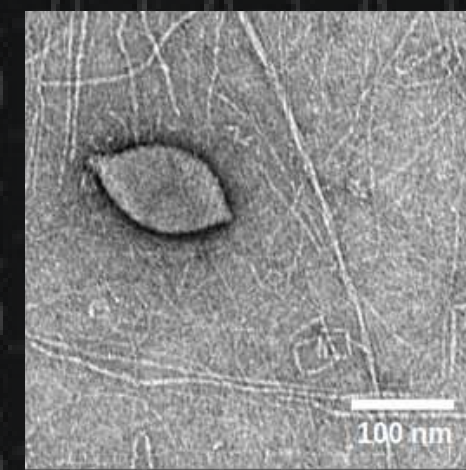
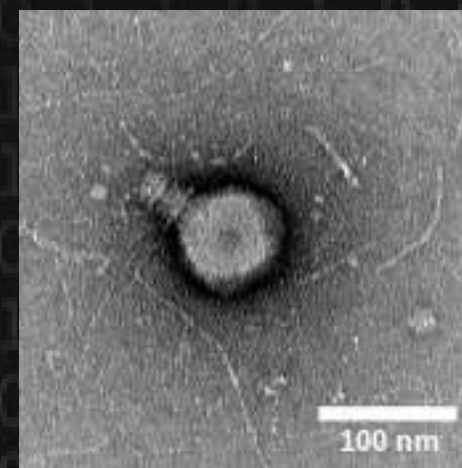
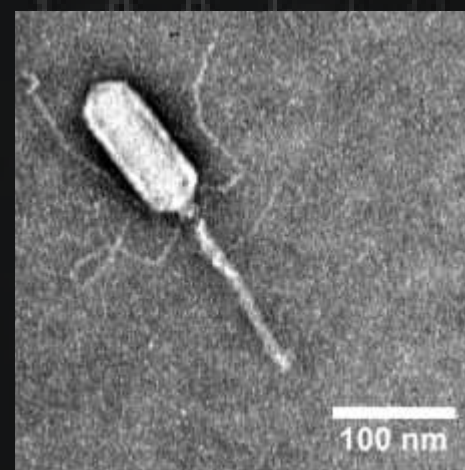
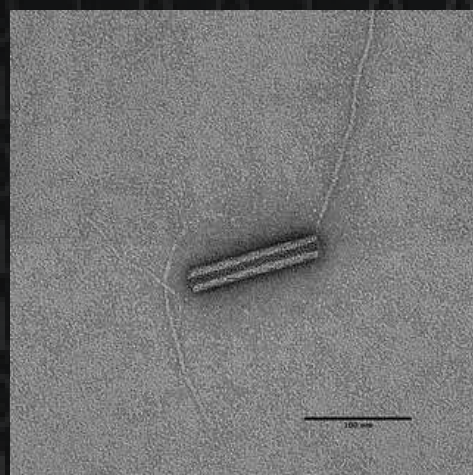


# VIROMICS Sergio Sánchez Carrillo



24/10/2024 Curso Aplicación de Herramientas -ómicas en Acuicultura

Gabinete de Formación CSIC

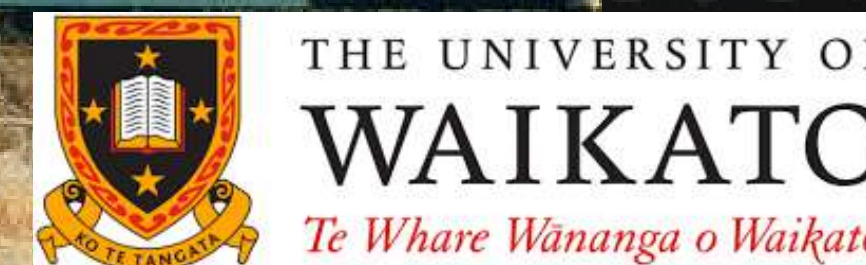
# Outline

1. Alpine and polar viral metagenomics
2. Some theory about Viromics
3. Wetlab workflow
4. Drylab workflow
5. “Hands on” sea cucumbers dataset



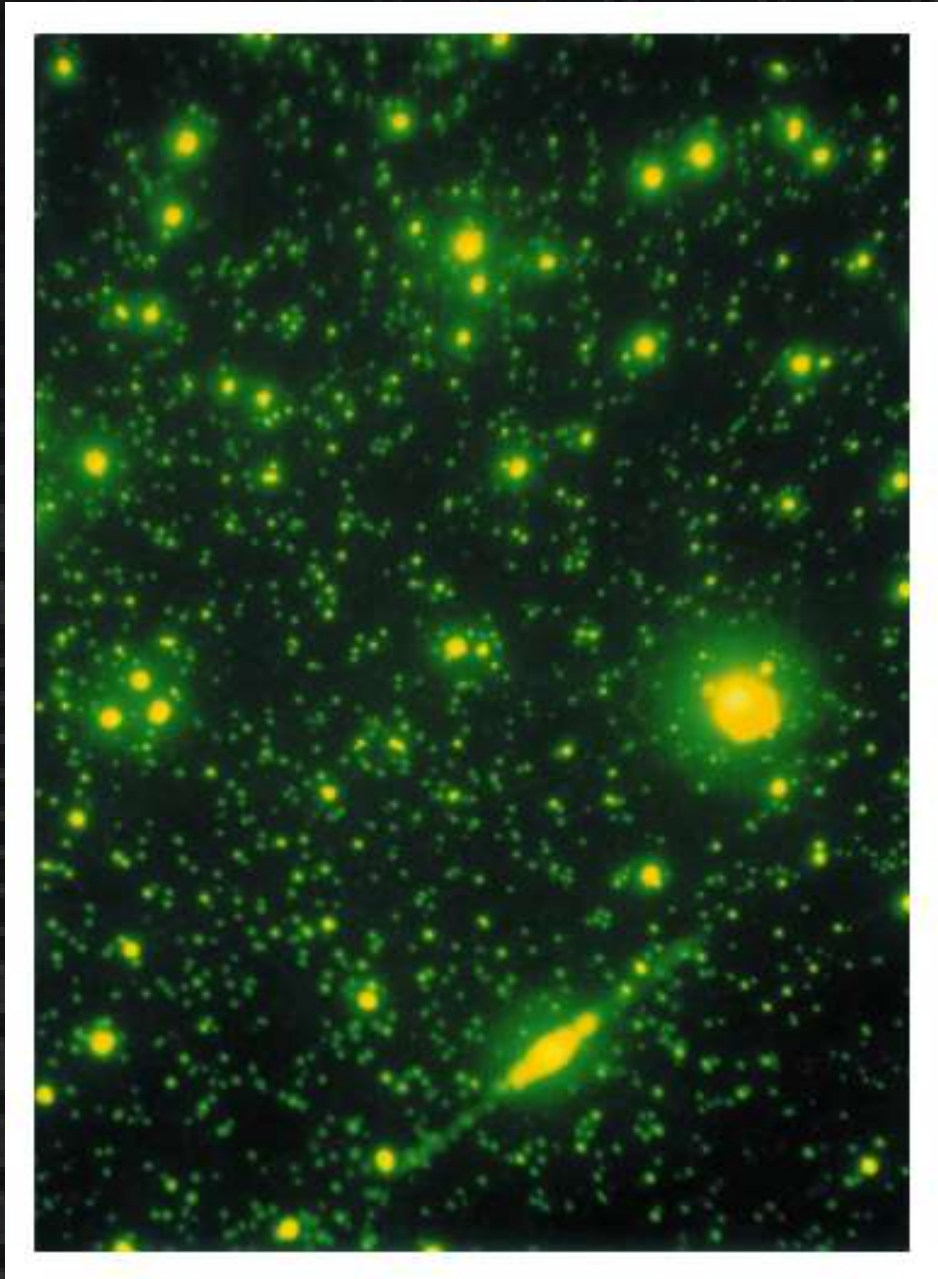
# 1. Alpine and polar viral metagenomics







# Environmental Viruses: Why?

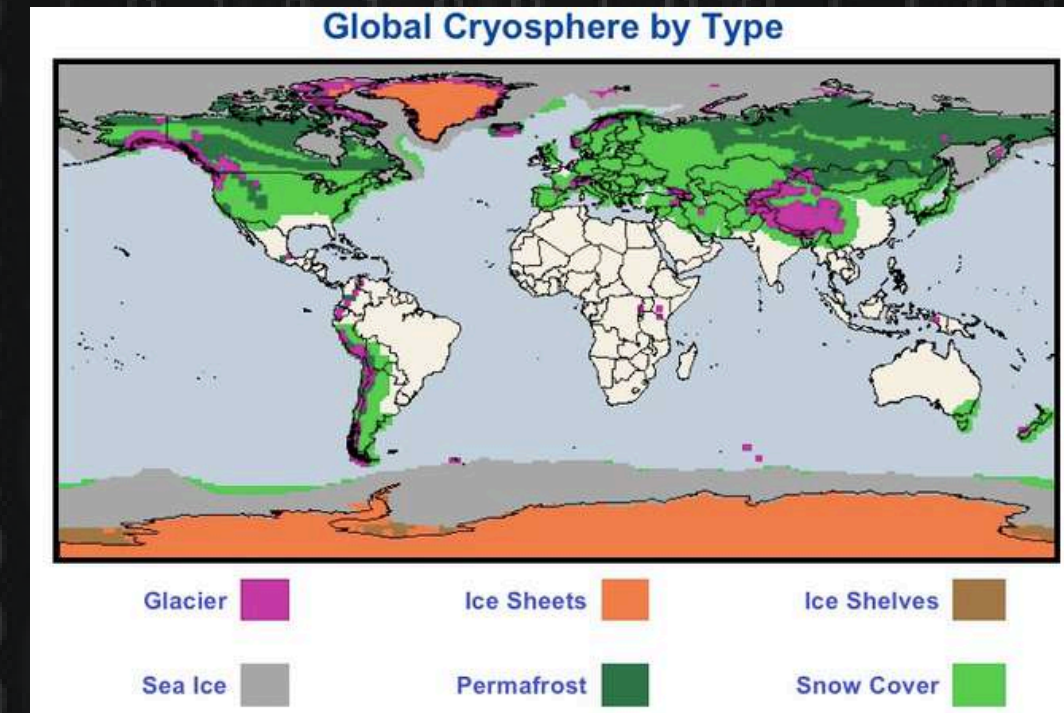


(Fuhrman, 1999)

