

# NGS: technologies and challenges

Olga Vinnere Pettersson, PhD  
Project Coordinator  
NGI/SciLifeLab, Uppsala Node



@OlgaVPettersson

# Outline

- INTRO: Sequencing service at NGI-SciLifeLab
- NGS general knowledge: current technologies
- NGS Challenges:
  - Sequencing artefacts
  - NGS sample quality requirements
- Next global challenge: the Earth Biogenome Project



SciLifeLab

# Operational principles of NGI

## User community

- Open to all Swedish academic scientists on equal terms.
- Consultation and introduction of new protocols.
- Workshops, courses, seminars.

## Cost basis

- Academic users of NGI only cover agent cost.
- Staff salaries at NGI covered by SciLifeLab, VR, and host universities.
- Premium and service contracts covered by SciLifeLab, VR, KAW and host universities.
- Capital equipment covered by KAW, VR, SciLifeLab.

## Quality

- Emphasis on data quality and needs of the users.
- Illumina sequencing and genotyping processes accredited by SWEDAC, ISO/IEC 17025
- Ion and PacBio: accreditation pending

We are non-profit  
We have technology and knowledge  
We want to help you to do GREAT  
research  
We do not want co-authorship  
Let us help YOU



# NGI Support

## Pre-sequencing

- **Project design** via discussions with users
- **Advise** in sample collection and preparation
- Case-to-case **DNA extraction service**

## Post-sequencing:

- Control over produced data: making sure data meet our **high standards** in terms of quality and yield.
- Primary **analysis of human genomes** is enabled
- **Genome assembly** of PacBio data is offered as a service
- Data is delivered to **UPPMAX** (Uppsala Multidisciplinary Center for Advanced Computational Science)

## Collaborative projects

 for technology and method development

## Education

# Current Technologies



Current leader on the NGS market

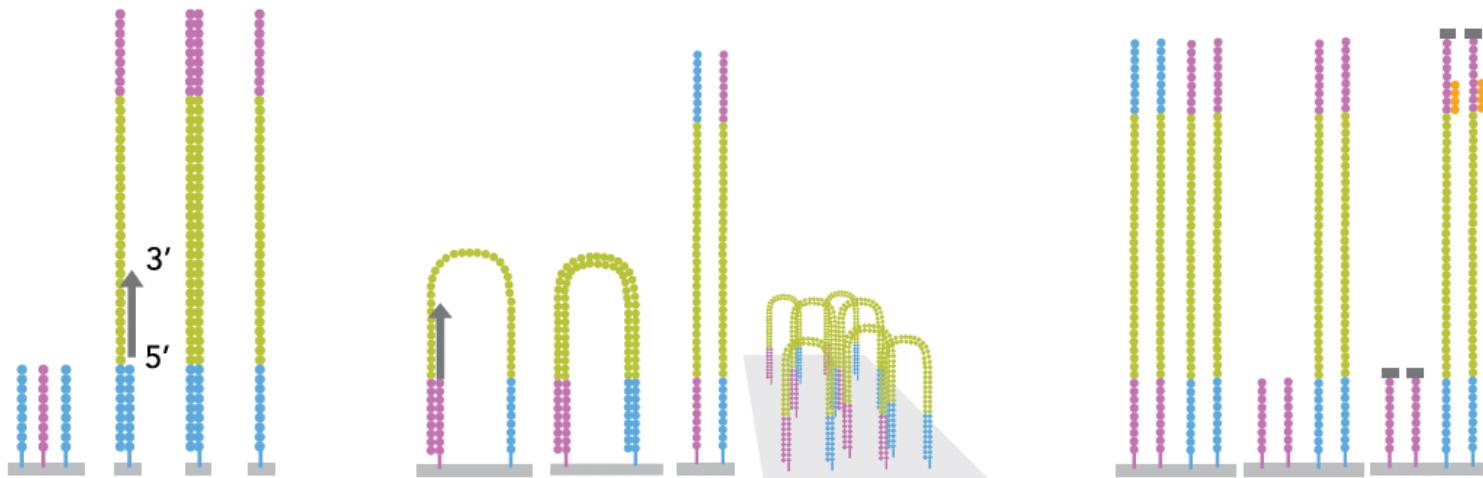
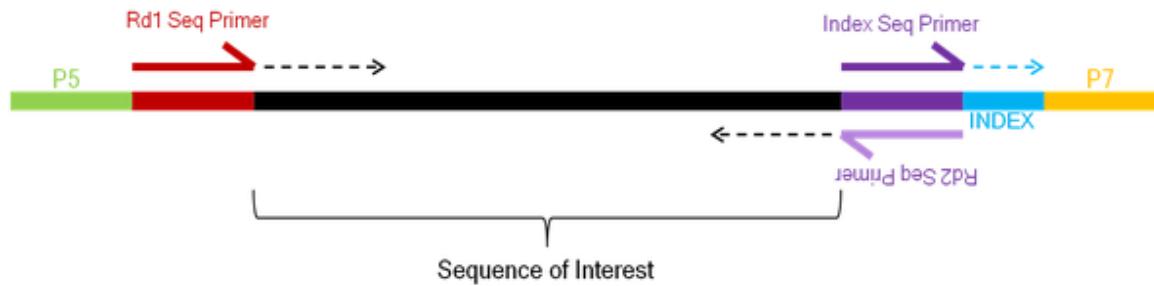
Instrument	Run time	Max output	Max reads/run	Max read length
iSeq	9.5 – 19 hrs	1.2 Gb	4 mln	PE 150
MiniSeq	4-24 hrs	7.5 Gb	25 mln	PE 150
MiSeq	4-55 hours	15 Gb	25 mln	PE 300
NextSeq series	12-48 hours	120-300 Gb	0.4 – 1 bln	PE 150
NovaSeq 6000	13-44 hours	6 Tb	20 bln	PE 250

RIP: HiSeq 2500 & HiSeq X

Used for everything



# Illumina: bridge amplification





## Ion S5 XL

Chip:	Run time	Output	Max reads/run	Max read length
510	2.5-4 hrs	0.3 - 0.5 Gb	2-3 mln	SE 400 bp
520	2.5-4 hrs	0.6-2 Gb	3-6 mln	SE 600 bp
530	2.5-4 hrs	3-8 Gb	15-20 mln	SE 600 bp
540	2.5-4 hrs	10-15 Gb	60-80 mln	SE 400 bp
550	2.5-4 hrs	18-20 Gb	100-130 mln	SE 200 bp

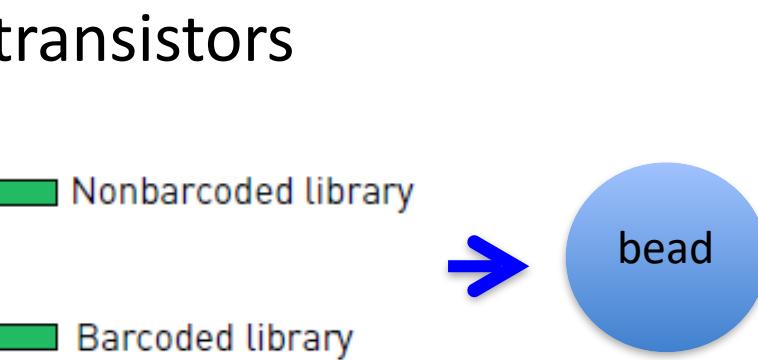
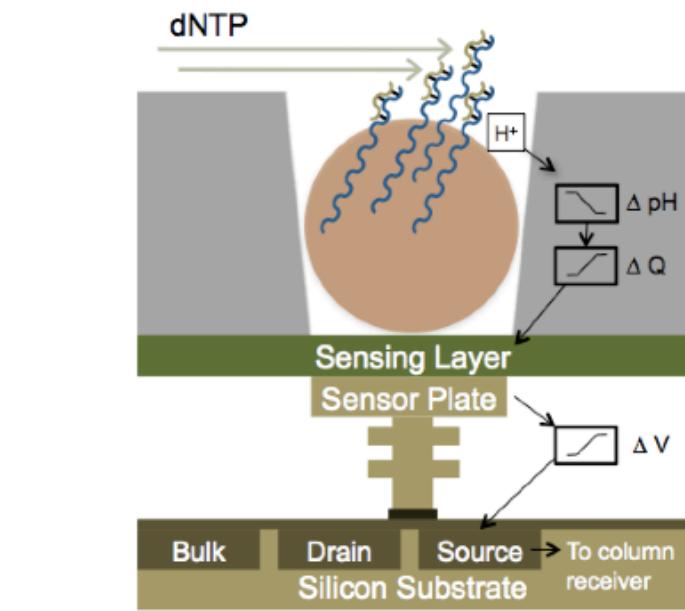
RIP: IonTorrent PGM, IonProton

Clinical applications mainly  
Standard analysis directly on the instrument

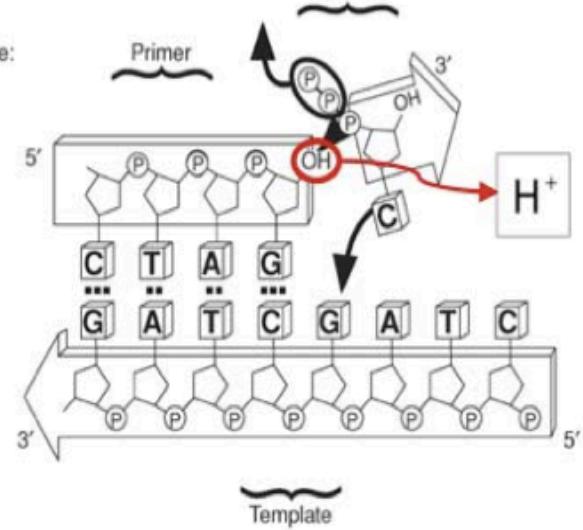
Multiplex-PCR panels



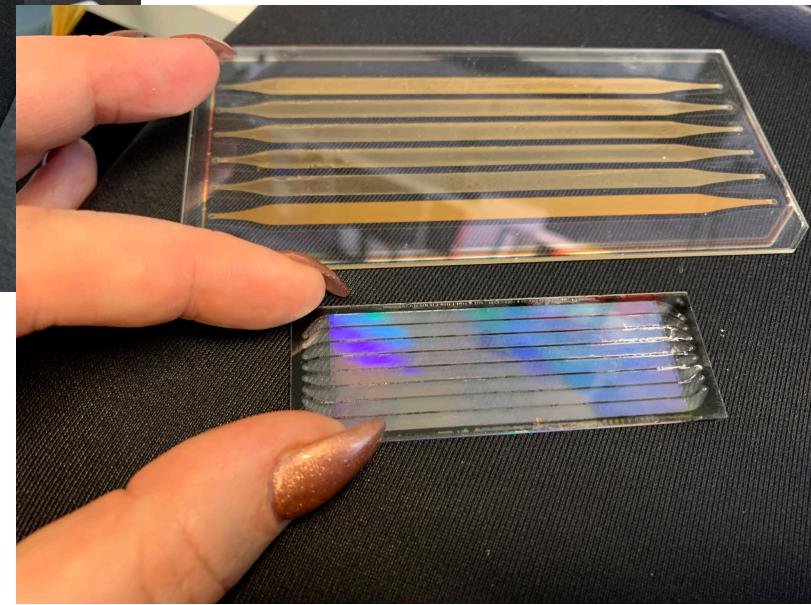
# Ion Torrent: H<sup>+</sup> ion-sensitive field effect transistors



