



Molecular & Genetic Surveillance

Dr. Rachel Sippy
University of St Andrews

Objectives

- Discuss the methods used to collect and analyze molecular data
- Describe the components and infrastructure to support molecular surveillance
- Discuss pathogens commonly studied with molecular surveillance
- Learn about examples of molecular surveillance in practice

Post Questions in the Chat!

(we will have breaks to answer these during the workshop)

Workshop Schedule

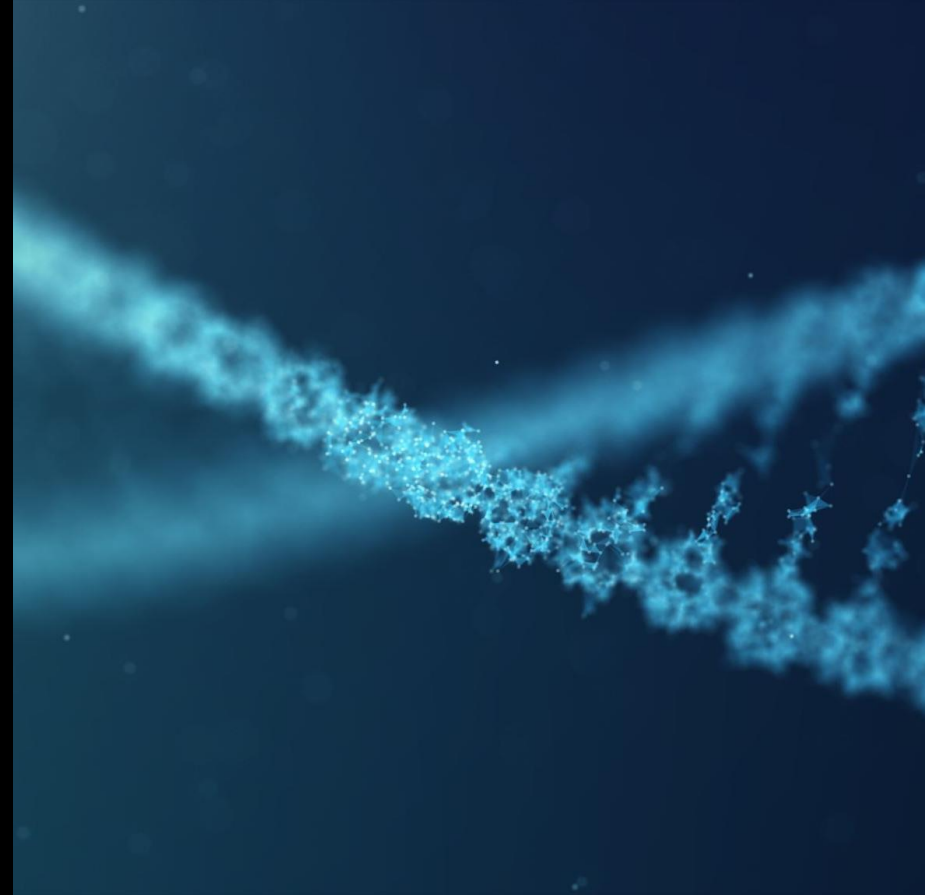
Time	Topics
2:00–2:30 pm	Molecular Surveillance & Infrastructure
2:30–2:45 pm	Pathogen Targets
2:45–3:05 pm	Examples of Molecular Surveillance in Use
3:05–3:15 pm	Break
3:15–4:00 pm	Tidyverse

Molecular & Genetic Surveillance

Data Generation and Infrastructure

Molecular Epidemiology & Surveillance

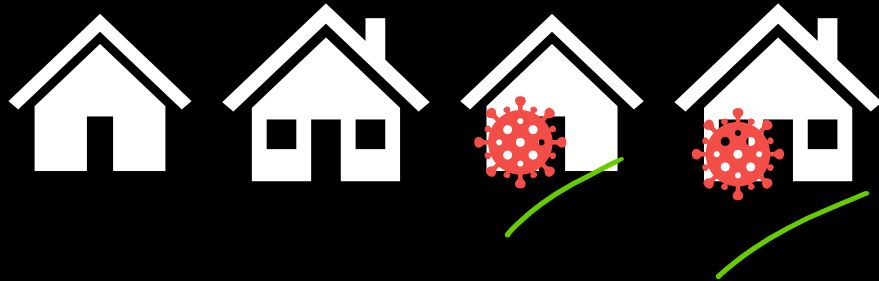
- Use of molecular information combined with traditional epidemiologic approaches
- Supplement surveillance programs and investigations
- Helps to:
 - detect antimicrobial resistance
 - identify emerging pathogens
 - find outbreaks/sources



Molecular Data for Investigations



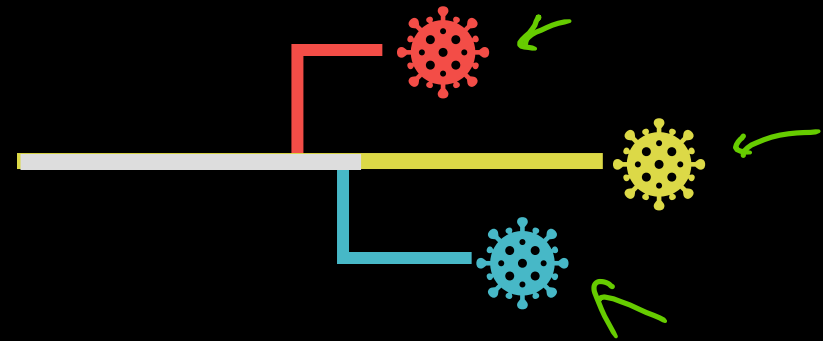
contact tracing



hotspots
super spreader events

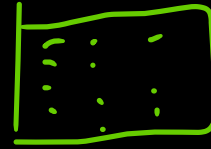


introductions of strains

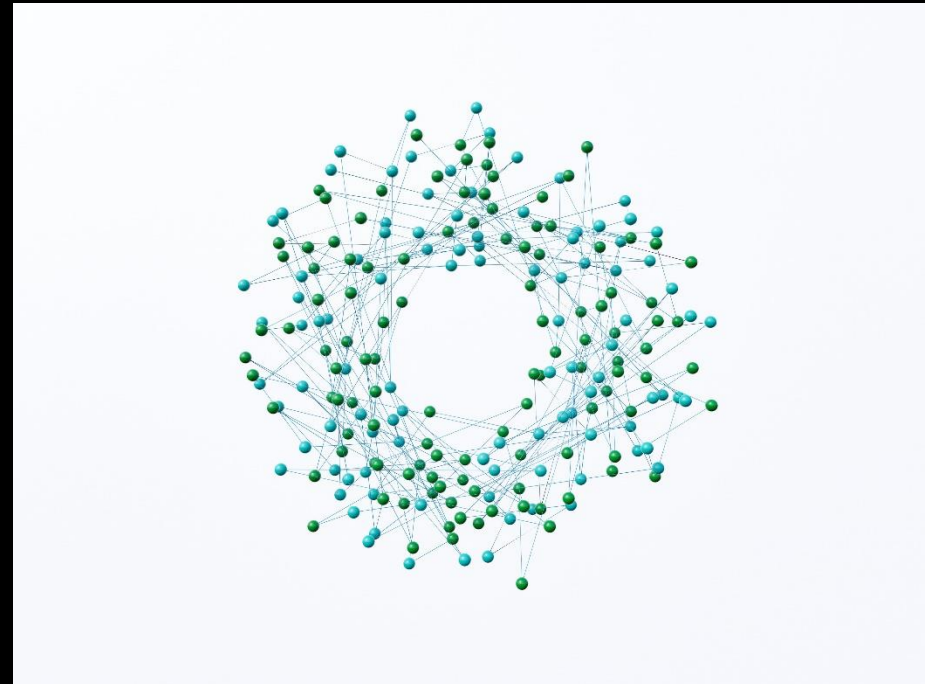


Molecular Data Types

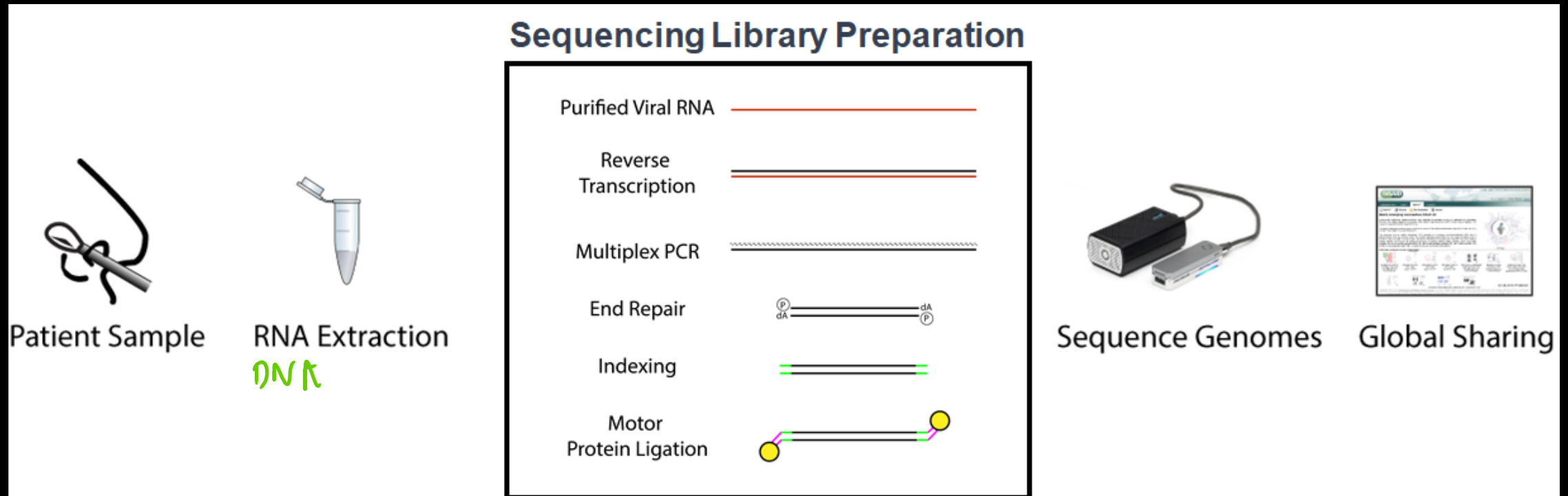
- Genetic profiles generated from sequencing or other laboratory analyses
 - whole or partial genomes
 - single or multiple genes
- Genetic profile is like a fingerprint for that pathogen
- Comparison of these profiles allows us to find relationships between individual samples