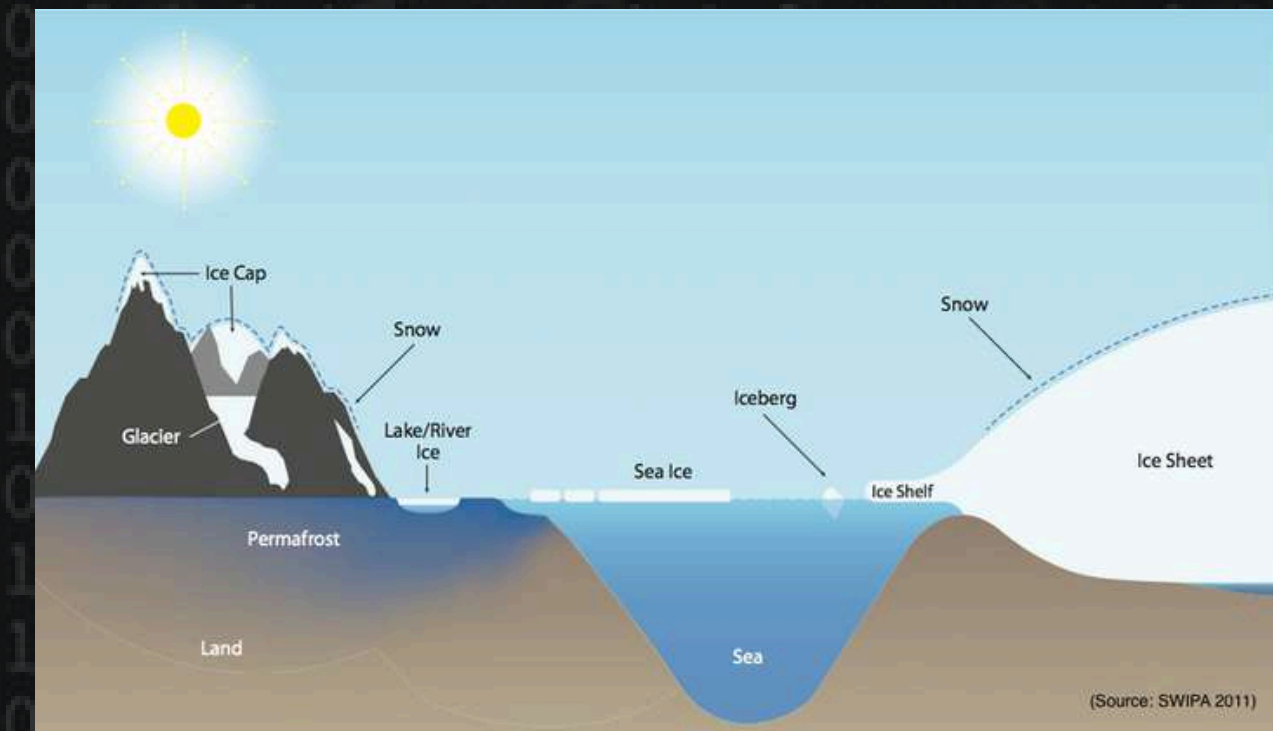
- Most abundant and genetically diverse bio-entities on Earth
- Reservoir of unknown genetic diversity
- Paramount role in microbial ecology and evolution (population limiters, metabolic reprogramming, evolutionary drivers)
- Key in biogeochemical cycles

# Cryosphere: Why?

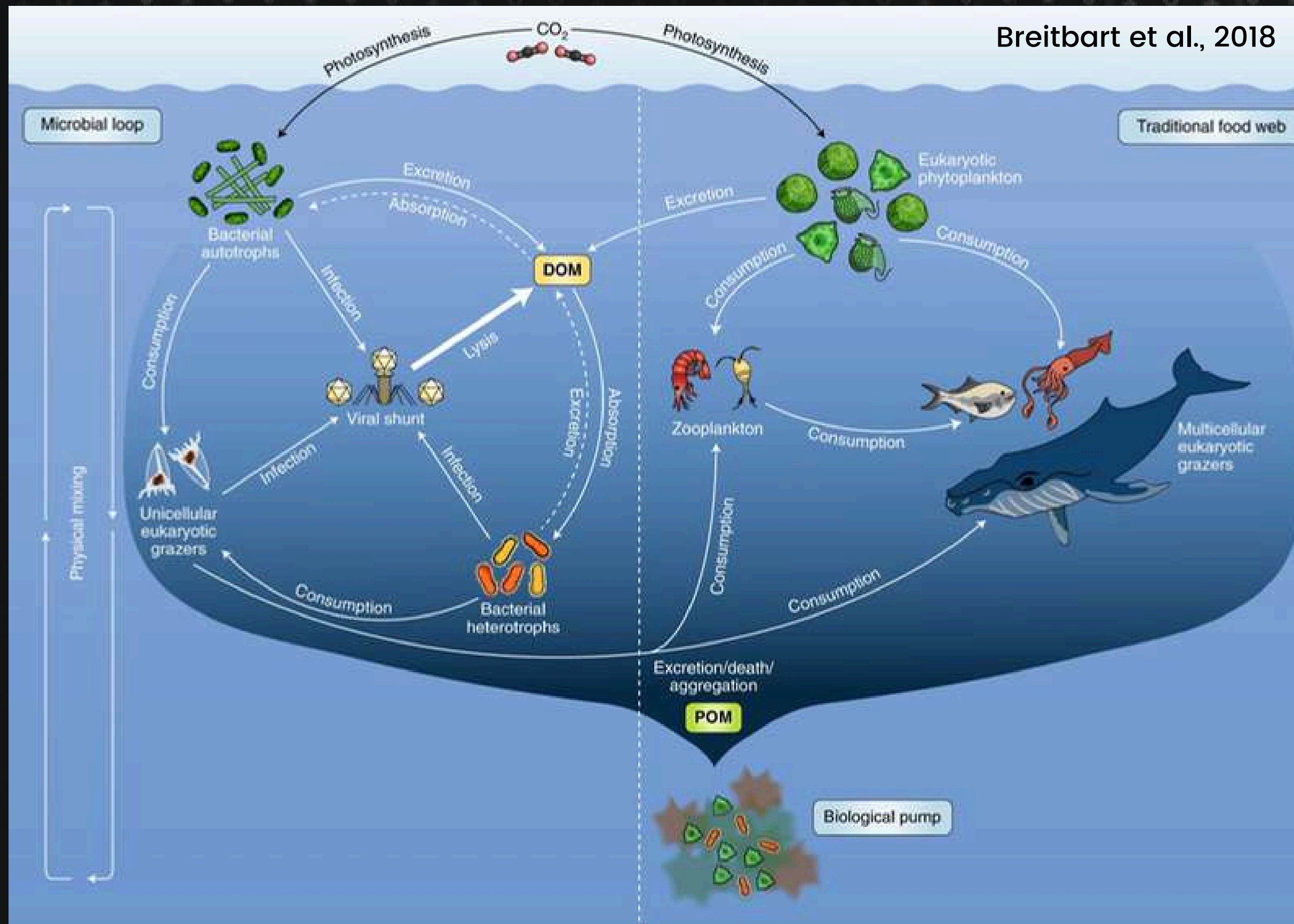
- Widely distributed across Earth
- Key regulator of global climate (albedo, Global Ocean Circulation)
- Sustains highly diverse ecosystems
- Paramount role in global biochemical cycles
- Water supply for billions of people
- Highly susceptible to global change
- Relatively little known, especially in the case of viruses
- Exoplanet system analogues



(<https://globalcryospherewatch.org>)