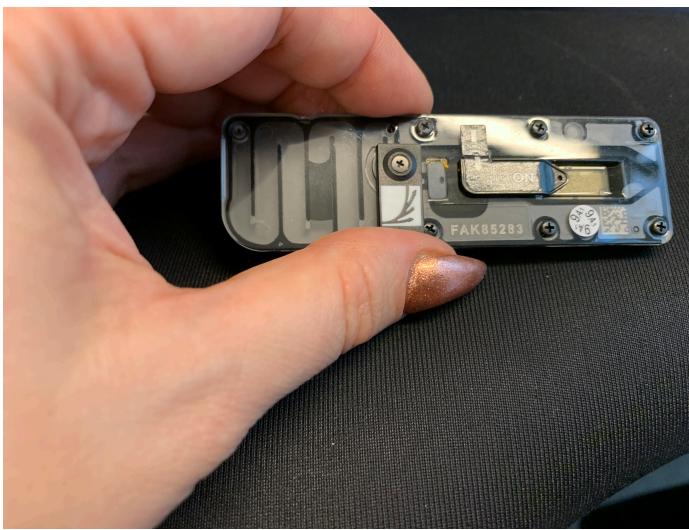
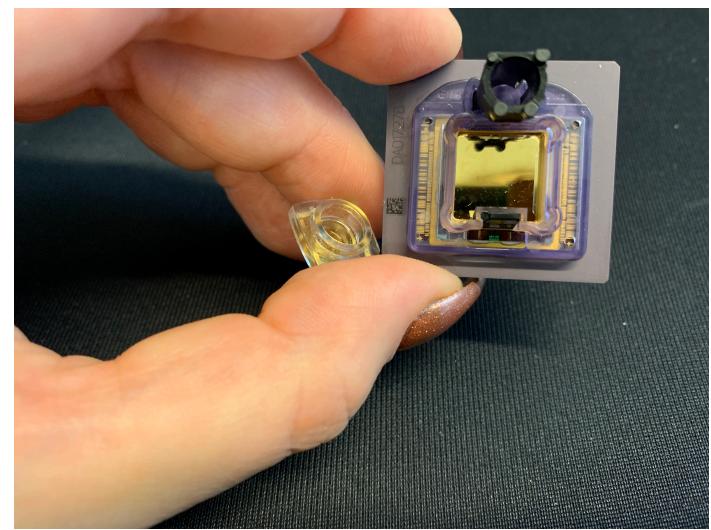
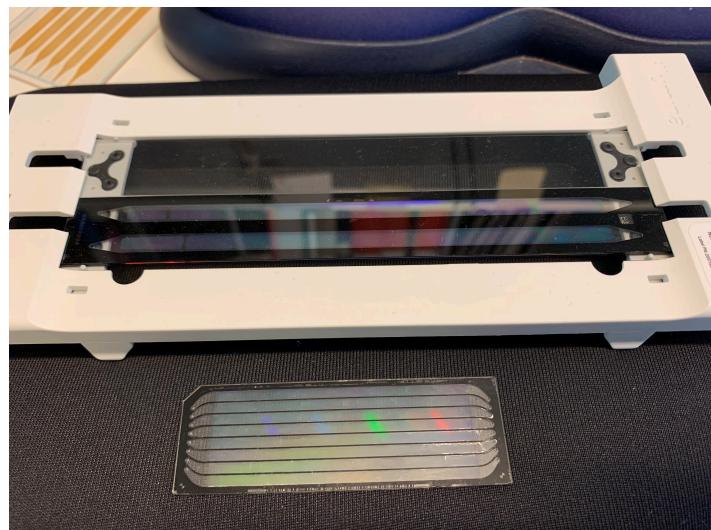
Example:



# How it looked yesterday



# How it looks like now





PACBIO®



Instrument	Run time /SMRT	Output /SMRT	Max reads / SMRT	Max read length*
RSII	30 min – 6 hrs	500 Mb – 2 Gb	50 000	40 kb
Sequel	30 min – 20 hrs	2 – 35 Gb	200 000	60 kb
Sequel II				
HiFi	30 hrs	320 Gb	4 mln	20 kb
CLR	15 hrs	300 Gb	3 mln	120 kb

Single Molecule Real Time sequencing: SMRT

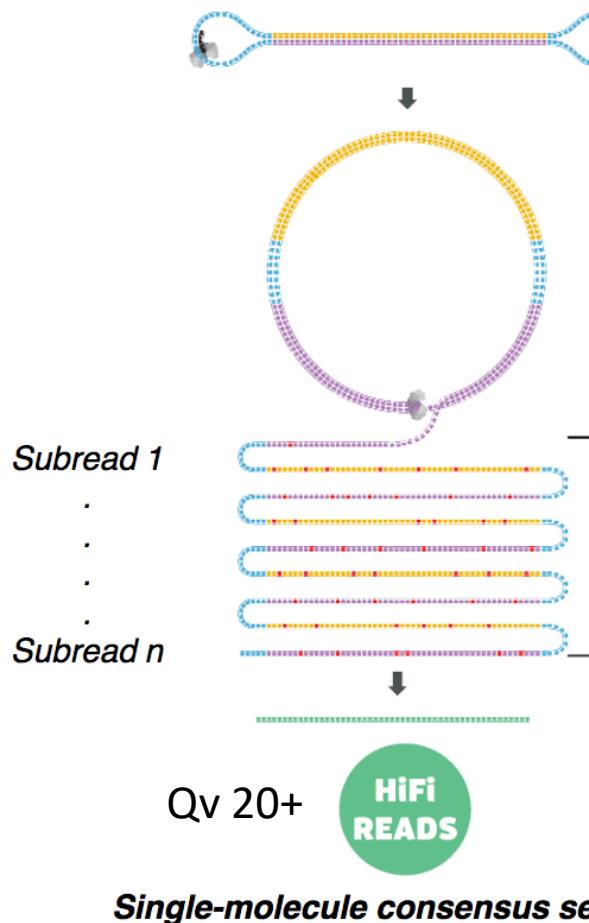




## TWO MODES OF SMRT SEQUENCING

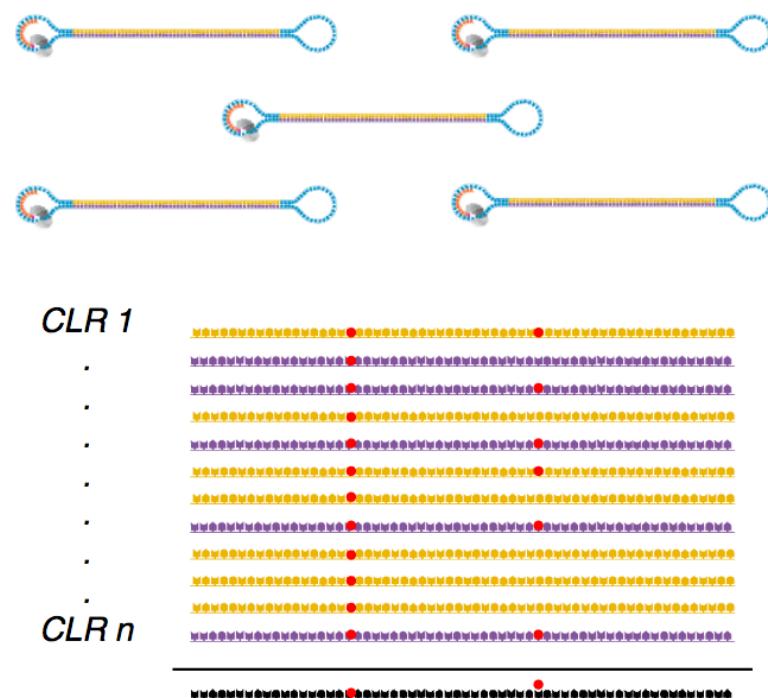
### Circular Consensus Sequencing (CCS) Mode

