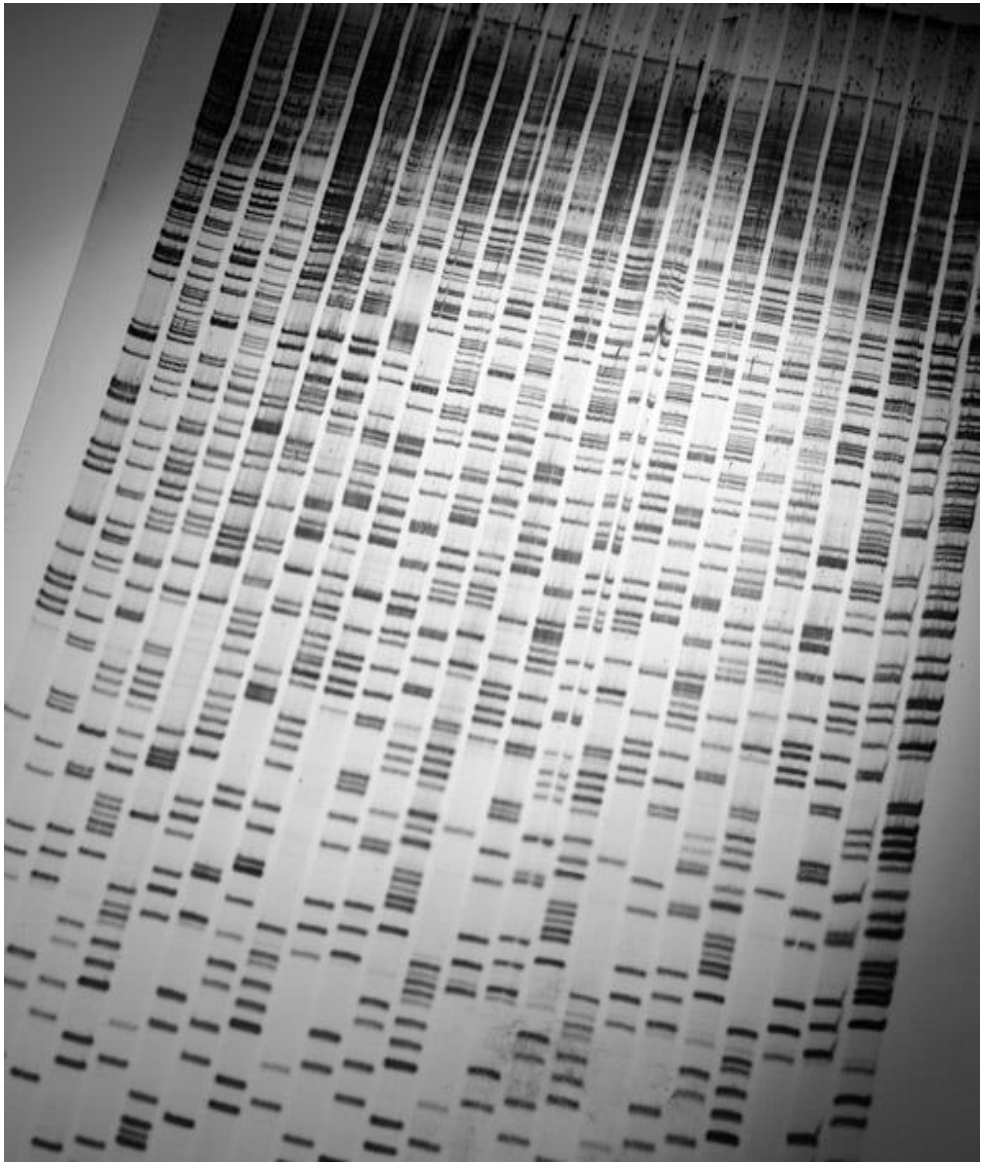


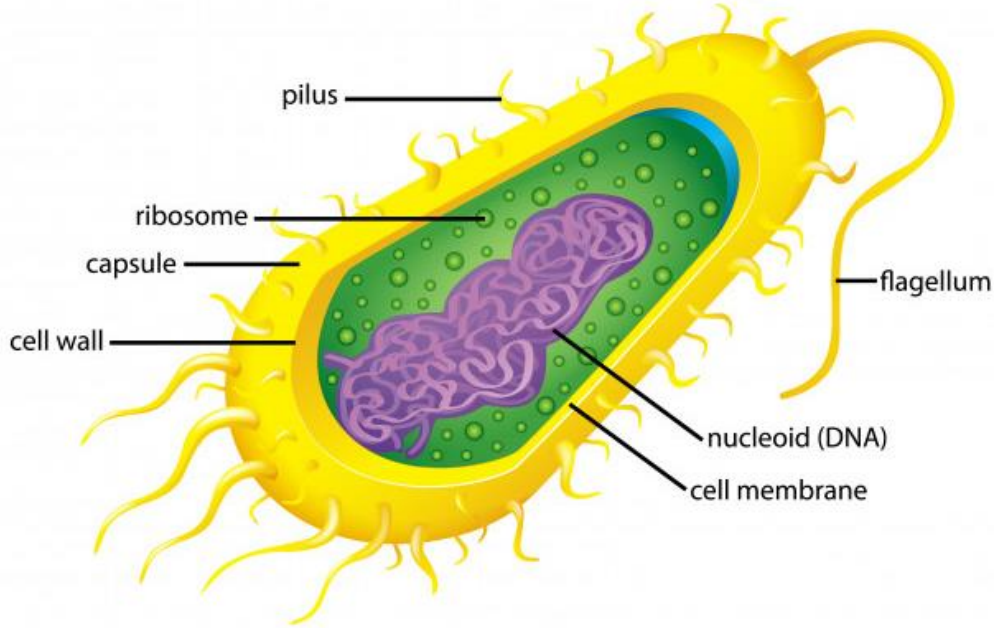
Autoradiograph - MRC Laboratory of Molecular Biology

Next-Generation Sequencing 101

Nicholas Florek, PhD, MPH

Allen Bateman, PhD, MPH, D(ABMM)

Microbial Genomes



most bacteria have a singular circular chromosome ranging from 160,000 to 12.2 million base pairs

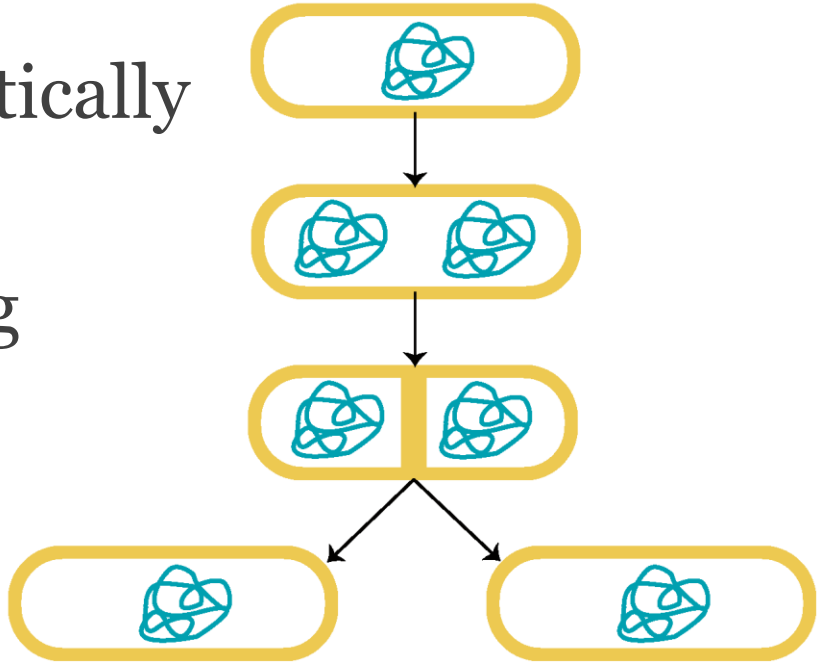
many can also carry plasmids

Basics of Microbial Evolution

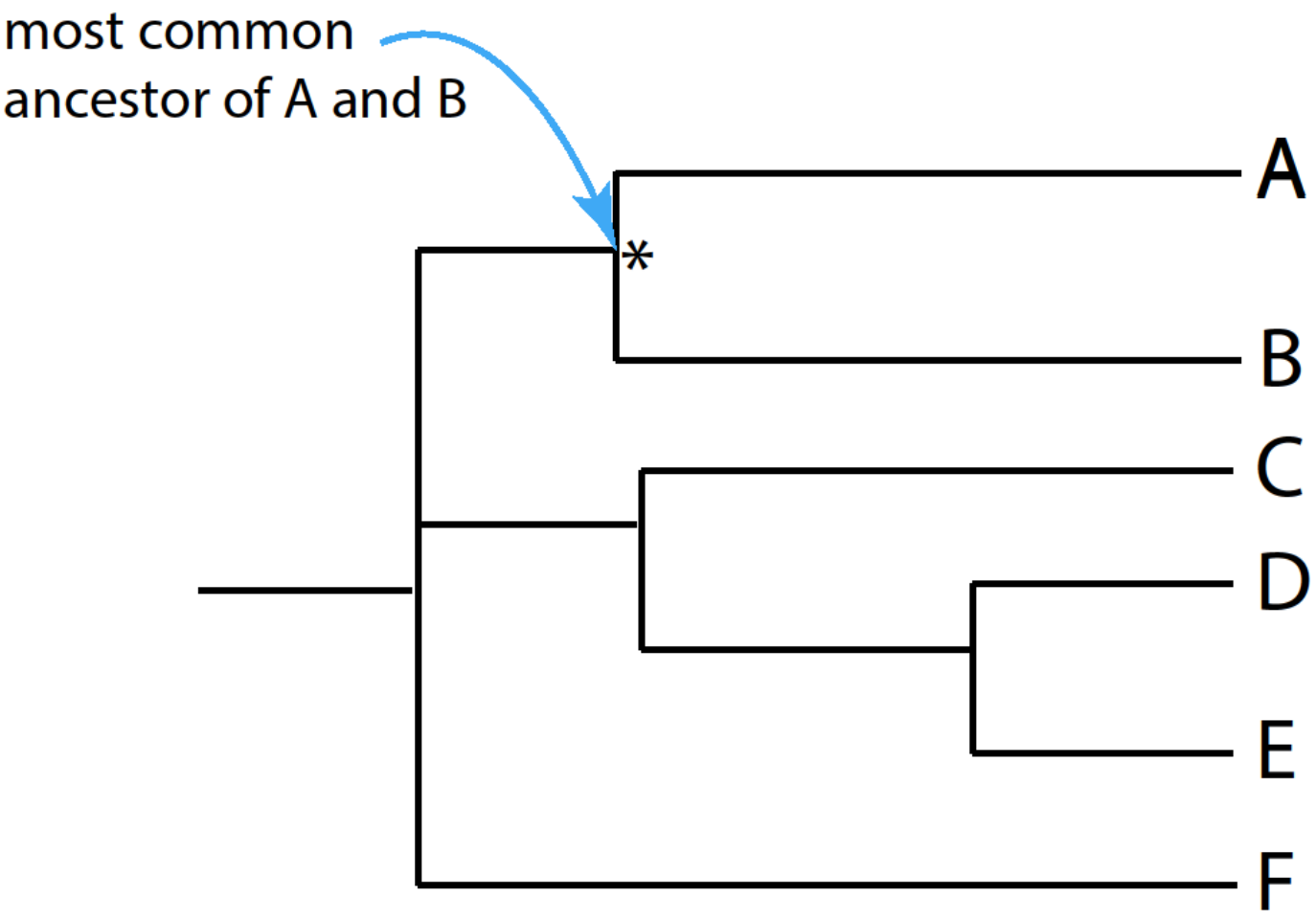
bacteria reproduce asexually through binary fission, each daughter cell is clonal to the parent