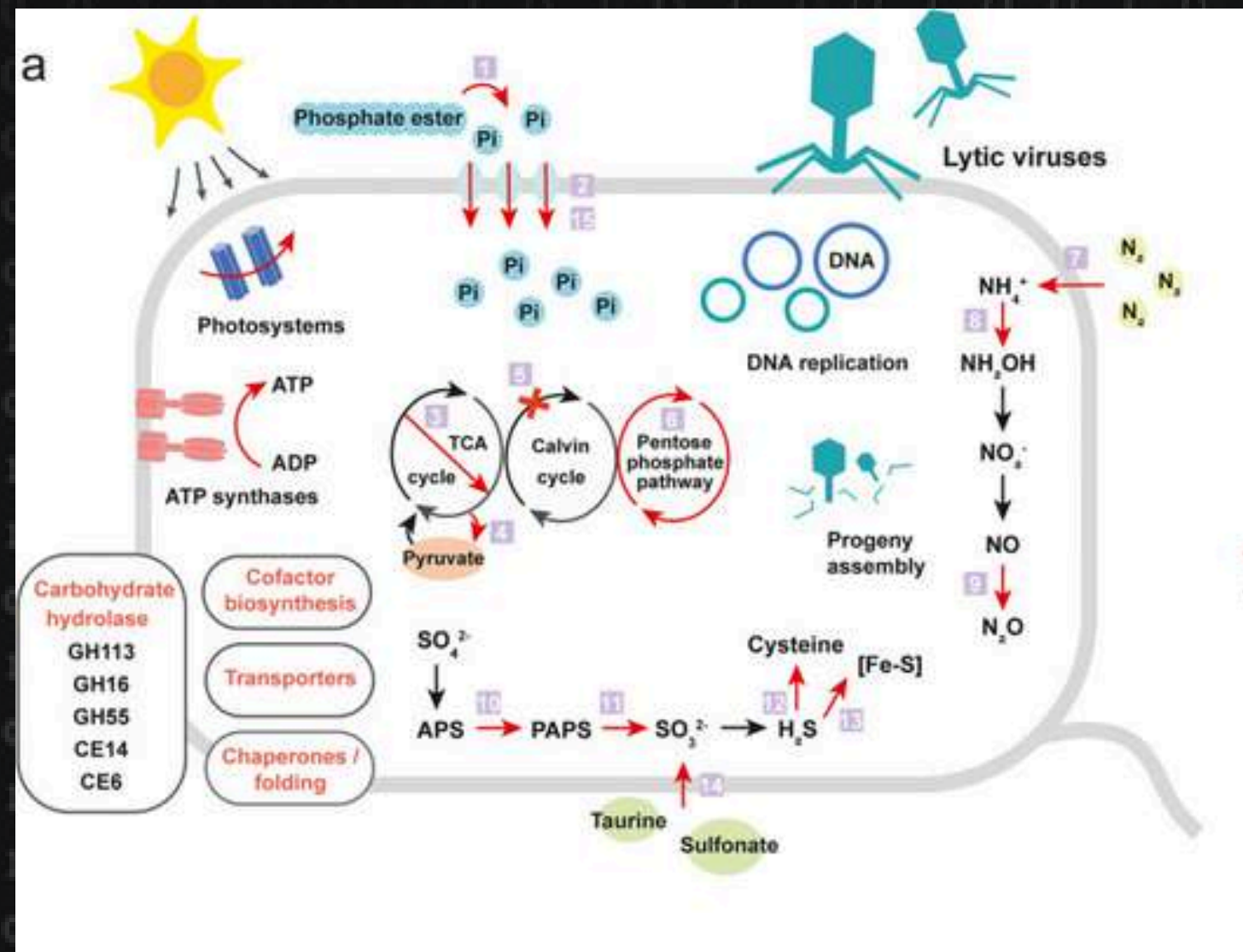






Microbial populations control  $\longrightarrow$  biogeochemical cycles

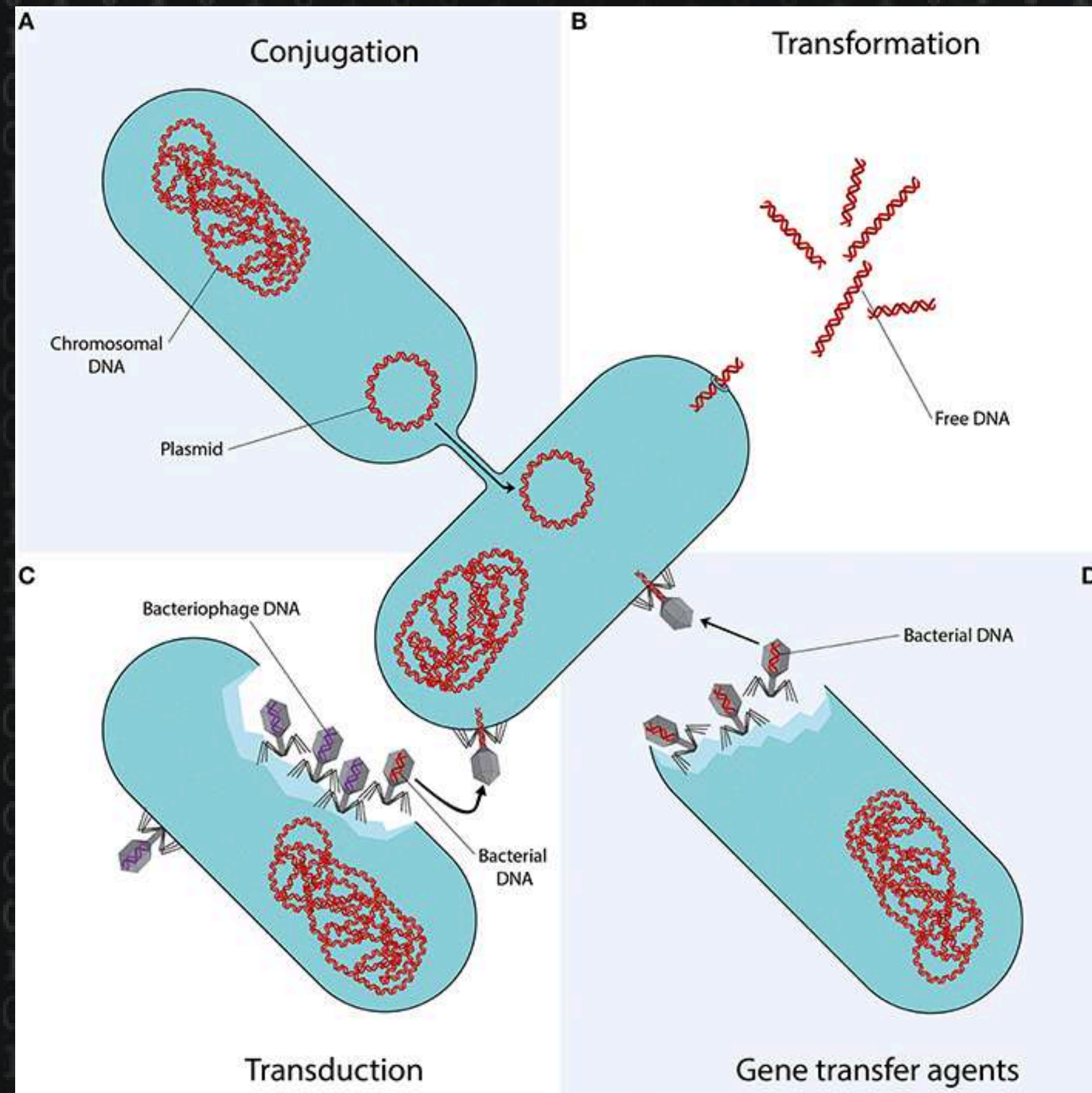
# AMGs $\longrightarrow$ Host metabolic reprogramming



Wang et al., 2022

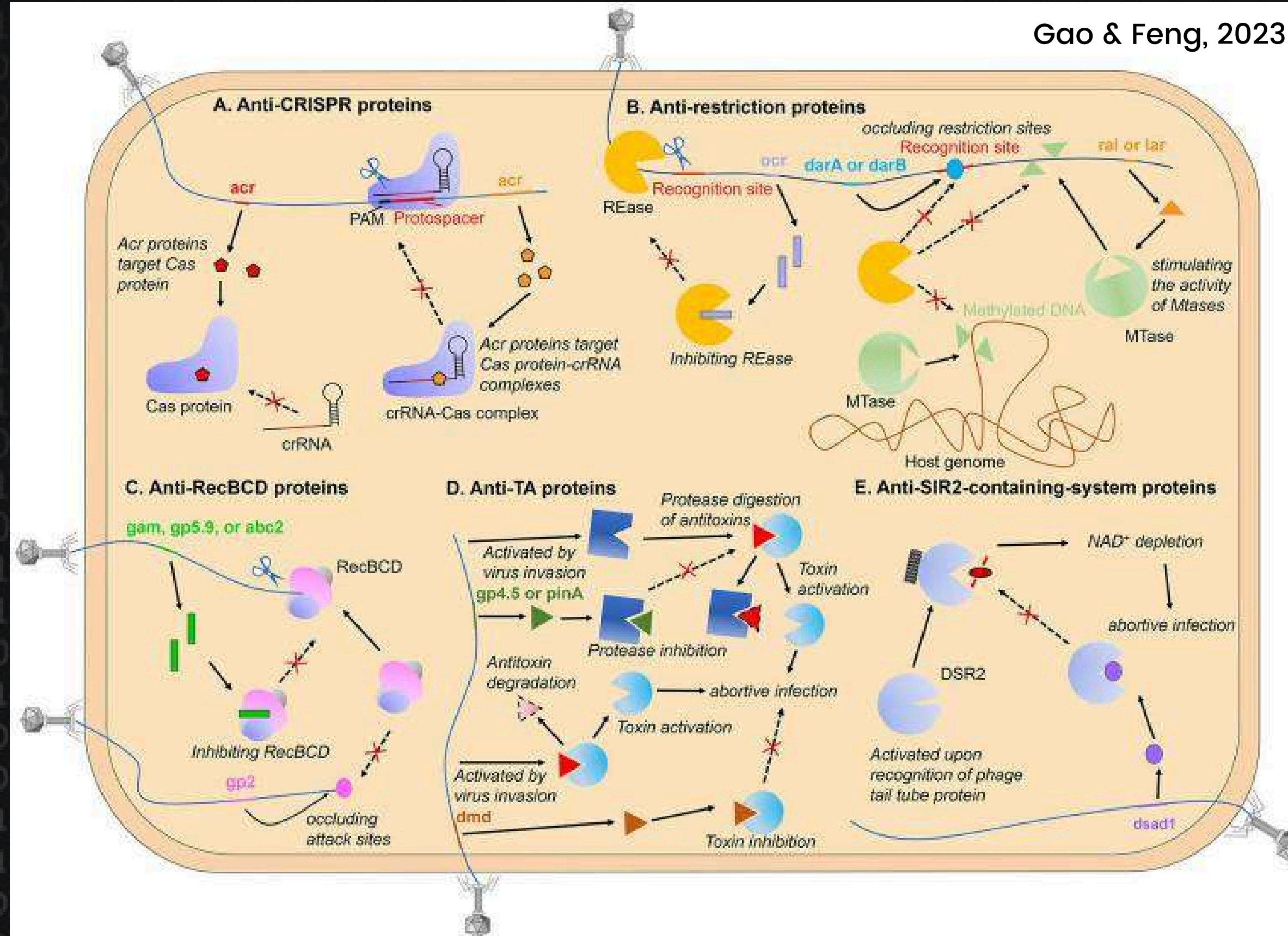


# HGT → Host and virus evolution

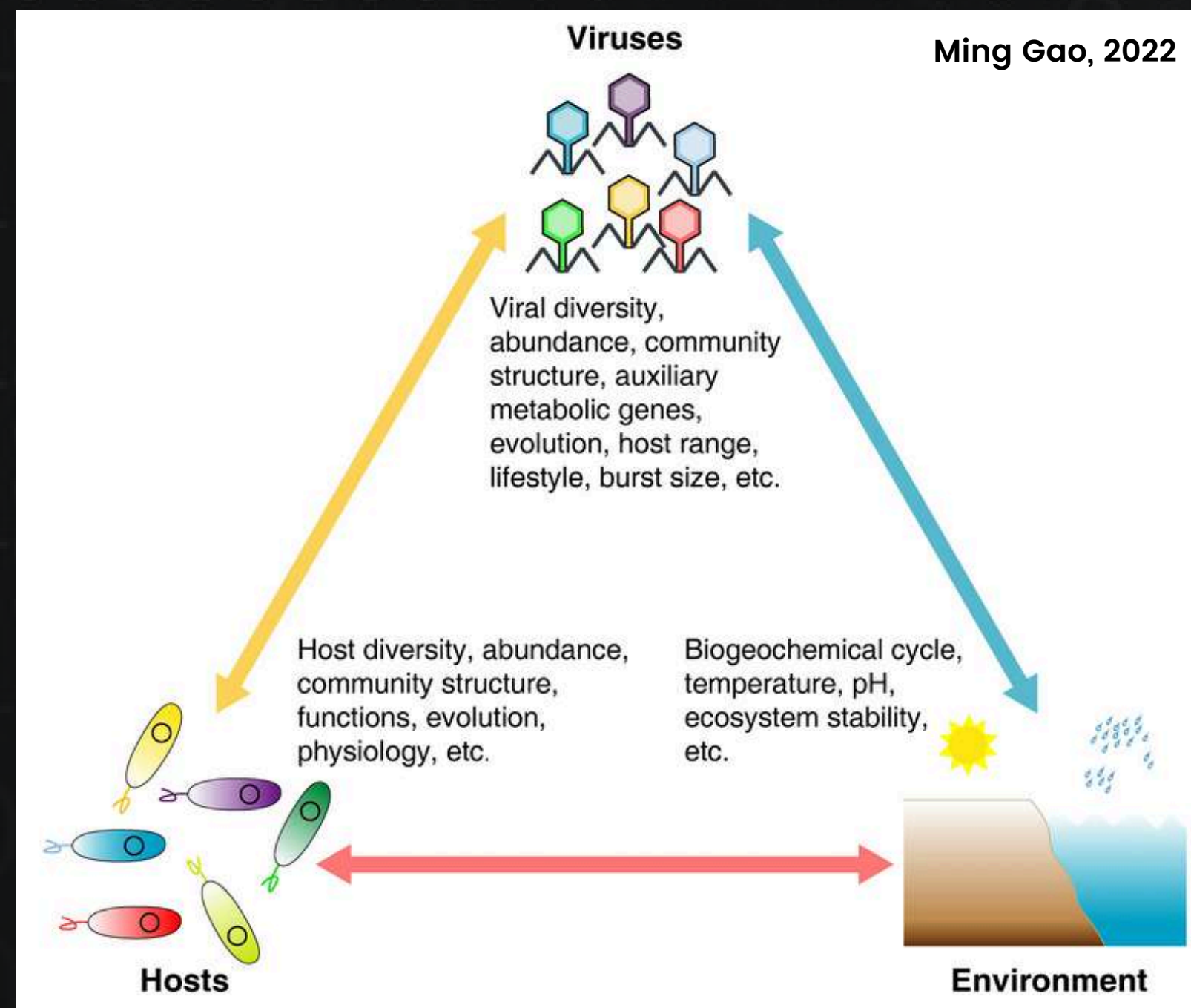
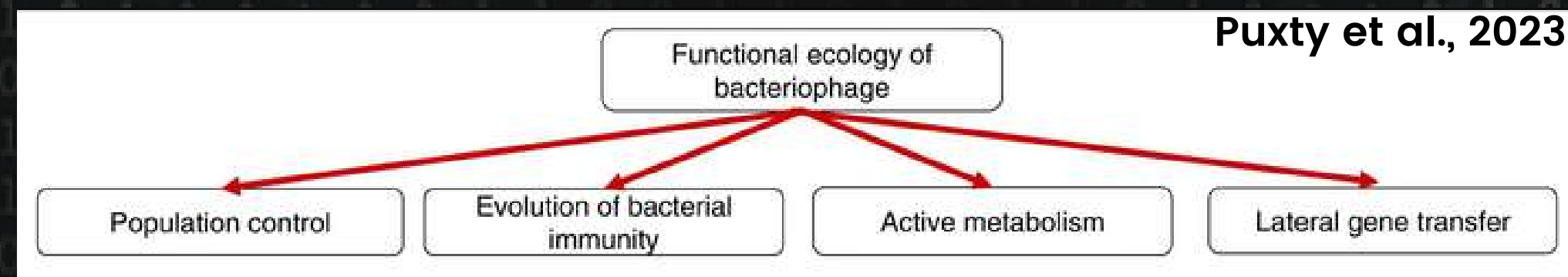