RNA/DNA *TGGG...-



RNA or DNA Whole Genome Sequencing



First Case

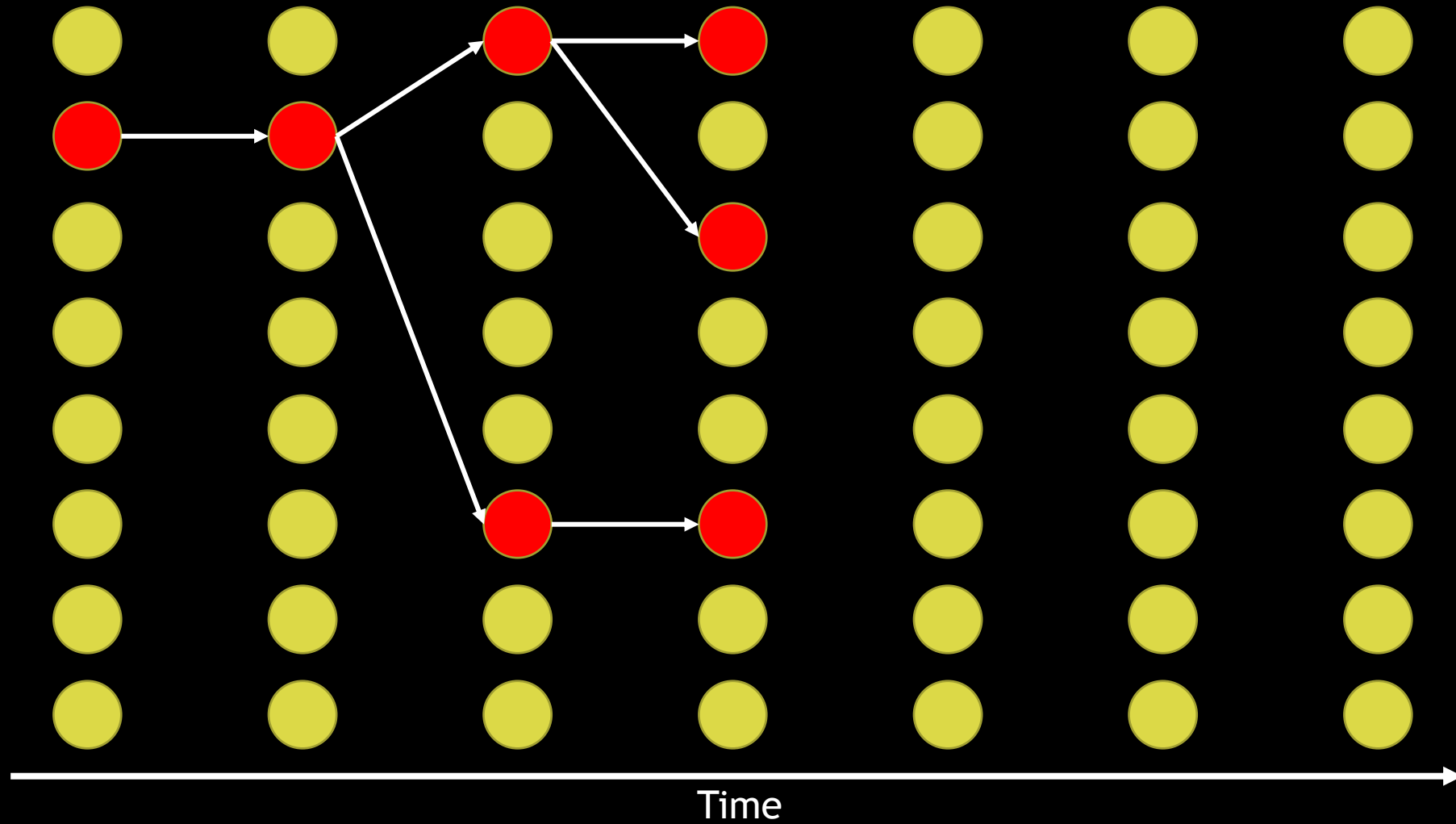
infected →



uninfected ↙

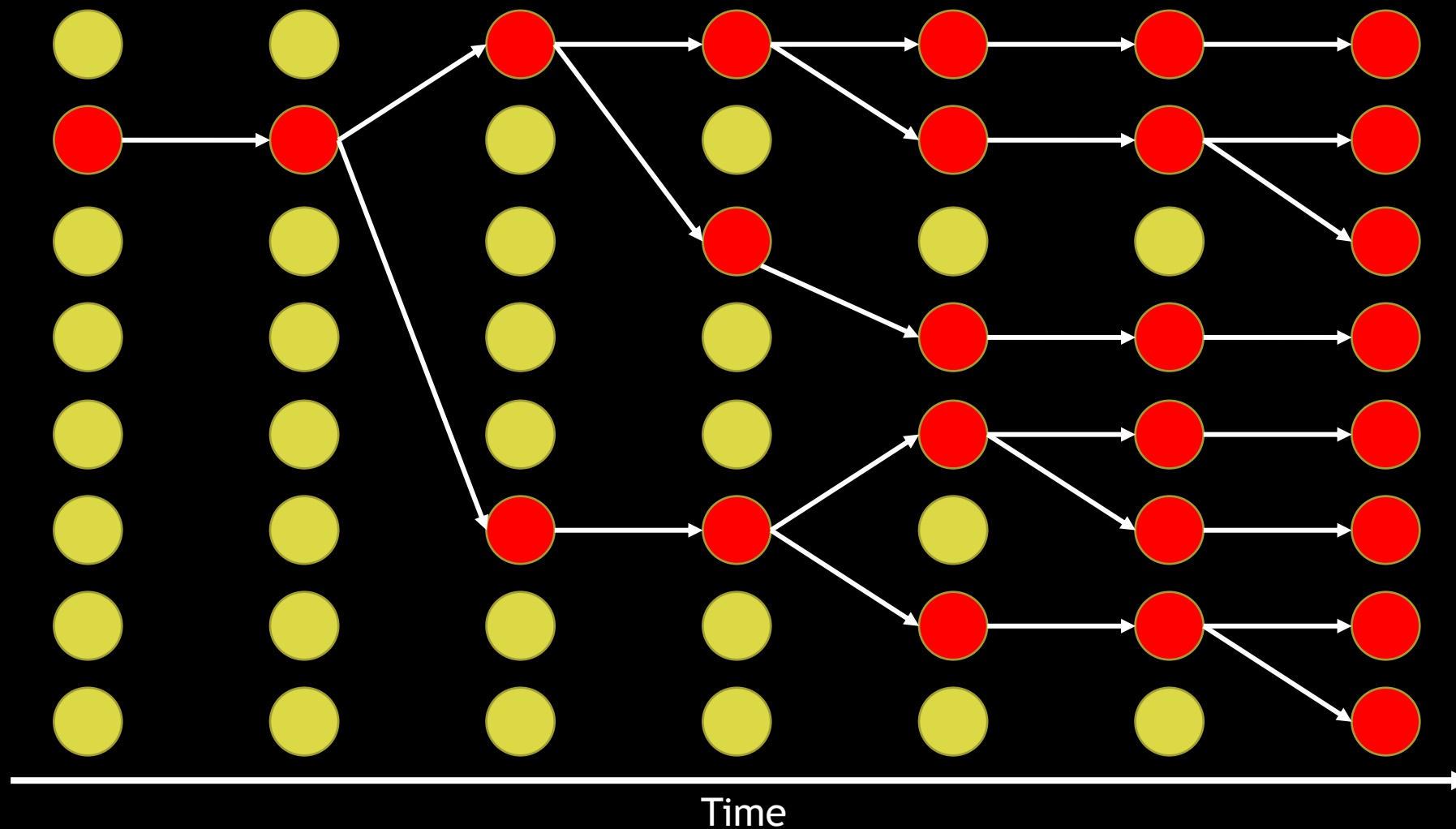
Time

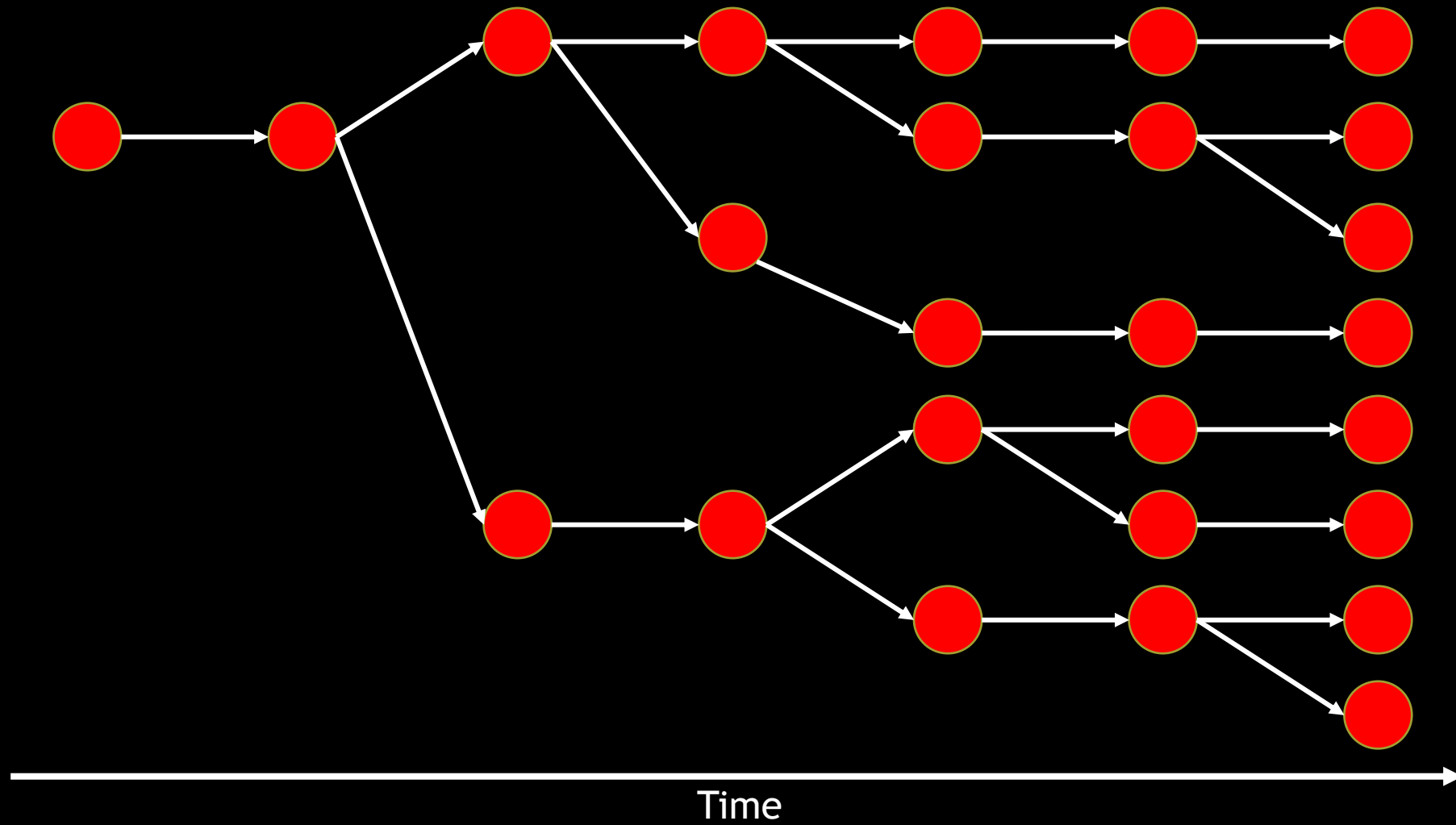
Spread

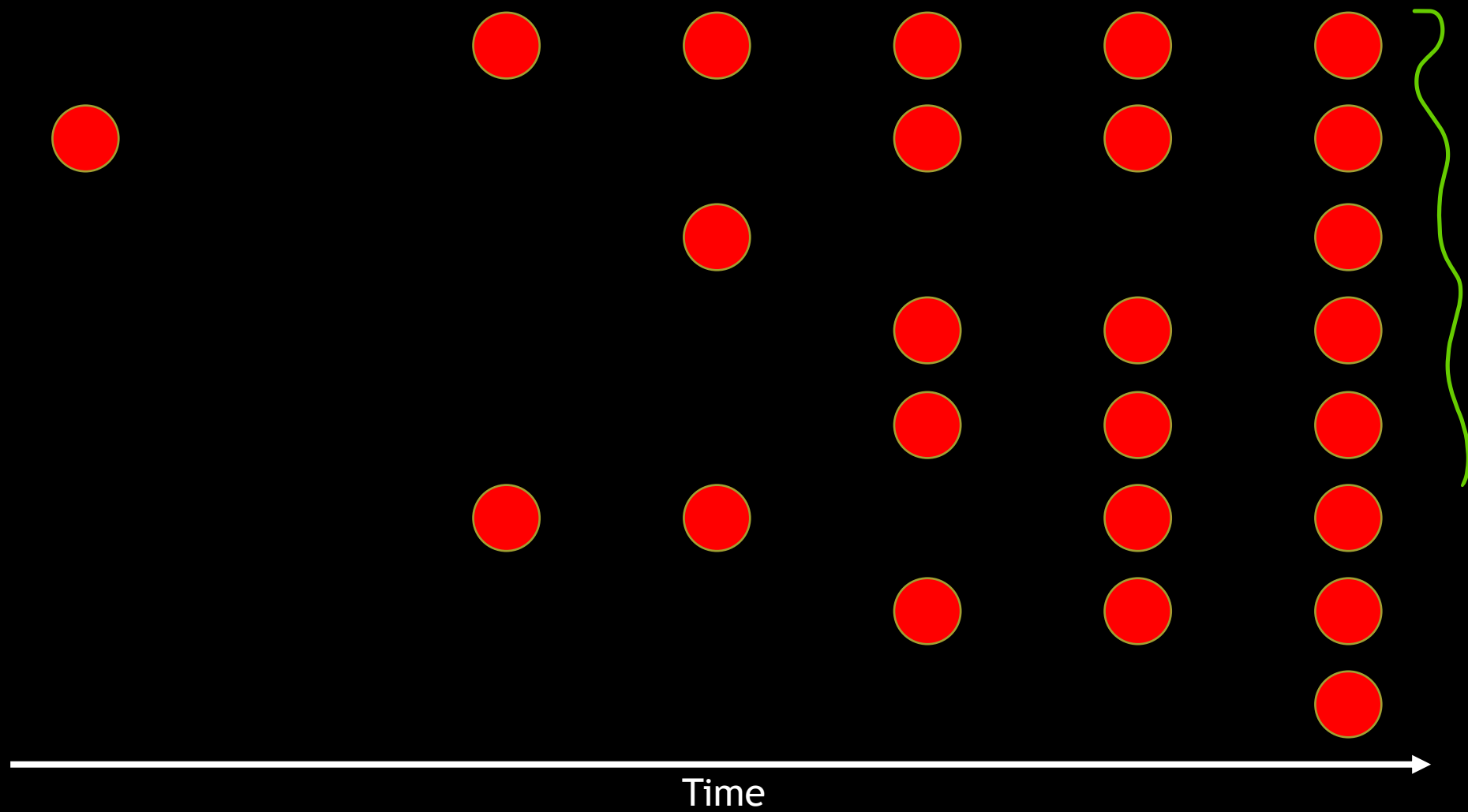


Spread

True