DNA is transferred vertically

mutations occur during replication



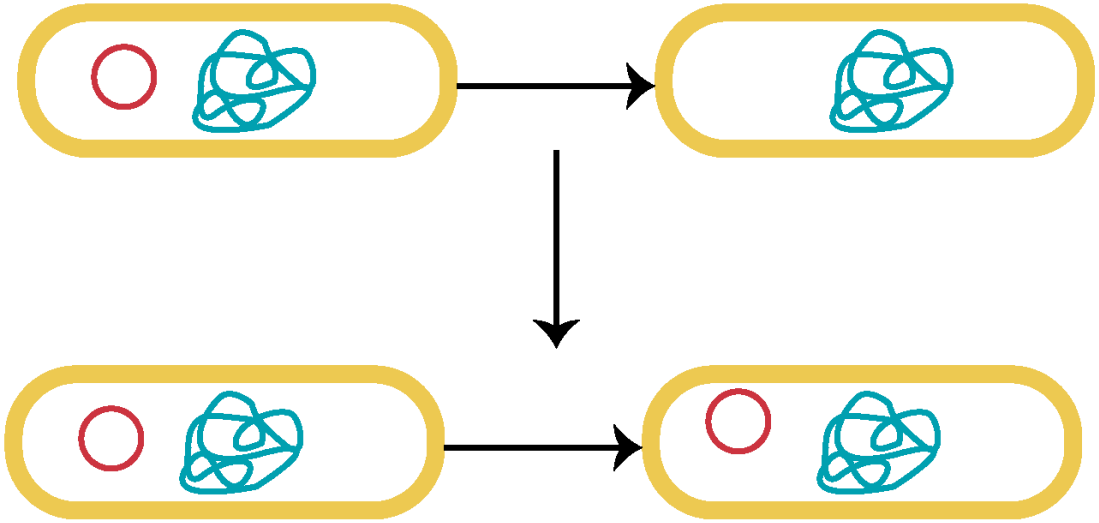
Basics of Microbial Evolution



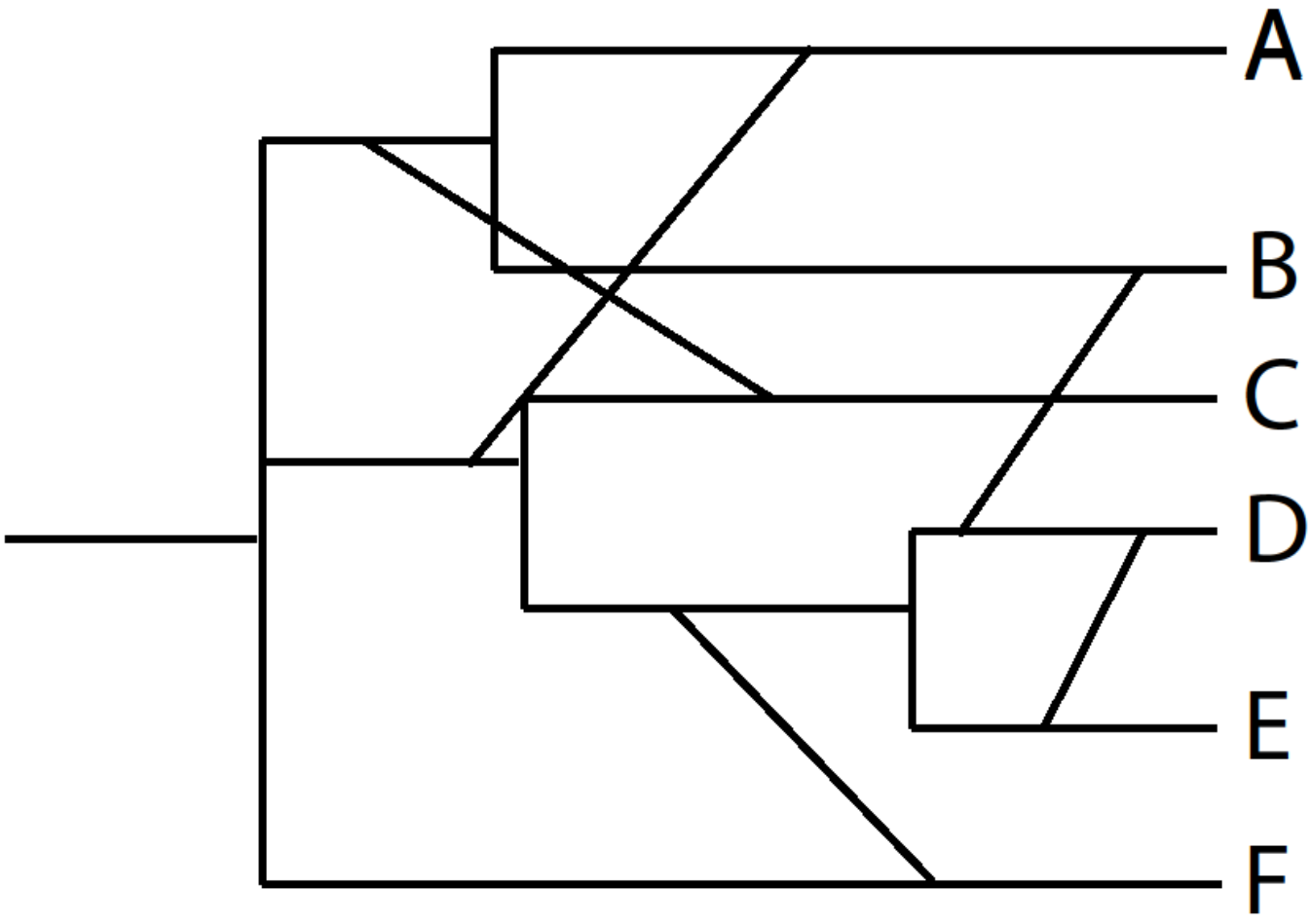
Basics of Microbial Evolution

DNA can also be transferred horizontally

- uptake from environment – transformation
- transferred between cell contact – conjugation
- transferred through phage – transduction



Basics of Microbial Evolution



Sequencing: insight into evolution

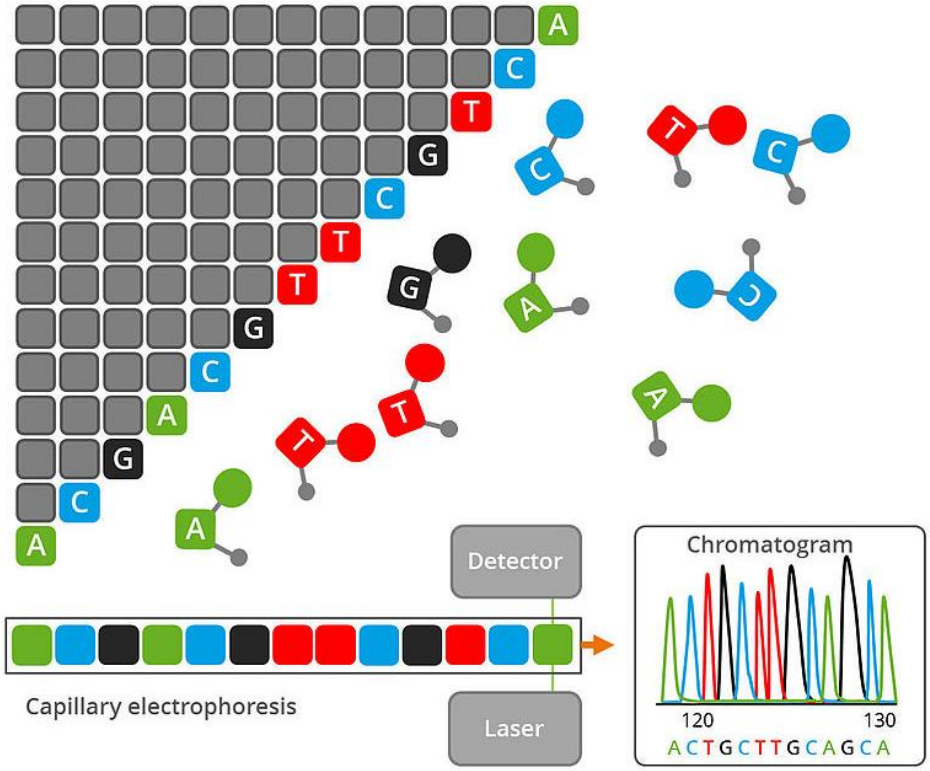
provides greater level of detail on the organism

- phenotypic traits can vary
  - difficulty in identification
  - difficulty in resistance detection
  - limited information about mechanism
- insight into relationships with other isolates
- lots of information can be determined from sequence
  - antimicrobial resistance (AR) genes
  - virulence factors
  - serotype …