Inserts 10-20 kb

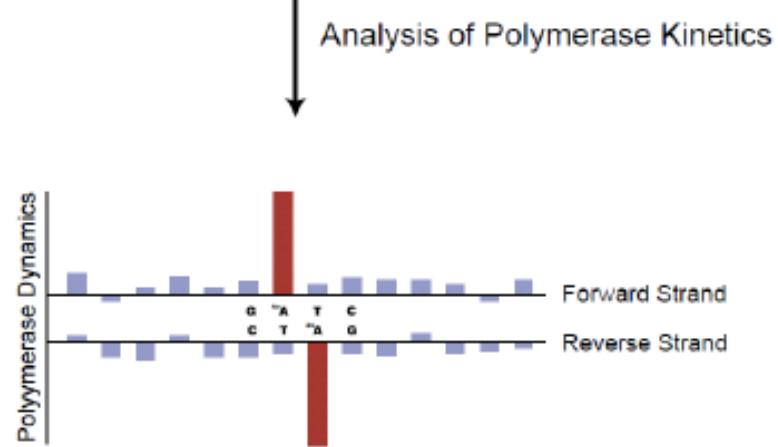
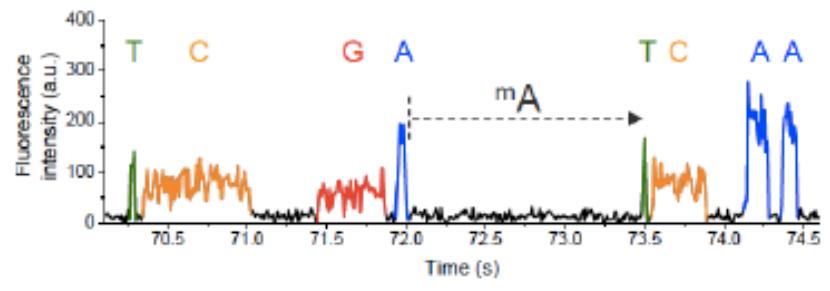
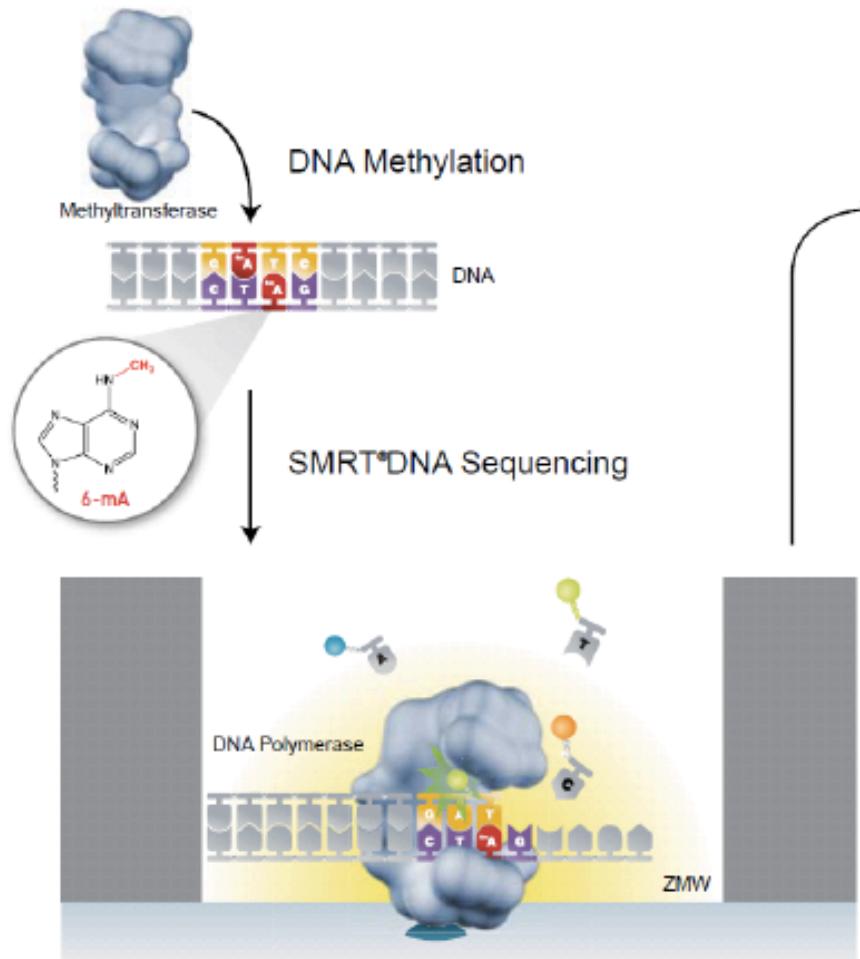


### Continuous Long Read (CLR) Sequencing Mode

Inserts >25 kb, up to 175 kb



# Base Modification: Discover the Epigenome



Detect base modifications using the kinetics of the polymerization reaction during normal sequencing



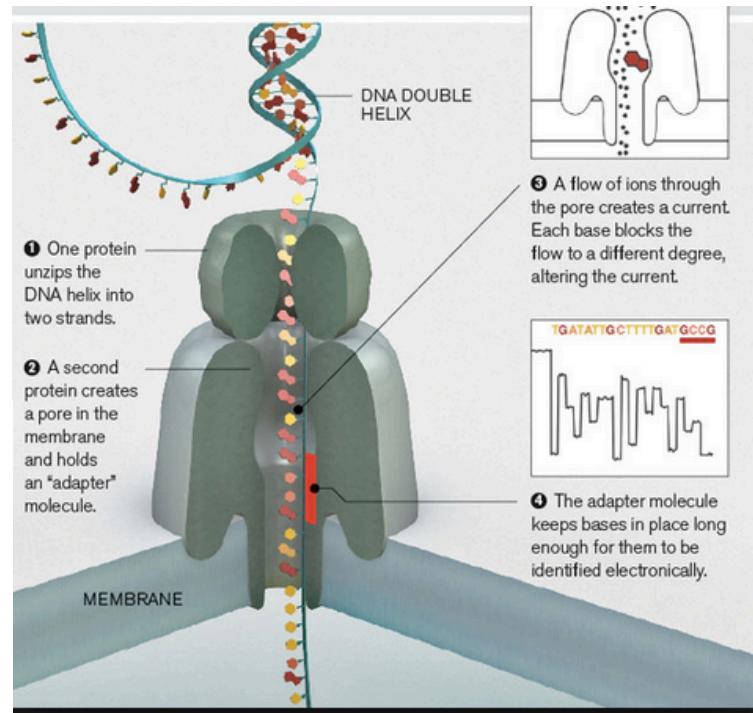
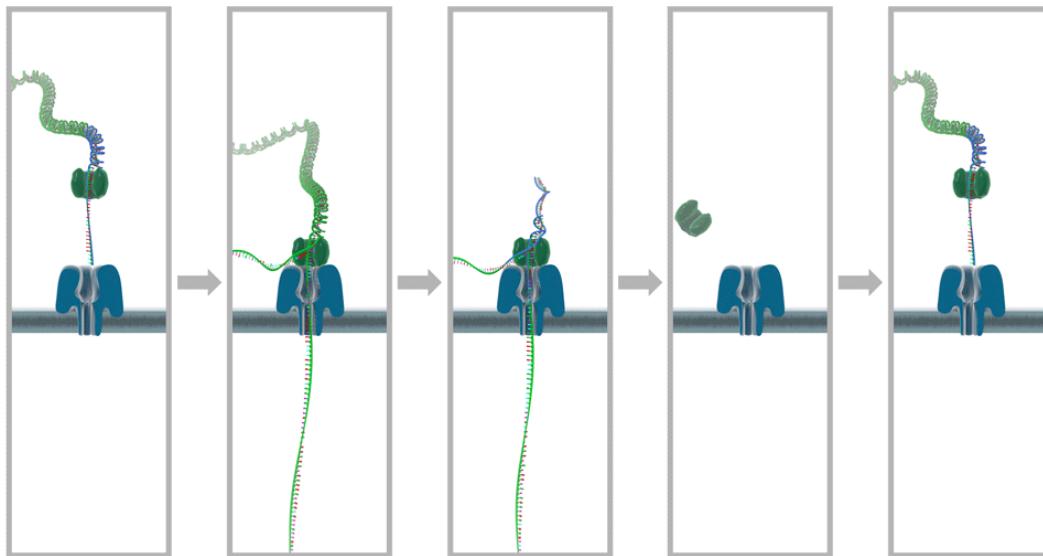
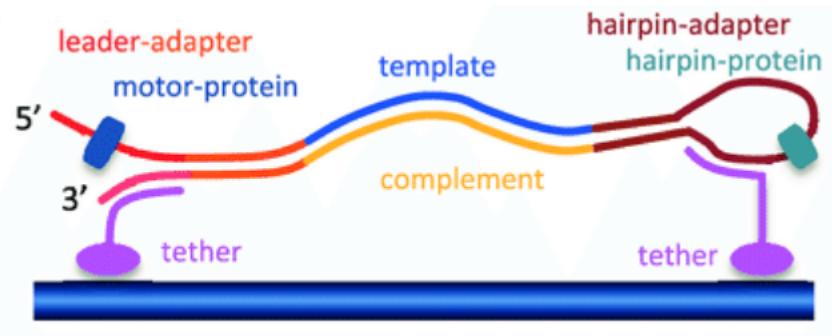
Instrument	Run time / FC	Output / FC	Nr of pores	Max read length
Flongle	16 hrs	1 Gb	126	1 Mb
MinION	24 hrs	2-15 Gb	512	1 Mb
GridION	24 hrs	2-15 Gb	512	1 Mb
PromethION	72 hrs	10 – 150 Gb	3 000	2 Mb

Q&A: "*It depends*"...





# ONT: DNA + Motor + Pore



# Main advantages of ONT: SPEED and PORTABILITY

## Rapid Confirmation of the Zaire Ebola Virus in the Outbreak of the Equateur Province in the Democratic Republic of Congo: Implications for Public Health Interventions

Placide Mbala-Kingebeni, Christian-Julian Villabona-Arenas, Nicole Vidal, Jacques Likofata, Justus Nsio-Mbeta, Sheila Makiala-Mandanda, Daniel Mukadi, Patrick Mukadi, Charles Kumakamba, Bathe Djokolo ... Show more

Clinical Infectious Diseases, Volume 68, Issue 2, 15 January 2019, Pages 330–333, <https://doi.org/10.1093/cid/ciy527>

Published: 29 June 2018 Article history ▾

ORIGINAL ARTICLE BRIEF REPORT

## A Novel Coronavirus from Patients with Pneumonia in China, 2019

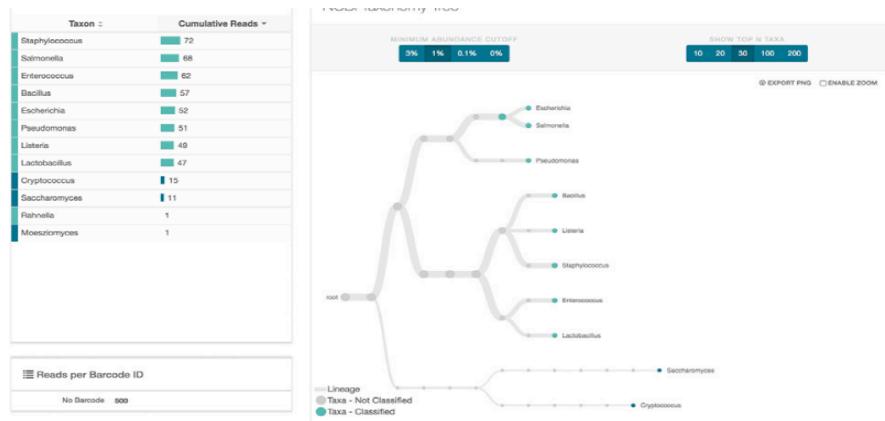
Na Zhu, Ph.D., Dingyu Zhang, M.D., Wenling Wang, Ph.D., Xinwang Li, M.D., Bo Yang, M.S., Jingdong Song, Ph.D., Xiang Zhao, Ph.D., Baoying Huang, Ph.D., Weifeng Shi, Ph.D., Roujian Lu, M.D., Peihua Niu, Ph.D., Faxian Zhan, Ph.D., et al., for the China Novel Coronavirus Investigating and Research Team

