

Introduction to Bioinformatics
NARL, Kawanda
10 - 14, November, 2014

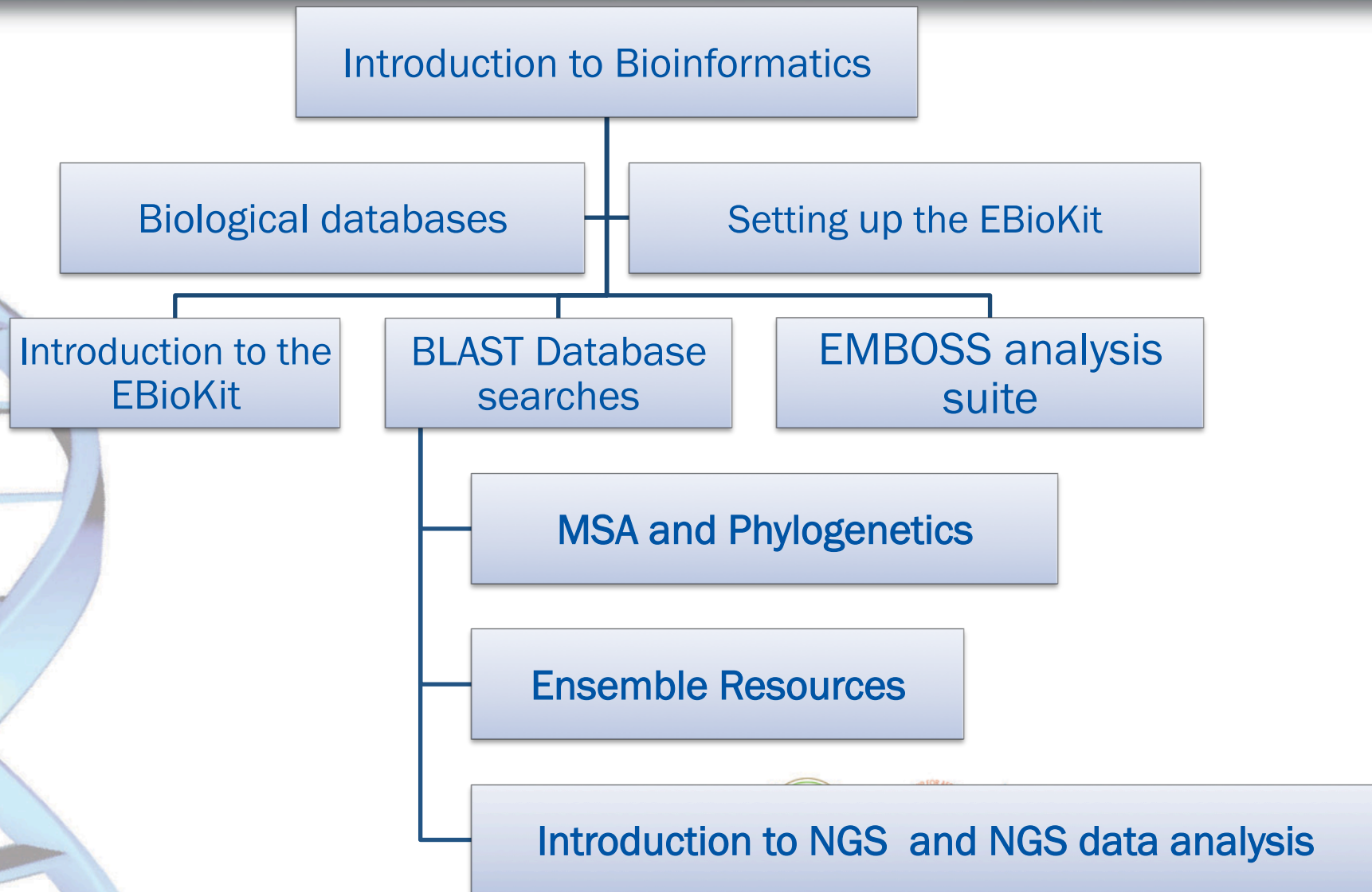
Joyce Njoki Nzioki
BecA-ILRI Hub, Nairobi, Kenya
<http://hub.africabiosciences.org/>
<http://www.ilri.org/>
j.n.njuguna@cgiar.org