Sanger method of sequencing

1977 – Frederick Sanger

- dideoxy chain-termination method



https://www.gatc-biotech.com/en/expertise/sanger-sequencing.html

Limitations of Sanger sequencing

limited to ~800bp for each run

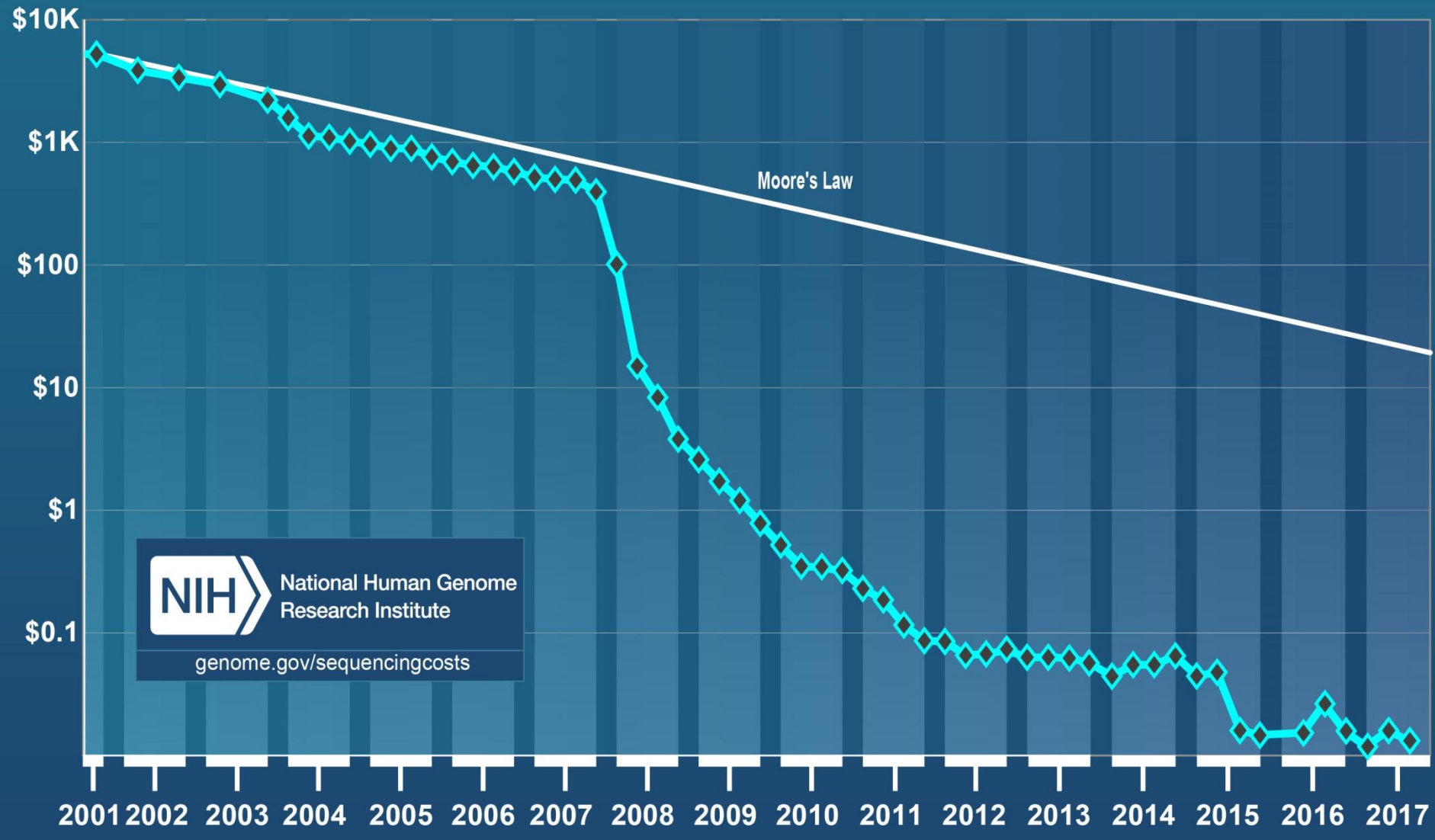
cost per bp is high

large amount of time for each base

highest possible accuracy

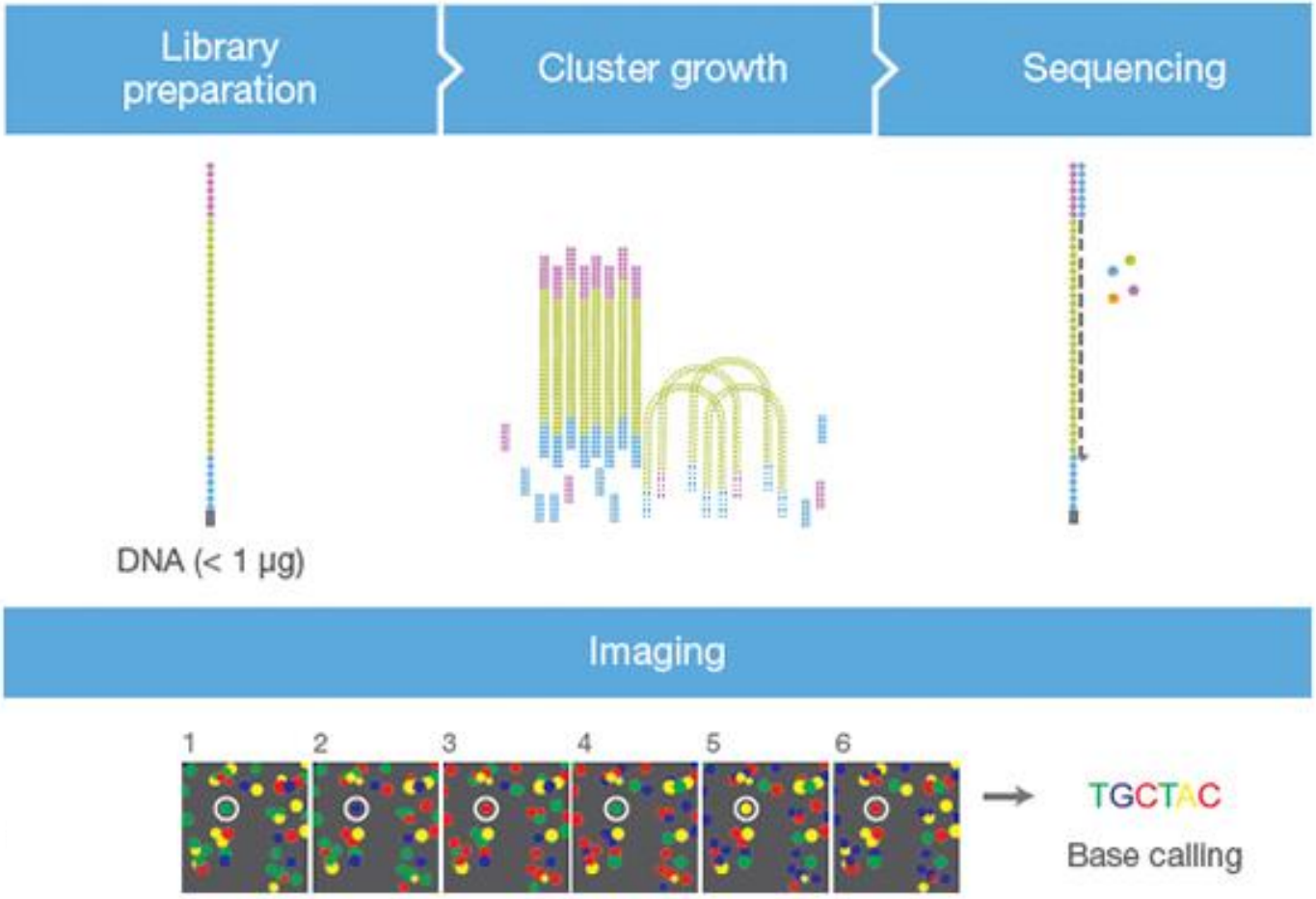
https://www.nature.com/news/human-genome-project-twenty-five-years-of-big-biology-1.18436

Cost per Raw Megabase of DNA Sequence



https://www.illumina.com

Illumina sequencing



https://www.illumina.com

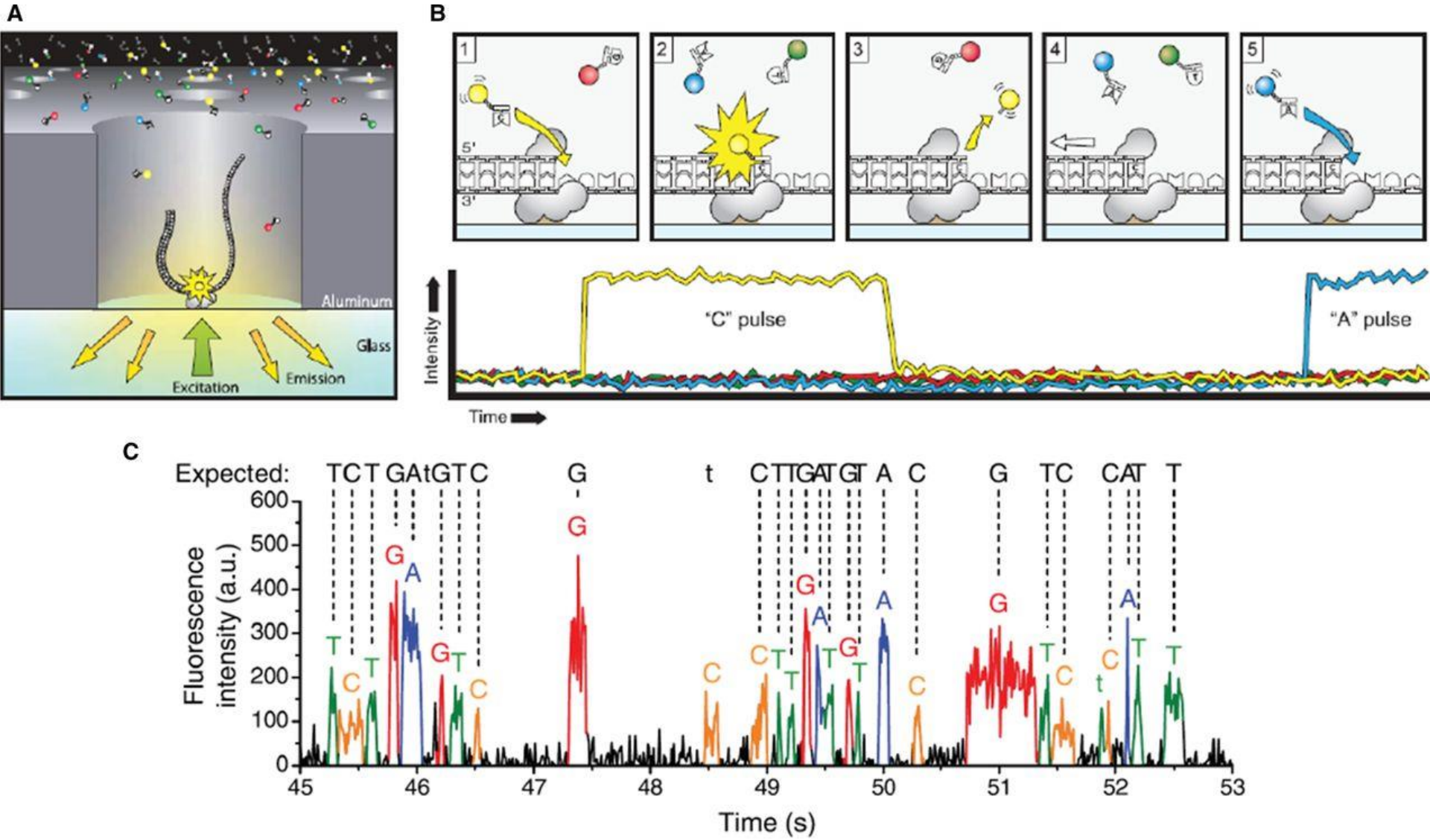
Next generation sequencing

- short read
  - 100bp to 600 bp reads
  - 1000s to millions of reads
  - 99.9999% accuracy
  - low cost per sample
  - unable to resolve larger structures