Reality: complete information



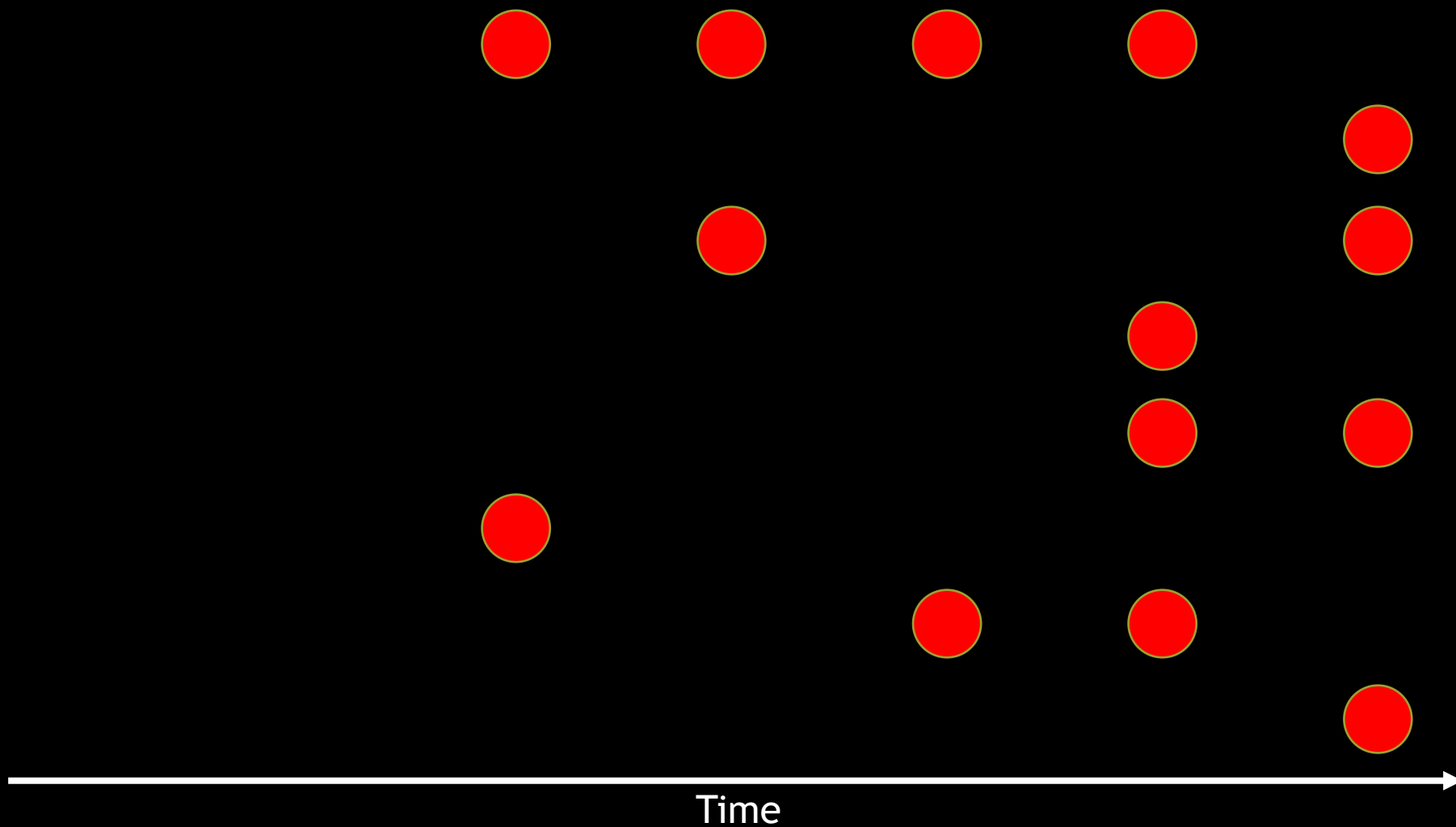




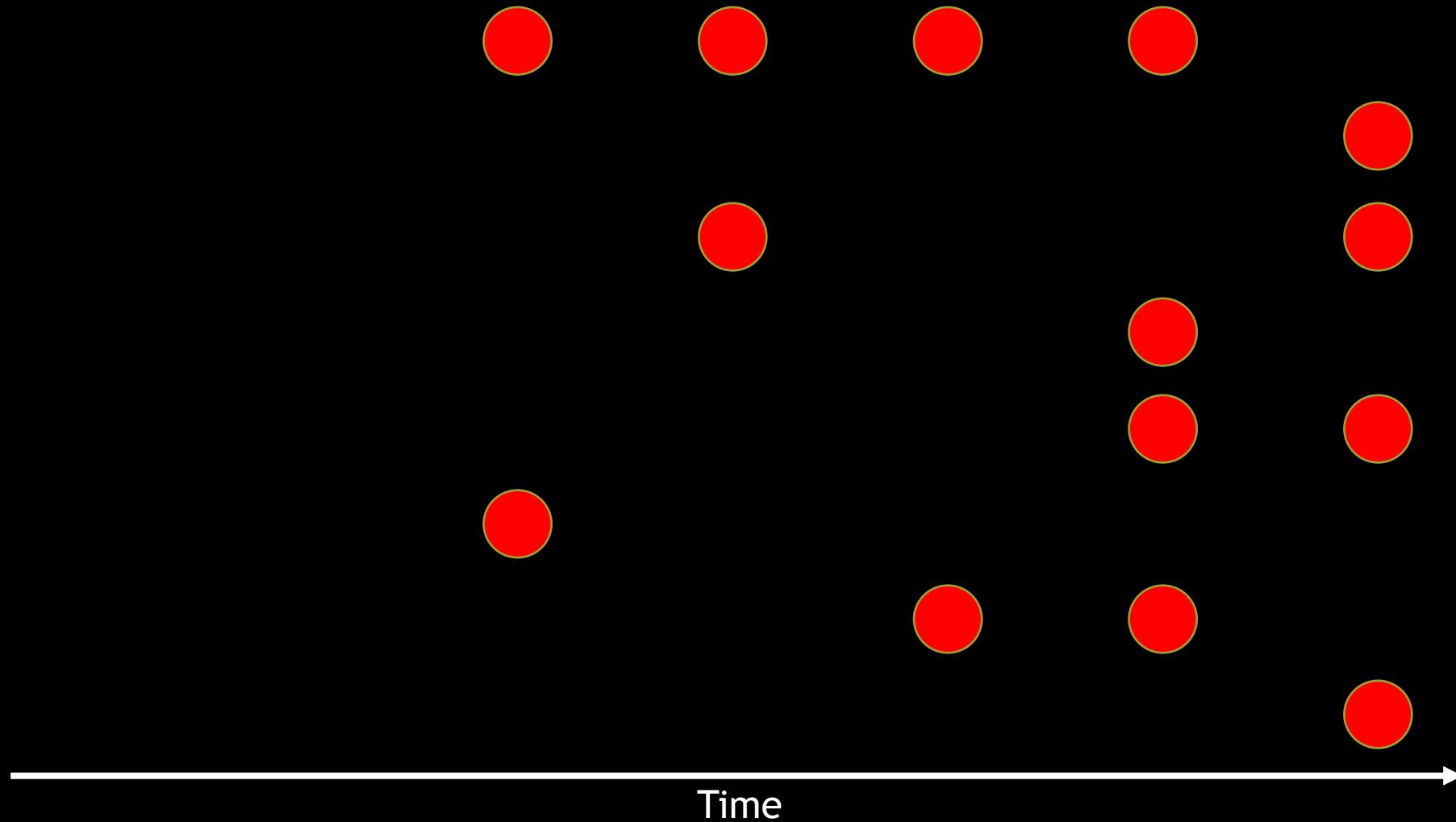
Sampling

- asymptomatics
- don't go to clinic

- clinic
sampling
limitations



How to Reconstruct the Epidemic?

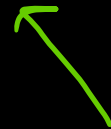


How to Reconstruct the Epidemic?