ILRI
INTERNATIONAL
LIVESTOCK RESEARCH
INSTITUTE



biosciences
eastern and central africa

Plan for the week



Bioinformatics - definition

Bio informatics

Bio – Biology, Life Sciences

Informatics – computational sciences

BIOINFORMAICS

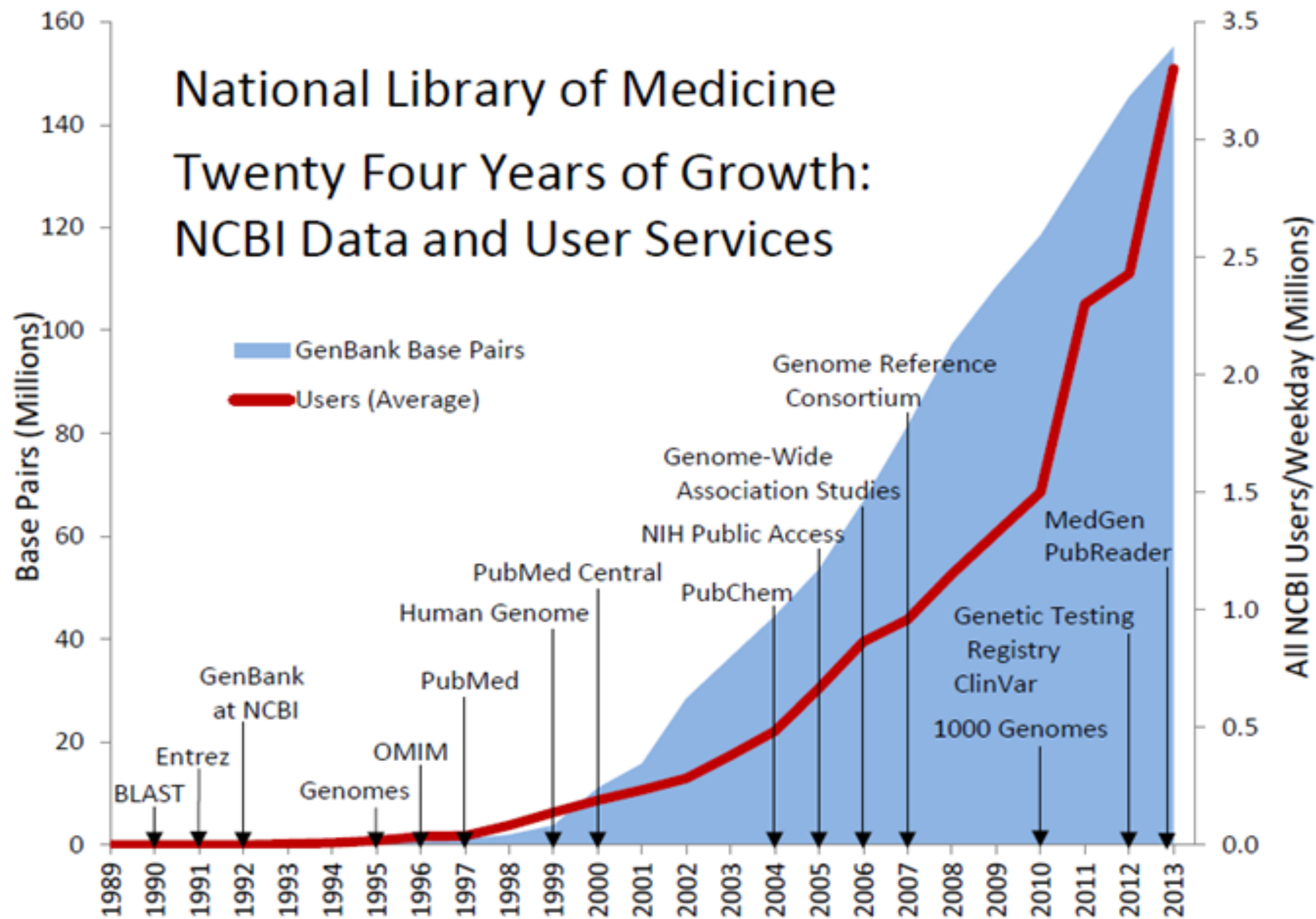
What is Bioinformatics

Bioinformatics is an interdisciplinary science that develops and improves on methods of storing, retrieving, organizing and analyzing biological data.