RESEARCH ARTICLE |  Full Access |

## Semi-quantitative characterisation of mixed pollen samples using MinION sequencing and Reverse Metagenomics (RevMet)

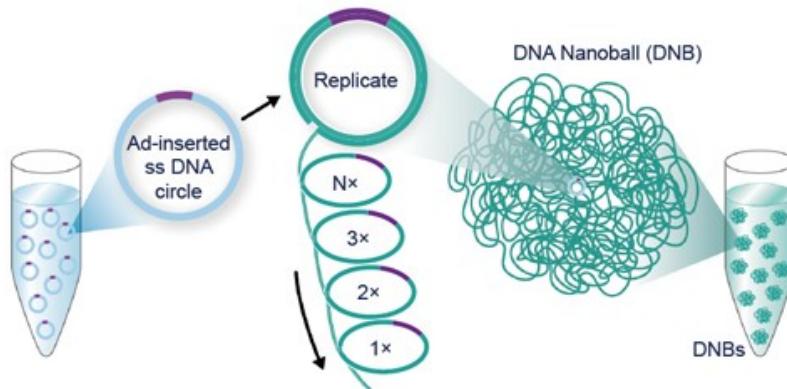
Ned Peel, Lynn V. Dicks, Matthew D. Clark, Darren Heavens, Lawrence Percival-Alwyn, Chris Cooper, Richard G. Davies, Richard M. Leggett, Douglas W. Yu 

First published: 15 July 2019 | <https://doi.org/10.1111/2041-210X.13265>



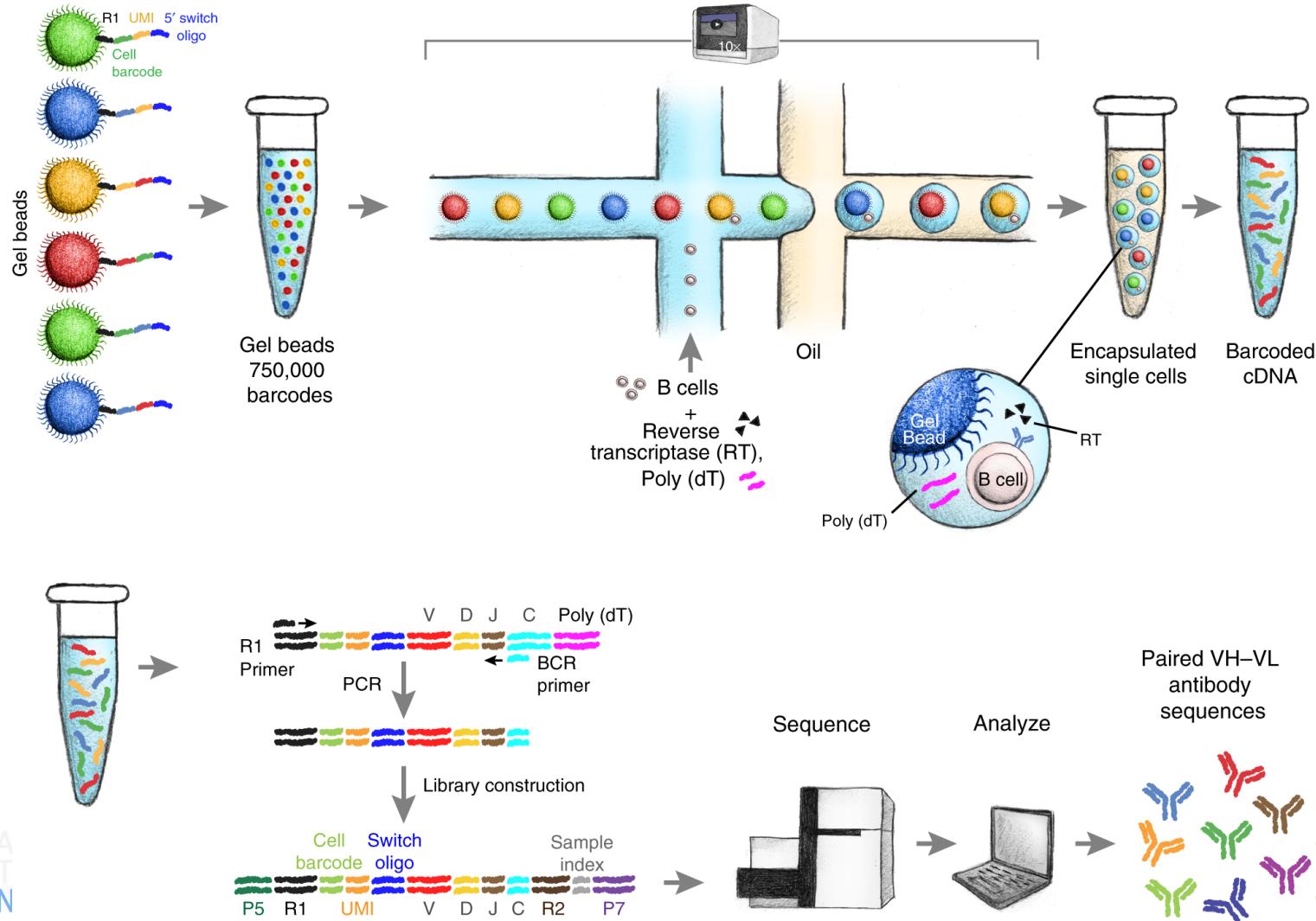
# Other technologies & methods

New Sequencing technology coming to NGI for *evaluation*:

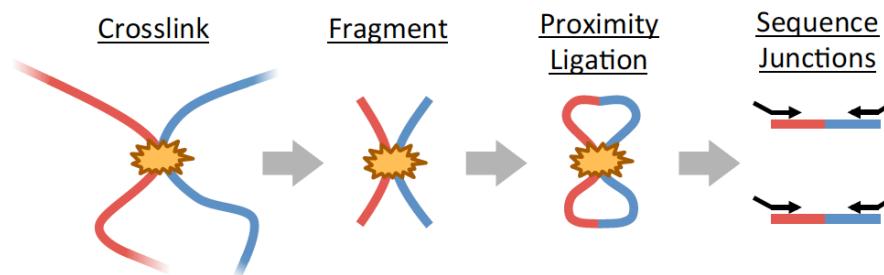
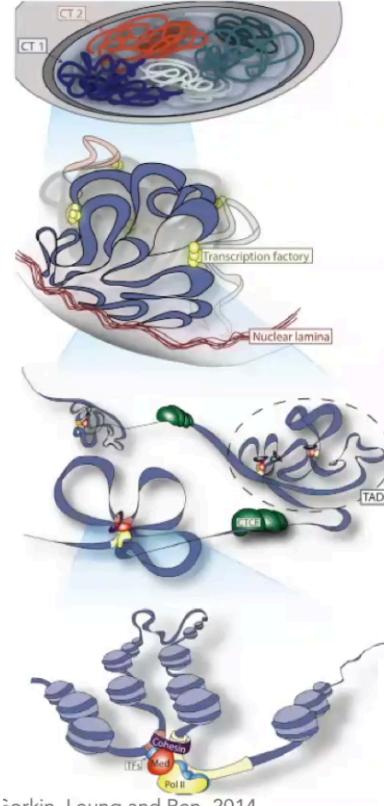


Output similar to Illumina

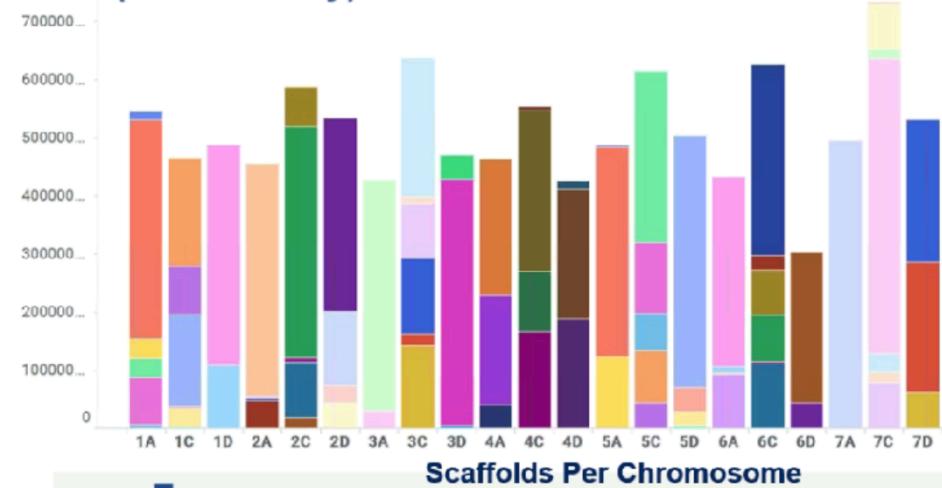
# 10x Genomics (Chromium)



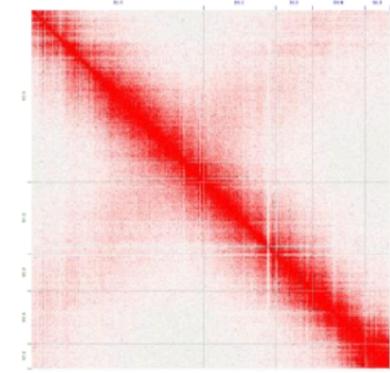
# Hi-C / OmniC: linking reads to chromosomes



(v1 assembly)



Chr 5C (612 Mb)



Start with a tissue!