Next generation sequencing

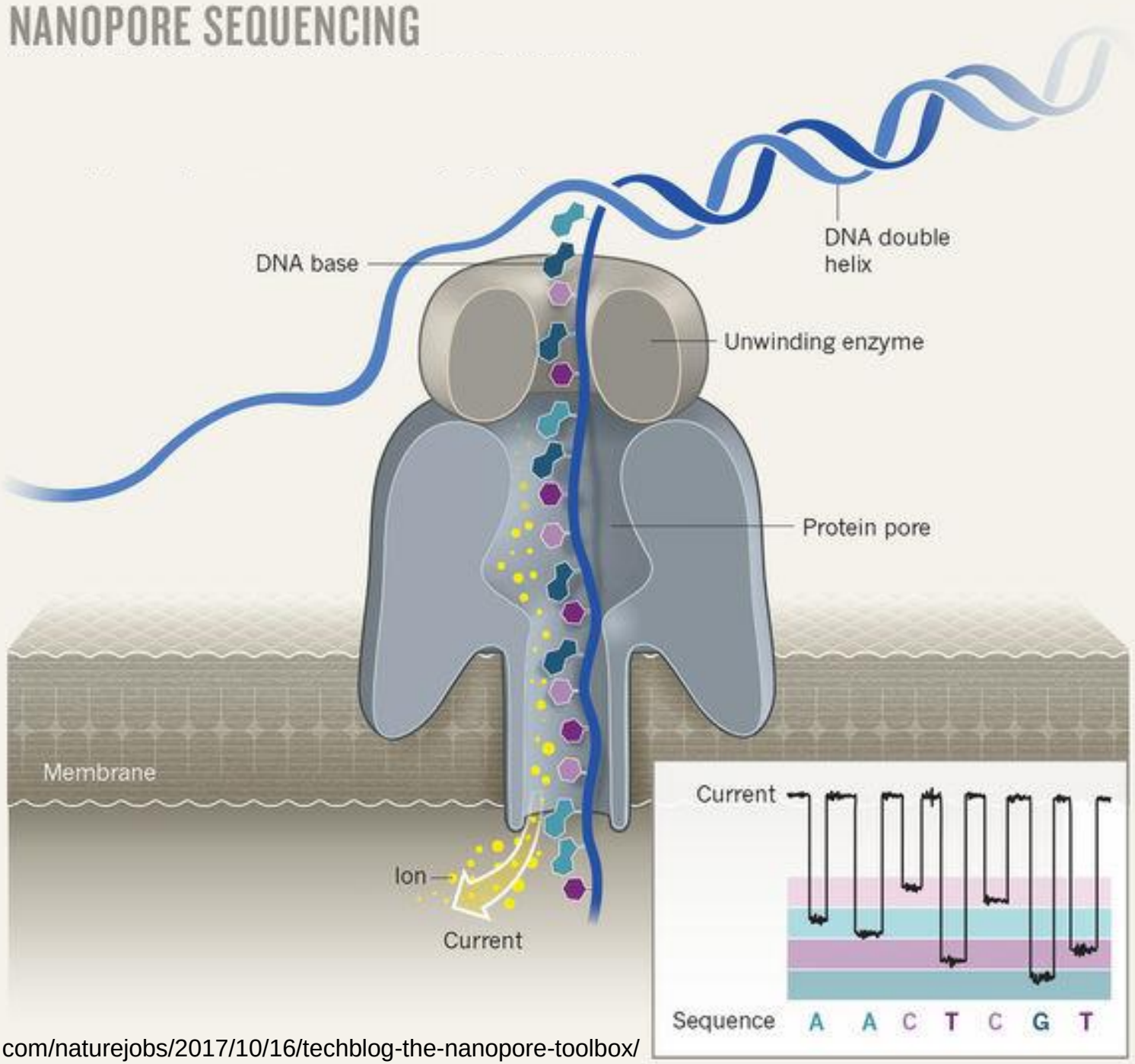
- short read
  - 100bp to 600 bp reads
  - 1000s to millions of reads
  - 99.9999% accuracy
  - low cost per sample
  - unable to resolve larger structures
- long read
  - 6kb to 2.3mb reads
  - thousands of reads
  - 80% to 90% accuracy
  - high cost per sample
  - can resolve larger structures