Wintersdorff et al, 2016

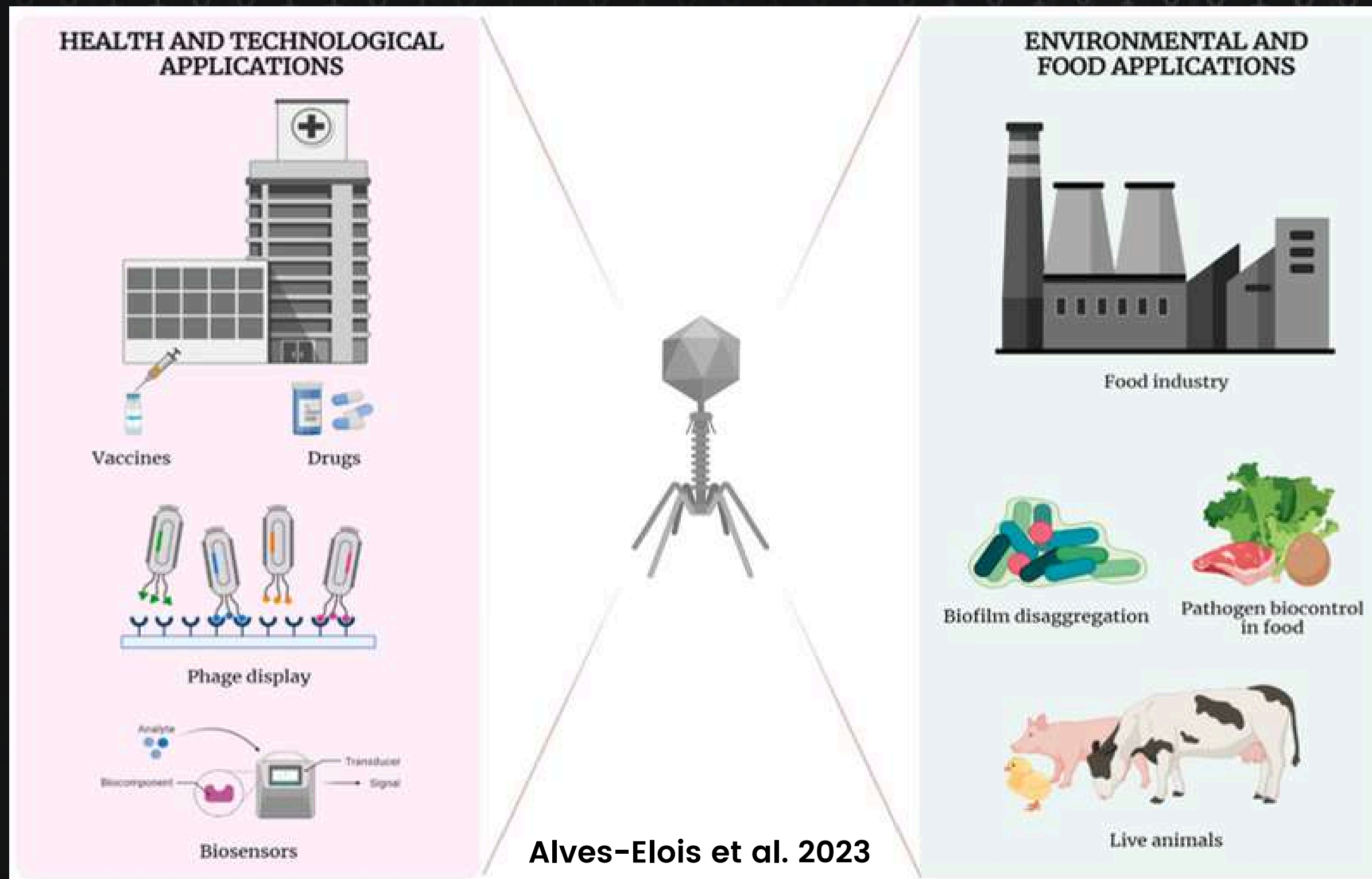
# Overcoming bacterial antiviral response systems → Arms race







And even more...









# MICROORDESA: Biodiversidad microbiana y viral en el P.N. de Ordesa y Monte Perdido

BBVA



Miguel Bartolomé  
Blas Valero-Garcés  
Ana Moreno  
J. Ignacio López-Moreno

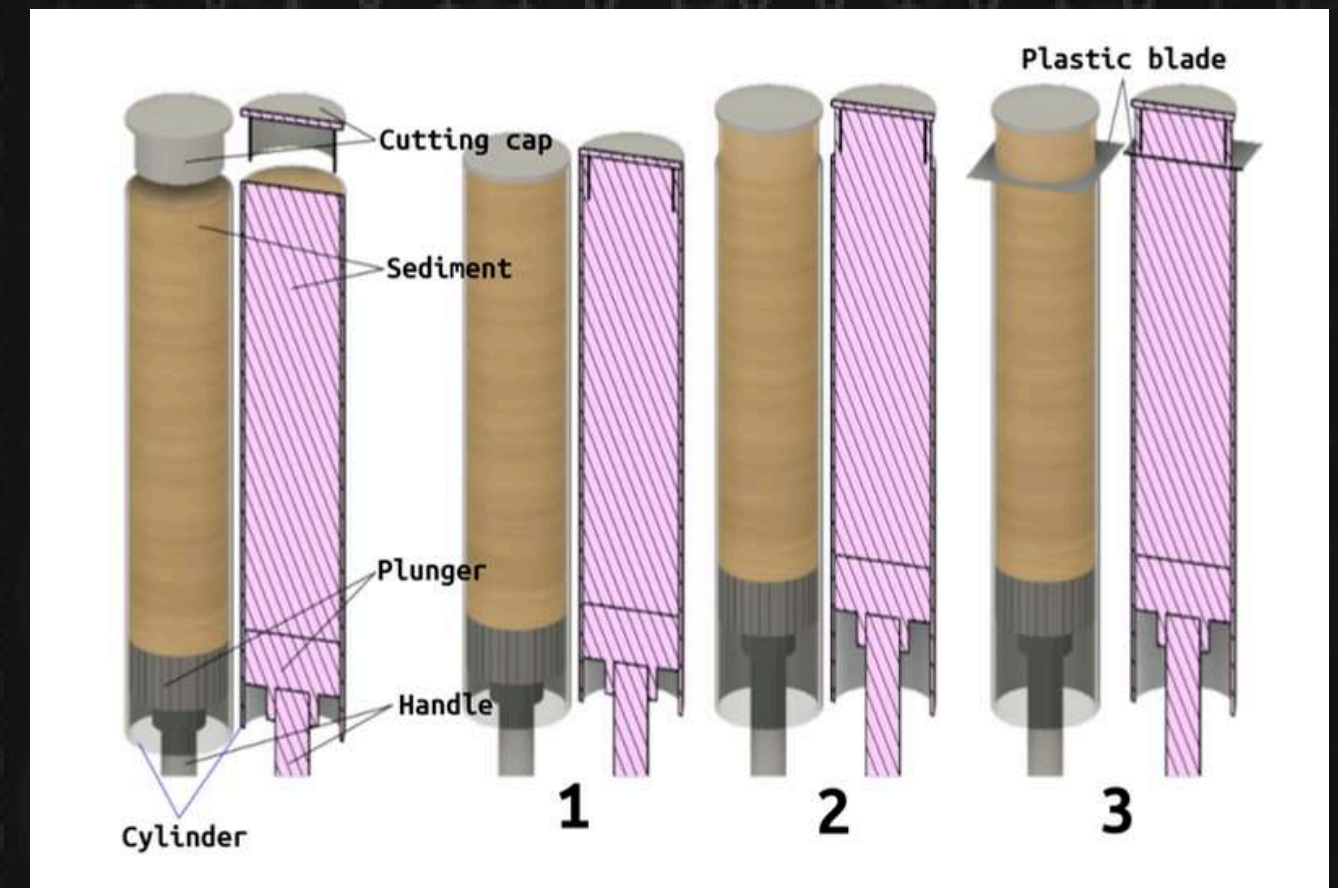




# DNA viruses of polar and alpine lake sediments



Name	Latitude	Longitude	Elevation (m asl)	Catchment area (Km2)	Lake area (ha)	Catchment/lake area ratio	Max Depth (m)	Thermal stratification	Trophic status
Escondido	62°37' S	61°0.4' W	92	8	1.6	5.0	5.3	Cold monomictic	Ultra-oligotrophic
Limnopolar	62°37' S	61°0.6' W	65	58	2.2	26.4	5.5	Cold monomictic	Ultra-oligotrophic
Marboré	42°41' N	0°2' E	2612	137	14.3	9.6	30	Dimictic	Ultra-oligotrophic
Tenndammen	78°06' N	15°02' E	5	210	15	14.0	2.5	Cold monomictic	Ultra-oligotrophic