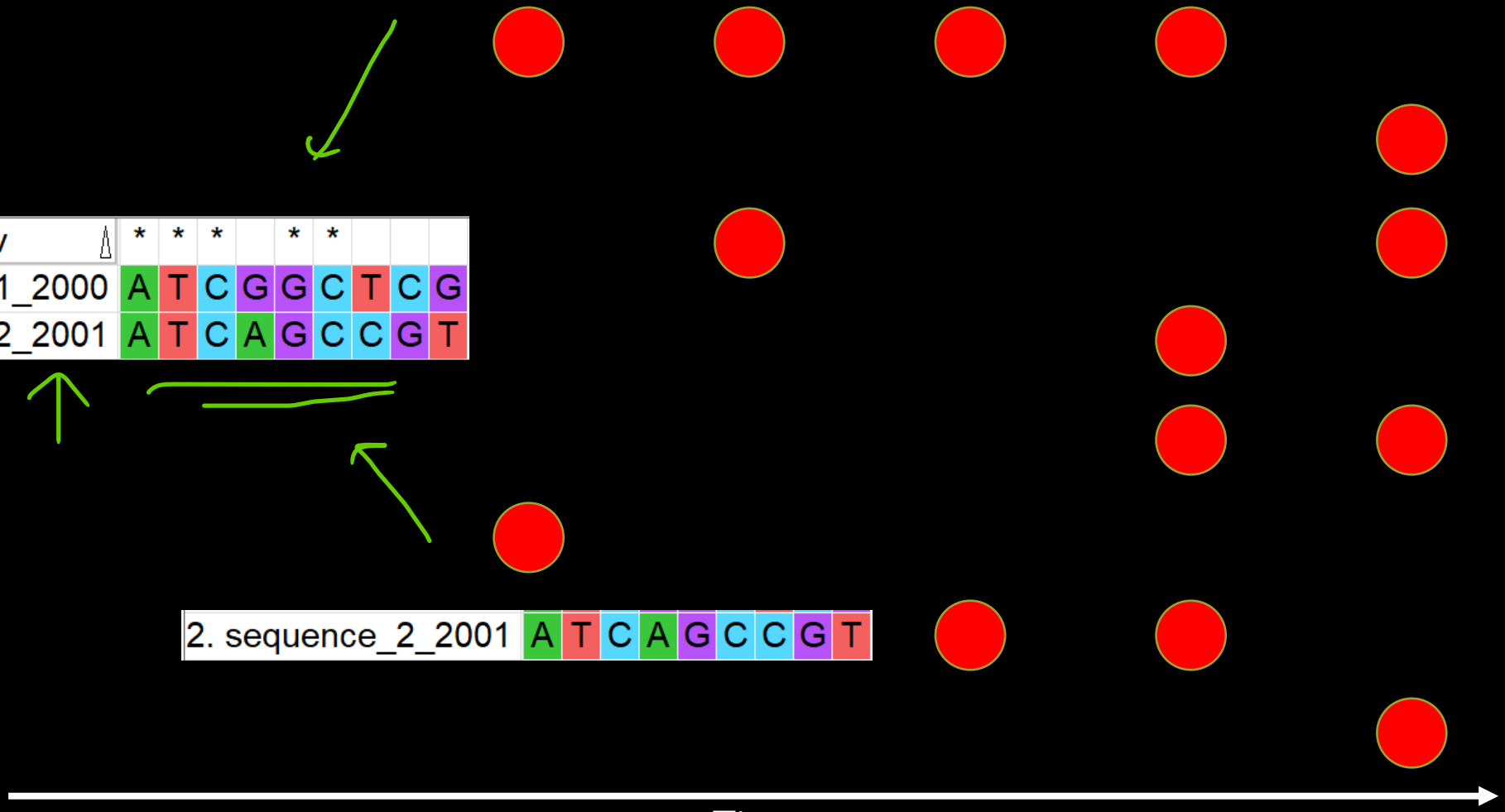
1. sequence_1_2000 A T C G G C T C G



Species/Abbrv	⚐	*	*	*		*	*			
1. sequence_1_2000	A	T	C	G	G	C	T	C	G	
2. sequence_2_2001	A	T	C	A	G	C	C	G	T	

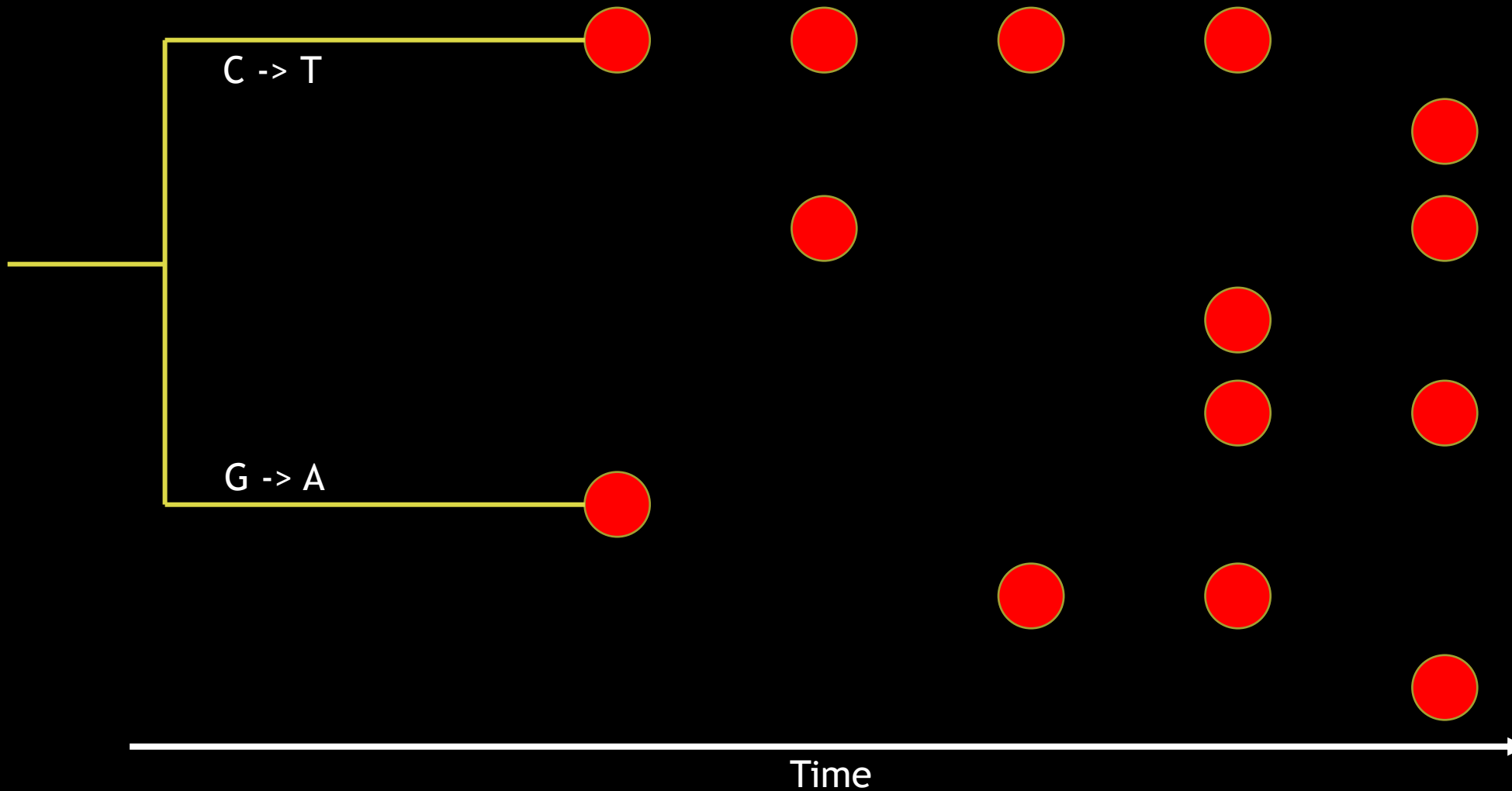


2. sequence_2_2001 A T C A G C C G T

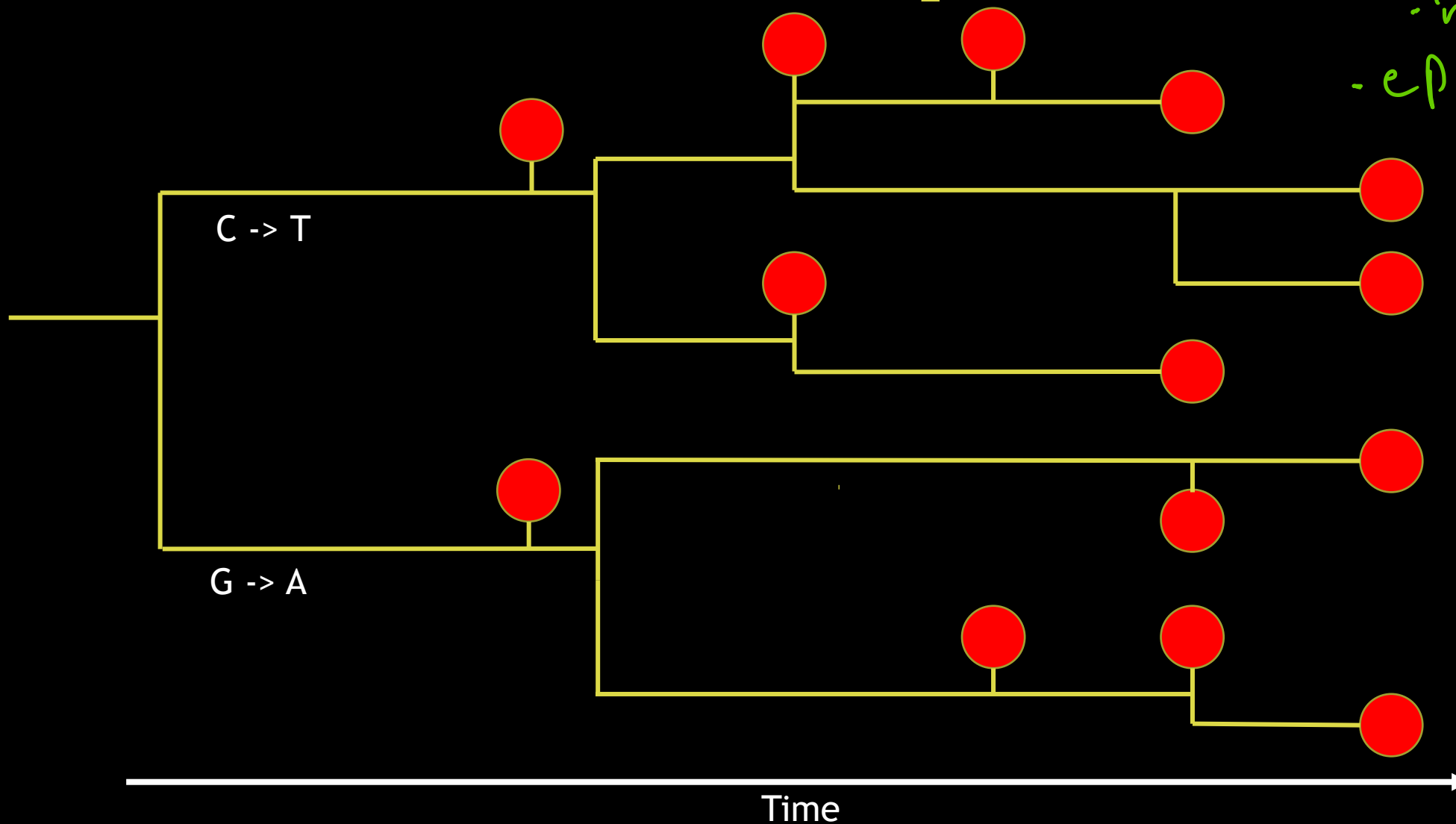


Time

How to Reconstruct the Epidemic?

