



Week 1 – Welcome and Introduction

Sean Wang DVM, Ph.D

PROTECTING, MAINTAINING AND IMPROVING THE HEALTH OF ALL MINNESOTANS



Sean Wang, DVM, PHD

Sequencing and Bioinformatics
Unit Supervisor

Sean.wang@state.mn.us



Jake Garfin

Bacteriology Laboratory Specialist

Jake.Garfin@state.mn.us



Angie Taylor, MPH, MB(ASCP)^{CM}

Bacteriology Laboratory Specialist

Angie.taylor@state.mn.us



Matt Plumb, MS

Bacteriology Laboratory Specialist

Matt.plumb@state.mn.us



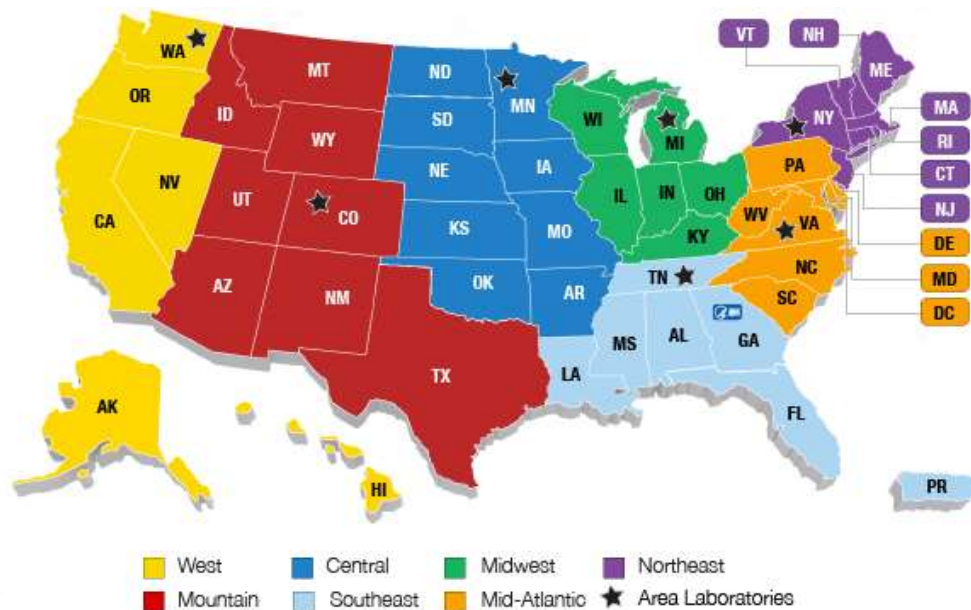
Kelly Pung

Bacteriology Laboratory Specialist

Kelly.pung@state.mn.us

Central Region BRR and WFD

Bioinformatics Regional Resource and Workforce Development Regions



- Ad-doc bioinformatics analyses for central region PHLs
- Onsite & in-house bioinformatics training
- Consultation with region PHL's laboratorians and IT staff members
- "Office hours" to ensure continuous remote bioinformatics support and consultation
- Serve as bioinformatics knowledge exchange and problem solving hub
- Building the "Stratus" cloud computing bioinformatics infrastructure, deploy VMs to region PHLs (**Central Region**)

Regions are the same with PulseNet and ARLN

Recordings and Slides: StaPH-B YouTube Channel and StaPH-B website

<http://www.staphb.org/>



MN, CO, VA and UT

StaPH-B
State Public Health Bioinformatics

- ✓ StaPH-B gen
- ✓ StaPH-B dev
- ✓ StaPH-B Steering



Monthly Webinar Series
Slack Channel
APHL, CDC and State Public Health
Laboratories communication