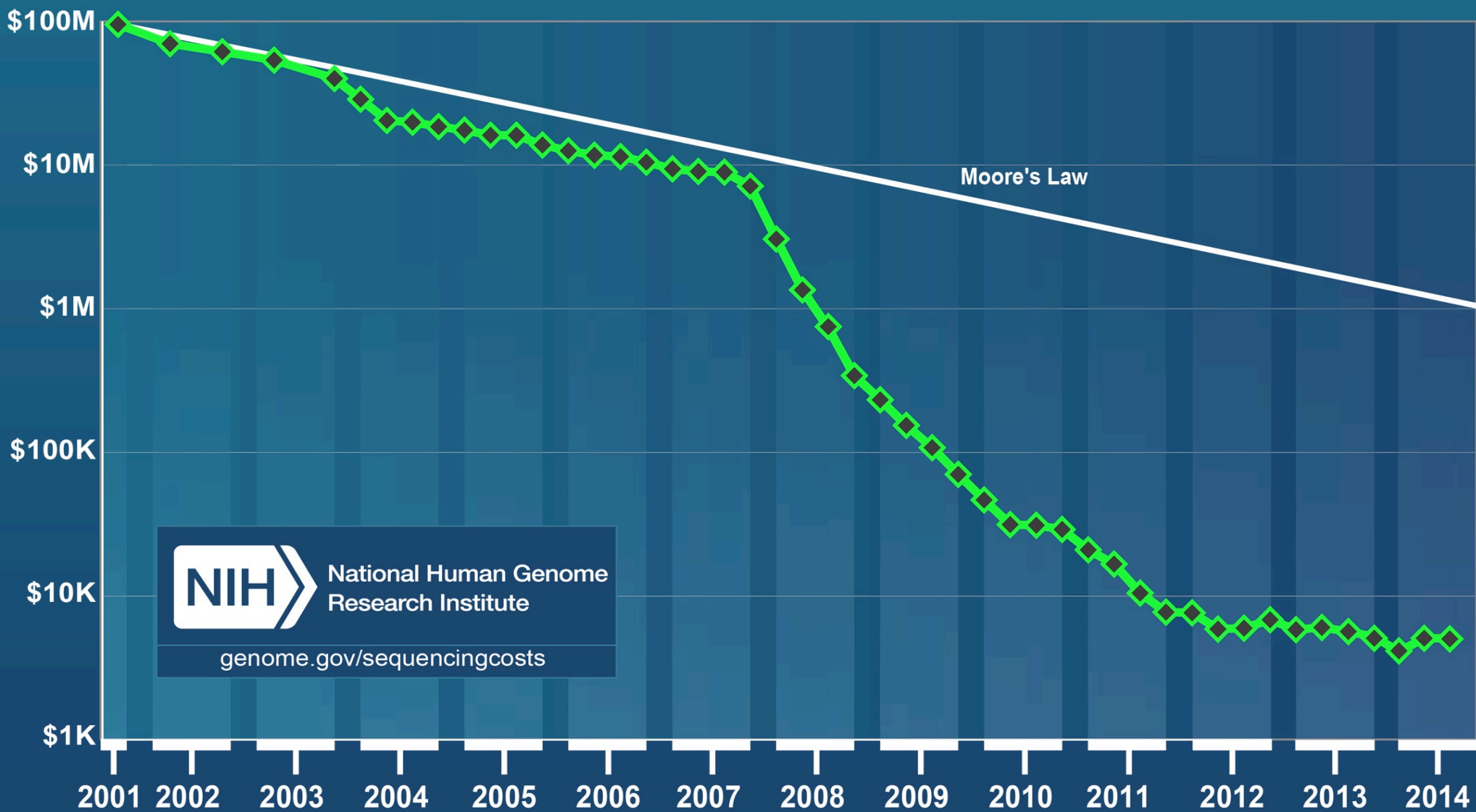
This is in order to solve biological problems and discover the wealth of biological information hidden in biological data.