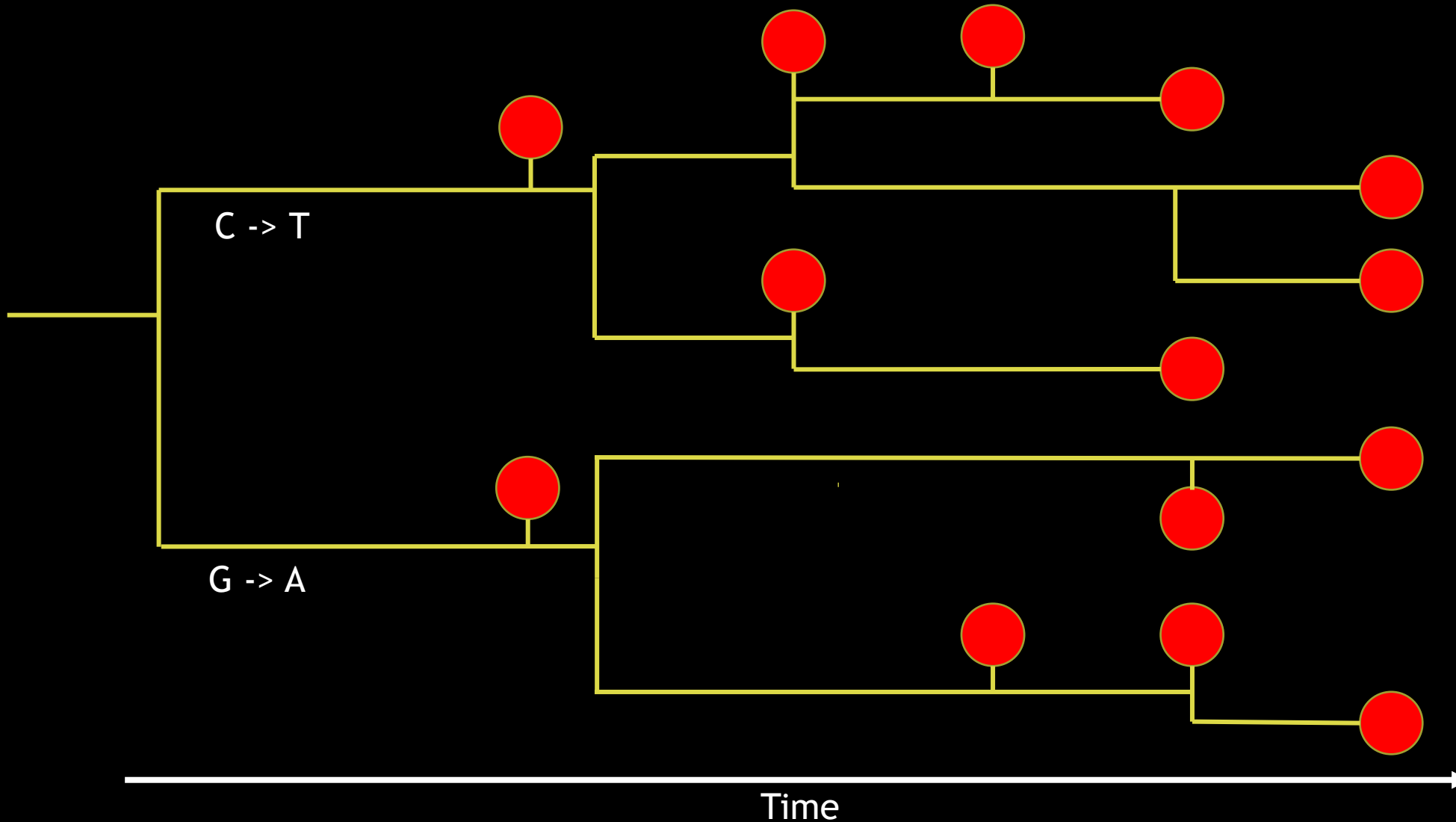


How to Reconstruct the Epidemic?



- sequence
- relationship
- timing
- epi data
- relationship
- timing

How to Reconstruct the Epidemic?



Important parameters

- Distance between sequences
- Tree model

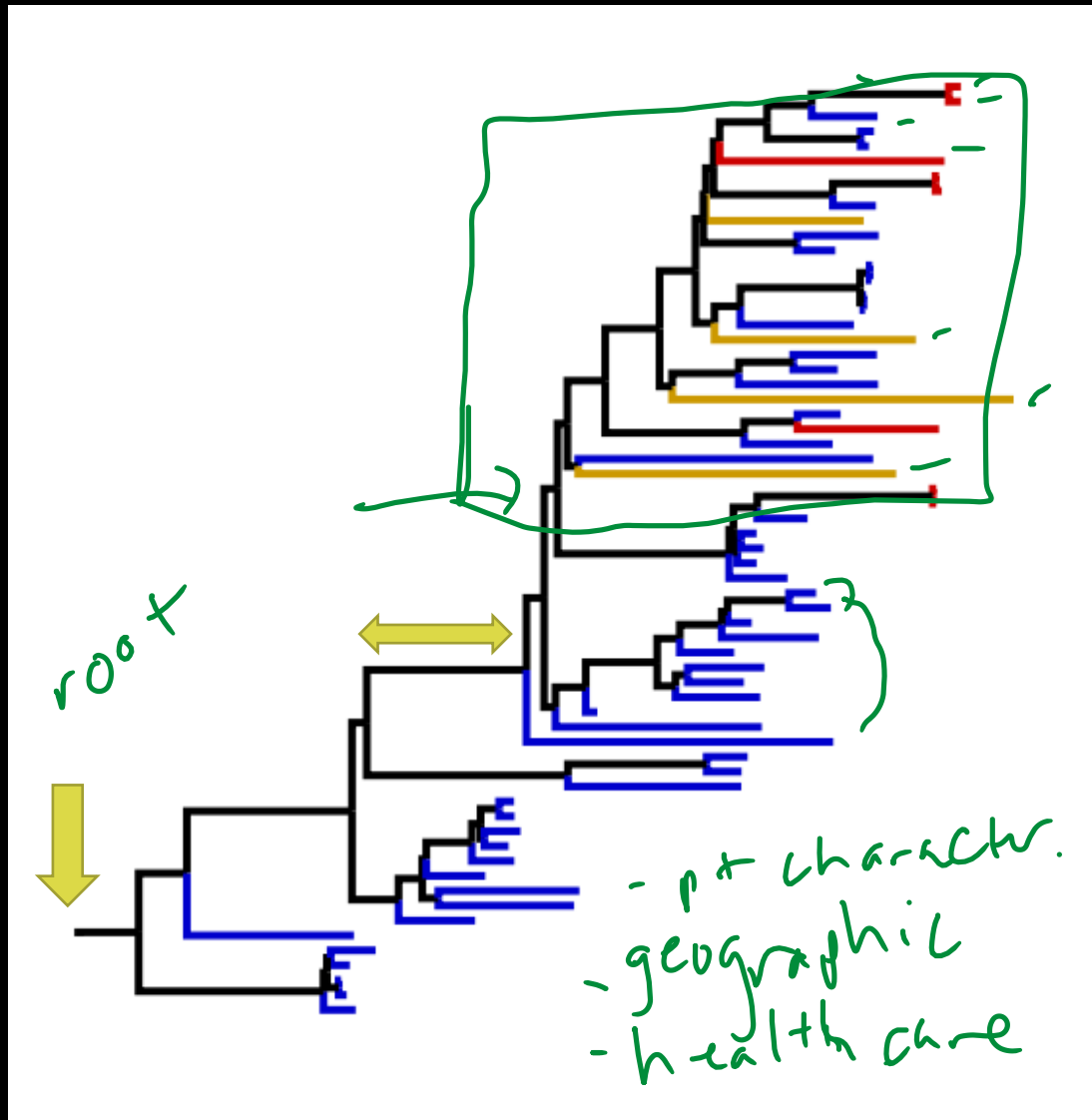
Sequencing Platforms



- more accurate, short reads
- high throughput
- expensive equipment
- widely used in public health systems



- less accurate, short reads
- rapid data generation
- low cost
- portable
- used in remote or limited resource settings

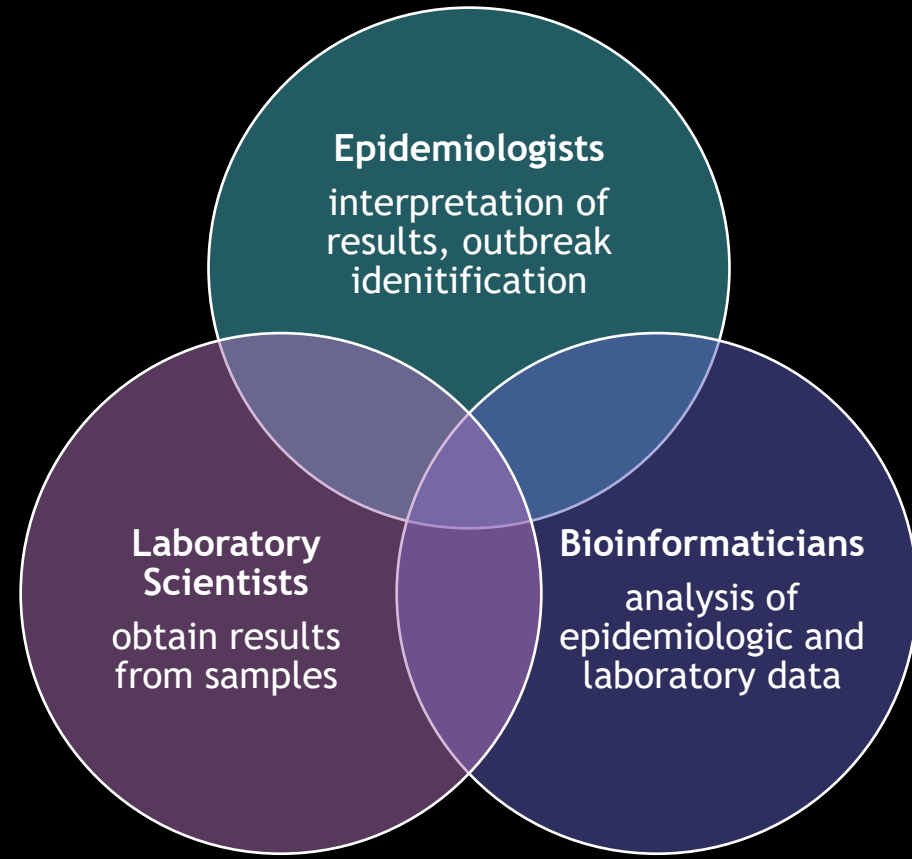


Data Analysis

- Use data to construct trees demonstrating relationships between samples
- phylogenetic trees
- combine with epidemiologic data to gain insights