Webinar and Training Overview

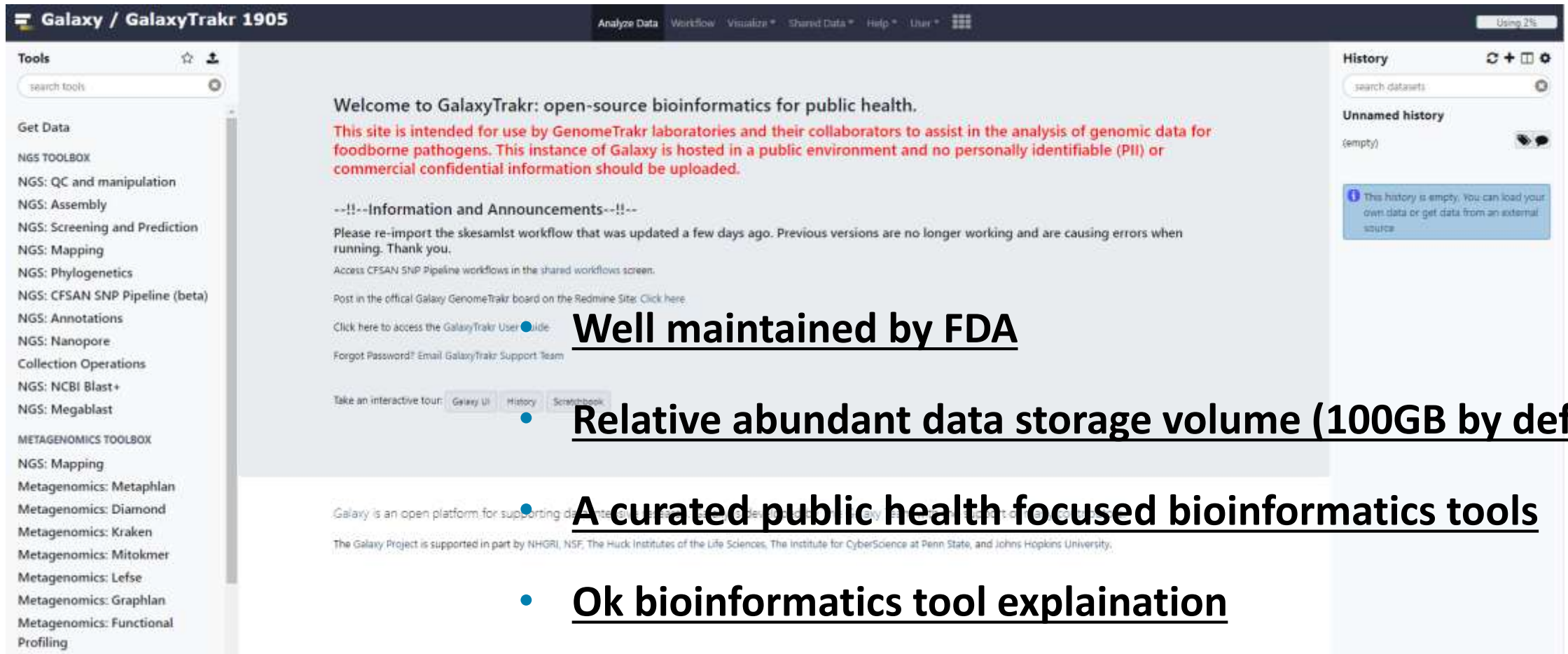
Week	Dates	Webinar Topic	Hands-on Topic
1	March 10th & 12th	Why Next-Generation Sequencing?	Introduction to GalaxyTrakr and the Galaxy interface.
2	March 17 th & 19 th	Turning Microbes into Megabases: How Next-Generation Sequencing Works	Understanding bioinformatics file types and sharing in Galaxy
3	March 24 th & 26 th	From Raw Data to Genomes: QC, Mapping, and Assembly	Assessing read quality and De Novo assembly using Shovill
4	March 31 & April 2 nd	Mysterious Microbes: NGS for Bacterial Identification	Using MicroRunQC and Kraken to identify unknowns.
5	April 7 th & 9 th	Predicting Phenotype from Genotype	Antimicrobial gene detection using Abricate and CARD
6	April 14 th & 16 th	Reading the Tree Leaves: Cluster Detection with Whole Genome Sequencing (WGS)	Running CFSAN SNP Pipeline on GalaxyTrakr
7	April 21st	NGS Public Health Applications	N/A

Webinar training goals

- Develop understanding of a selection of bioinformatics topics and theories
- Get hands-on bioinformatics analyses experience using the GalaxyTrakr Platform
- Become more familiar with bioinformatics terminology and platforms
- Be able to execute analyses and interpret bioinformatics results within the GalaxyTrakr independently

***We encourage participation commitment but also fully understand the gravity of ongoing COVID-19 situation**

Why GalaxyTrakr?



Galaxy / GalaxyTrakr 1905

Analyze Data Workflow Visualize Shared Data Help User

Tools

search tools

Get Data

NGS TOOLBOX

NGS: QC and manipulation

NGS: Assembly

NGS: Screening and Prediction

NGS: Mapping

NGS: Phylogenetics

NGS: CFSAN SNP Pipeline (beta)

NGS: Annotations

NGS: Nanopore

Collection Operations

NGS: NCBI Blast+

NGS: Megablast

METAGENOMICS TOOLBOX

NGS: Mapping

Metagenomics: Metaphlan

Metagenomics: Diamond

Metagenomics: Kraken

Metagenomics: Mitokmer

Metagenomics: Lefse

Metagenomics: Graphlan

Metagenomics: Functional Profiling

Welcome to GalaxyTrakr: open-source bioinformatics for public health.

This site is intended for use by GenomeTrakr laboratories and their collaborators to assist in the analysis of genomic data for foodborne pathogens. This instance of Galaxy is hosted in a public environment and no personally identifiable (PII) or commercial confidential information should be uploaded.

--!!--Information and Announcements--!!--

Please re-import the skesam1st workflow that was updated a few days ago. Previous versions are no longer working and are causing errors when running. Thank you.

Access CFSAN SNP Pipeline workflows in the shared workflows screen.

Post in the official Galaxy GenomeTrakr board on the Redmine Site. [Click here](#)

[Click here to access the GalaxyTrakr User Guide](#)

[Forgot Password? Email GalaxyTrakr Support Team](#)

Take an interactive tour: [Galaxy UI](#) [History](#) [Scratchbook](#)

History

search datasets

Unnamed history

(empty)

This history is empty. You can load your own data or get data from an external source.

Galaxy is an open platform for supporting distributed genomic data analysis. Galaxy is supported by the following organizations:

The Galaxy Project is supported in part by NHGRI, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Johns Hopkins University.