Capture DNA bound to the same nucleosome

Make a library and sequence on Illumina NovaSeq

# Technologies and Applications at NGI



## NGS technologies

### Short read NGS



### Long-read NGS



Whole genome re-sequencing  
RNA-seq  
Exome  
Targeted re-seq  
Panels  
Amplicons up to 600 bp

*De novo* genome sequencing  
Whole-transcript sequencing  
Structural variant resolving  
Allele phasing  
Targeted re-seq  
Amplicons up to 13 kb

## Research and development

# NGS Technologies: SUMMARY

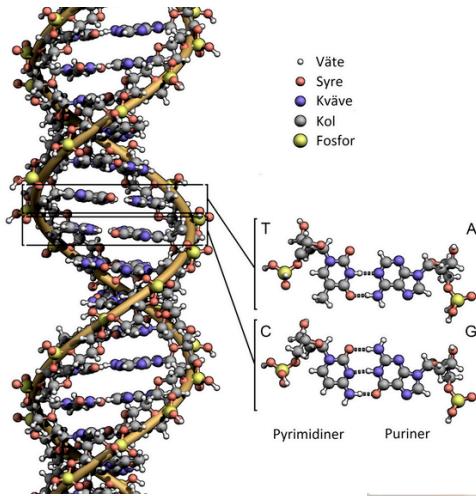
- Development goes VERY FAST
- All technologies have their PROs and CONs
- One technology does not suit all the applications
- In some projects, several technologies should be combined

# Making sense of genomics data:

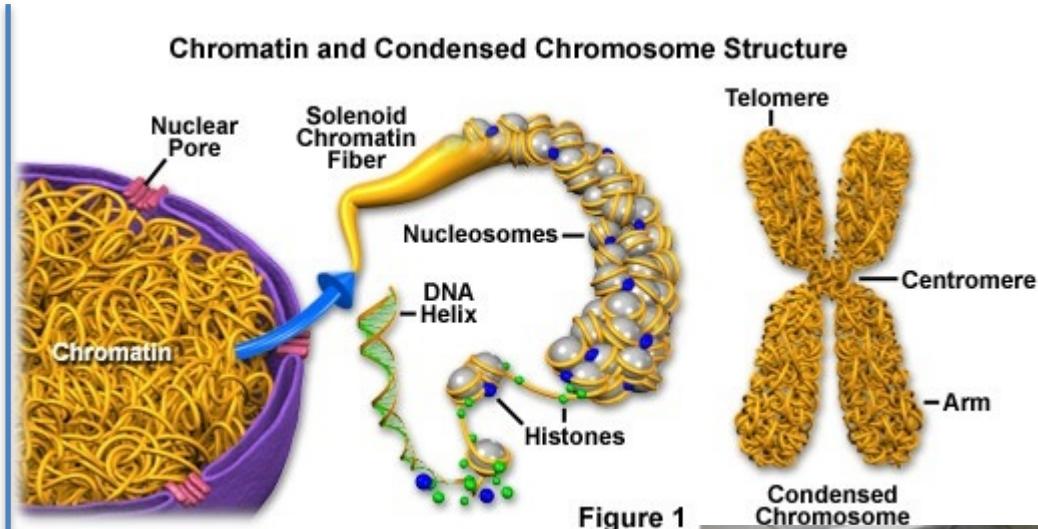
## **Understanding sequencing bias**

# Sequencing artefacts: what are they?

Sequencing a representative, completely randomized subsample:  
it starts with input material



What textbook  
tells you



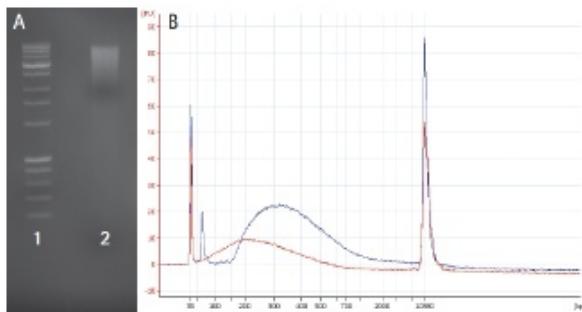
Brutal reality

*Do not forget:  
DNA in solution behaves differently*

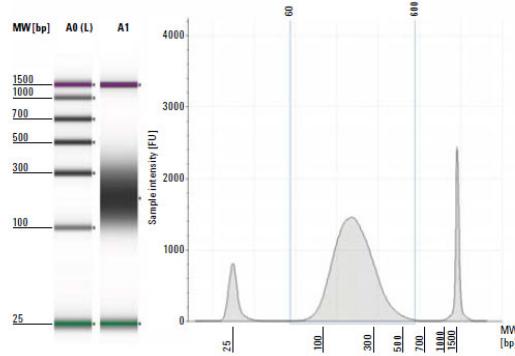


# Sequencing artifacts: what are they?

Sequencing a representative, completely randomized subsample:  
continues with library preparation

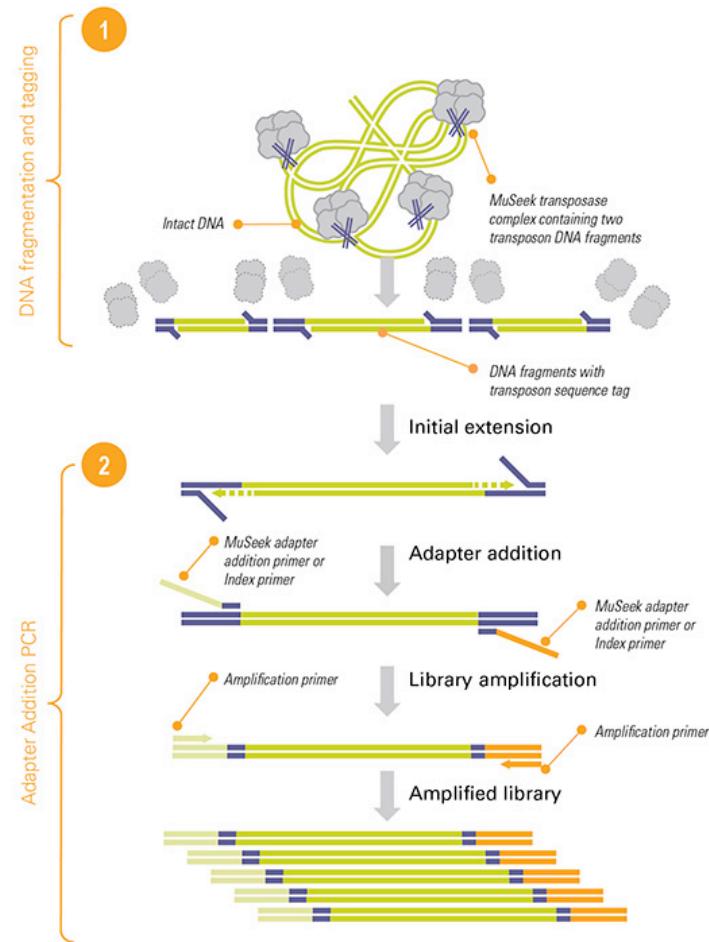


Input sample



Shearing and size-selection

Loosing molecules all the way



Less material -> more amplification cycles

# Sequencing artefacts: what are they?

## PCR bias – important source of sequencing artefacts

**PCR steps involved in any NGS but PacBio and Oxford Nanopore:**

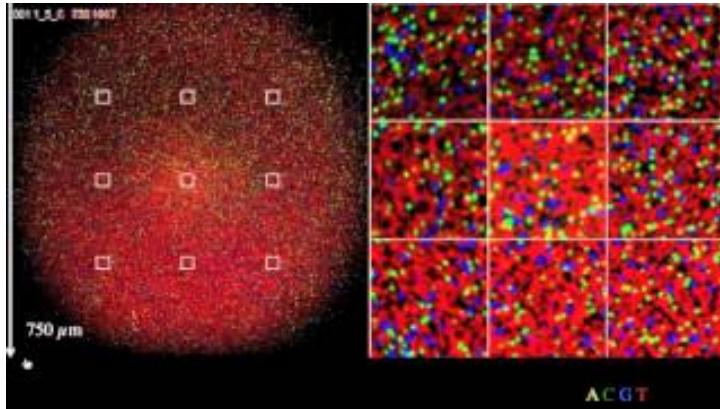
1. Library amplification
2. Amplification during templating (Illumina – on glass; Ion – emPCR)

**Main PCR bias:**

1. Size: shorter fragments amplify faster -> higher sequencing signal and coverage
2. Polymerase errors
  - slippage in low complexity regions
  - incorporation of erroneous bases & indels
3. GC-bias (fragments with high GC diminish to 1/10<sup>th</sup> from initial amount)

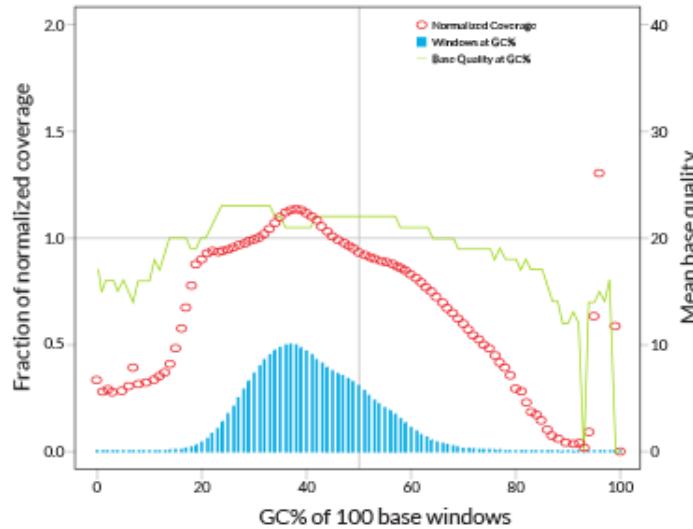
# Sequencing artefacts: what are they?