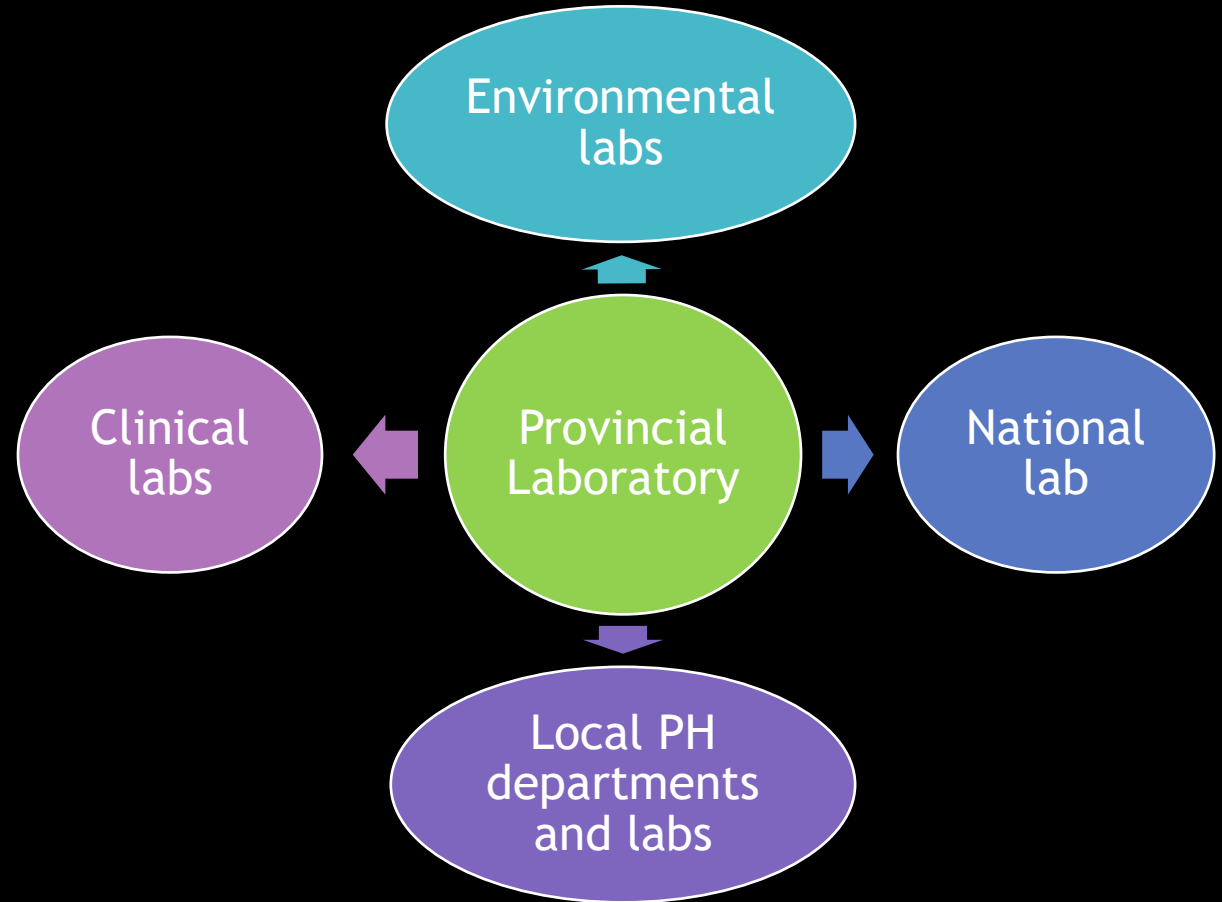
Collaboration among Disciplines

- Molecular results require collaborations, especially with laboratory networks or a national reference laboratory
- Need for investment in expansion of molecular testing capabilities
 - training of personnel
 - availability of equipment and software for analysis
- rapid & open data sharing



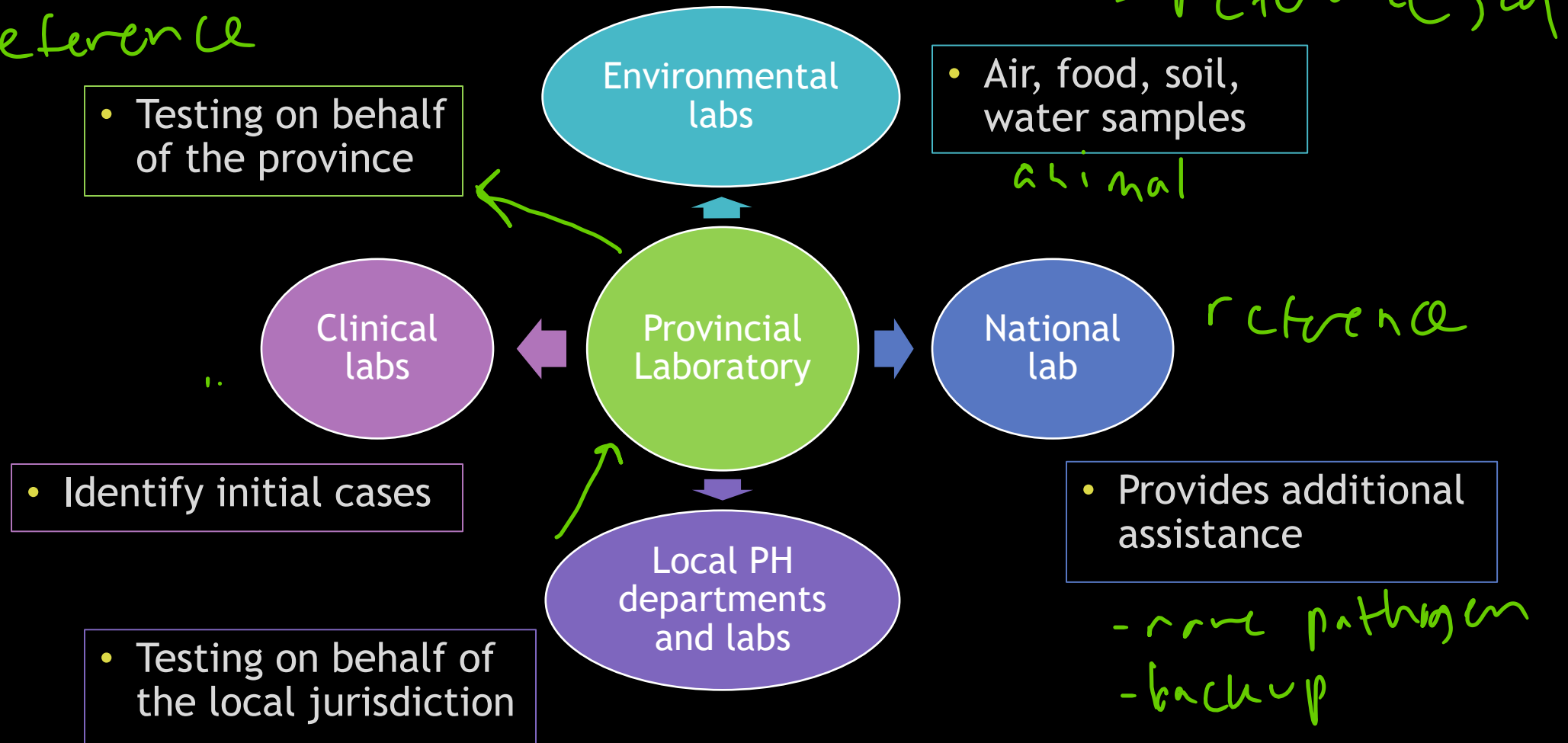
Public Health Laboratories

- Easy and rapid sharing of samples and data between these components
- Interaction is vital for addressing public health issues
- Additional network may exist between provincial/national laboratories
- Structure of networks and collaborations depends on availability of funding and facilities



Public Health Laboratories

reference



Online Databases

- Searchable public databases containing sequence or other genetic profiles *+ epi. data*
- Data submitted by researchers or laboratory personnel, along with accompanying epidemiologic information
- Samples taken from outbreaks, single cases, or healthy samples

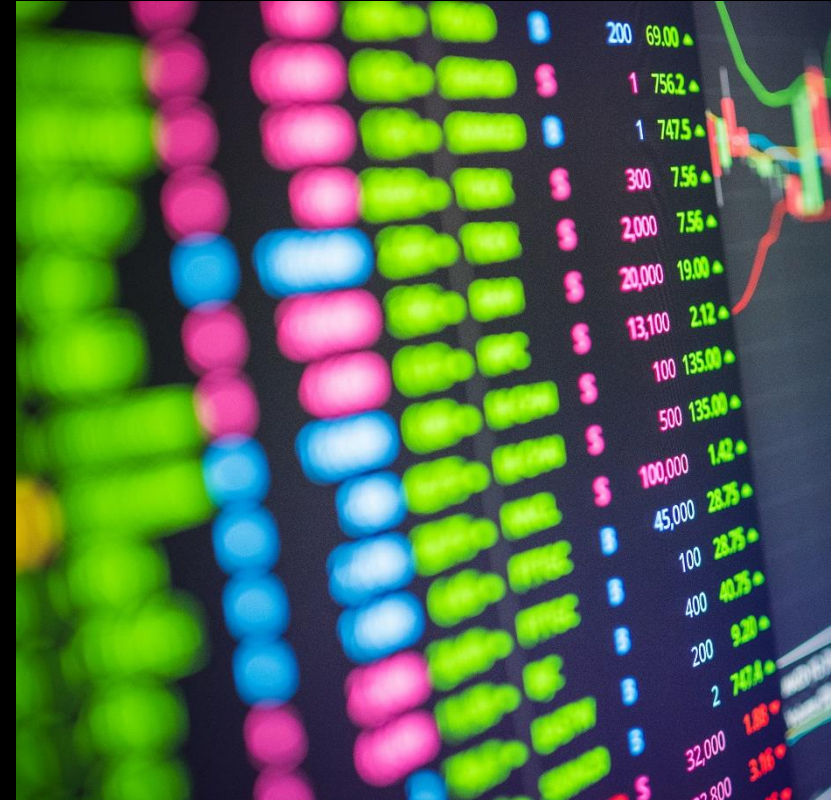
- cases
- environment
- food



Online Databases

- reference sequences

- PulseNet: PFGE profiles, WGS, bacterial pathogens
- GHOST: Global Hepatitis Outbreak and Surveillance Technology
- ViPR: Virus Pathogen Resource
- VEuPathDB: Eukaryotic Pathogen, Vector and Host Informatics Resource
- Global Salm-Surv: *Salmonella* serotypes
- EpiFlu: influenza viruses
- MicrobeNet: rare pathogens



- phenotype (antimicrobial) - genomes - genes
- genetic profiles - epi

Suitable Pathogens/Diseases: AMR

- detection of antimicrobial resistant pathogens
- testing for genes that confer resistance
 - • single gene: PCR, sequencing
 - • multiple genes: sequencing
- Example: penicillin-binding protein type in *Streptococcus pneumoniae*



Suitable Pathogens/Diseases: HAIs

- hospital-acquired infections occur in healthcare or long-term care settings
- detailed molecular data can help:
 - understand/detect outbreaks
 - identify rare/emerging pathogens
 - detect antimicrobial resistance
 - determine risk factors



Suitable Pathogens/Diseases: Foodborne

- identification of foodborne pathogens may require molecular techniques
- subtyping or genomic data of these pathogens can link cases
 - identify outbreaks
 - determine source of outbreaks
 - determine risk factors



Suitable Pathogens/Diseases: VPD

- Vaccine preventable diseases may be under selective pressure
- Molecular data can allow tracking of pathogen evolution over time and as vaccines are updated
- This information can also be used to make decisions about updating vaccines
- Example: shift/drift of influenza virus is monitored



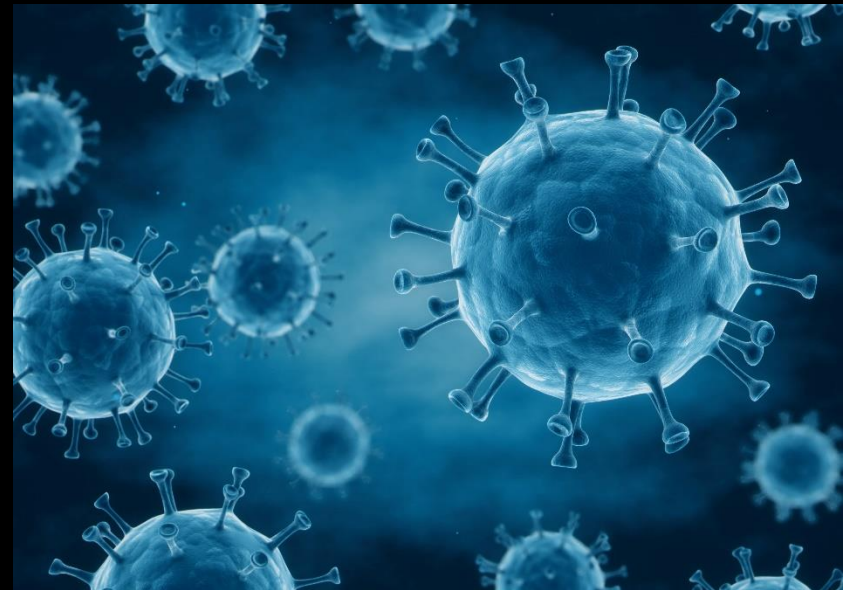
Suitable Pathogens/Diseases: VBD

- Vectorborne diseases are controlled through treatment of human/animal cases and through reduction of vector populations
- Molecular data can help:
 - detect antimicrobial/drug resistance in pathogens
 - monitor insecticide resistance in vector populations