Alejandra Vicente de Vera  
&  
Blas Blas Valero-Garcés





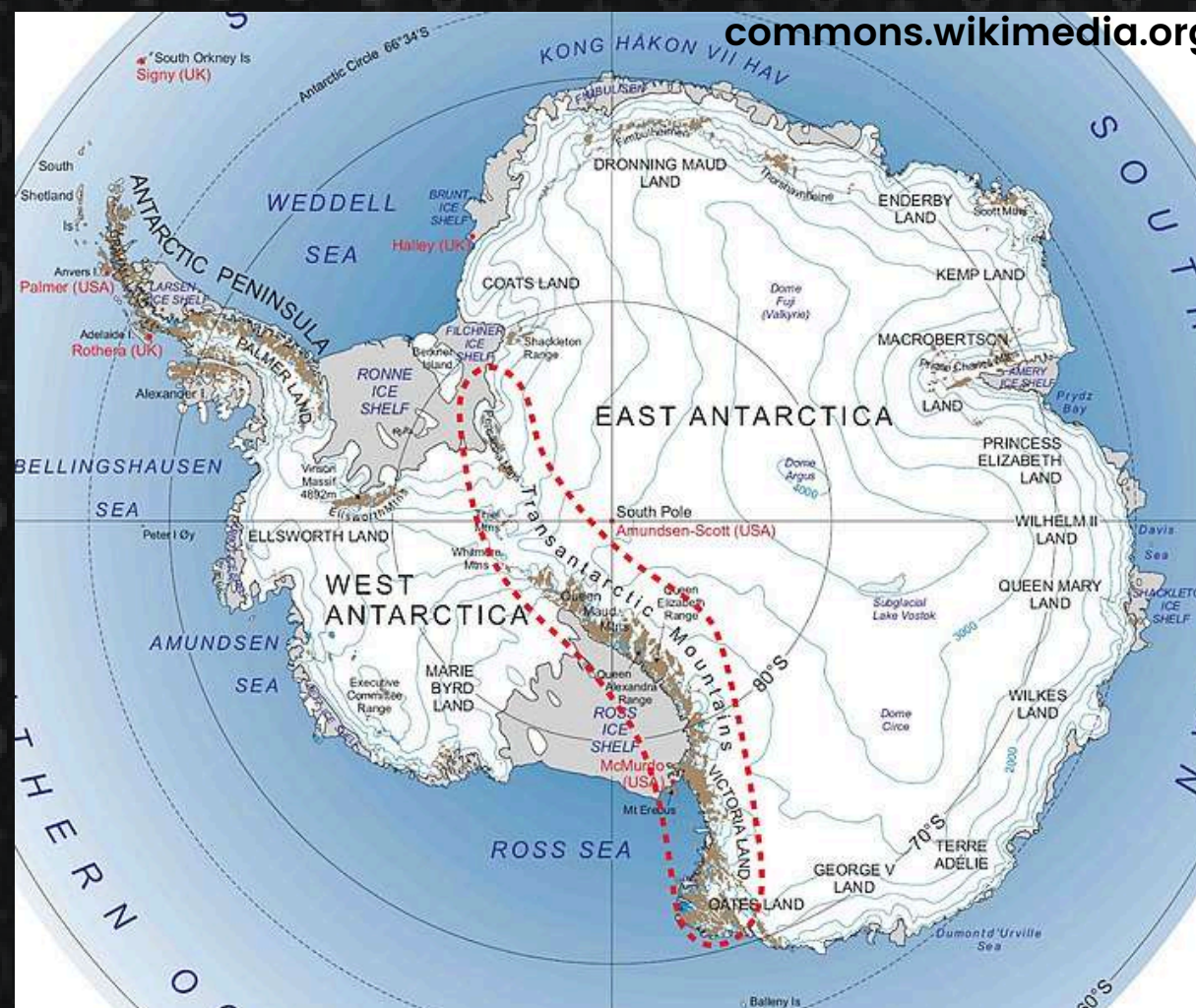


## Diversity of DNA and RNA soil viruses in Antarctica

Ian R. McDonald & S. Craig Cary



## Viral soil communities in ice-free soils in Antarctica





## 2. Some theory about Viromics

# Viromics

Study of viral genomes and their interactions within a specific environment using high-throughput sequencing and bioinformatics techniques, focusing on understanding the diversity, structure, and function of viruses in the ecosystems. Viromics enables scientists to explore:

- Viral diversity
- Virus–host interactions
- Viral evolution
- Ecological roles

chatGPT

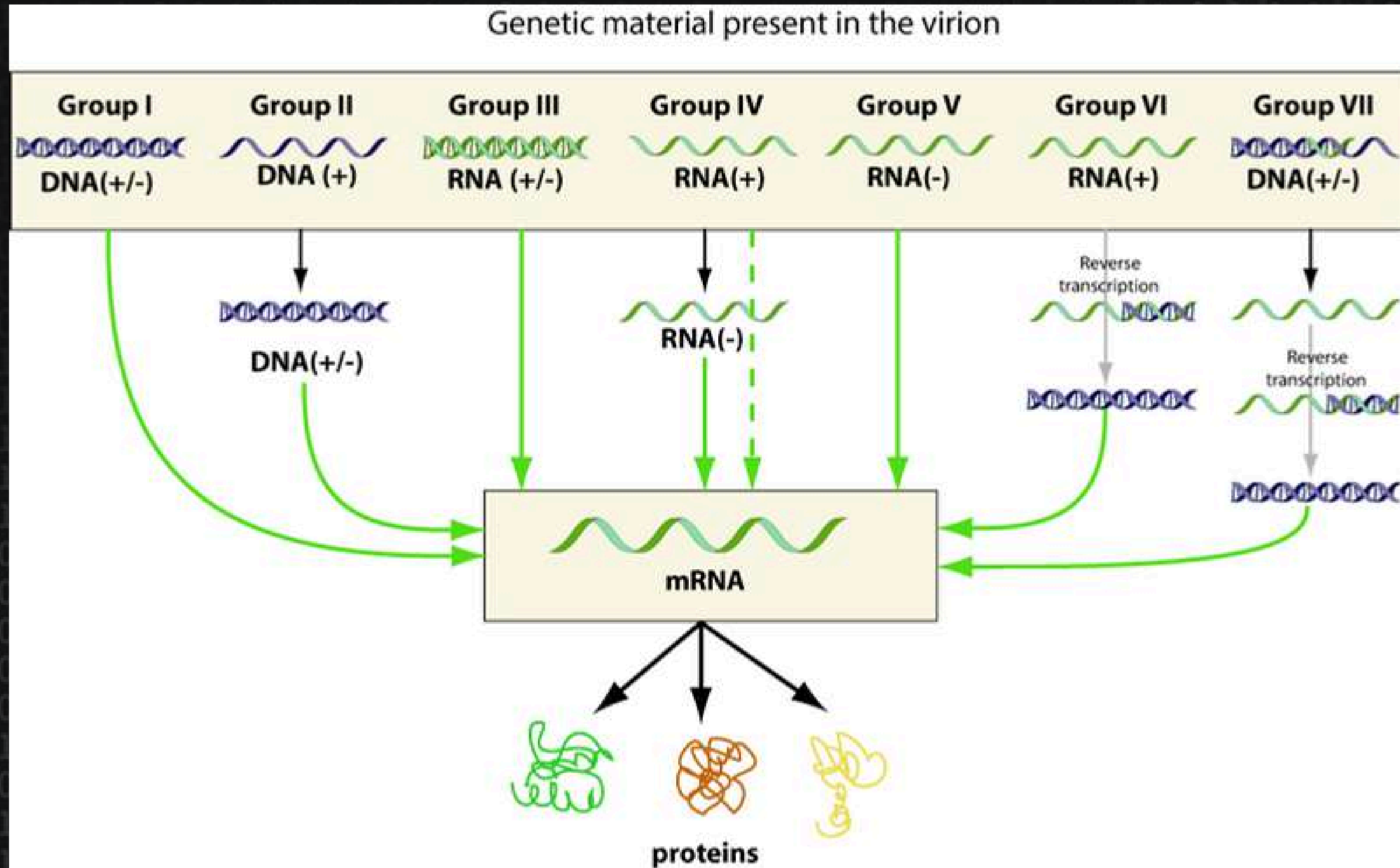
VLP enrichment  
before sequencing?

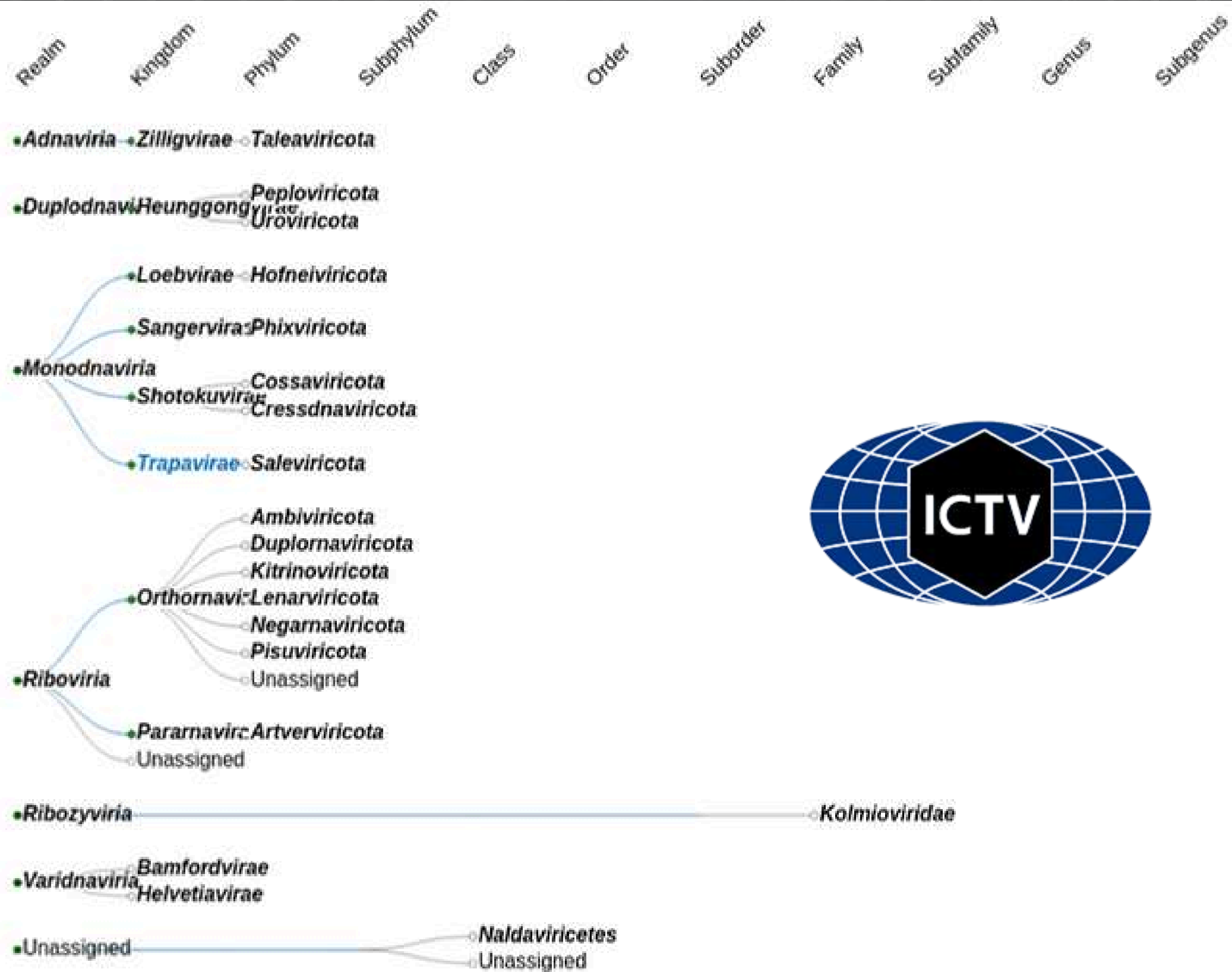


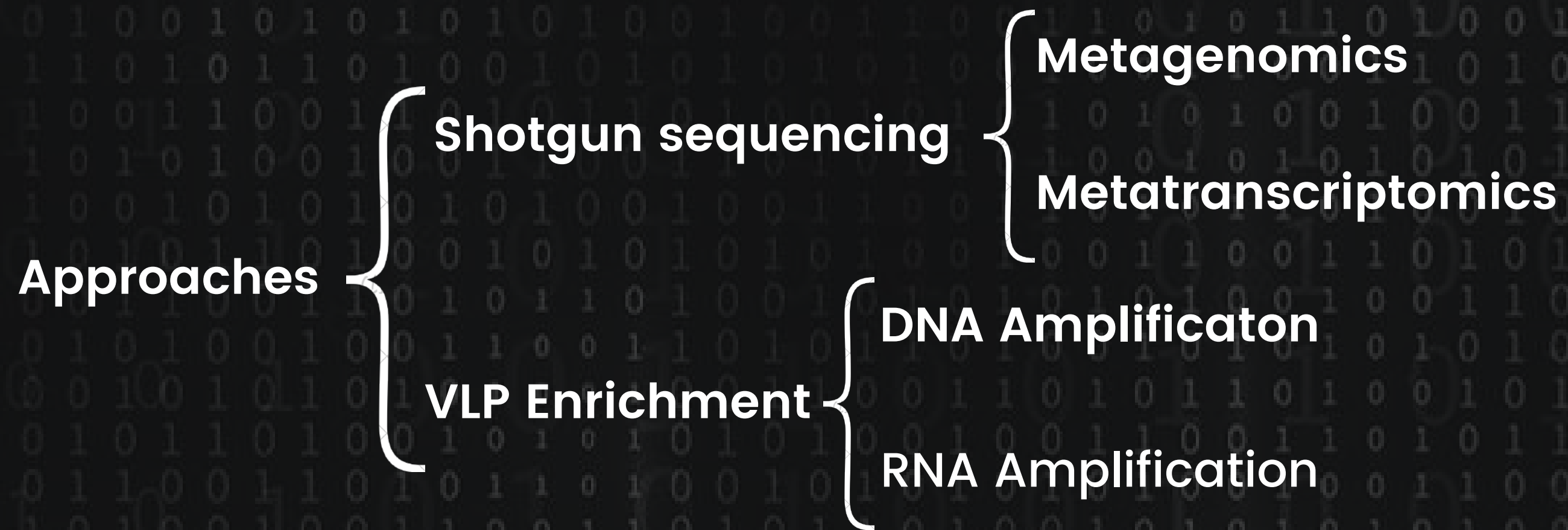
shotgun  
sequencing?