National Library of Medicine

Twenty Four Years of Growth: NCBI Data and User Services



Cost per Genome

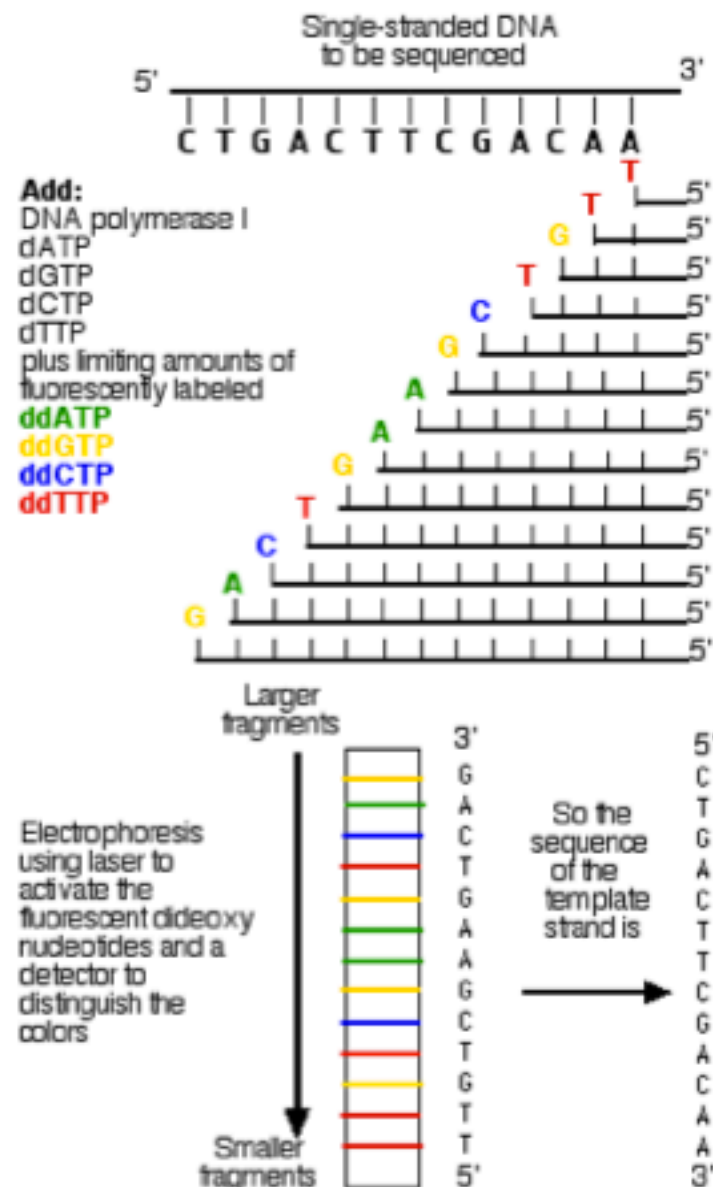
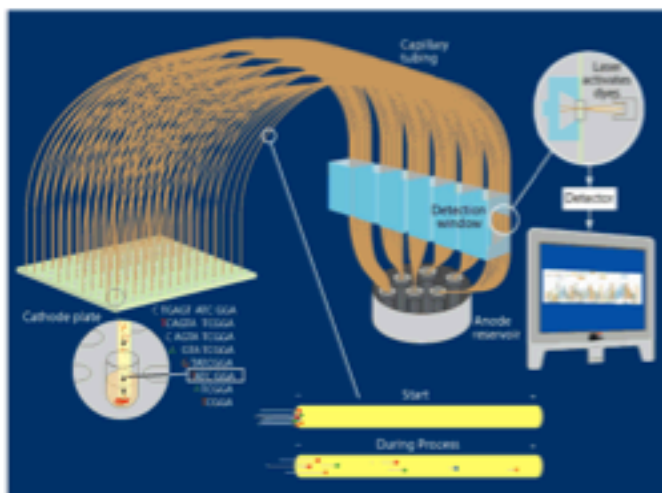


Current Sequencing Technologies

- Sanger
- (Roche) 454
- Illumina
- PacBio
- Ion Torrent
- (Illumina) Moleculo
- Oxford Nanopore