## PCR bias – important source of sequencing artefacts

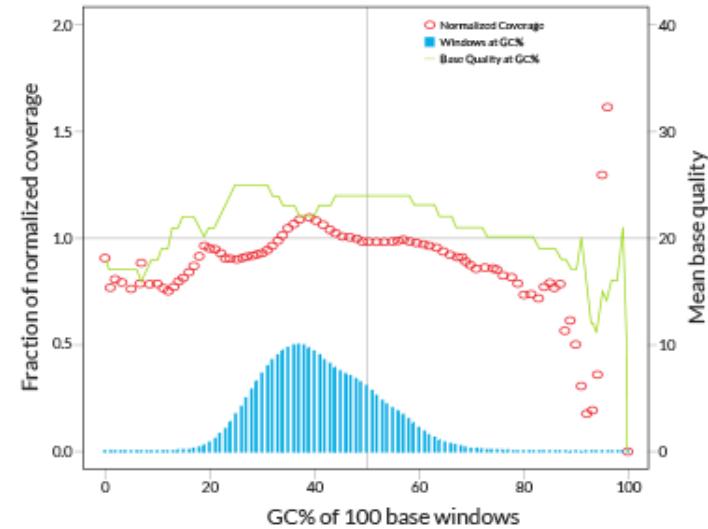


Clusters with shorter fragments grow faster -> quality signal from smaller clusters worsens

### GC bias & genome coverage



Heavily amplified library



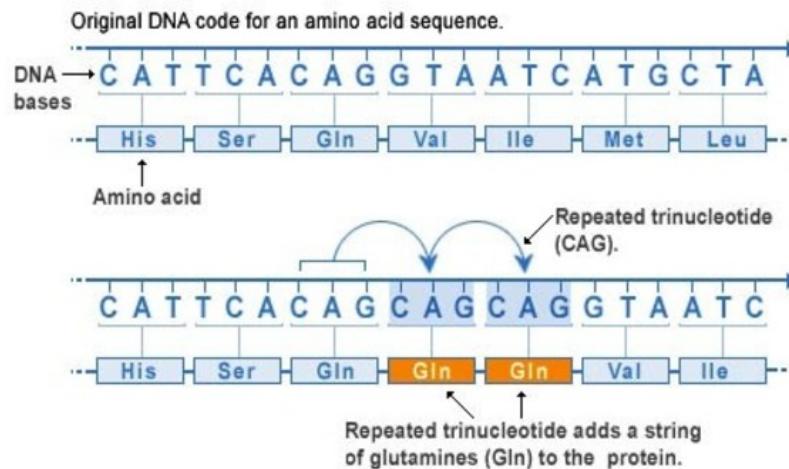
PCR-free library

# Sequencing artefacts: what are they?

## PCR bias – important source of sequencing artefacts

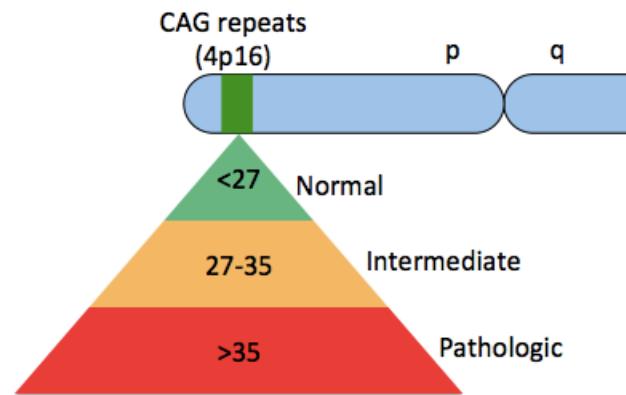
Polymerase slippage – low complexity regions

Repeat expansion mutation



U.S. National Library of Medicine

Huntington's Disease



**Huntington's disease:**

- Inherited disorder resulting in brain cell death
- Decline of motoric and cognitive functions
- Common onset: 30-50 years of age
- No cure
- Causative genetic variant: CAG-repeat expansion in *HTT* gene

# Sequencing bias: SUMMARY

- Keep in mind that they are there
- Coverage varies across the genome
- One technology does not suit all the applications

# BREAK

# SAMPLE QUALITY REQUIREMENTS

**Garbage in – garbage out:**

Sequencing success always depends on the  
**sample quality.**

**NGS-quality DNA and  
PCR-quality DNA  
are two completely different things.**

# THE NUCLEIC ACIDS

Chemistry and Biology

Edited by

ERWIN CHARGAFF  
*Department of Biochemistry  
Columbia University  
New York, N. Y.*

J. N. DAVIDSON  
*Department of Biochemistry  
University of Glasgow  
Glasgow, Scotland*

Volume I

## a. Extraction with Strong Salt Solution, Deproteinization with Chloroform

(1) *Sodium Deoxyribonucleate of Calf Thymus.*<sup>98</sup> Fresh frozen calf thymus glands (54.5 kg.) were minced and suspended in 0.9% sodium chloride (54 l.) and milled to produce a fine suspension. This suspension was centrifuged (6500 r.p.m.) and the solid material resuspended in 0.9% sodium chloride (45.5 l.) and milled and centrifuged as before. The tissues, which were now free of material containing pentose, were suspended in 10% sodium chloride (214 l.) with vigorous mechanical stirring at 0°. At this stage the viscosity of the solution increased considerably. After extraction at 0° for 48 hours, the insoluble material was removed by centrifuging (6300 r.p.m.) and the deoxypentose nucleoprotein precipitated from the resultant solution (pH 6.5) by the addition of an equal volume of industrial methanol. The precipitated solid was washed with 70%, then 100% industrial methanol and dried in a vacuum at room temperature. Yield, 1.69 kg. of a very slightly yellow fibrous solid.

A general method for isolation of high molecular weight DNA from eukaryotes

---

Nikolaus Blin and Darrel W. Stafford

---

Department of Zoology, University of North Carolina, Chapel Hill, NC 27514, USA

---

Received 24 June 1976

---

### ABSTRACT

A new method for isolation of high molecular weight DNA from eukaryotes is presented. This procedure allows preparation of DNA from a variety of tissues such as calf thymus or human placenta and from cells which were more difficult to lyse until now (e.g. *Cryptocodinium cuhnii*, a dinoflagellate). The DNA obtained in such a way has an average molecular weight of about  $200 \times 10^6$  d and contains very few, if any, single strand breaks.

### INTRODUCTION

Isolation of large quantities of nick-free, high molecular weight DNA from eukaryotic organisms has heretofore presented considerable technical difficulties. DNA prepared by conventional techniques has been a heterogeneous population of molecules ranging in molecular weight from  $10 \times 10^6$  to  $20 \times 10^6$  d (1, 2). The single strand molecular weight was often around

## THE PREPARATION OF DEOXYRIBONUCLEIC ACIDS BY THE $\phi$ -AMINOSALICYLATE-PHENOL METHOD

K. S. KIRBY

*Chester Beatty Research Institute, Institute of Cancer Research,  
Royal Cancer Hospital, London (Great Britain)*

(Received February 17th, 1959)

# 1983: PCR



Journal of Microbiological Methods

Volume 19, Issue 3, March 1994, Pages 167-172



Protocol | Published: November 1990

## A rapid and inexpensive method for isolation of total DNA from dehydrated plant tissue

Thomas H. Tai & Steven D. Tanksley

*Plant Molecular Biology Reporter* 8, 297-303(1990) | [Cite this article](#)

1176 Accesses | 183 Citations | 3 Altmetric | [Metrics](#)

## A general method for the extraction of DNA from bacteria

Michael W Lema, Arnold Brown , Jo H Calkins

Show more

[https://doi.org/10.1016/0167-7012\(94\)90066-3](https://doi.org/10.1016/0167-7012(94)90066-3)

## A simple, rapid, inexpensive and widely applicable technique for purifying plant DNA

S Gilmore, PH Weston and JA Thomson

*Australian Systematic Botany* 6(2) 139 - 148

Published: 1993

## Simple, Efficient, and Nondestructive DNA Extraction Protocol for Arthropods

Aloysius J. Phillips, Chris Simon

*Annals of the Entomological Society of America*, Volume 88, Issue 3, 1 May 1995,

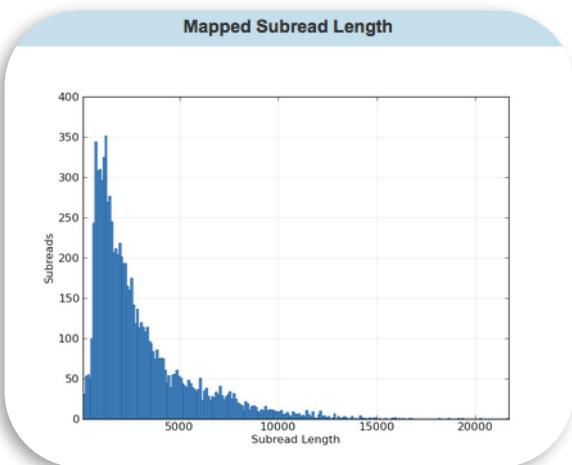
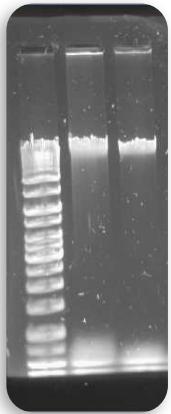
Pages 281–283, <https://doi.org/10.1093/aesa/88.3.281>

Published: 01 May 1995 [Article history](#) ▾

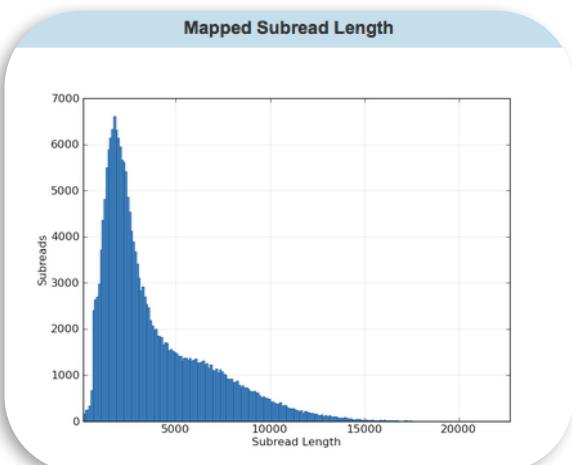


# 2013: a wake-up call

---



Polished Contigs	223	Max Contig Length	36,298
N50 Contig Length	2,932	Sum of Contig Lengths	480,087



Polished Contigs	9	Max Contig Length	1,508,929
N50 Contig Length	1,353,702	Sum of Contig Lengths	7,813,244



For Long Reads one needs to have *long and pure DNA*

SciLifeLab

# DNA quality and inhibition of sequencing

Short-read technologies: PCR inhibition

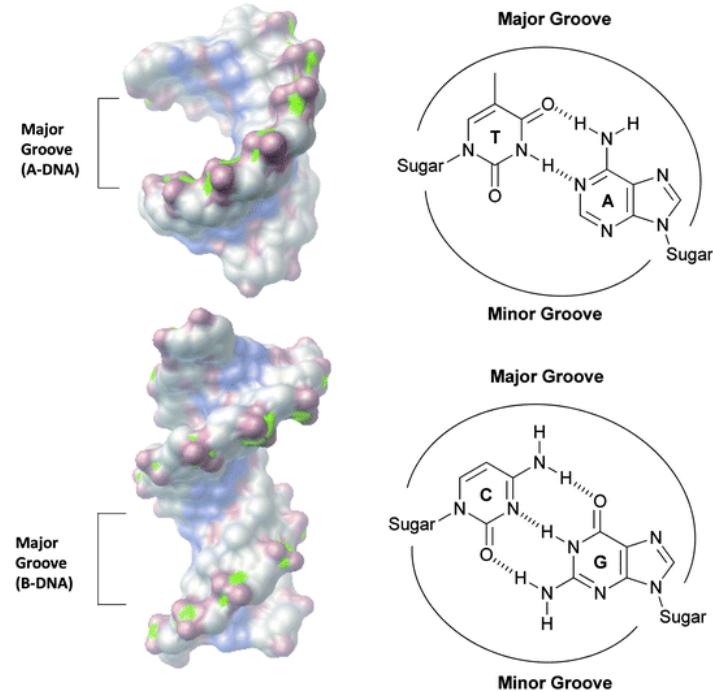
Long-read technologies are PCR-free, but one sequences native DNA “as is”.