# Baltimore Classification







**Is there any viral gene analogous to the universal gene 16S in Bacteria which we can amplify to study viruses?**

**Is viral metabarcoding possible?**



**RdRP in RNA viruses?**

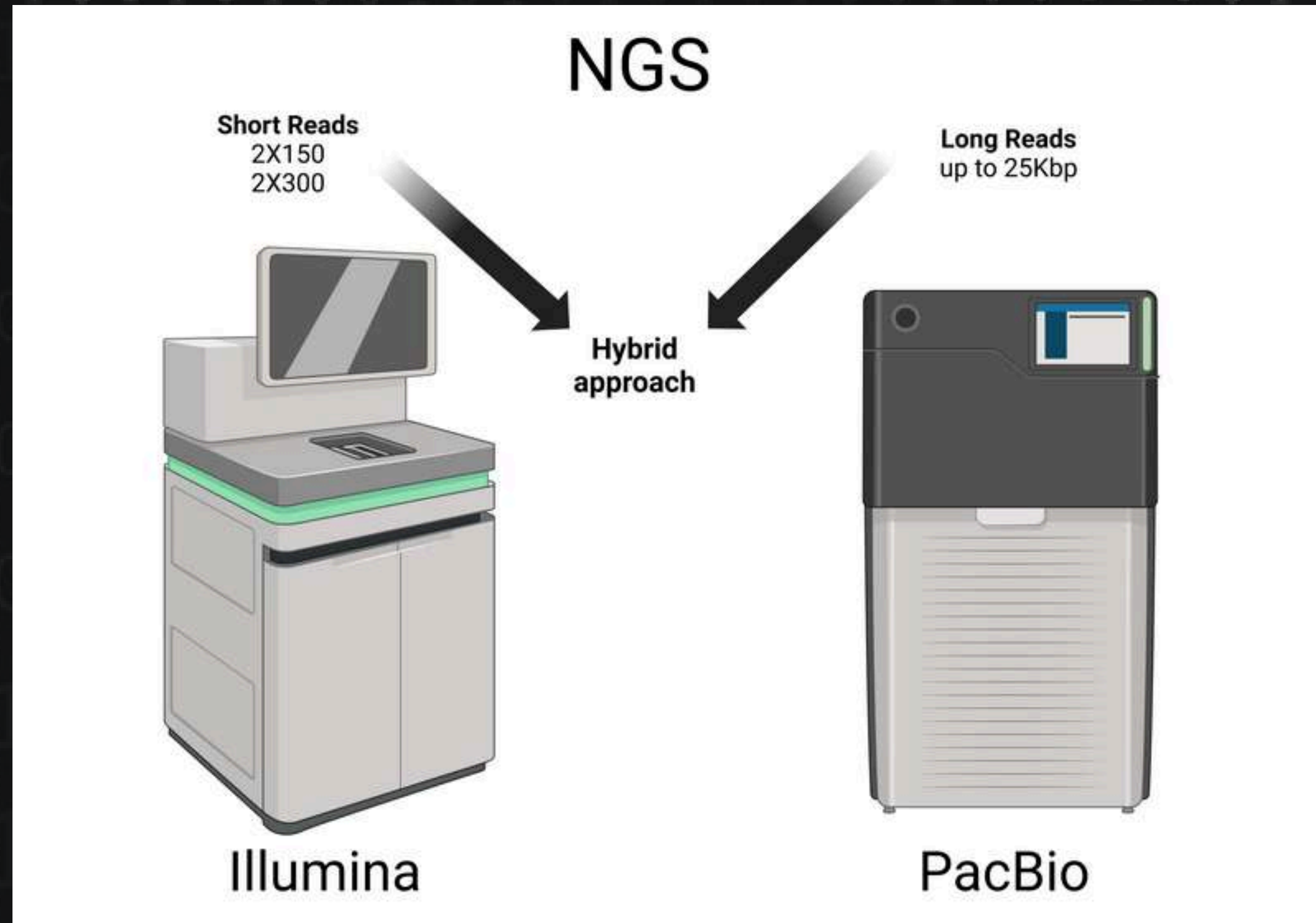


### Pros:

1. High Accuracy
2. High Throughput
3. Cost-Effective
4. Established Bioinformatics Tools

### Cons:

1. Fragmentation of Genomes
2. Difficulty in Resolving Repeats and Complex Regions
3. Recombination and Diversity Detection



### Pros:

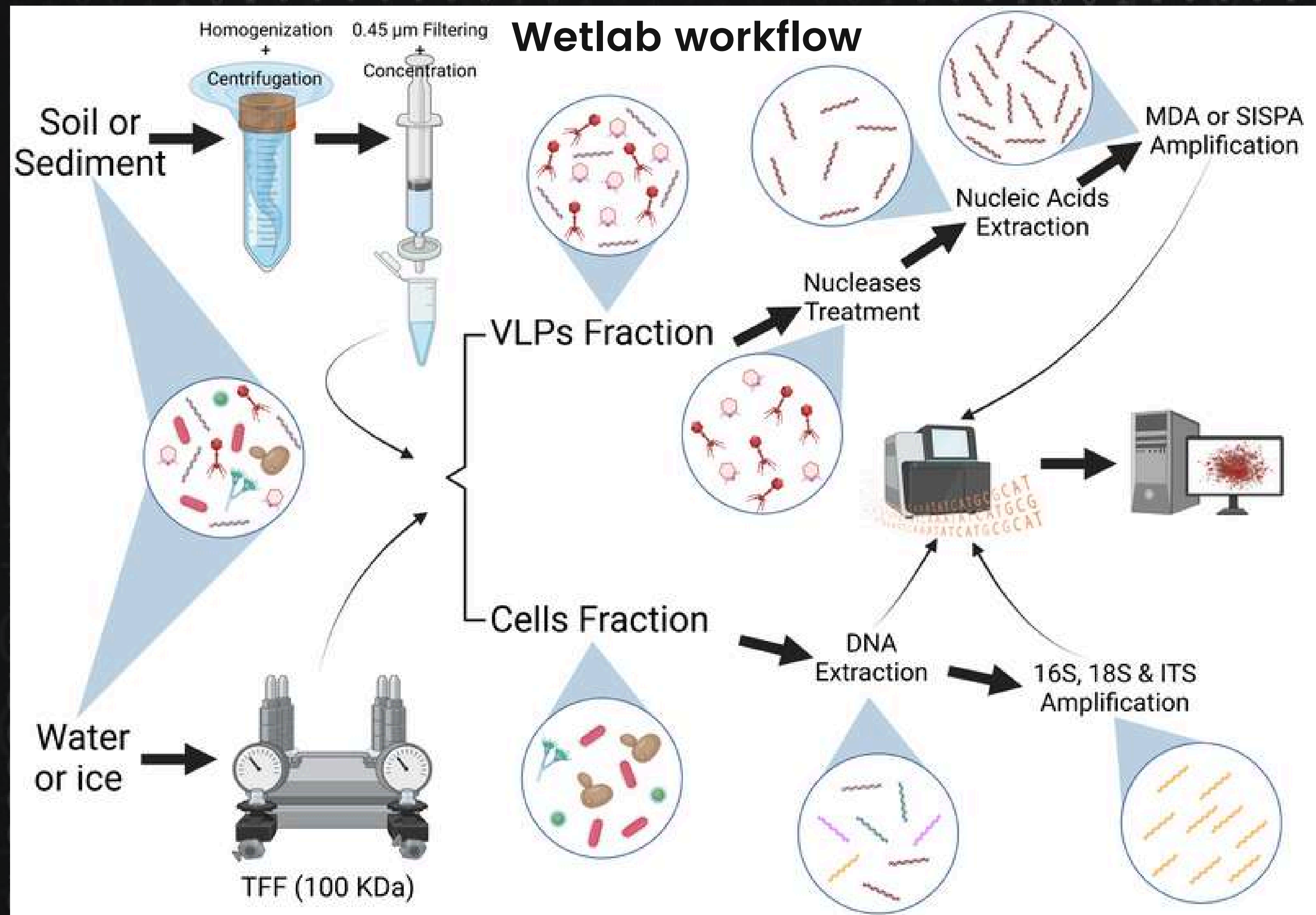
1. Full Genome Assembly
2. Detection of Structural Variations.
3. Resolving Repetitive and Conserved Regions
4. Real-Time Sequencing

### Cons:

1. Higher Error Rates
2. Lower Throughput
3. Cost Per Read
4. Bioinformatics Complexity

### 3. Wetlab workflow





# Issues

The filtering pore diameter matters

0.45  $\mu\text{m}$  VS 0.22  $\mu\text{m}$

0.45  $\mu\text{m}$ : Prokaryotes contamination  
0.22  $\mu\text{m}$ : giant viruses lost



16S PCR check

PROs:

- Normally provides higher resolution of viral diversity

CONS:

- No Prokaryotes information
- Bias in frequencies

Amplification = Bias

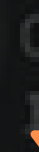
SISPA (RNA):

- low abundance viruses
- Small RNA viruses
- SSRNA Viruses
- Highly secondary structured
- High GC content
- Segmented viruses