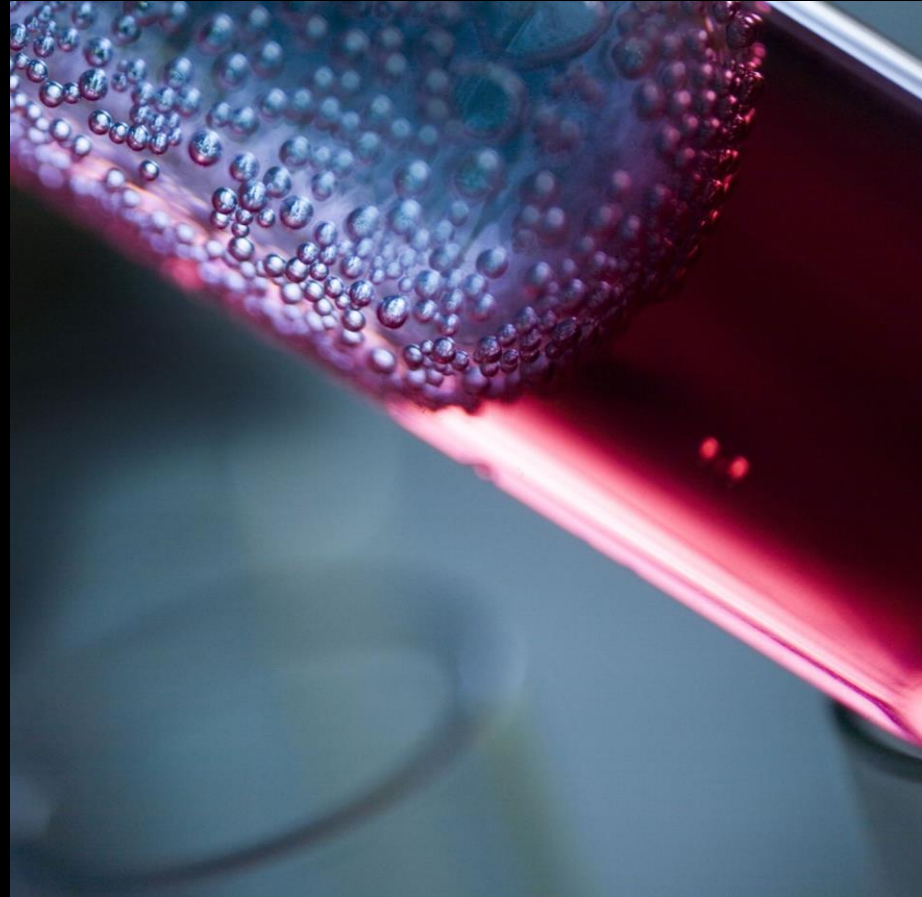
Suitable Pathogens/Diseases: Emerging

- Emerging or previously unknown pathogens can only be identified and understood through molecular/genetic information
- Example: SARS-CoV-2 identification and its relationship to previous epidemic coronaviruses



Suitable Pathogens/Diseases: HIV/TB

- Serious notifiable diseases affecting populations globally
- molecular surveillance helps:
 - tracing of outbreaks
 - recreate transmission networks
 - assess resistance
- Example: better understanding when HIV infection occurred according to genetic profile



Example: Emergence of Highly Pathogenic *Campylobacter jejuni*

Sahin *et al.*, 2012, Molecular evidence for zoonotic transmission of an emergent, highly pathogenic *Campylobacter jejuni* clone in the United States. *J Clin Micro.* 50(3): 680-687.

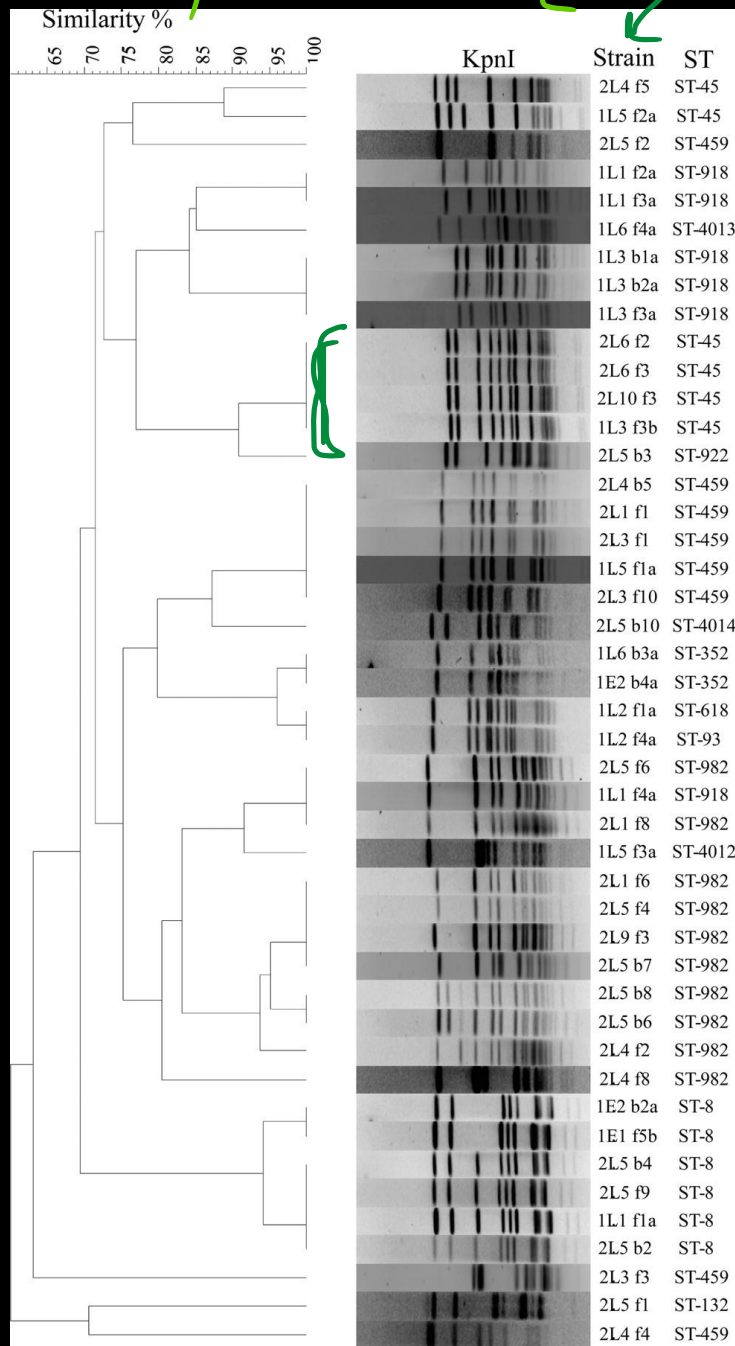
Campylobacter jejuni

genetic profile
antibiotic resistance

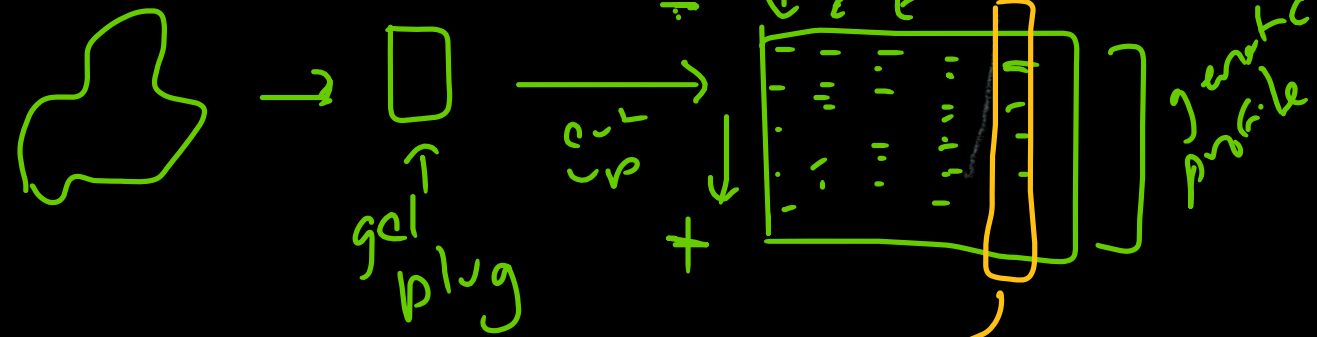
- Bacterial pathogen
 - foodborne illness in humans
 - spontaneous abortion in sheep, cattle, goats
 - commensal in chickens
- Veterinary reference laboratory conducts testing to identify pathogens from sheep samples submitted by veterinarians
- Research group conducted further testing to obtain genetic profiles of *C. jejuni*



relatedness



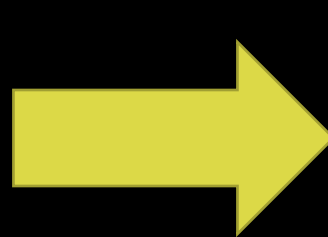
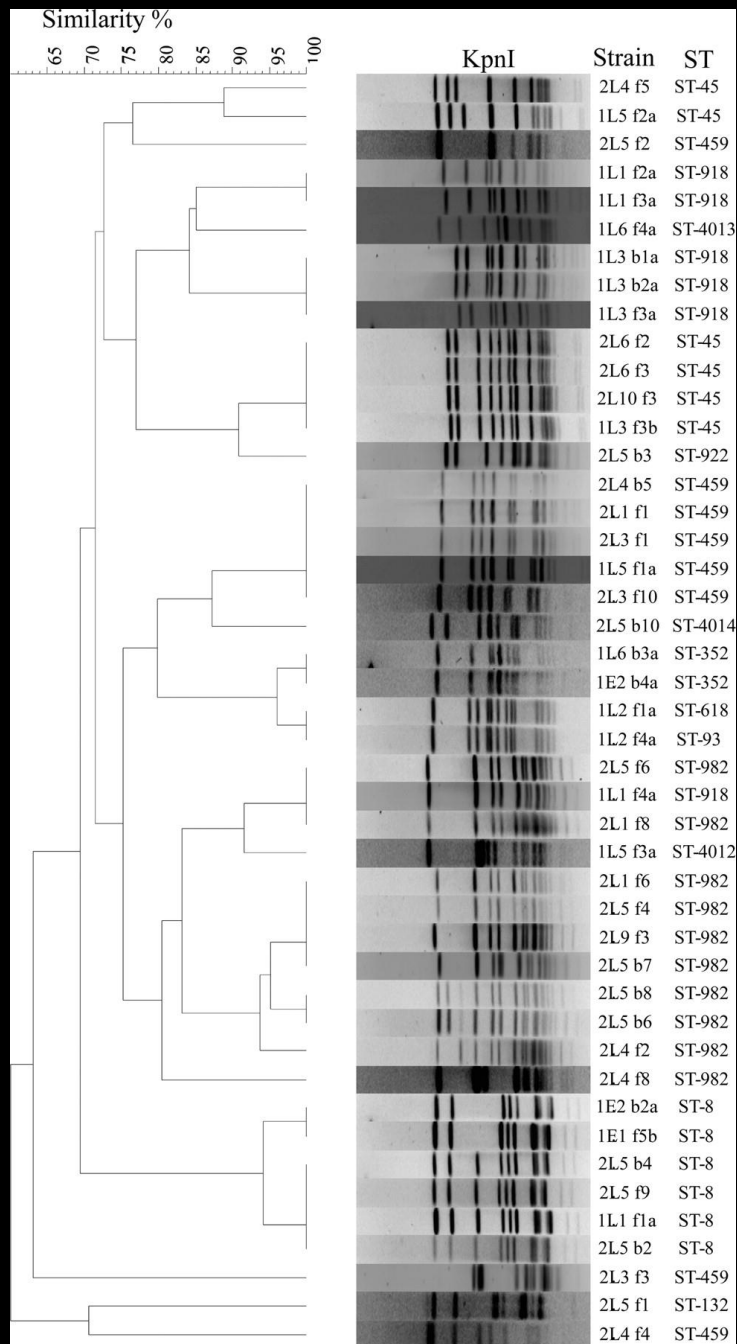
pulse genetic profile



restriction enzyme

KpnI AGGTTCA
cut
the genome

profile
larger genome piece at top
smaller at bottom



• PulseNet

Table 1 Human *C. jejuni* isolates indistinguishable by PFGE from clone SA identified in the PulseNet National Campylobacter database at the CDC through 2010

