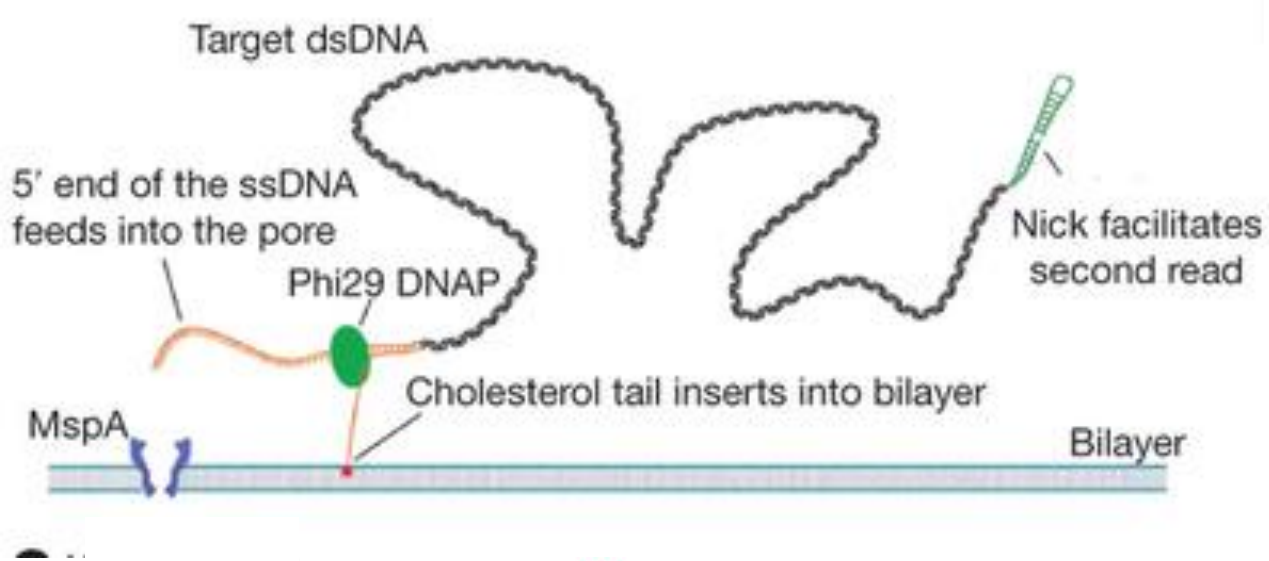
* The list view for each taxonomy node shows only the ne
* Unclassified/unassigned names are written in copper

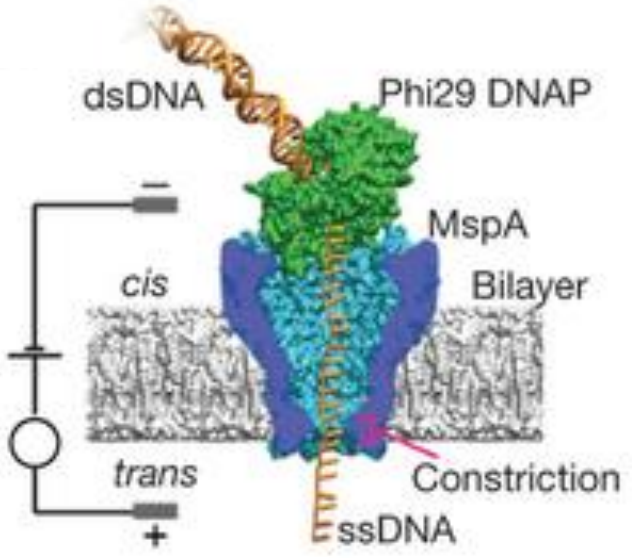
Sort the genomes list by: taxonomy Filter by host: all hosts