Well maintained by FDA

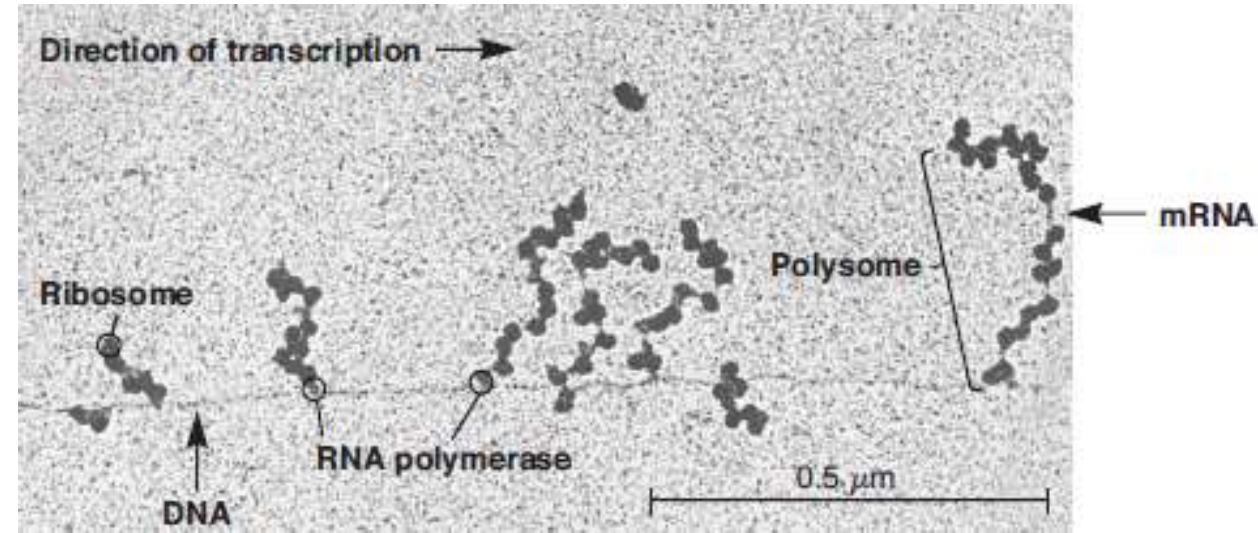
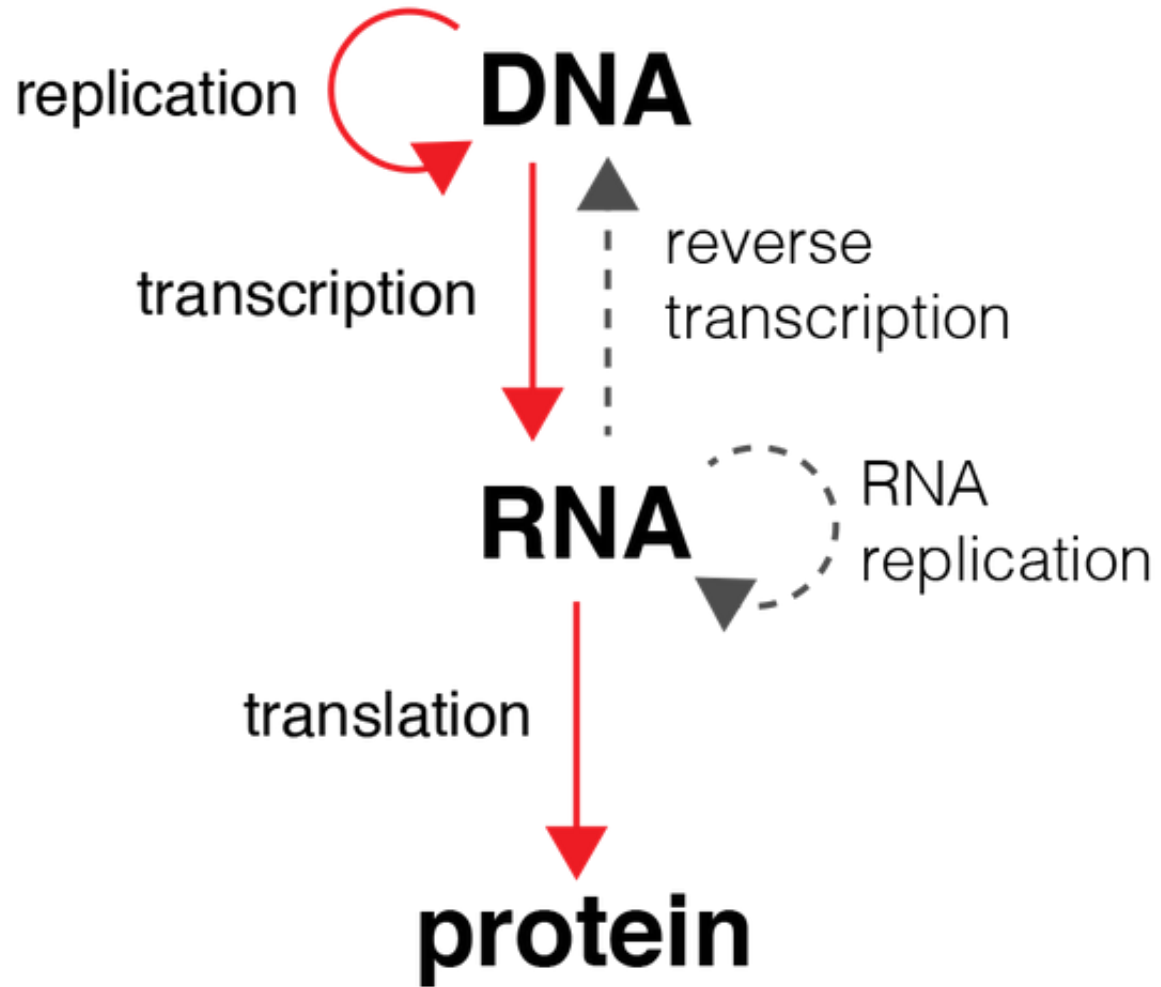
Relative abundant data storage volume (100GB by default)

A curated public health focused bioinformatics tools

Ok bioinformatics tool explanation

FREE!