https://www.pacb.com/



Hill et al. Biocehmical Society Transactions 2017

https://nanoporetech.com/



http://blogs.nature.com/naturejobs/2017/10/16/techblog-the-nanopore-toolbox/

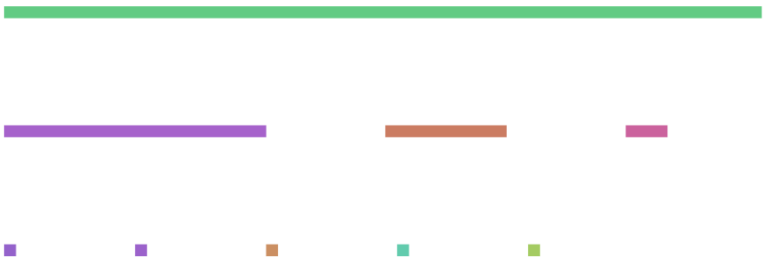
Why read length matters

short read

Why read length matters

short read

long read



Sequencing at the WSLH

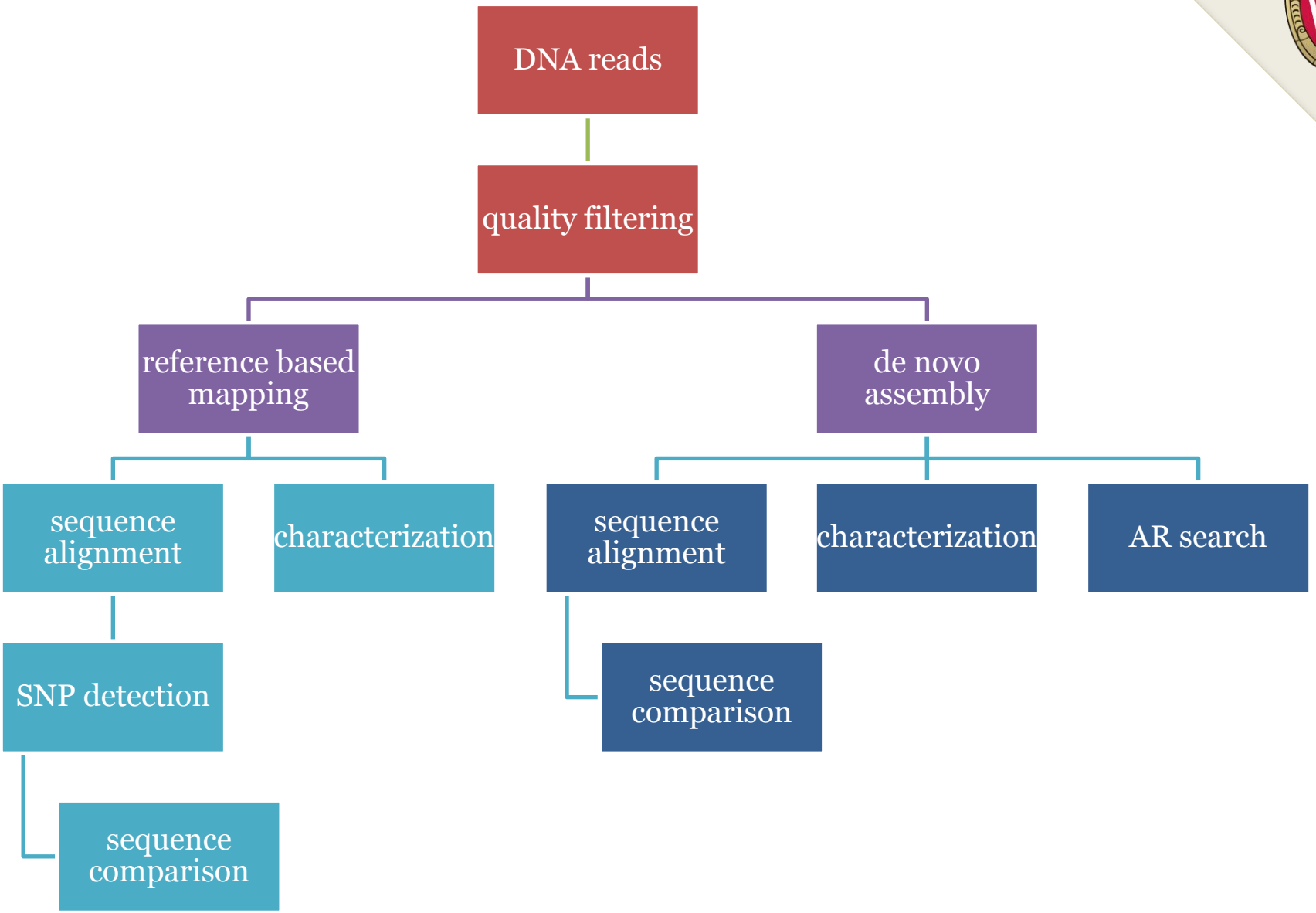
short read – illumina

divided into 2 components

- wet lab
- computational

sample processing and sequencing – 5 days

computational analysis – hours to weeks

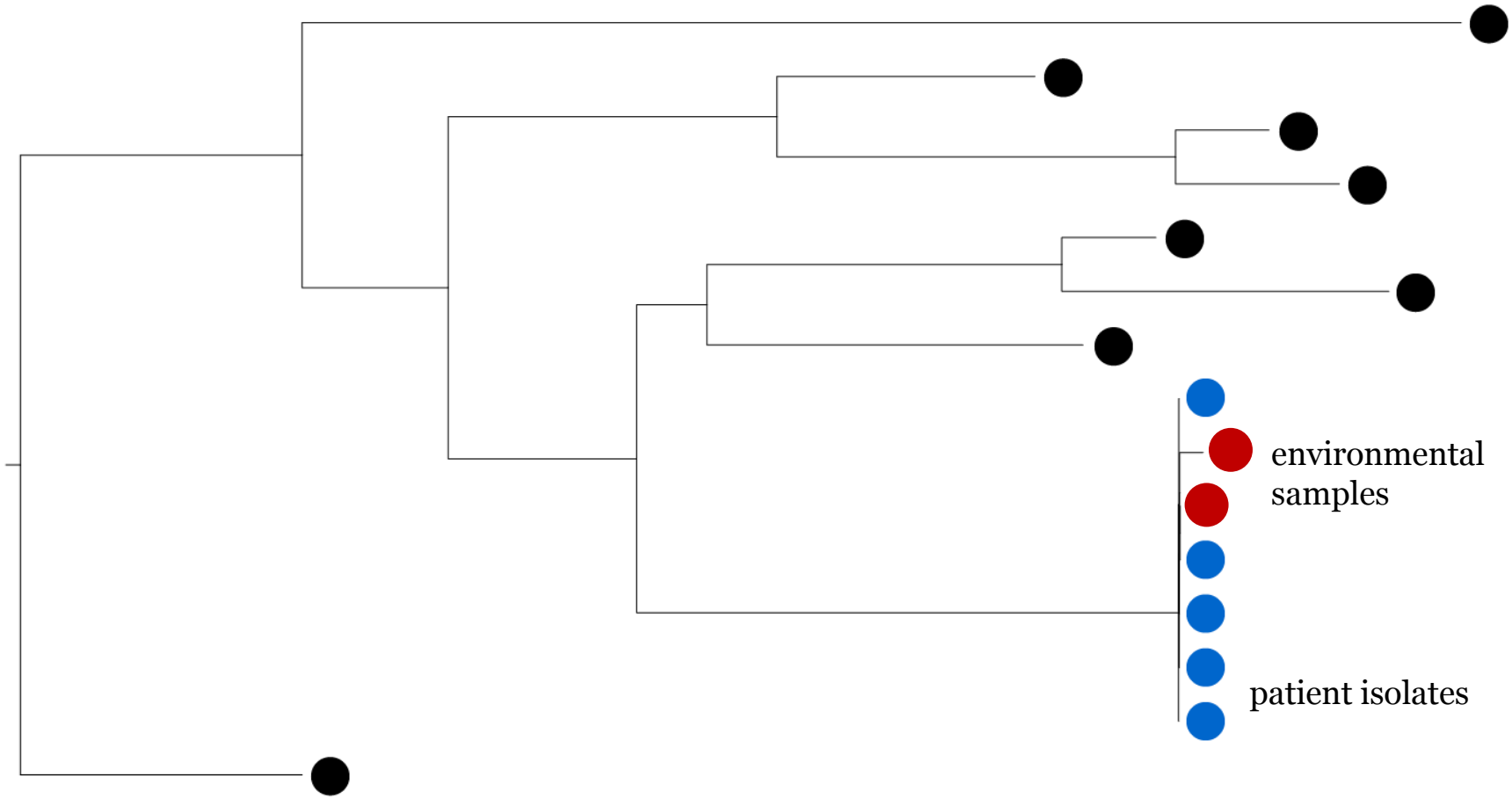

DNA reads
quality filtering
reference based mapping
de novo assembly
sequence alignment
characterization
sequence alignment
characterization
AR search
SNP detection
sequence comparison
sequence comparison

Single nucleotide polymorphisms (SNPs)

Reference	AATGCACCGATCGTCGATCGCTGCTAGCC
Sample 1	AATGCACCGATCGTCGATTGCTGCTAGCC
Sample 2	AATGCACCGTTCGTCGATTGCTGCTAGCC
Sample 3	AATGCACCGATCGTCGATCGCTGCTAGCC
Sample 4	AATGCACCGGTCGTCGATCGCTGCTAGCC
Sample 5	AATGCACCGATCGTCGATCGCTGCTAGCC

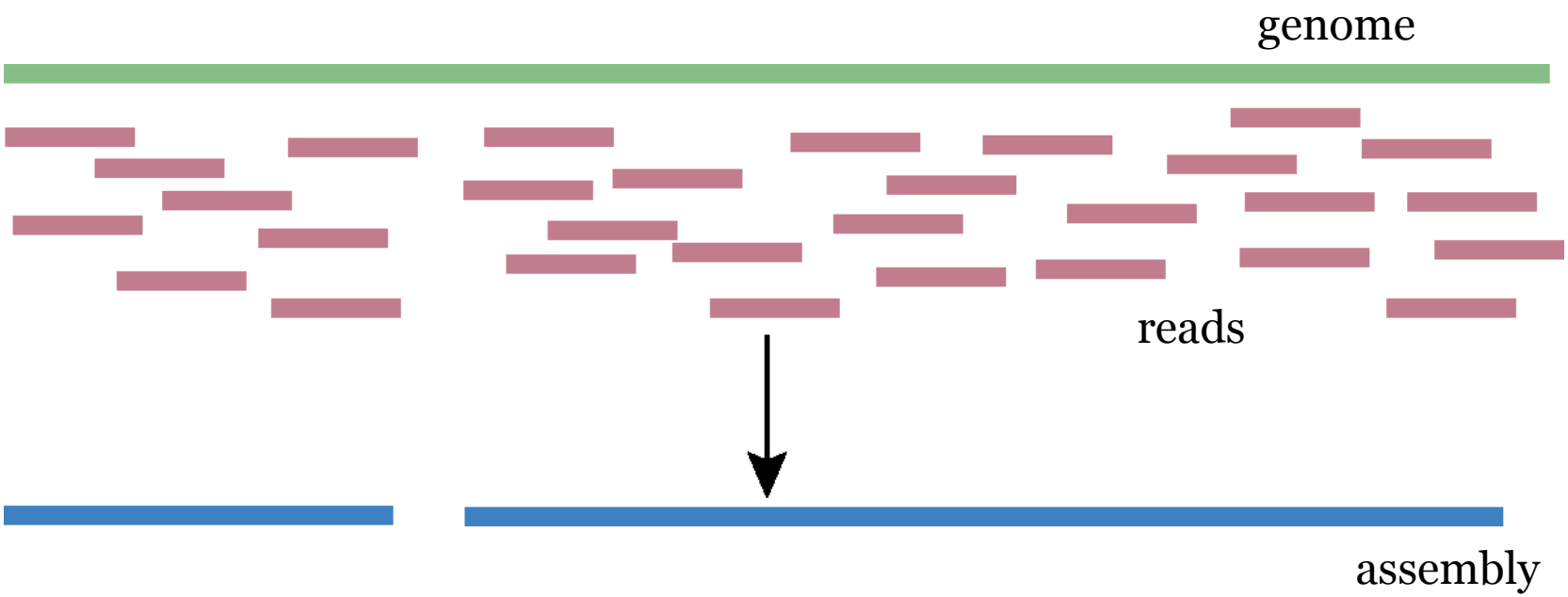
differences in SNPs allow for comparisons between organisms

Phylogenetic comparison of SNPs



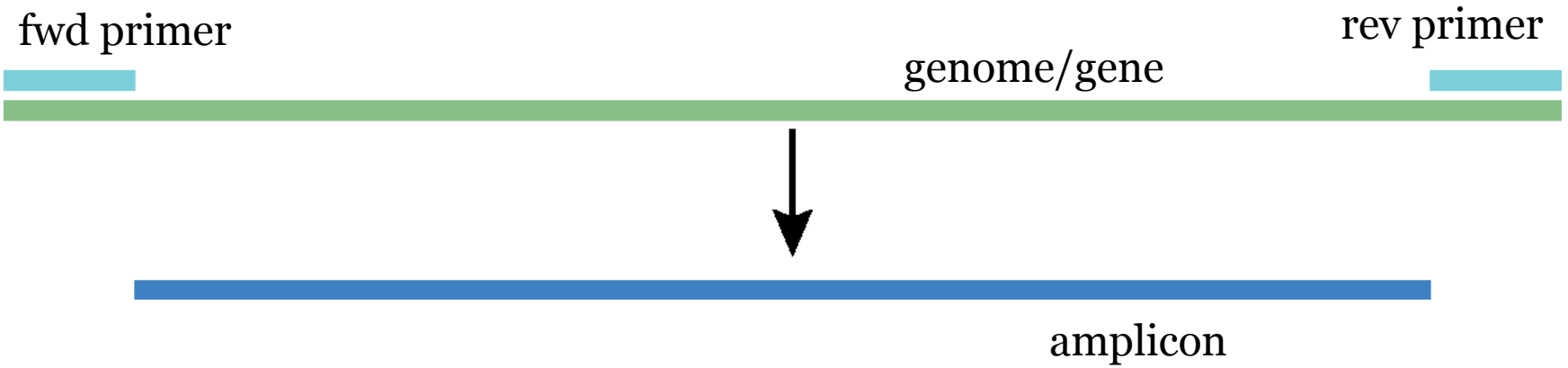
Methods of sequencing

Shotgun whole genome



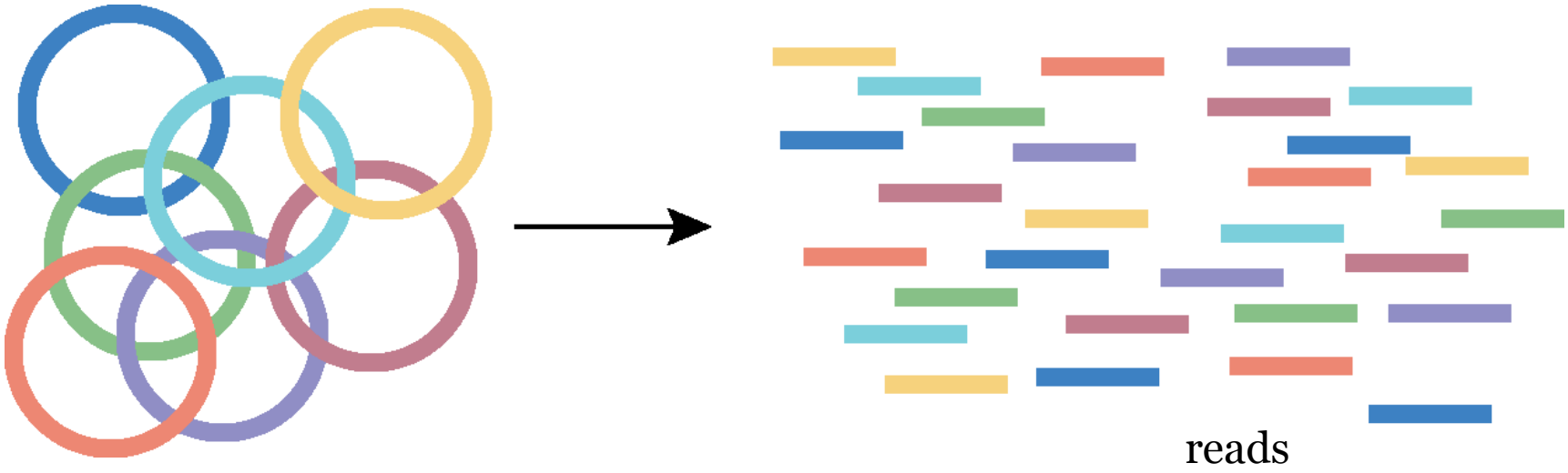
Methods of sequencing

Amplicon



Methods of sequencing

Metagenomic



Applications of next-generation sequencing

(clinical and public health microbiology)