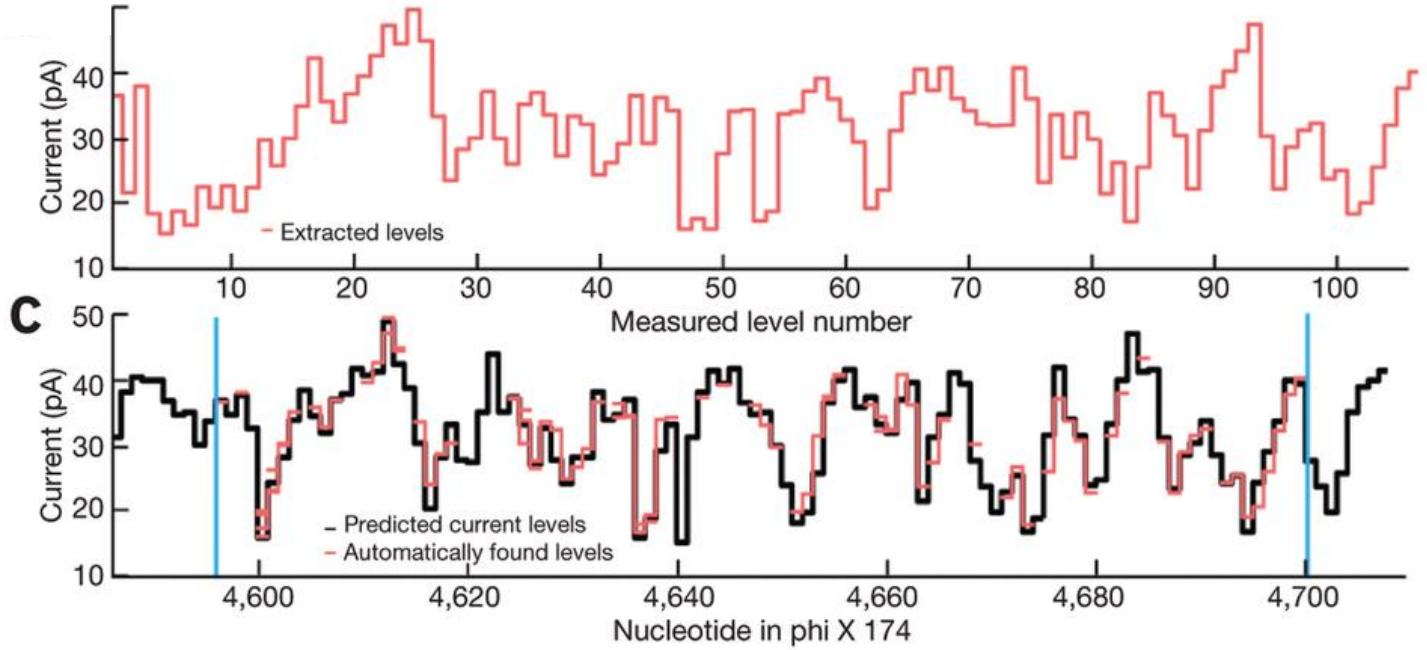
Genome	Accession	Source information	Segm	Length	Protein	Neighbors	Host
Bundibugyo virus	NC_014373	isolate:Bundibugyo virus/H.sapiens-tc/UGA/2007/Butalya-811250	-	18940 nt	9	6	vertebrates, human
Reston ebolavirus	NC_004161	isolate:Reston virus/M.fascicularis-tc/USA/1989/Philippines89-Pennsylvania	-	18891 nt	8	7	vertebrates, human
Sudan ebolavirus	NC_006432	isolate:Sudan virus/H.sapiens-tc/UGA/2000/Gulu-808892	-	18875 nt	8	12	vertebrates, human
Tai Forest ebolavirus	NC_014372	isolate:Tai Forest virus/H.sapiens-tc/CIV/1994/Pauleoula-CI	-	18935 nt	9	1	vertebrates, human
Zaire ebolavirus	NC_002549	isolate:Ebola virus/H.sapiens-tc/COD/1976/Yambuku-Mayinga	-	18959 nt	9	1103	vertebrates, human

Real-time, portable genome sequencing for Ebola surveillance.