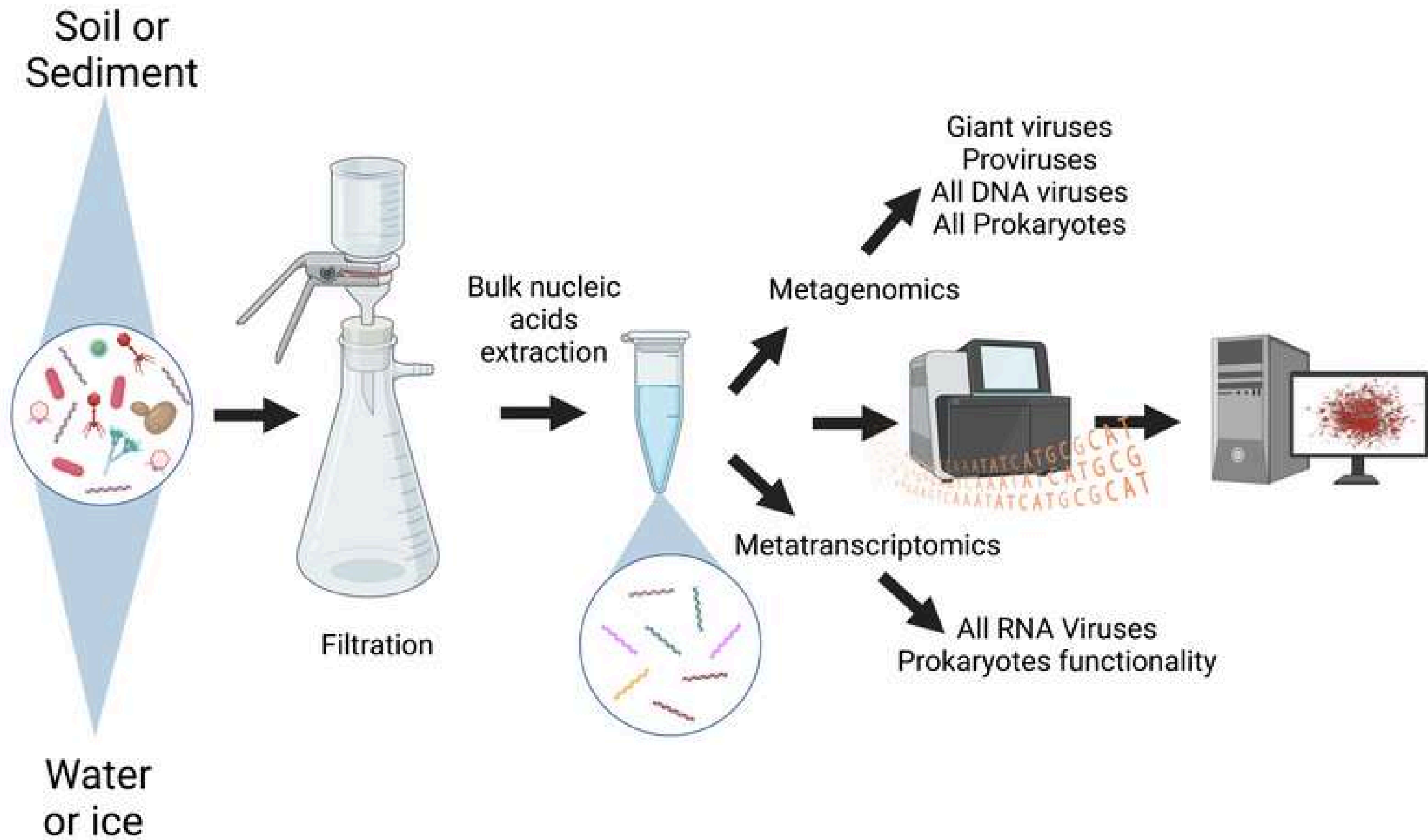
MDA (DNA):

- ssDNA
- cDNA
- Extreme GC content
- Secondary structured
- Repetitive regions
- low abundance viruses





# Wetlab workflow



# Issues

## PROs:

- Normally provides higher resolution in terms of viral frequencies distribution
- Abundant information about Prokaryotes

## CONS:

- Low proportion of viral sequences
- Lost of viral diversity
- Requires more bioinformatics (“needle in a haystack”)



# The Big Issue



**Avoid contamination!!!**

- Sterile conditions (lab, tools, etc)
- Sample cleaning
- Controls

## 4. Drylab workflow



## Common workflow:

1. Reads quality inspection
2. Quality filtering and trimming
3. Reads quality re-inspection
4. Assembling
5. Virus discovery
6. Binning
7. Quantification
8. Checking contigs quality
9. Functional annotation
10. Taxonomic annotation
11. Host prediction
12. Many others...

## 5. “Hands on” sea cucumbers dataset



# sersancar/ **Viromics\_Vigo**



Viromics day within the course "APLICACIÓN DE HERRAMIENTAS -ÓMICAS EN ACUICULTURA"

1

Contributor

0

Issues

0

Stars

0

Forks



**sersancar/Viromics\_Vigo: Viromics day within the course "APLICACIÓN DE HERRAMIENTAS -ÓMICAS EN...**

Viromics day within the course "APLICACIÓN DE HERRAMIENTAS -ÓMICAS EN ACUICULTURA" - sersancar/Viromics\_Vigo

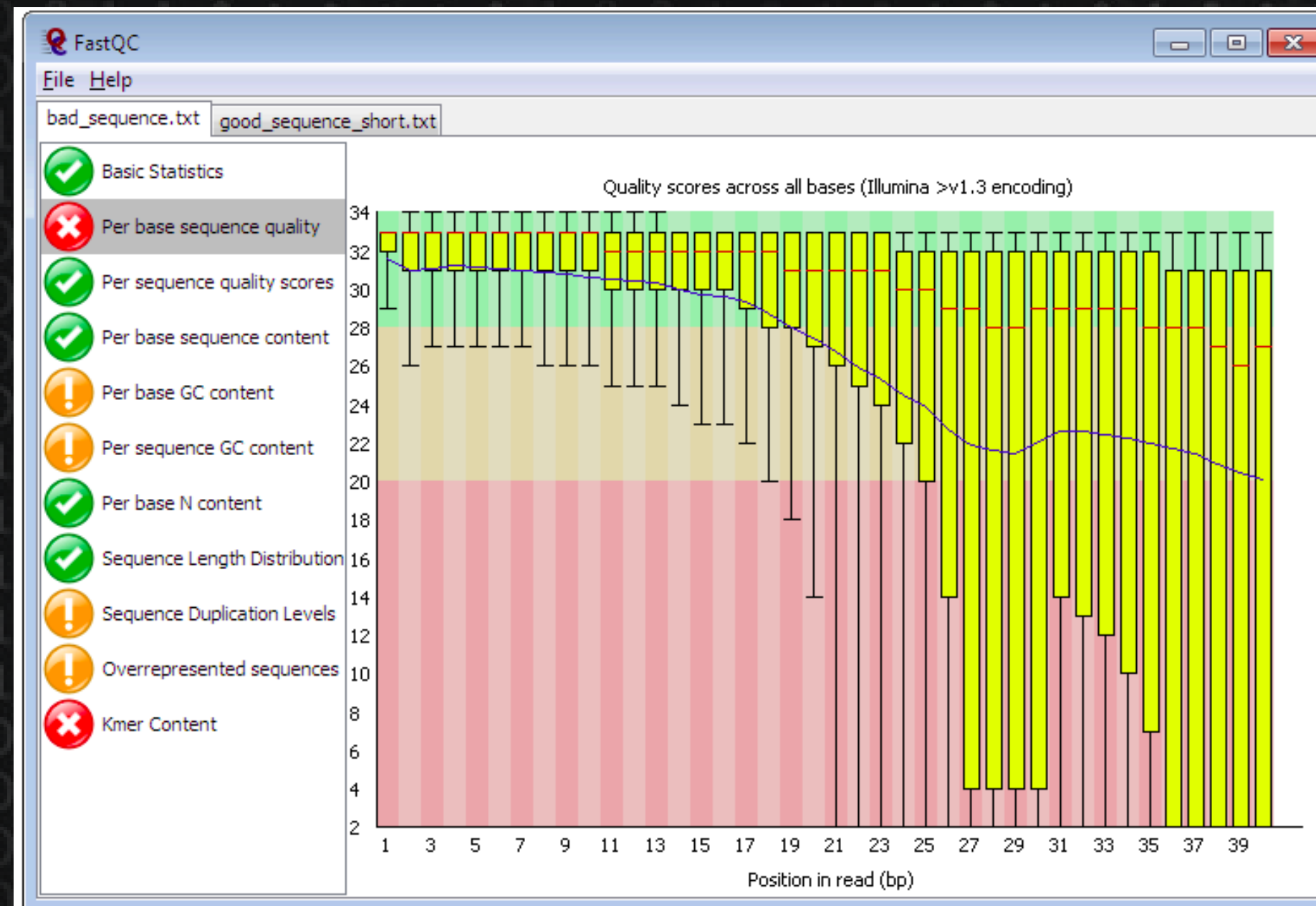


[https://github.com/sersancar/Viromics\\_Vigo](https://github.com/sersancar/Viromics_Vigo)