Clinical microbiology

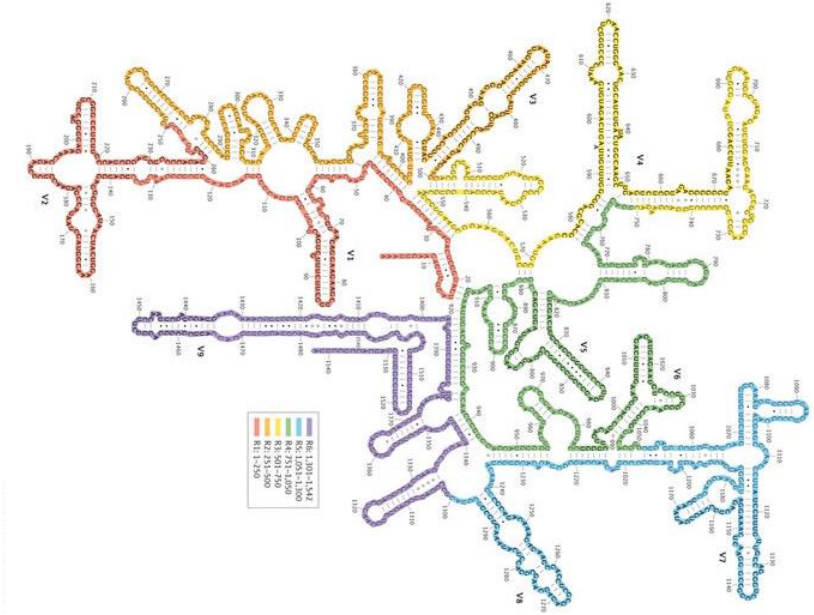
- 16S deep sequencing for mixed infections
- Metagenomic CSF sequencing
- Viral resistance testing (HIV)
- Hospital outbreaks

Public health microbiology

- Foodborne disease outbreak detection
- Influenza surveillance

16S deep sequencing for mixed infections

- 16S rRNA, informative region for evolutionary relatedness of bacteria
- Highly variable region of DNA, flanked on either side by highly conserved regions (for primer binding)



https://www.nature.com/articles/nrmicro3330

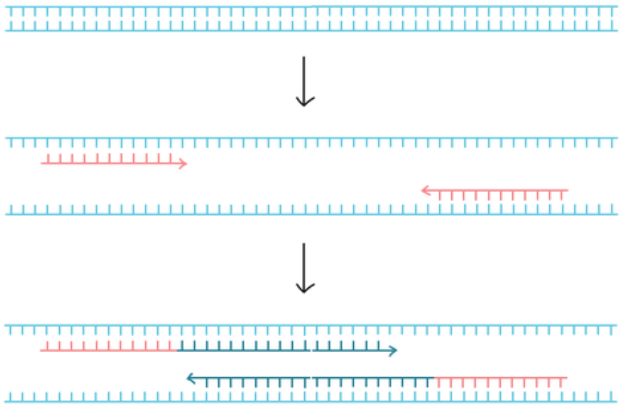
Clarridge JE, Clin Micro Rev '04. Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases

16S deep sequencing for mixed infections

- 16S rRNA, informative region for evolutionary relatedness of bacteria
- Highly variable region of DNA, flanked on either side by highly conserved regions (for primer binding)

<u>Steps in Sanger 16S sequencing</u>

1. purify DNA
2. PCR amplify the 16S rRNA gene
3. Sequence the PCR product
4. Compare sequence to database
5. Identify bacterial species



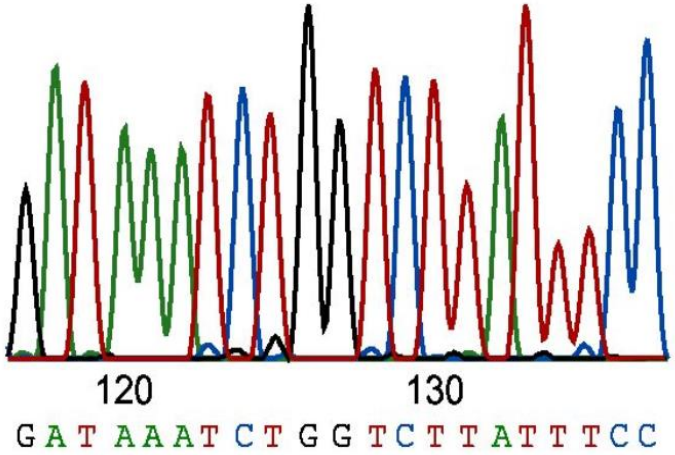
https://www.khanacademy.org/science/biology/biotech-dna-technology/dna-sequencing-pcr-electrophoresis/a/polymerase-chain-reaction-pcr

16S deep sequencing for mixed infections

- 16S rRNA, informative region for evolutionary relatedness of bacteria
- Highly variable region of DNA, flanked on either side by highly conserved regions (for primer binding)

<u>Steps in Sanger 16S sequencing</u>

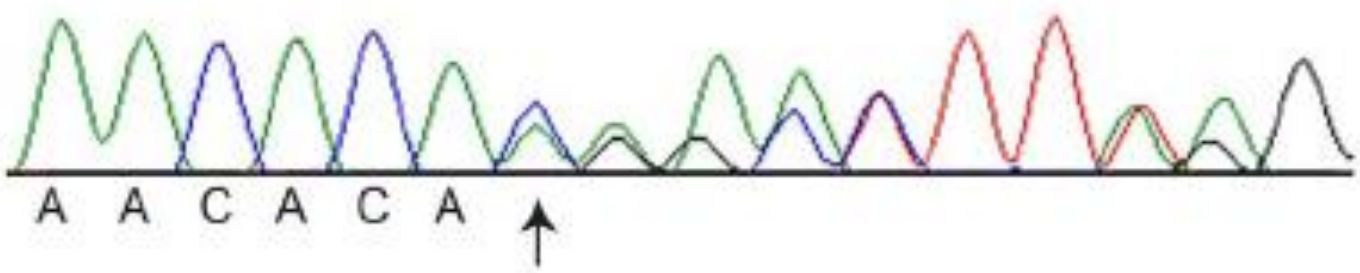
1. purify DNA
2. PCR amplify the 16S rRNA gene
3. Sequence the PCR product
4. Compare sequence to database
5. Identify bacterial species



https://bitesizebio.com/27985/sanger-sequencing-genome-won/

16S deep sequencing for mixed infections

- Sanger 16S sequencing, mixed infections



Lipska M, PLoS ONE '13. A Frameshift Mutation in the Cubilin Gene (CUBN) in Border Collies with Imerslund-Gräsbeck Syndrome

OPEN ACCESS Freely available online

PLOS | ONE

Rapid 16S rRNA Next-Generation Sequencing of Polymicrobial Clinical Samples for Diagnosis of Complex Bacterial Infections

Stephen J. Salipante[1,2*], Dhruba J. Sengupta[1], Christopher Rosenthal[1], Gina Costa[4], Jessica Spangler[4], Elizabeth H. Sims[3], Michael A. Jacobs[3], Samuel I. Miller[3], Daniel R. Hoogestraat[1], Brad T. Cookson[1,3], Connor McCoy[5], Frederick A. Matsen[5], Jay Shendure[2], Clarence C. Lee[4], Timothy T. Harkins[4], Noah G. Hoffman[1*]

Steps in NGS 16S sequencing

1. purify DNA
2. PCR amplify the 16S rRNA gene
3. Sequence the PCR product
   (using next-generation sequencing)
4. Compare sequences to database
   Need a bioinformatics pipeline
5. Identify bacterial species

16S deep sequencing for mixed infections

JCM
Journals.ASM.org

CASE REPORT

Molecular Diagnosis of *Actinomadura madurae* Infection by 16S rRNA Deep Sequencing

Stephen J. Salipante,[a] Dhruba J. SenGupta,[a] Daniel R. Hoogestraat,[a] Lisa A. Cummings,[a] Bronwyn H. Bryant,[a] Catherine Natividad,[a] Stephanie Thielges,[a] Peter W. Monsaas,[a] Mimosa Chau,[a] Lindley A. Barbee,[c,d] Christopher Rosenthal,[a] Brad T. Cookson,[a,b] Noah G. Hoffman[a]

- 50 y.o. diabetic woman, 12-year history of gradually enlarging right foot with multiple draining lesions
- Punch biopsy: acute and chronic inflammation, filamentous structures consistent with aerobic actinomycetes
- Biopsy cultures overwhelmed with overgrowth of other organisms (*S. aureus*)
- FFPE block of biopsy, 16S rRNA Sanger sequencing: *S aureus*
- Performed 16S NGS analysis on FFPE specimen

TABLE 1 Deep-sequencing results

No. of reads	% of reads	Classification[a]
13,000	35.94	*Staphylococcus aureus**
4,402	12.17	*Alcaligenes faecalis*
3,049	8.43	*Acinetobacter bereziniae/A. guillouiae**
1,449	4.01	*Comamonas testosteroni*/C. thiooxidans**
1,084	3.00	*Acinetobacter lwoffii**
935	2.59	*Acinetobacter lwoffii*/A. psychrotolerans*
924	2.55	*Pseudomonas geniculata/P. hibiscicola; Stenotrophomonas maltophilia*
748	2.07	*Actinomadura madurae**
747	2.07	*Flavobacterium lindanitolerans*
738	2.04	*Escherichia coli*/E. fergusonii*; Shigella dysenteriae*/S. flexneri**
654	1.81	*Campylobacter concisus*
476	1.32	*Enterobacter cowanii; Escherichia coli*/E. hermannii; Shigella boydii/S. flexneri/S. sonnei*
458	1.27	*Staphylococcus epidermidis*
417	1.15	*Ochrobactrum anthropi*/O. cytisi*/O. lupini**
377	1.04	*Enterobacter asburiae*/E. cancerogenus*/E. cloacae/E; cowanii; Leclercia adecarboxylata*
4,120	11.39	≤99.0% match to a reference strain

- *S. aureus* reads were highly prevalent
- Many other organisms detected
- One of the classical agents of actinomycotic mycetoma, and consistent with the organism visualized histologically (filamentous, basophilic)
- *Actinomadura madurae* implicated as the cause of mycetoma; treated with TMP-SMX

16S deep sequencing for mixed infections

UNIVERSITY OF WASHINGTON DEPARTMENT OF LABORATORY MEDICINE
Molecular Diagnosis Microbiology Section

BROAD-RANGE PCR

___ AFB

___ Bacteria (reflex to NGS16S when multiple templates are present)

___ Standard Bacterial PCR only

___ Fungi

Sanger first; NGS if multiple templates present

Sanger only

Applications of next-generation sequencing

(clinical and public health microbiology)

Clinical microbiology

- 16S deep sequencing for mixed infections
- Metagenomic CSF sequencing
- Viral resistance testing (HIV)
- Hospital outbreaks

Public health microbiology

- Foodborne disease outbreak detection
- Influenza surveillance

Metagenomic CSF sequencing

- 14 y.o. boy with severe combined immunodeficiency
- Presented with fever and headache 3 times over 4 months
- Progressed to hydrocephalus and status epilepticus
  - Necessitated medically-induced coma
- Extensive diagnostic workup (including brain biopsy) unrevealing

Table S1. Diagnostic testing for potential microbial causes of the patient's meningoencephalitis*.