Nanopore strand sequencing







Laszlo et al. Nature Biotechnology **2014.**

Human genome

nature biotechnology

OPEN

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

Miten Jain[1,13], Sergey Koren[2,13], Karen H Miga[1,13], Josh Quick[3,13], Arthur C Rand[1,13], Thomas A Sasani[4,5,13], John R Tyson[6,13], Andrew D Beggs[7], Alexander T Dilthey[2], Ian T Fiddes[1], Sunir Malla[8], Hannah Marriott[8], Tom Nieto[7], Justin O'Grady[9], Hugh E Olsen[1], Brent S Pedersen[4,5], Arang Rhie[2], Hollian Richardson[9], Aaron R Quinlan[4,5,10], Terrance P Snutch[6], Louise Tee[7], Benedict Paten[1], Adam M Phillippy[2], Jared T Simpson[11,12], Nicholas J Loman[3] & Matthew Loose[8]

We report the sequencing and assembly of a reference genome for the human GM12878 Utah/Ceph cell line using the MinION (Oxford Nanopore Technologies) nanopore sequencer. 91.2 Gb of sequence data, representing ~30× theoretical coverage, were produced. Reference-based alignment enabled detection of large structural variants and epigenetic modifications. *De novo* assembly of nanopore reads alone yielded a contiguous assembly (NG50 ~3 Mb). We developed a protocol to generate ultra-long reads (N50 > 100 kb, read lengths up to 882 kb). Incorporating an additional 5× coverage of these ultra-long reads more than doubled the assembly contiguity (NG50 ~6.4 Mb). The final assembled genome was 2,867 million bases in size, covering 85.8% of the reference. Assembly accuracy, after incorporating complementary short-read sequencing data, exceeded 99.8%. Ultra-long reads enabled assembly and phasing of the 4-Mb major histocompatibility complex (MHC) locus in its entirety, measurement of telomere repeat length, and closure of gaps in the reference human genome assembly GRCh38.

© Springer Nature. All rights reserved.