## DNA-binders:

- Proteins
- Polyphenols
- Secondary metabolites (e.g. toxins)
- Pigments
- Polysaccharides

## Polymerase inhibitors:

- Salts
- Phenol
- Alcohols



*Hamilton & Arya, Nat. Prod. Rep., 2012, 29, 134-143*

## Physical inhibiting factors – debris

# What do absorption ratios tell us?

## Pure DNA 260/280: 1.8 – 2.0

< 1.8:

Too little DNA compared to other components of the solution; presence of organic contaminants: proteins and phenol; glycogen - **absorb at 280 nm**.

> 2.0:

High share of RNA.

## Pure DNA 260/230: 2.0 – 2.2

<2.0:

Salt contamination, humic acids, peptides, aromatic compounds, polyphenols, urea, guanidine, thiocyanates (latter three are common kit components) – **absorb at 230 nm**.

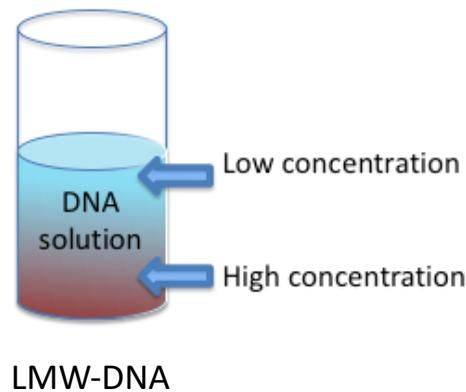
>2.2:

High share of RNA, very high share of phenol, **high turbidity**, dirty instrument, wrong blank.

*Photometrically active contaminants:  
phenol, polyphenols, EDTA, thiocyanate, protein,  
RNA, nucleotides (fragments below 5 bp)*

# How to make a correct DNA measurement

- Thaw DNA completely
- Mix gently (**never vortex!**)
- Put the sample on a thermoblock: 37°C, 15-30 min
- Mix gently
- **Dilute 1:100 (if HMW)**
- Mix gently
- Make a measurement with an appropriate blank
  
- **NANODROP is Bad. Point.**
- Use Qubit, or PicoGreen.



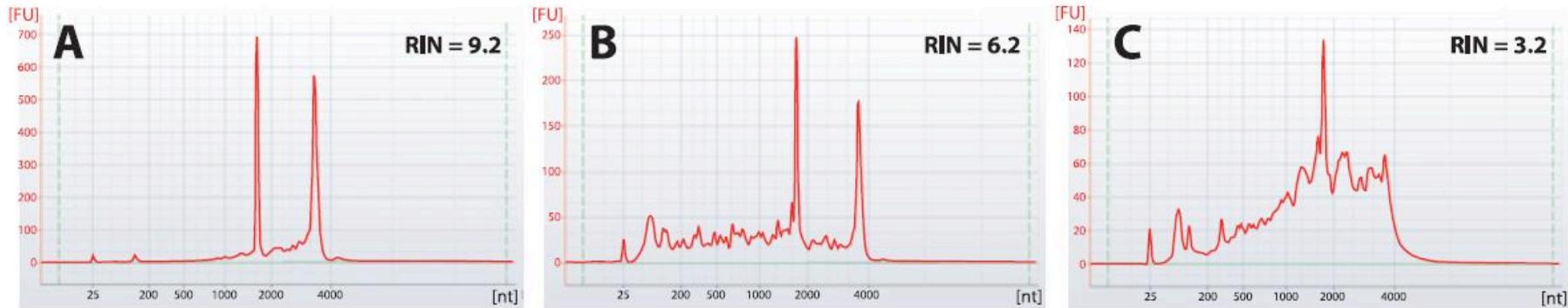
# Sample prep: RNA

mRNA degrades FAST

Freeze sample or place it in RNA-later within 30 sec (*if possible*)

Chose a correct kit for your particular application!  
Always treat samples with DNase

Differential expression, miRNA – **RIN value over 8.0**  
Aim for 4 biological replicates

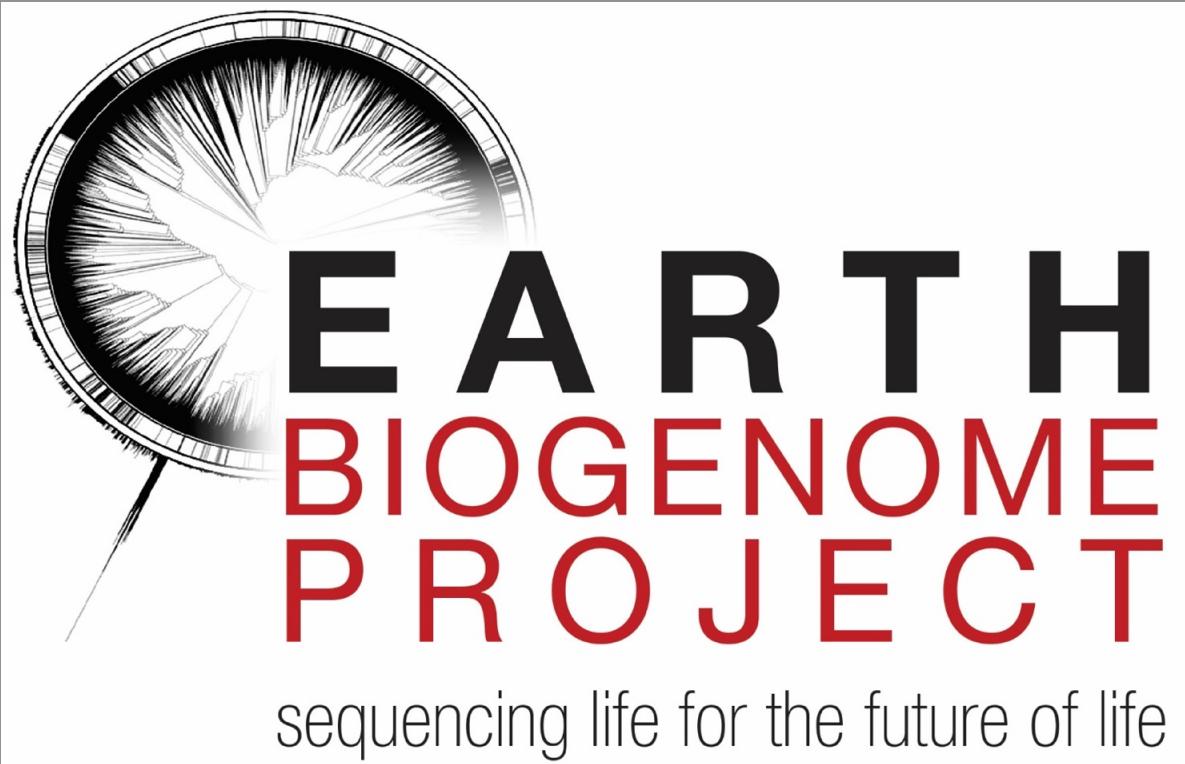


# Sample prep: SUMMARY

- Sequencing success depends on the sample quality, DNA or RNA
- DNA quality is **essential** for PacBio and ONT sequencing  
... as well as PCR-free Illumina libraries & linked reads!

# NGS and its challenges: SUMMARY

- Technologies develop VERY FAST.
- Beware of sequencing bias.
- Sequencing result depends on sample quality.
- Consult experts when it comes to experimental design and technology choice.



# EARTH BIOTRUST BIOTRUST

sequencing life for the future of life



We choose to go to the moon in this decade and do the other things, not because they are easy, but because they are hard, because that goal will serve to organize and measure the best of our energies and skills.

(John F. Kennedy)

[izquotes.com](http://izquotes.com)

**USA Moon program cost, 1962-1967:**

**28 bln USD (283 bln)**



**Hubble space telescope cost 1986-2010:**  
**10 bln USD**



**Manned mission to Mars, 2024-2030:**

**6 bln USD start-up + 4 bln USD per launch**

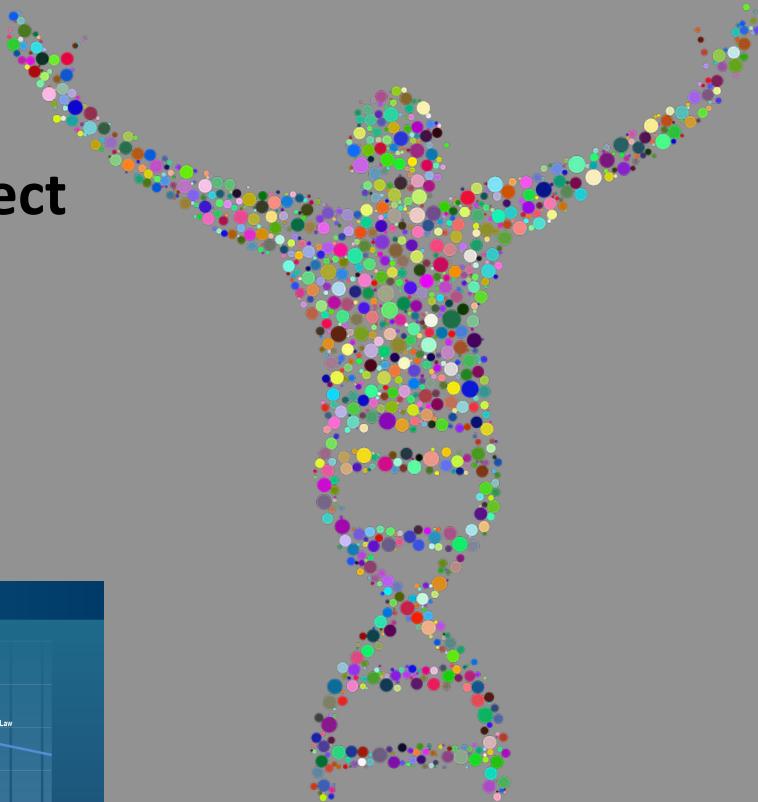
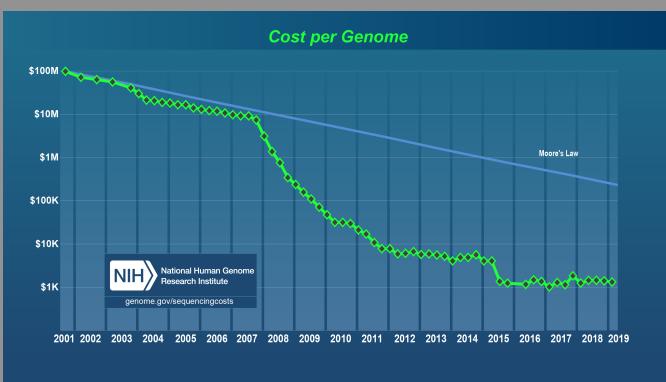