	Sample Type	First Hosp	Second Hosp	Third Hosp
M. pneumoniae PCR	CSF		—	
Histoplasma / *Blastomyces* antigen	CSF		—	
Bartonella PCR	CSF		—	
VZV / HHV-6 / HHV8 PCR	CSF		—	
Borrelia burgdorferi PCR	CSF		—	
Adenovirus / CMV / EBV PCR	CSF		—	—
HSV-1,2 PCR	CSF		—	—
Enterovirus PCR	CSF		—	—
Cryptococcal antigen	CSF		—	—
Mycobacterial culture	CSF		—	—
West Nile Virus IgG/IgM and PCR	CSF		—	—
CEV / EEEV / WEEV / SLEV IgG/IgM	CSF		—	—
Bacterial / fungal culture	CSF		—	—
Toxoplasma gondii PCR	CSF			—
Powassan virus PCR	CSF			—
Aspergillus antigen	CSF			0.11 (nl <0.5)
JCV / BKV / HHV7 PCR	CSF			—
Viral culture (including mumps culture)	CSF			—
16S bacterial rRNA PCR	CSF			— (x2)
Bacterial culture	Brain			—
Adenovirus / CMV / EBV / VZV PCR	Brain			—
Enterovirus PCR	Brain			—
Epstein-Barr virus PCR	Brain			—
Varicella zoster virus PCR	Brain			—
EBV / CMV PCR	Plasma	—		
Bacterial culture	Blood	—	—	—
Enterovirus PCR	Serum	—	—	—
Parvovirus B19 / HHV7 / BKV / JCV PCR	Serum			—
Blastomyces / *Histoplasma* / *Cryptococcus* antigen	Serum / Urine			—
BKV PCR**	Urine			+
Toxoplasma gondii PCR	Serum			—
Adenovirus / HTLV-1,II / HIV / HSV-1,2 PCR	Serum			—
16S bacterial rRNA PCR	Blood			—
Influenza A, B / RSV PCR	NP swab	—	—	
Respiratory Viral Panel (Luminex)**	NP swab		—	+(rhinovirus)
Mycoplasma pneumoniae PCR	OP swab			—
Bacterial culture**	Sputum			+ (MRSA)
Enterovirus PCR	Stool		—	—

Metagenomic CSF sequencing

- 14 y.o. boy with severe combined immunodeficiency
- Presented with fever and headache 3 times over 4 months
- Progressed to hydrocephalus and status epilepticus
  - Necessitated medically-induced coma
- Extensive diagnostic workup (including brain biopsy) unrevealing
- CSF subjected to unbiased next-generation sequencing
  - No PCR amplification
  - Extract and sequence all DNA from CSF
  - Most is human DNA (human cells)
  - A small amount could be a pathogen
- Over 3 million reads total; 475 (0.016%) reads were *Leptospira*
- *Leptospira-specific* PCR also positive
- Patient treated with 7-day course of intravenous penicillin G
- Gradually recovered over the next 7 days, with resolution of status epilepticus, normalization of CSF, and resolution of leptomeningitis
- Discharged 14 days after completing treatment

Metagenomic CSF sequencing

The NEW ENGLAND JOURNAL of MEDICINE

BRIEF REPORT

Actionable Diagnosis of Neuroleptospirosis by Next-Generation Sequencing

Michael R. Wilson, M.D., Samia N. Naccache, Ph.D., Erik Samayoa, B.S., C.L.S., Mark Biagtan, M.D., Hiba Bashir, M.D., Guixia Yu, B.S., Shahriar M. Salamat, M.D., Ph.D., Sneha Somasekar, B.S., Scot Federman, B.A., Steve Miller, M.D., Ph.D., Robert Sokolic, M.D., Elizabeth Garabedian, R.N., M.S.L.S., Fabio Candotti, M.D., Rebecca H. Buckley, M.D., Kurt D. Reed, M.D., Teresa L. Meyer, R.N., M.S., Christine M. Seroogy, M.D., Renee Galloway, M.P.H., Sheryl L. Henderson, M.D., Ph.D., James E. Gern, M.D., Joseph L. DeRisi, Ph.D., and Charles Y. Chiu, M.D., Ph.D.

CHANNEL3000.com
WISC-TV • NEWS 3 • MADISON

10 Day Fo

WEATHER

HOME NEWS WEATHER TRAFFIC SPORTS ENTERTAINMENT LIFESTYLE OBITS MADISON M

WEATHER ALERT There is 1 area under alert.

Local News

Experimental DNA test helps diagnose Cottage Grove teen

Next-generation genomic sequencing tests for viruses, bacteria, fungi

By: Mary Ola

Posted: Jun 12, 2014 02:47 PM CDT Updated: Nov 15, 2016 10:07 PM CST

COTTAGE GROVE, Wis. - This spring Cottage Grove teenager Joshua Osborn will most likely be found on his trampoline, which is a welcome change from last summer when the 15-year-old was hospitalized and in a coma.

https://www.channel3000.com/news/local-news/experimental-dna-test-helps-diagnose-cottage-grove-teen_20161116040713183/162917534

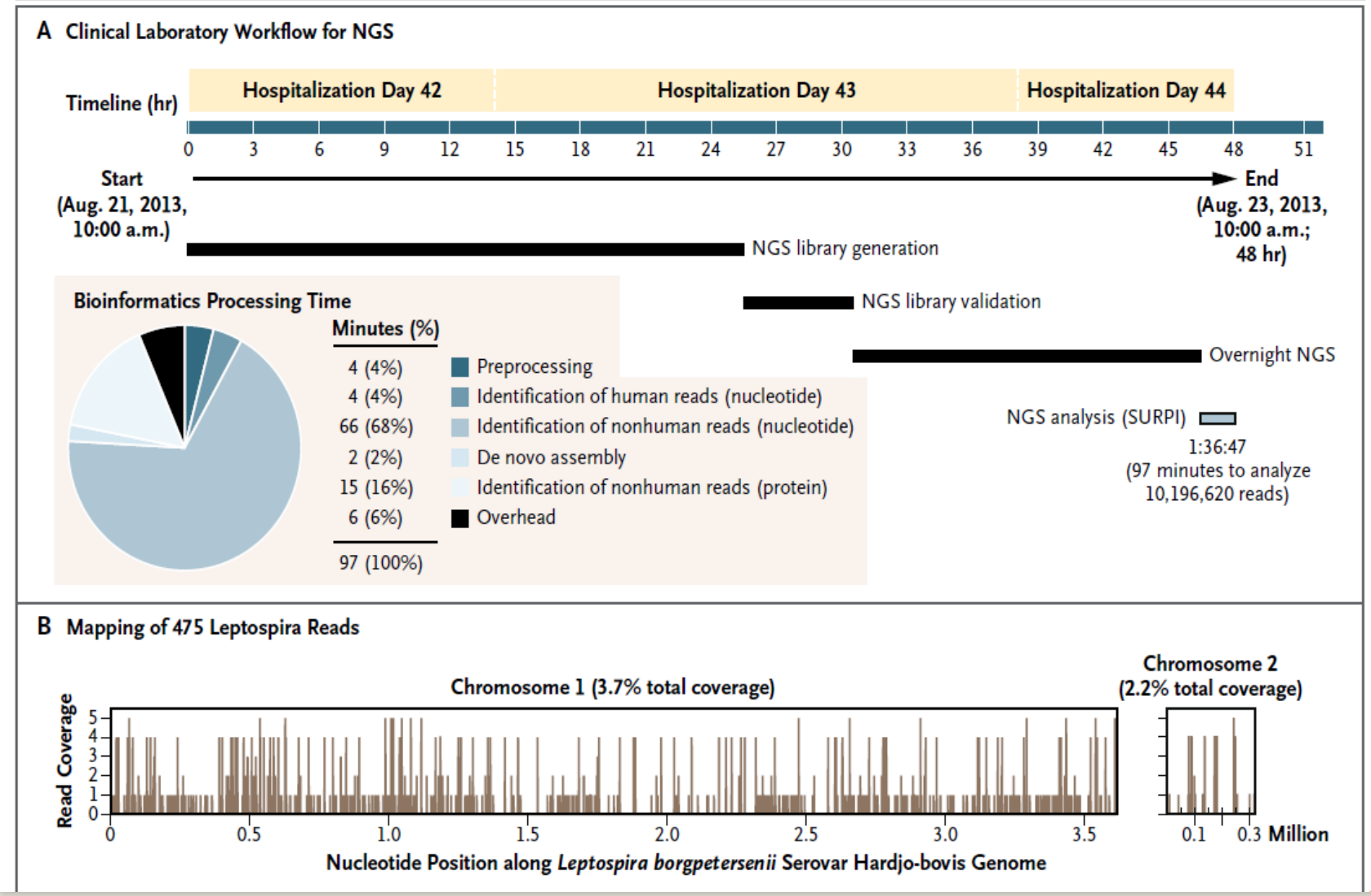

A Clinical Laboratory Workflow for NGS
Timeline (hr)
Hospitalization Day 42
Hospitalization Day 43
Hospitalization Day 44
0 3 6 9 12 15 18 21 24 27 30 33 36 39 42 45 48 51
Start (Aug. 21, 2013, 10:00 a.m.)
End (Aug. 23, 2013, 10:00 a.m.; 48 hr)
NGS library generation
NGS library validation
Overnight NGS
NGS analysis (SURPI)
1:36:47 (97 minutes to analyze 10,196,620 reads)
Bioinformatics Processing Time
Minutes (%)
4 (4%) Preprocessing
4 (4%) Identification of human reads (nucleotide)
66 (68%) Identification of nonhuman reads (nucleotide)
2 (2%) De novo assembly
15 (16%) Identification of nonhuman reads (protein)
6 (6%) Overhead
97 (100%)
B Mapping of 475 Leptospira Reads
Chromosome 1 (3.7% total coverage)
Chromosome 2 (2.2% total coverage)
Read Coverage
0 1 2 3 4 5
0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
0.1 0.3 Million
Nucleotide Position along Leptospira borgpetersenii Serovar Hardjo-bovis Genome

Table 1. Confirmatory Diagnostic Testing for Neuroleptospirosis.*

Assay†	Testing Site	Sample Type	Before Diagnosis‡	After Diagnosis‡	Date of Test Result
16S rRNA bacterial PCR assay	UW	CSF	Negative		July 14, 2013
16S rRNA bacterial PCR assay	UW	CSF	Negative		Aug. 12, 2013
16S rRNA bacterial PCR assay	UW	Serum		Negative	Aug. 24, 2013
Leptospira PCR assays targeting *lipL32*, *lipL41*, *ompA*, *rpoB*, and *secY*	UCSF	CSF		Positive	Aug. 28, 2013
Leptospira PCR assays targeting *lipL32*, *ompA*, and *secY*	UCSF	Serum		Negative	Aug. 28, 2013
Leptospira culture	CDC	CSF		Negative	Oct. 15, 2013
Leptospira PCR assay targeting *lipL32* with the use of a clinically validated assay[11]	CDC	CSF		Negative	Oct. 15, 2013
16S rRNA bacterial PCR assay	CDC	CSF		Negative	Oct. 15, 2013
Leptospira PCR assay targeting *lipL32* with the use of a clinically validated assay[11]	CDC	Serum		Negative	Oct. 15, 2013
Leptospira IgM antibody with the use of dot blot ELISA	CDC	Serum		Negative	Oct. 15, 2013
Leptospira IgM antibody with the use of dot blot ELISA	CDC	Serum		Negative (sample obtained on Oct. 9, 2013)	Oct. 17, 2013
Leptospira PCR assays targeting *lipL32*, *ompA*, and *secY*	UCSF	Brain		Negative	Oct. 17, 2013
Leptospira PCR assays targeting *lipL32*, *ompA*, and *secY*	UCSF	Serum		Negative	Oct. 31, 2013
Leptospira PCR assay targeting *lipL32* with the use of a clinically validated assay and a change in the amplification mix[11]	CDC	CSF		Positive	Jan. 16, 2014
Leptospira IgM antibody with the use of latex agglutination ELISA[12]	CDC	Serum		Positive (sample obtained on Oct. 9, 2013)	Feb. 6, 2014
Leptospira PCR assay targeting *lipL32* with the use of a clinically validated assay and a change in the amplification mix[11]	CDC	CSF		Negative (sample obtained on Feb. 5, 2014)	Feb. 24, 2014