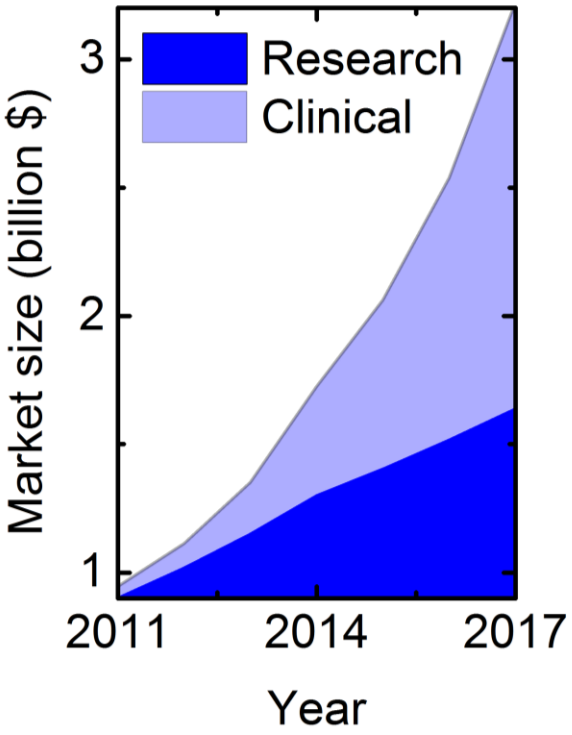
Sequencing applications

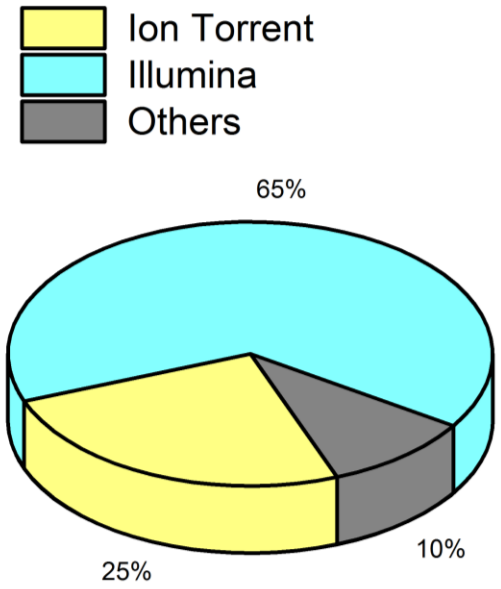
Personalized medicine- drug development

Cancer genomics

Microbial genetics – (Human microbiome)

Population genetics – evolutior





Steinbock and **Radenovic**, *Nanotechnology* **26** 074003, **2015**

Emerging nanopore applications beyond DNA sequencing

Protein Biomarker Analysis

Characterization, identification, and counting of individual protein molecules or nanoparticles

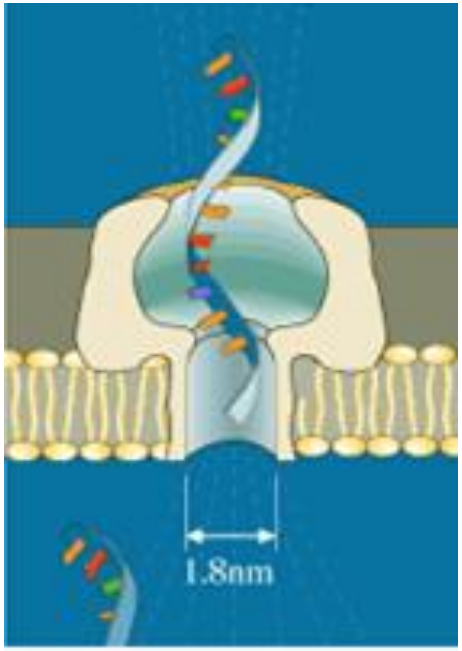
Peptide Sequencing

Water Desalination

Power Generation

Why Synthetic Pores? Disadvantages of protein pores

α hemolysin



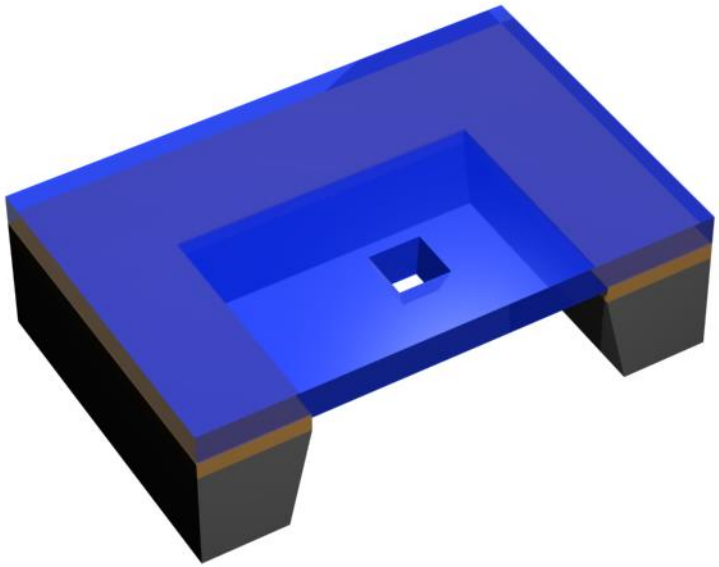
Fixed diameter

Typical velocity:
~1 base/μs=0.3 mm/s

Short life time

Some protein pore display t self gaiting

Synthetic



Tunable diameter

Typical velocity:
~3 base/μs=10 mm/s

Usable up to several days

Why Synthetic Pores?