Sanger Sequencing

- ddNTP's (with fluorescent labels) incorporated (along with unlabeled dNTP's) in amplification step, resulting in some molecules terminated *at every position*
- Gel / capillary electrophoresis orders molecules by length
- Fluorescent label (color) indicates terminal base identity at each position
- Read colors, in order, to derive sequence



Current Sequencing Technologies

Roche 454 GS FLX Titanium



Illumina HiSeq 2000 / 2500

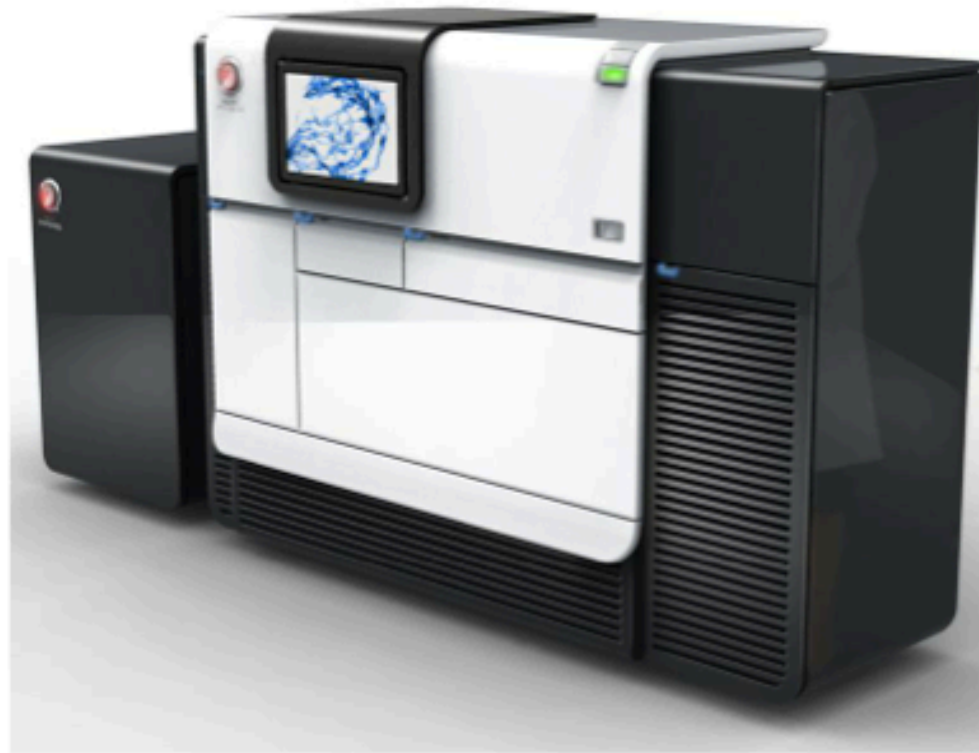


Illumina MiSeq



Current Sequencing Technologies

PacBio RS



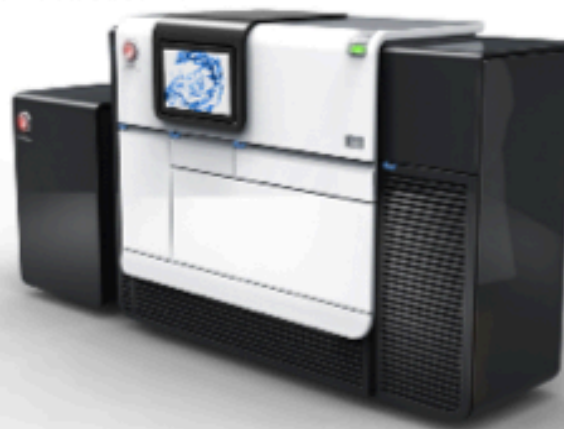
Ion Torrent
(Life Technologies)
Ion Proton



Ion Torrent
(Life Technologies)
Ion PGM



Tech Comparison



Feature	HiSeq2000	MiSeq	PacBio RS	Roche/454 FLX+
Number of reads	187 m/lane	15-18 m/lane	~40 K reads/SMRT cell	900-1500k/PTP
Read length	2 x 100 bp	2 x 250 bp	~ 3-10kb (120 min movie)	600-800 bp
Yield per run (PF data)	~37.5 Gb	~8.5 Gb	~Up to 0.2 Gb	~0.9 Gb
Pricing per run	\$2,040	\$1,179	\$250	\$6,800
Pricing per Gb	\$54	\$138	\$1,250	\$7,555
In Development	2 x 150 bp	2x300 bp	0.5G, 1Gb, 2Gb	???