Applications of next-generation sequencing

(clinical and public health microbiology)

Clinical microbiology

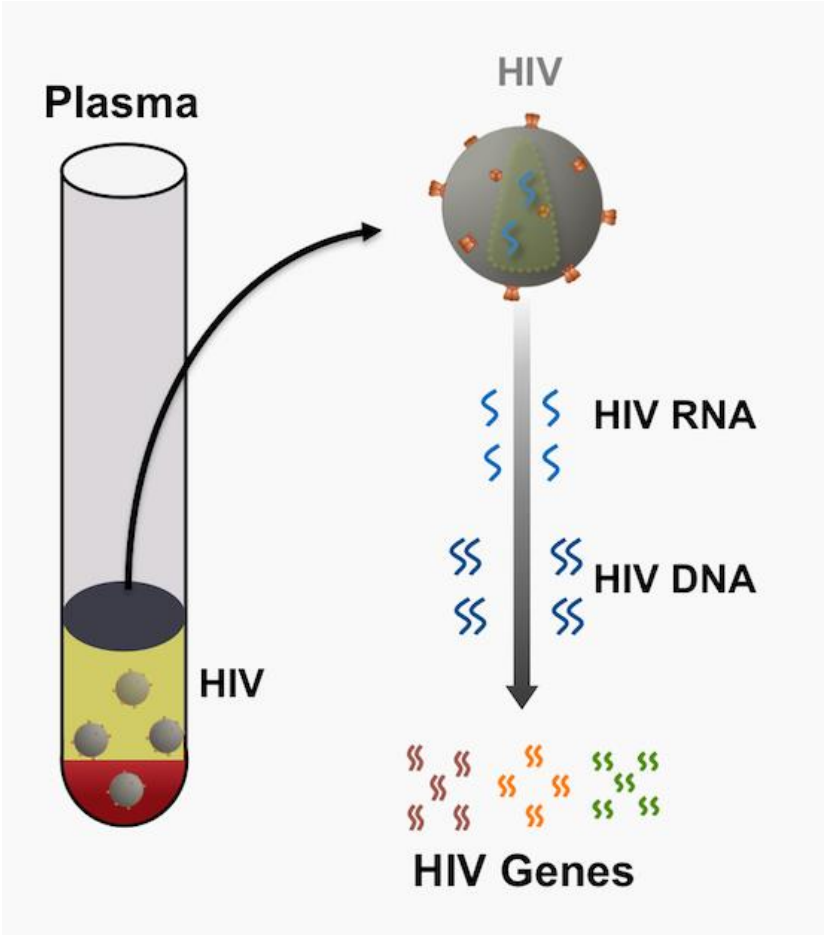
- 16S deep sequencing for mixed infections
- Metagenomic CSF sequencing
- Viral resistance testing (HIV)
- Hospital outbreaks

Public health microbiology

- Foodborne disease outbreak detection
- Influenza surveillance

Viral resistance testing

- Most common HIV resistance testing
  - Extract RNA from sample
  - PCR amplify certain regions (pol)
  - Sanger sequence the amplified regions
  - Use a database to identify mutations associated with drug resistance



https://www.hiv.uw.edu/go/antiretroviral-therapy/evaluation-management-virologic-failure/core-concept/all

Viral resistance testing

Sanger sequencing	Next-generation sequencing
Manual process (no multiplexing)	Multiplexing readily available
High volume, more expensive	High volume, can be cheaper
Identifies mutations where majority of virus has mutated	Can identify low-prevalence mutations (<10%)

Applications of next-generation sequencing

(clinical and public health microbiology)

Clinical microbiology

- 16S deep sequencing for mixed infections
- Metagenomic CSF sequencing
- Viral resistance testing (HIV)
- Hospital outbreaks

Public health microbiology

- Foodborne disease outbreak detection
- Influenza surveillance

Hospital outbreaks

- Whole-genome sequencing of bacterial isolates
- Unparalleled resolution
  - Every A, T, C, G visible
- Can replace pulsed-field gel electrophoresis (PFGE) for investigating outbreaks

PLOS | GENETICS

RESEARCH ARTICLE

A Year of Infection in the Intensive Care Unit: Prospective Whole Genome Sequencing of Bacterial Clinical Isolates Reveals Cryptic Transmissions and Novel Microbiota

Applications of next-generation sequencing

(clinical and public health microbiology)

Clinical microbiology

- 16S deep sequencing for mixed infections
- Metagenomic CSF sequencing
- Viral resistance testing (HIV)
- Hospital outbreaks

Public health microbiology

- Foodborne disease outbreak detection
- Influenza surveillance

PulseNet International:

On the path to implementing whole genome sequencing for foodborne disease surveillance

Why are PulseNet International members moving toward WGS?

- WGS is more precise and provides more detailed information about bacteria than traditional methods such as pulsed-field gel electrophoresis (PFGE) and multi-locus variable number tandem repeat analysis (MLVA).
- WGS will streamline workflow practices by replacing traditional methods for characterization of isolates, thereby making outbreak detection and characterization faster.
- WGS can be used for most foodborne disease organisms, including *Salmonella*, *Listeria*, *Campylobacter*, *Escherichia coli*, *Vibrio*, and *Shigella*.
- Data can be used across laboratories for routine surveillance, outbreak identification, source attribution, antimicrobial resistance prediction, and reference characterization.

https://www.cdc.gov/pulsenet/index.html

Applications of next-generation sequencing

(clinical and public health microbiology)

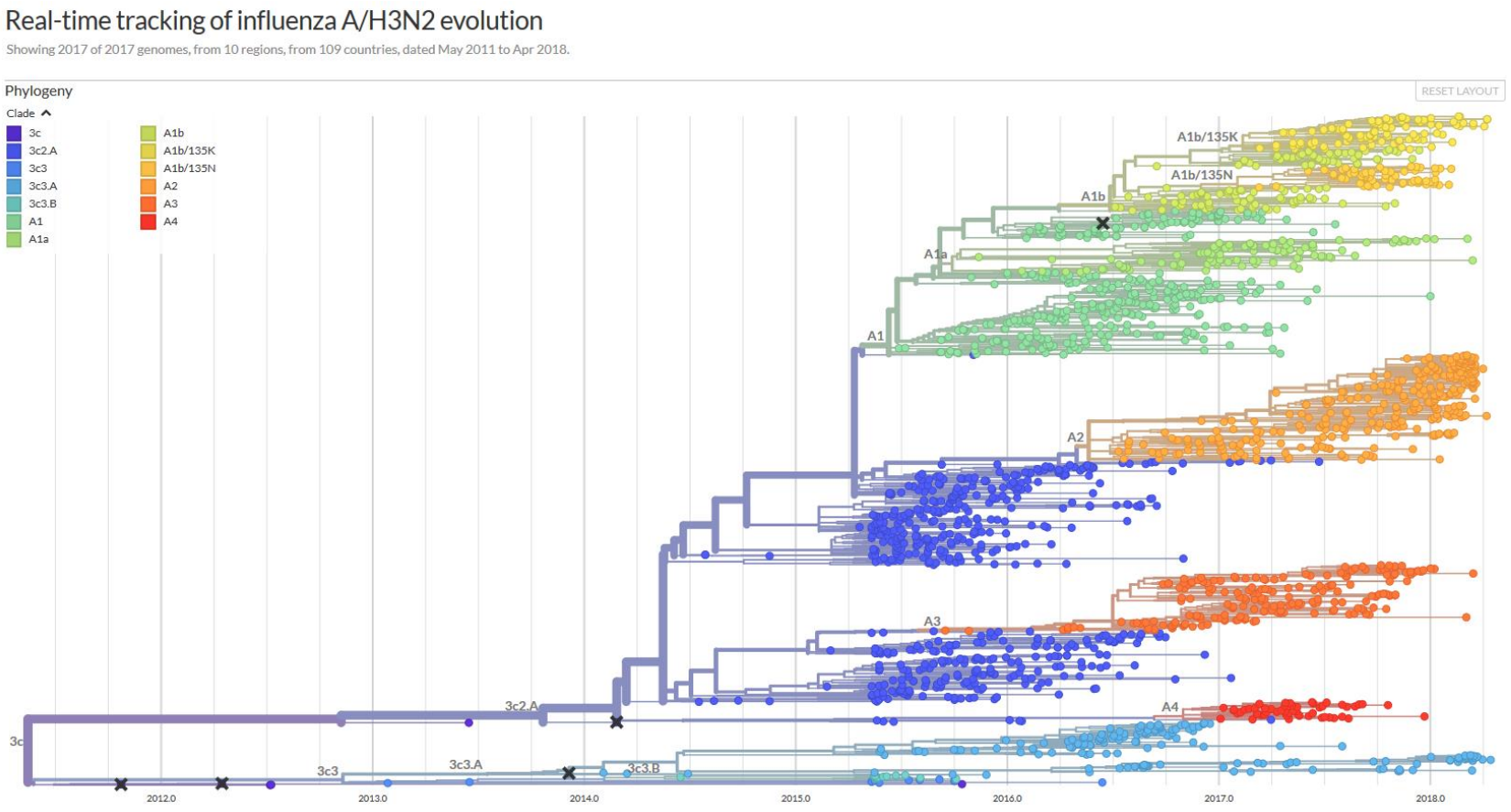
Clinical microbiology

- 16S deep sequencing for mixed infections
- Metagenomic CSF sequencing
- Viral resistance testing (HIV)
- Hospital outbreaks

Public health microbiology

- Foodborne disease outbreak detection
- Influenza surveillance

- nextflu.org uses all global influenza sequence information
- U.S. sequence data comes from NGS
  - PCR-based amplification of each influenza gene segment
  - Next-generation sequencing
  - Data uploaded through CDC bioinformatics pipeline
- Informing influenza vaccine strain selection and identifies drug resistant mutations



nextflu.org

Questions?