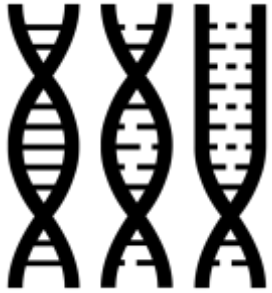
Why Sequencing? Central Dogma of Biology



Think about some molecular diagnostic assay in your laboratory

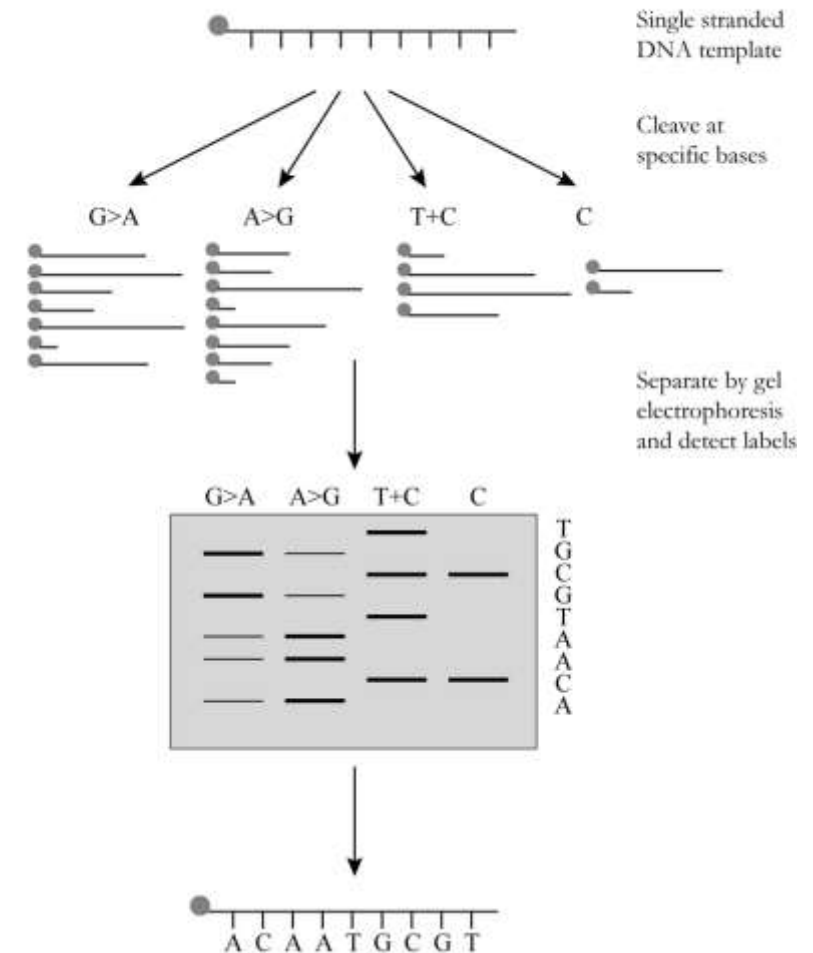
- PCR
- RT-PCR
- RT-qPCR
- Biofire

The history of DNA Sequencing



Maxam–Gilbert Sequencing

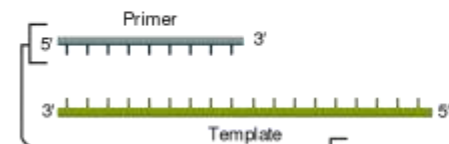
● label



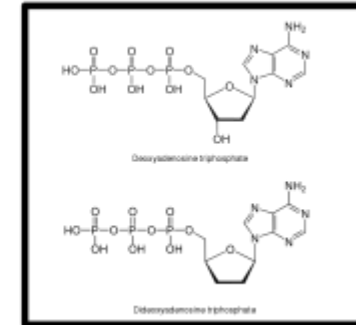
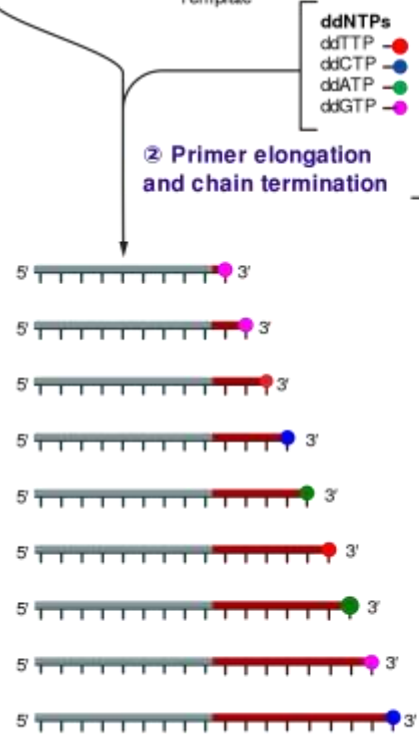
Chain-termination Sequencing

① Reaction mixture

- ▶ Primer and DNA template
- ▶ DNA polymerase
- ▶ ddNTPs with flourochromes
- ▶ dNTPs (dATP, dCTP, dGTP, and dTTP)



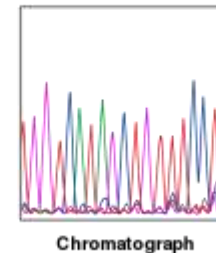
② Primer elongation and chain termination