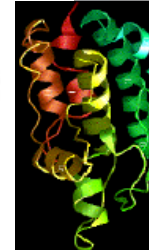
All this data ... Then what ...?

- Sequence with out knowledge connected to it is meaning less.
- Finding genes and regulatory elements in the sequences.
- Functional analysis of the genes – link between gene expression and cell/organ/tissue, function or dysfunction.
- Make sense of gene and protein data so as to assign function and understand the biological processes at the molecular level.

Scope of Bioinformatics

1. **Storage and retrieval of biological data.**
2. **Sequence analysis:** Sequence alignments, database searches, genome assemblies, motif detection.
3. **Structural analysis:** protein / nucleic acid structure, visualization and analysis, classification, prediction.
4. **Genomics:** annotation, comparative genomics
5. **Functional genomics:** Transcriptome, proteome, interactome
6. **Analysis of biochemical networks:** metabolic networks, regulatory networks
7. **Phylogeny**

Bioinformatics plays a role at each stage of the central dogma



- 1. DNA sequences determine protein sequences**
 - Genes and their features
 - Gene expression
 - Genome comparison
- 2. Protein sequences determine protein structure**
 - Sequence Analysis
 - Protein Structure
 - Phylogeny
- 3. Protein Structure determines function.**
 - Protein – protein interactions
 - Metabolic pathways

The integration of information learned about this three biological processes give insight on the biology of organisms.

Scope of Bioinformatics

1. Development of computational tools and databases:

- Writing software for analysis.
- Construction of biological databases

2. Application of tools and databases in generating biological knowledge.

- *Sequence analysis:*
 - Alignment, database searches, genome assembly, phylogeny, gene and promoter prediction
- *Structural analysis:*
 - Protein/ nucleic acid structure, protein structure prediction and comparison
- *Functional analysis:*
 - Protein-protein interactions.

Who works in Bioinformatics

- Biologists
- Chemists
- Biochemists
- Computer scientists
- Mathematicians
- Statisticians
- Physicists
- Engineers

Applications of Bioinformatics

Molecular Medicine

- Drug target identification
- Personalized medicine
- Diagnostics of cancer

Biotechnology

- Bioengineering
- Gene therapy
- Agriculture biotechnology

Microbial genome applications

- Waste cleanup
- Climate change

History of Bioinformatics

1950s and 1960s:

- First collection of protein sequences in the Protein information resource (PIR)
- Protein structure comparison

1970s:

- Protein Data bank published
- Sequence alignment
- Protein structure prediction

History of Bioinformatics

1980s:

- Sequence databases EMBL, GenBank, DDBJ
- Curated protein databases SwissProt
- Sequence searches in databases
- NCBI and EMBnet created

1990s:

- Genome annotation
- Comparative genomics
- Structural Bioinformatics
- Transcriptome analysis

Limitations of Bioinformatics

- **Bioinformatics is a science of inference hence:**
- **Quality of bioinformatics predictions depends on the quality of data and sophistication of algorithms.**
- **Sequence data may have errors which subsequently leads to errors in downstream analysis.**
- **Many exhaustive algorithms cannot be used due to computational limitations.**
- **Trade-off between specificity and sensitivity**

Why bioinformatics then ?

In most cases wet lab is needed to validate bioinformatic predictions

Bioinformatics can:

- Reduce data to a small set of testable predictions
- Assign a degree of confidence to each prediction

The biologist will often have to choose the appropriate degree of confidence, depending on:

- Cost of validating predictions.
- Benefit expected from the right predictions.

Bioinformatics as *in silico* biology:

- Allows for exploration of domains that cannot be addressed manually e.g study of past evolutionary events.

Contigs and conflicts

Acknowledgement for some of the slides

Thanks to Joe Fass (UC Davis, Bioinformatics Core)
Mark Wamalwa (Beca-ILRI Hub, Bioinformatics)

THE END