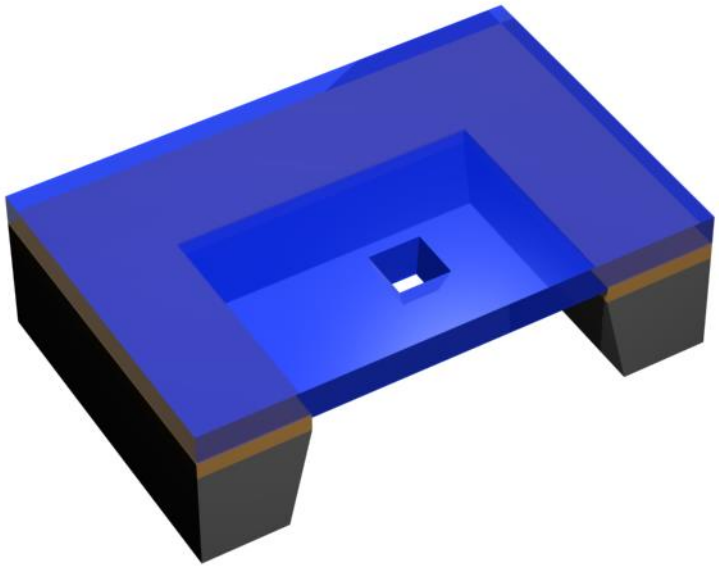
precise control over the location and chemical properties of the pores (pH, solvents, ionic strengths, oxidizers, etc.)

sufficiently **stiff** to permit high resolution distance detection and application of forces > 60 pN to the polymer (Physically robust (vibrations, pressure changes)

directly **compatible** with numerous detection schemes, including optical trapping, pA current measurements, and solid state detectors deposited near a pore

Tunable size and interface ,Fixed coordinates (pore position always the same)

Synthetic



Tunable diameter

Typical velocity:
~3 base/μs=10 mm/s

Usable up to several days

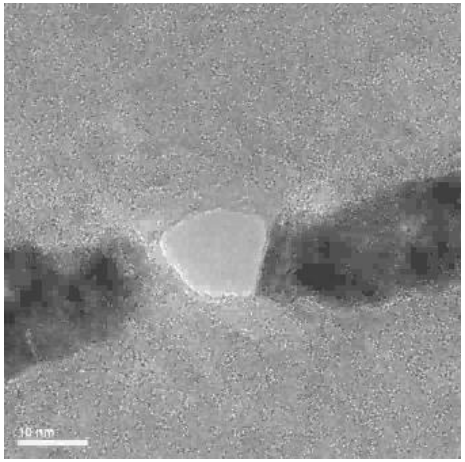
Nanopores - materi

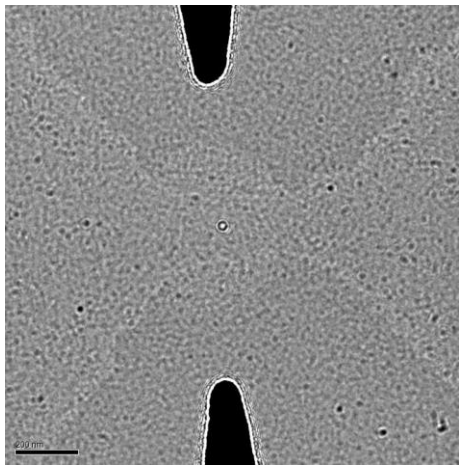
Nanopore material – dictates application