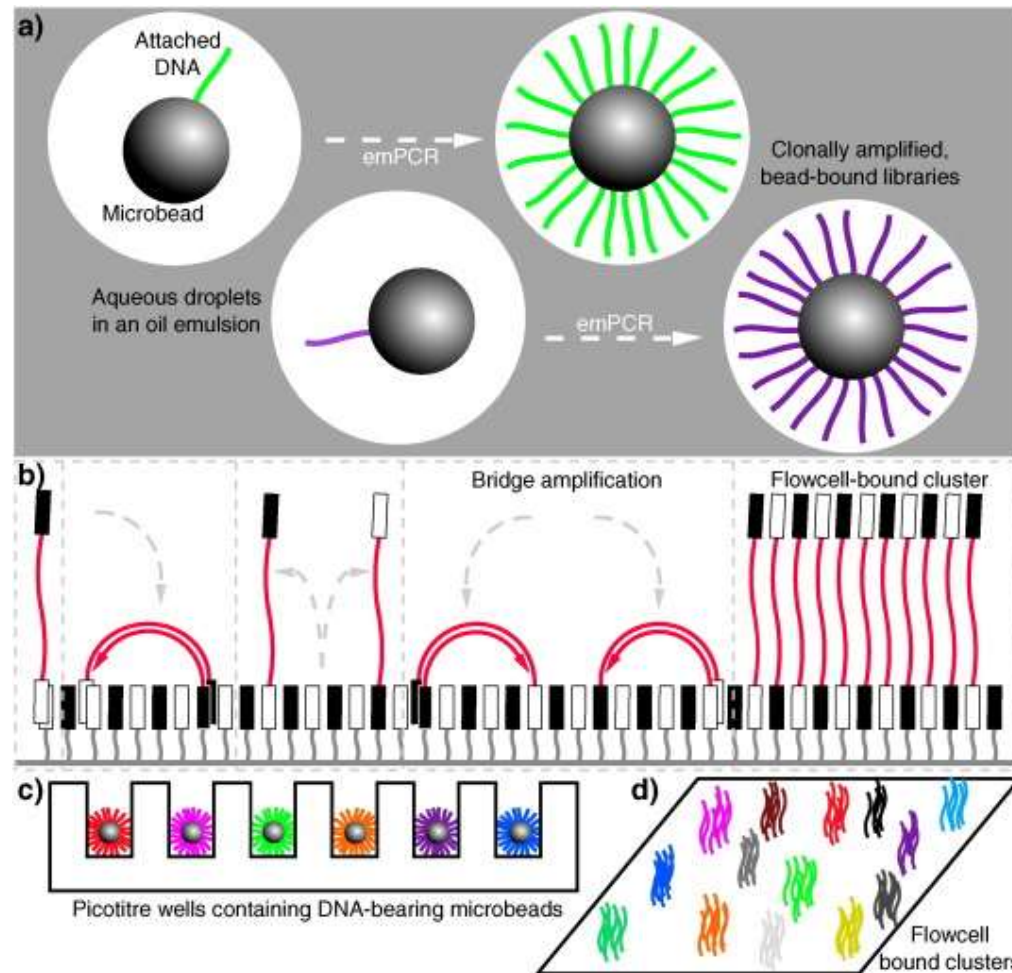
③ Capillary gel electrophoresis separation of DNA fragments



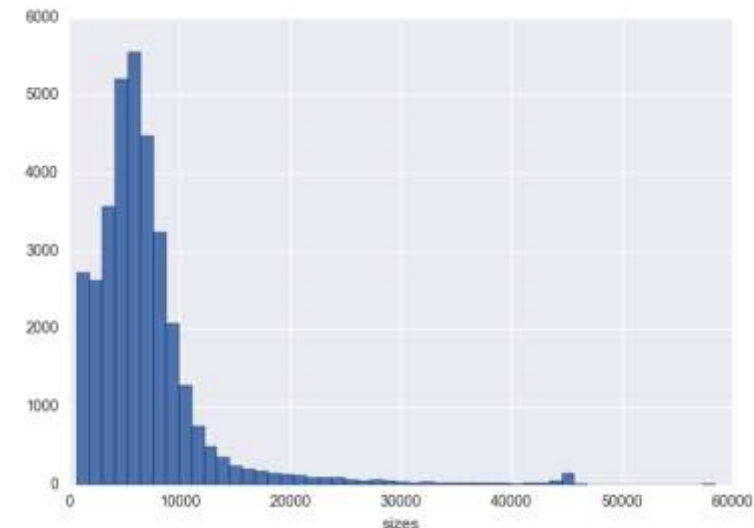
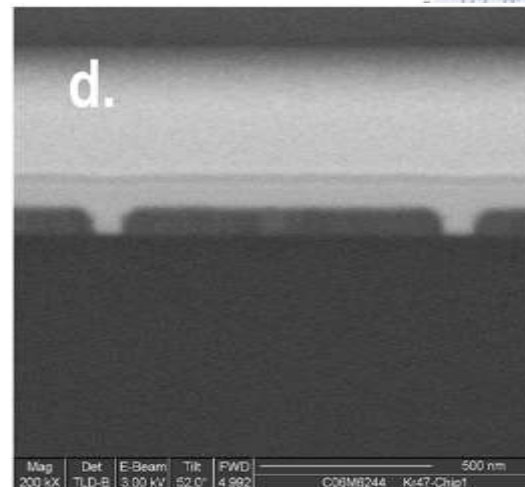