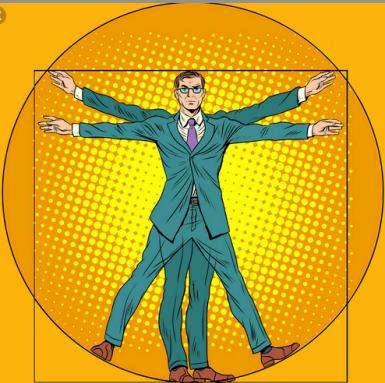
**Meanwhile on Earth: 6th Mass Extinction Event**

**WWF estimate: 1 species is lost every 5 minutes**

# Human Genome Project 1990-2003

Spent: USD 14.5 bln  
Output: USD 965 bln



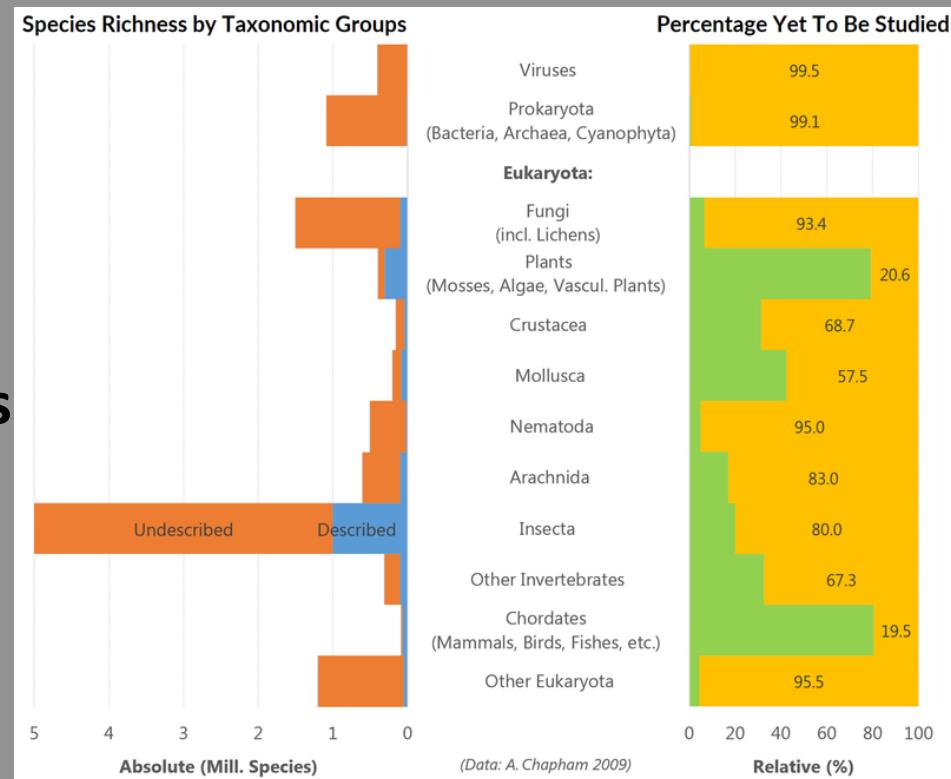


## *Homo sapiens*

N species = 1

## Global Biodiversity

N species = 1 trillion





HARRIS LEWIN  
UNIVERSITY OF  
CALIFORNIA, DAVIS  
*CHAIR*



JOHN KRESS  
SMITHSONIAN  
INSTITUTION  
*CO-CHAIR*



GENE ROBINSON  
UNIVERSITY OF  
ILLINOIS, URBANA-  
CHAMPAIGN  
*CO-CHAIR*

Our task now is to resynthesize biology; put the organism back into its environment; connect it again to its evolutionary past; and let us feel that complex flow that is organism, evolution, and environment united.

Carl R. Woese, *New Biology for a New Century*

PNAS Proceedings of the National Academy of Sciences of the United States of America

Keyword, Author, or

Home Articles Front Matter News Podcasts Authors

NEW RESEARCH IN Physical Sciences Social Sciences

Earth BioGenome Project: Sequencing life for the future of life

Harrie A. Lewin, Gene E. Robinson, W. John Kress, William J. Baker, Jonathan Coddington, Keith A. Crandall, Richard Durbin, Scott V. Edwards, Félix Forest, M. Thomas P. Gilbert, Melissa M. Goldstein, Igor V. Grigoriev, Kevin J. Hackett, David Haussler, Erich D. Jarvis, Warren E. Johnson, Aristides Patrinos, Stephen Richards, Juan Carlos Castilla-Rubio, Marie-Anne van Sluys, Pamela S. Soltis, Xun Xu, Huanning Yang, and Guojie Zhang

PNAS April 24, 2018 115 (17) 4329-4333; first published April 23, 2018 https://doi.org/10.1073/pnas.1720151115  
Edited by John C. Arise, University of California, Irvine, CA, and approved March 15, 2018 (received for review January 6, 2018)



## EBP committees:

- Sample prep
- Informatics
- Genome sequencing
- Phylogenetics
- Genome assembly
- Legislation
- Genome annotation

# EBP working groups

UT EBP

GOVERNANCE

COMMITTEES

GOALS

MEDIA

EVENTS

CONTACT

★SCIENTIFIC SUBCOMMITTEE: SAMPLE COLLECTION AND PROCESSING OVP

SCIENTIFIC SUBCOMMITTEE: SEQUENCING AND ASSEMBLY

SCIENTIFIC SUBCOMMITTEE: ANNOTATION

★SCIENTIFIC SUBCOMMITTEE: DATA ANALYSIS Chair: KLT

SCIENTIFIC SUBCOMMITTEE: IT AND INFORMATICS

ETHICAL, LEGAL, AND SOCIAL ISSUES

COMMUNICATIONS AND PUBLIC AFFAIRS

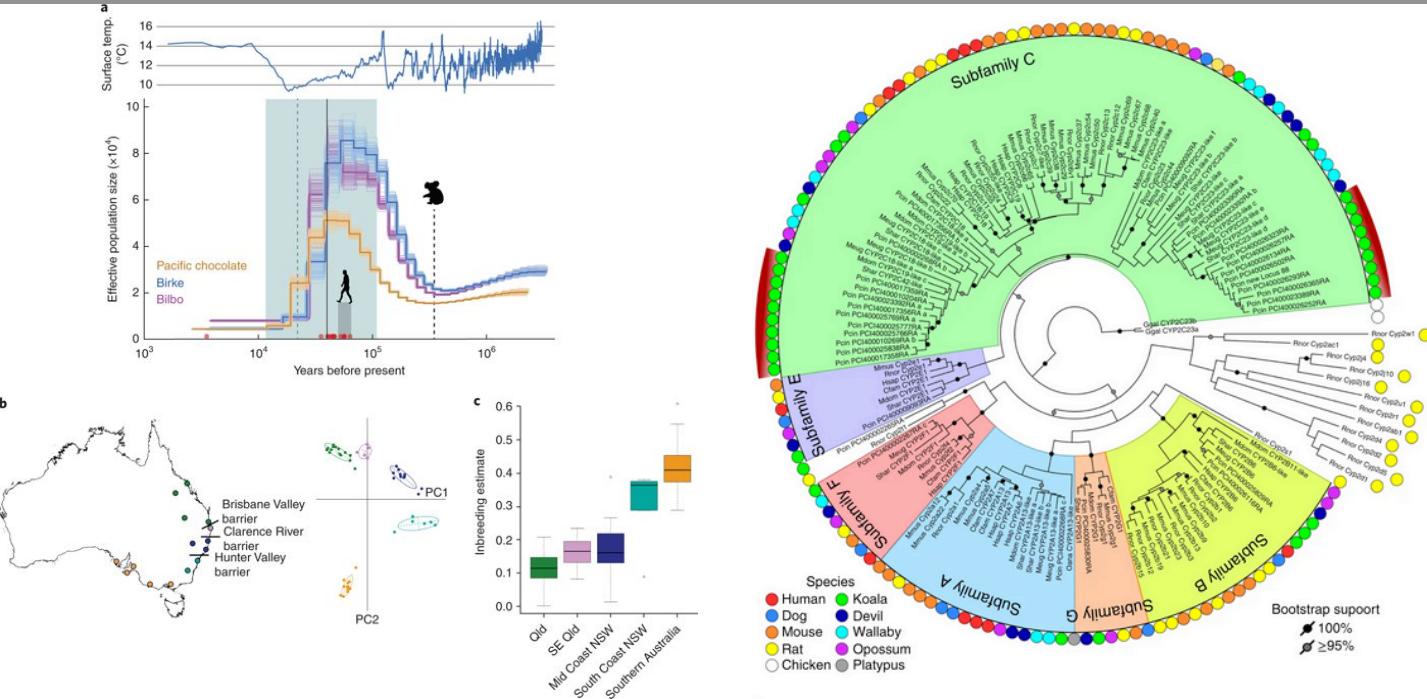
SY





## Adaptation and conservation insights from the koala genome

Rebecca N. Johnson , Denis O'Meally, [...] Katherine Belov



# Swe-EBP

- September 29 - first Swedish meeting
- Oct 5 - Steering board selected
- Next meeting: November 30
- TO BE CONTINUED!

# THANK YOU!



@OlgaVPettersson