Case	No. of isolates	State	Yr	Isolation source	Exposure
Outbreak 1	1	VT	2003	Unknown	Raw milk
Outbreak 2	4	SC	2007	Stool	Raw milk
Outbreak 3	16 ^a	PA	2008	Unknown	Raw milk
Outbreak 4	4	RI	2008	Unknown	Chicken
Outbreak 5	32	WI	2009	Stool	Raw milk
Outbreak 6	2	MA	2010	Stool/blood	Raw milk
Outbreak 7	7	MI	2010	Stool	Raw milk
Outbreak 8	1	MT	2010	Unknown	Well water
Outbreak 9	2	VT	2010	Stool	Raw milk
Sporadic ^b	56	Multiple	2004-2010	Stool	Unknown

online databases
gels with genetic profile

Example: Emergence of Novel Avian Influenza

Arafa et al., 2015, Emergence of a novel cluster of influenza A (H5N1) virus clade 2.2.1.2 with putative human health impact in Egypt, 2014/15. *Euro Surveill.* 20(13):p00-21085.

Highly Pathogenic Avian Influenza

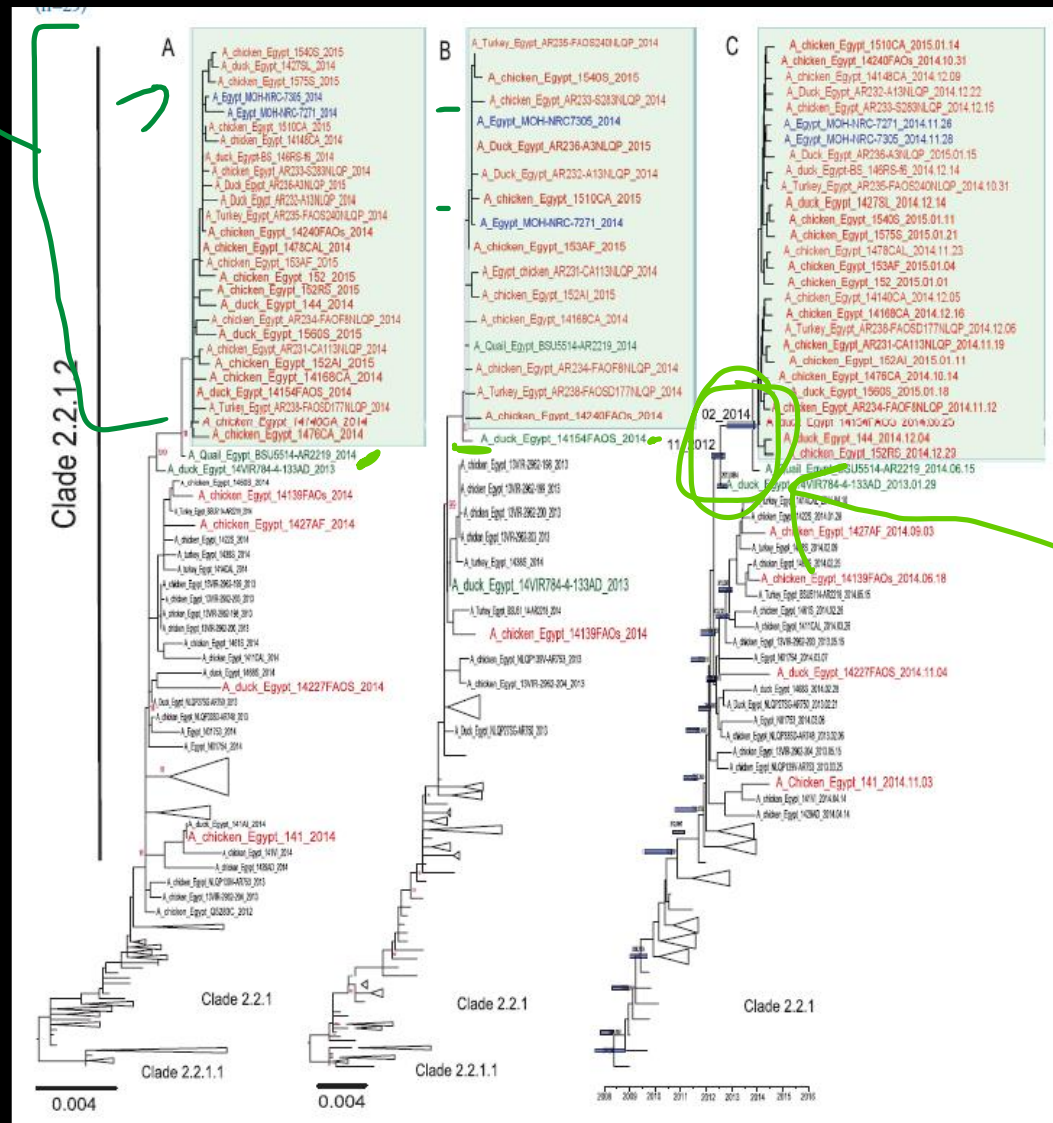
- Avian influenza circulates in poultry populations and cause sporadic human cases
- Influenza A (H5N1) present in Egypt since 2006
- Outbreak of human H5N1 cases in early 2015
- Examination of poultry H5N1 outbreaks, passive surveillance



cluster
most closely related

HA NA time-resolved

H5 H1
N1 N5
:



HA NA

H5 N1

Highly Pathogenic Avian Influenza

- HA and NA gene sequences
- emergence in February 2014
- identification of distinct mutations in cluster

black = reference
red = poultry
blue = human
green = roots

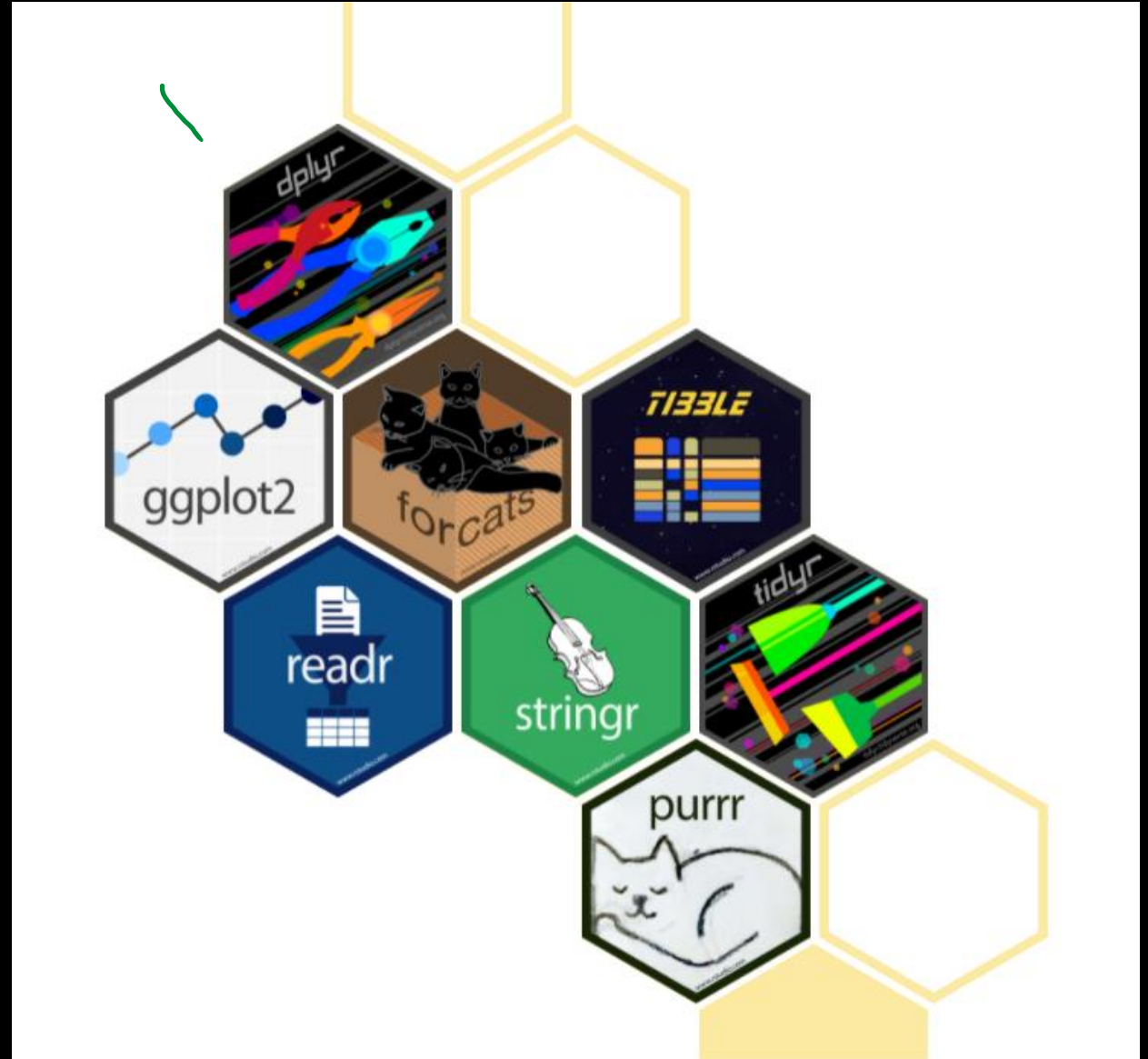
Workshop Schedule

Time	Topics
2:00–2:30 pm	Molecular Surveillance & Infrastructure
2:30–2:45 pm	Pathogen Targets
2:45–3:05 pm	Examples of Molecular Surveillance in Use
3:05–3:15 pm	Break
3:15–4:00 pm	Tidyverse

base R]

[tidyverse

- Collection of packages that work with each other
- Intended for complete data process: import, cleaning, manipulation, visualization, modeling, etc.



tidyverse

- Really good references/learning materials available online
- Very popular way to use R!



tidyverse

- We will use some tidyverse methods for data manipulation and data visualization



tidyverse data manipulation

- some common commands used in tidyverse for data manipulation

- subset data

- adding, deleting, changing data variables

- rearranging data

- filter()

- select()

- mutate()

- group_by()

- spread()

- arrange()

tidyverse
commands

[filter(,)]

dataframe

variable

data [3, 10]
↑ ↑

tidyverse data manipulation

- pipe operator
 - `%>%`
 - take the output from left side and put it into the next command or argument
 - easy to set up “layers” of commands

tidyverse data manipulation

- pipe operator

- %>%
- take the output from left side and put it into the next command or argument
- easy to set up “layers” of commands

```
data <- covid19_df %>%
```

```
  filter(location == "Bangladesh") %>%
```

```
  spread(data_type, value) %>%
```

```
  arrange(date) %>%
```

```
  mutate(deaths = deaths_new)
```