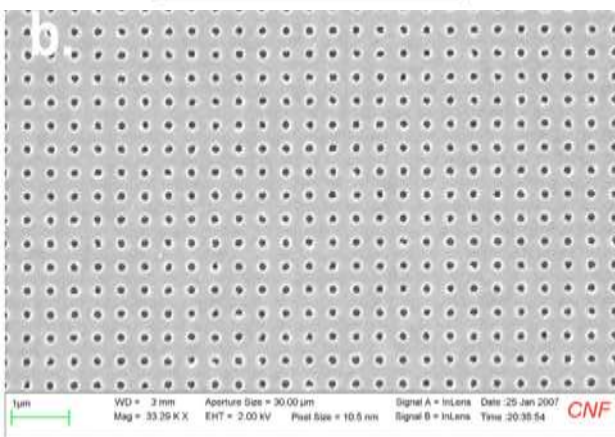
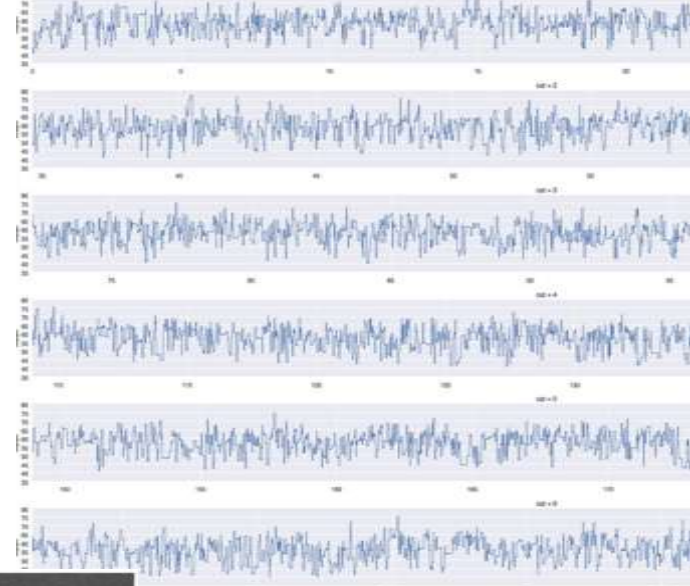
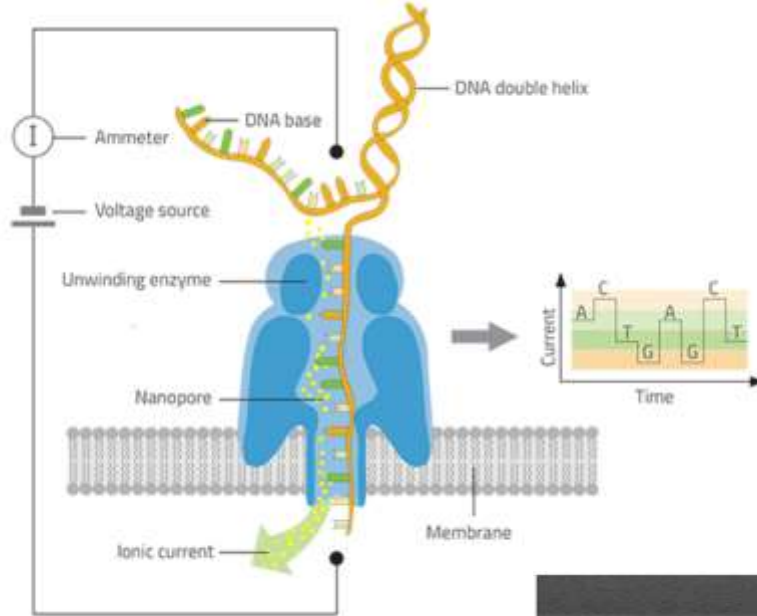
④ Laser detection of flourochromes and computational sequence analysis