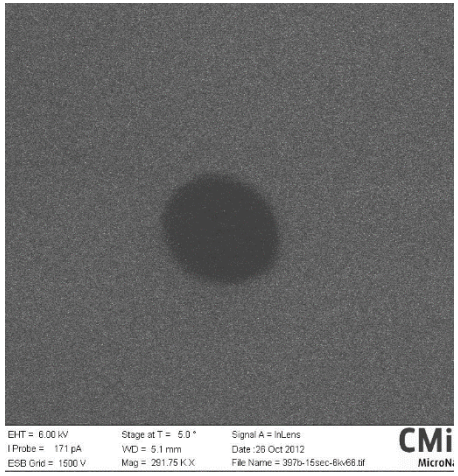
Silicon nitride nanopores – gold standard for solid state nanopores **5-20 nm thick**

2D material nanopores: graphene and molybdenum disulfide (MoS_2)-**0.3-0-7 nm thick**

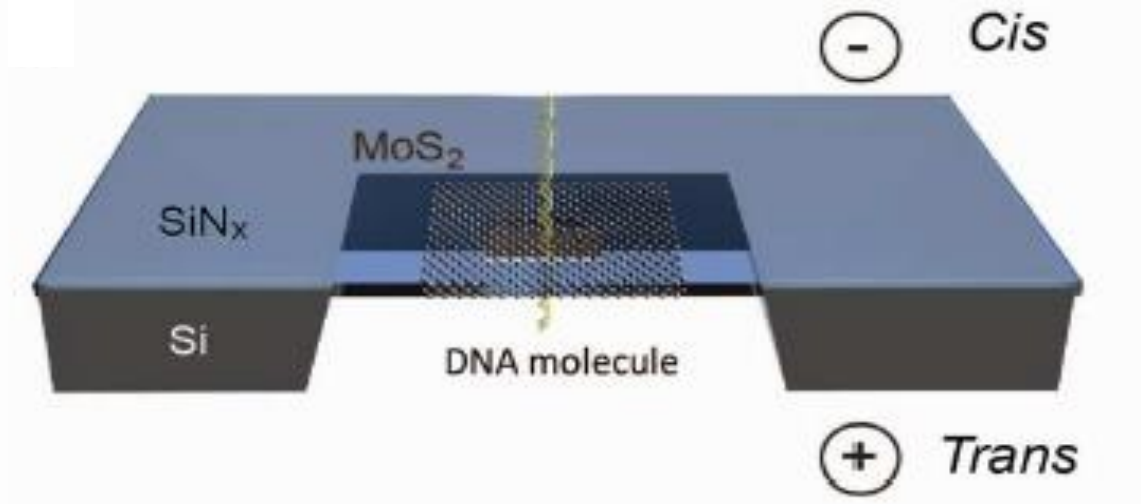
Glass/Quartz nanocapillaries







Nanopore sensitivity – 2D materials



- **Sensitivity** and **Selectivity:** Transverse current, monolayer (Graphene, **molybdenum disulfide MoS_2** other 2D materials or very thin nitride)
- **Statistics:** high throughput via multiplexing using concomitant detection with- for example integrated FETs
- **Speed** (time resolution) slowing down ultra fast translocation
- **Price** – low requires fabrication without TEM

Garaj S, et al. **2010** Nature, Merchant CA, et al. **2010** Nano Lett. Schneider GF, et al. **2010**. Nano Lett Liu, **Radenovic** et al. **2014** ACS Nano. Traversi ,**Radenovic** et al. Nature Nano. **2013**.

Research Groups

- D. Brantob, Jene Golovchenko , Harvard nanopore Group
- Dave Deamer, University of California at Santa Cruz
- Cees Dekker, Delft University of Technology
- Jiali Li, University of Arkansas
- Andre Marzialli, University of British Columbia
- Amit Meller Laboratory, Boston University
- Gregory Timp, Nano-Bio Group, University of Illinois, Urbana-Champaign
- Aleksei Aksimentiev, UIUC
  - Sequencing a DNA molecule using silicon integrated circuit nanopores.
- Rahid Bashir, Samir Iqbal ,Perdue University
- Jacob Schmidt , Bioengineering, UCLA