Next-Generation Sequencing (NGS) – Amplification is the King



Third Generation Sequencing: Long read, real time sequence read

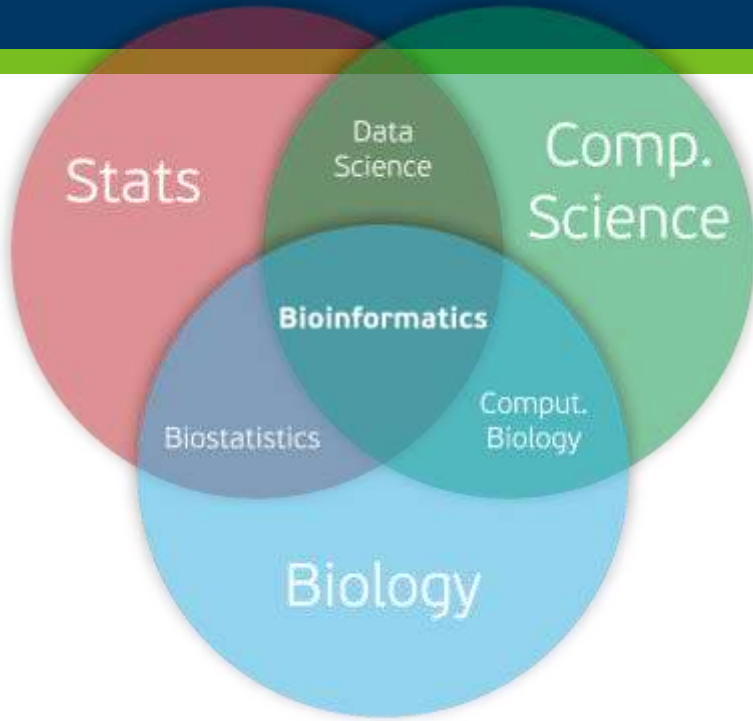



```
@M03999:11:000000000-ANHE8:1:1101:14560:1871 1:N:0:12
ATCAAGGATAGAGCGTGTCTTTCTCCTAAAGTAGCTGCAATAATGACATATTTACAAAGTATTCTTGCTCGCTCATGTCA
+
1>>1>111DB111B1E1EBGGGFGDAF1111D33BF11AAF11F22F1GFHGHBE211AGFHHFHHFDGF?EGFCGFHHF
@M03999:11:000000000-ANHE8:1:1101:14445:1900 1:N:0:12
ACTTAATTGAGTGCTGGCTTCATCTACTGTAGGTTCTTTATCATCTTCATCTTGTTTTAAACACTTTTATCCATTTGAGT
+
1>>A11DFF11DDDG11AFGF1GAF31D1D331FGFHHH3GA2FFGHF2GGGH2FGGH211DFAGHHHBGFE2GHHB2CG
@M03999:11:000000000-ANHE8:1:1101:18618:1914 1:N:0:12
ATCAAGCACAAACATCAACGCCTGCTTCAACTAAAGCATTACACGATCCATTTGCCCTACTCCTATAGCC
+
3AAA?4DFFFFBAFGCGGBGGCEEG44FGEDGGGHHFBE3GHGBGDFE?GHEBGGH5AFGHGHHHHHGHEBFG
@M03999:11:000000000-ANHE8:1:1101:17412:1919 1:N:0:12
CTTTTTATTTGTTTATAAGCTTAAGTGGGCTAGTGGAGTTTTTGCGCTTTTAAATTTTCAAATCACGGTTATTTATCA
+
ABBBBF@FFFFFFGGGGGCDDEGHCB5F5A22EE2EGS232FFHHG2AAAEGGHB35FHHHFFFDGDFECFGCHHHHFHFB
@M03999:11:000000000-ANHE8:1:1101:15228:1925 1:N:0:12
CGGAAAAGTCGAAACTCATAGTATATGGCTTTTATTTAAAAAATTGGTTTGAAAATGAAATGATTCCAAATACTAGAAAAAT
+
>1?1A?FBFDA?E1EGEBG3FGHG3F31AFHGHBBHHB2111BGFFGGHFB111DE111GFFHHHFBFBGGHHFFFFFHGAG
@M03999:11:000000000-ANHE8:1:1101:12304:1932 1:N:0:12
GTTTTTATTGATGAGATTGATAAAATTGCTAGAATGAGTGAAAATCGTTCTATTACTCGTGATGTAAGTGGAGAGGGCGTG
+
1AAAAA?FFF3DD3C1BF11FBD11DG1ADED11DA11D11211DA0FGGG2FH2AFA/G/0FFGFF2GGB1/1AE/AAEC
@M03999:11:000000000-ANHE8:1:1101:14974:1937 1:N:0:12
CCATGATTTAAATCTTTTATAAATCTTCCATCAATCTCACTTCGTTCTTTATTAATGTAAAGTGGCGTCATATCCTAGC
+
AAAA?4FFFFFFGGGGGGFGFFFGFHDF5FGDBGHHFBFGHHHDHGHGHHBGHEF5GBHFHFFHGEFCEGGHHHHHHHGHF
@M03999:11:000000000-ANHE8:1:1101:16924:1960 1:N:0:12
GCTTGATAGTGTTATTACCGCAGTAGGAATTGCTCAAGATGTAACCATCATGATTATCGCTGTTATTATCGCAGTAGCTGT
+
A?ABB>DDFFFFGBGGGDDE?2EEB5445BF45AFE3F3FFGCDFFDFAAFB5FGGGD2AHEGHHHHHFFA?FFGHGHHE
```



Data Generation

What is Bioinformatics?



- ❖ An interdisciplinary study which combine Statistics, Computer Science and Biology
- ❖ “Big data” of Genomics
- ❖ Translate meaningless millions of short sequences into meaningful and actionable biological results
- ❖ Population level infectious disease surveillance will require massive computing power to process the large amount of biological data

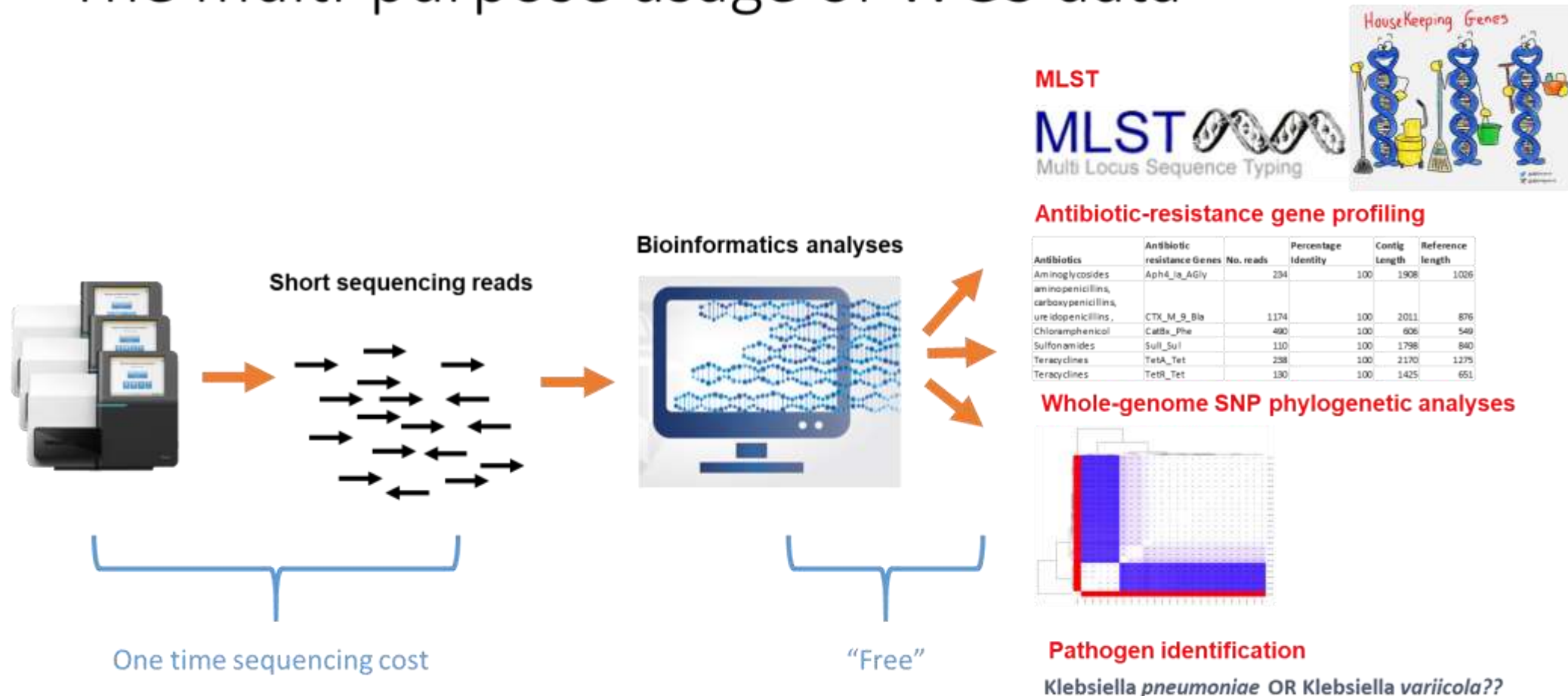


Three main analyses platforms

- Webpage-based single purpose/small collection of bioinformatics analyses
 - CGE, **GalaxyTrakr**, Web version of Seqsero
- Commercially/non-profit available bioinformatics software
 - DNASTar
 - Bionumerics
 - CLC Genomic Workbench
 - Geneious
- Linux/Unix based platform
 - Mac OS
 - Linux OS (Desktop, HPC, Cloud etc.)

Benefit of Sequencing

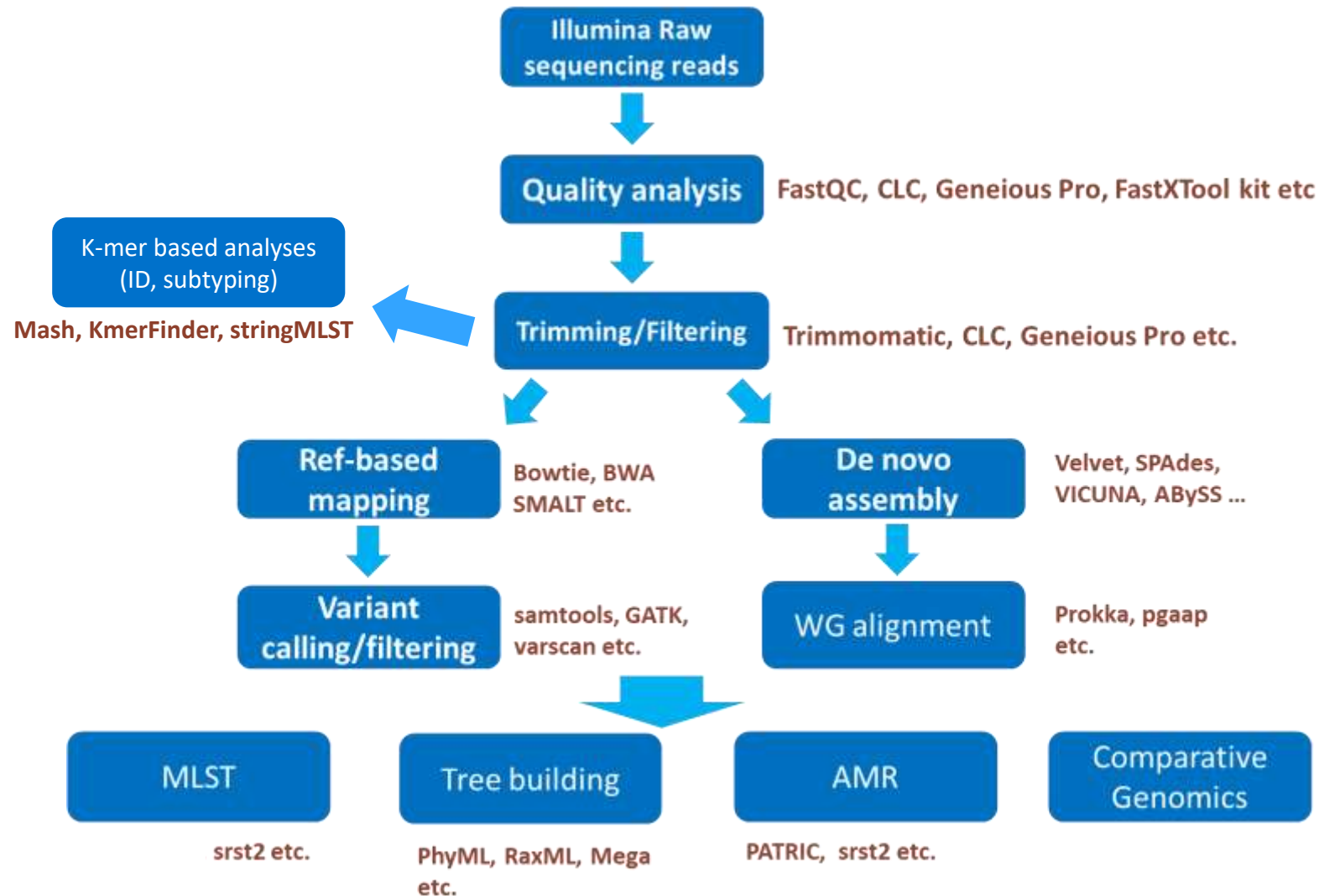
The multi-purpose usage of WGS data



Sequencing and Bioinformatics Public Health Application

- ❖ Species discovery and identification
 - COVID-19 (<https://nextstrain.org/ncov>)
- ❖ Infectious pathogen characterization
 - Serotyping and genotyping (i.e. Salmonella, Strep pneumo)
 - Antibiotic resistance gene profiling (i.e. mcr-1, mcr-2)
 - Virulence marker/gene detection and mutation monitoring (i.e. SH AA mutation may lead to increased RSV virulence)
- ❖ Genetic relatedness
 - PFGE -> WGS
 - HCV “Ghost” program

General workflow from raw NGS yield to biological results



Questions?