# Reads quality inspection

## FastQC

<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

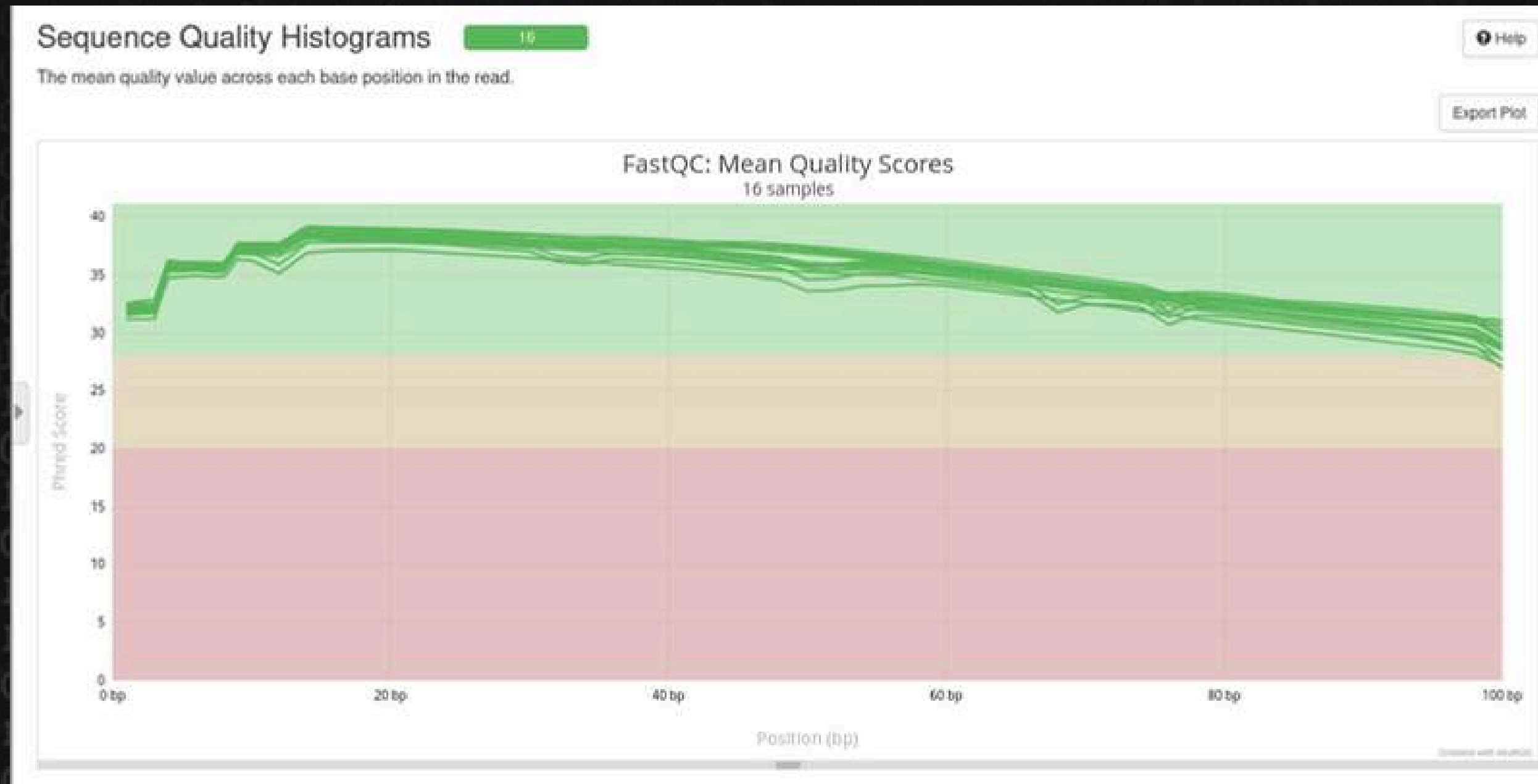




## Reads quality inspection

# MultiQC

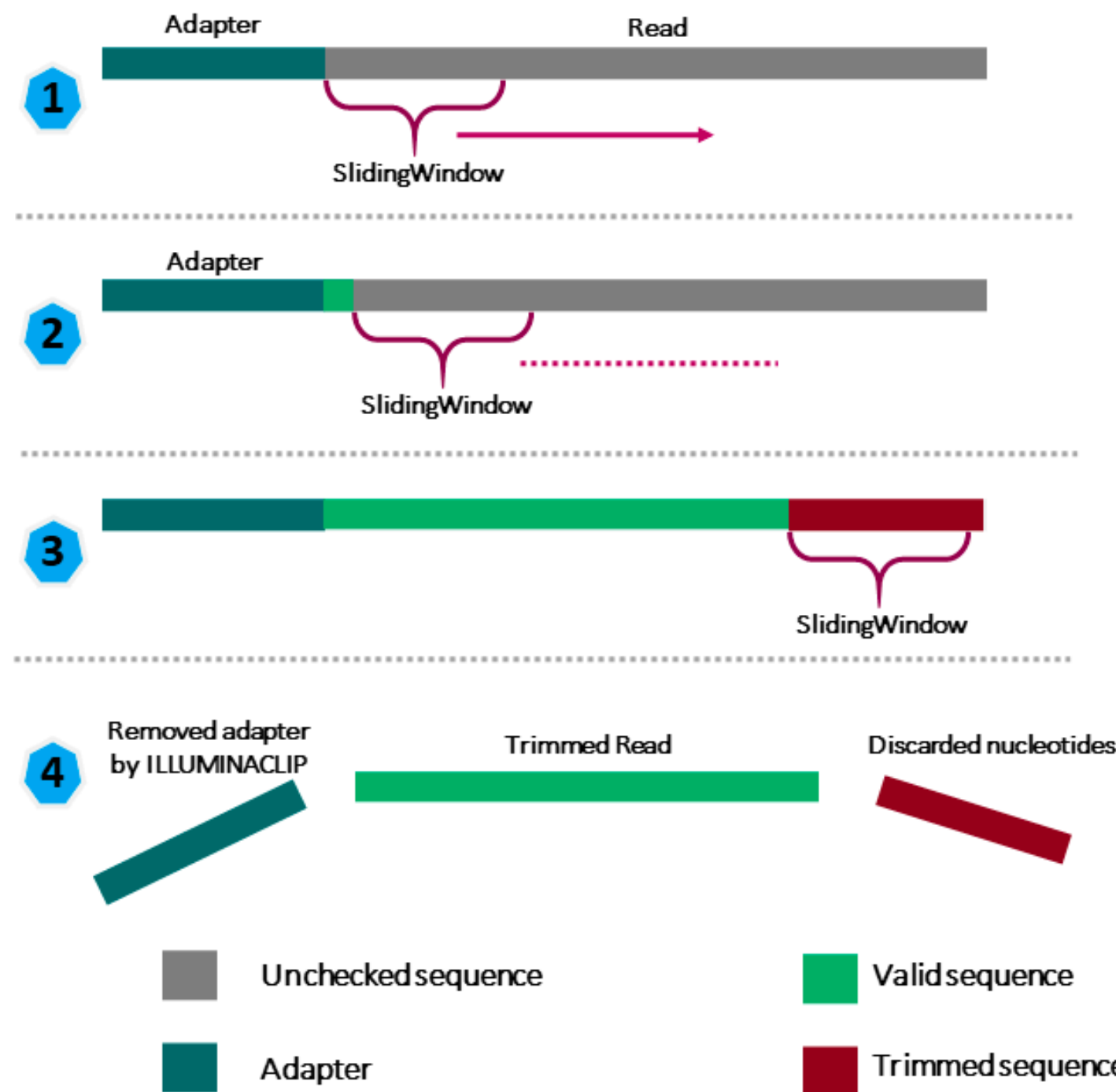
Ewels et al 2016



## Quality filtering and trimming

# Trimmomatic

Bolger, Lohse & Usadel 2014





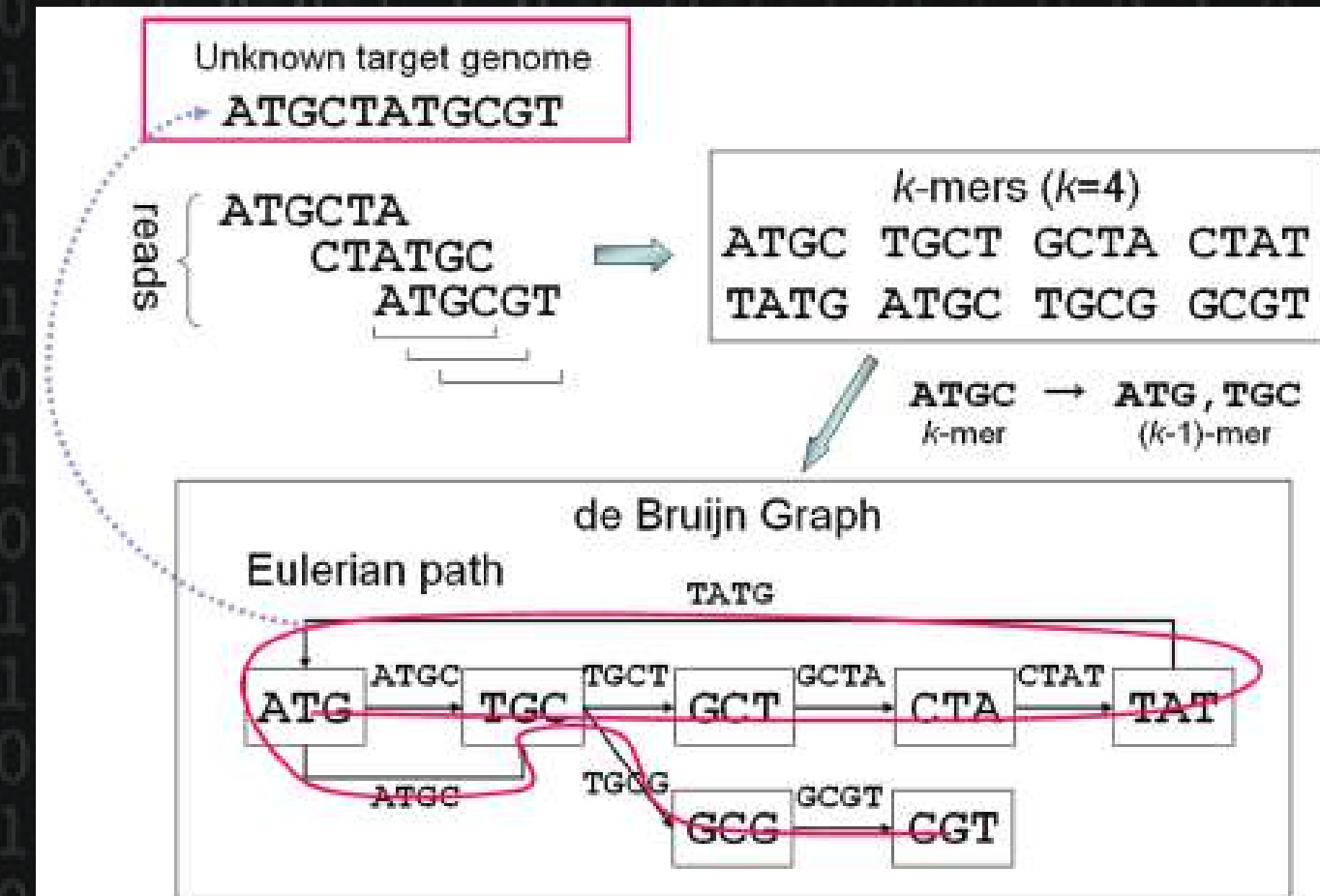
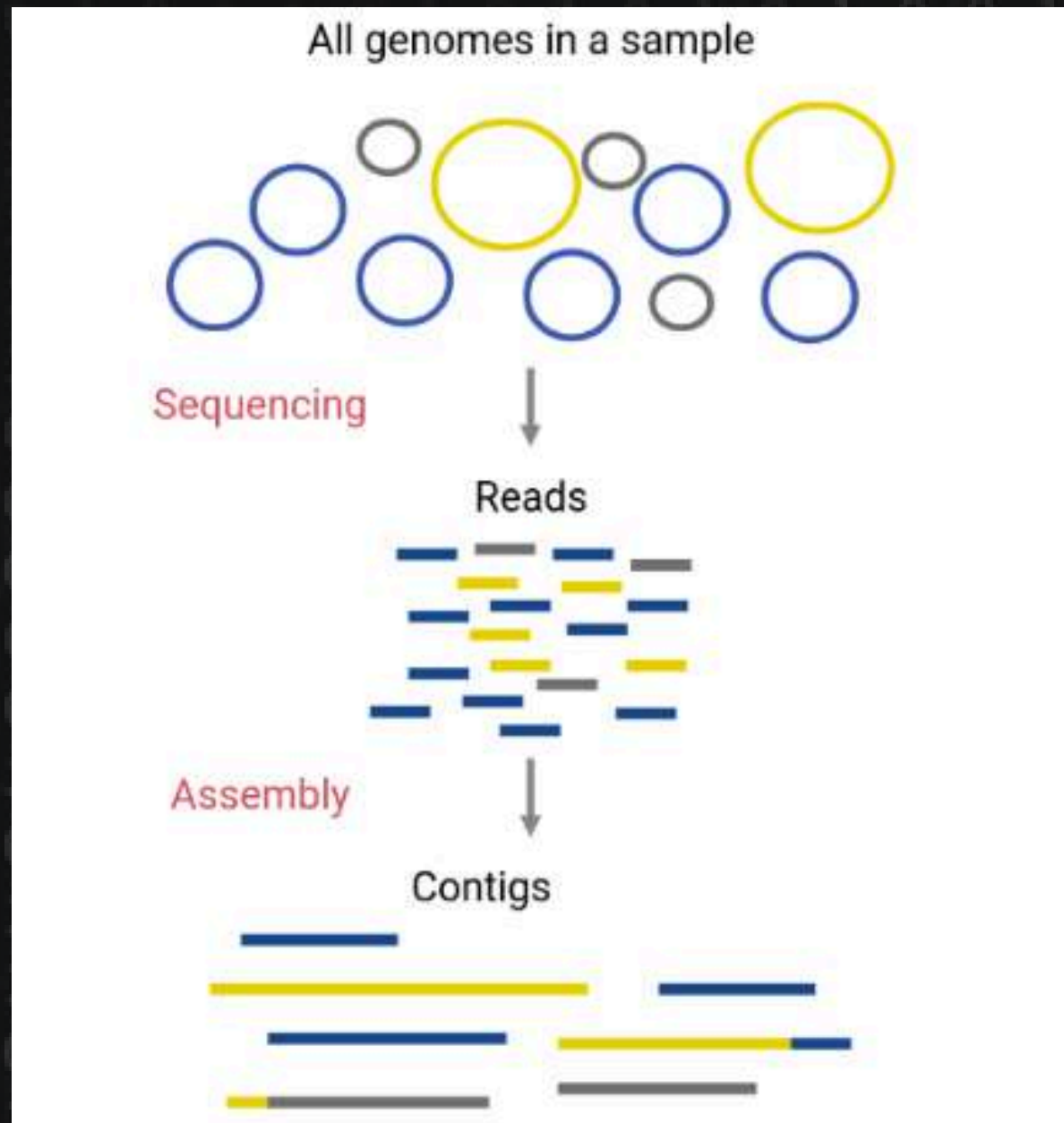
## Assembling

SPAdes

Prjibelski et al. 2020

MEGAHIT

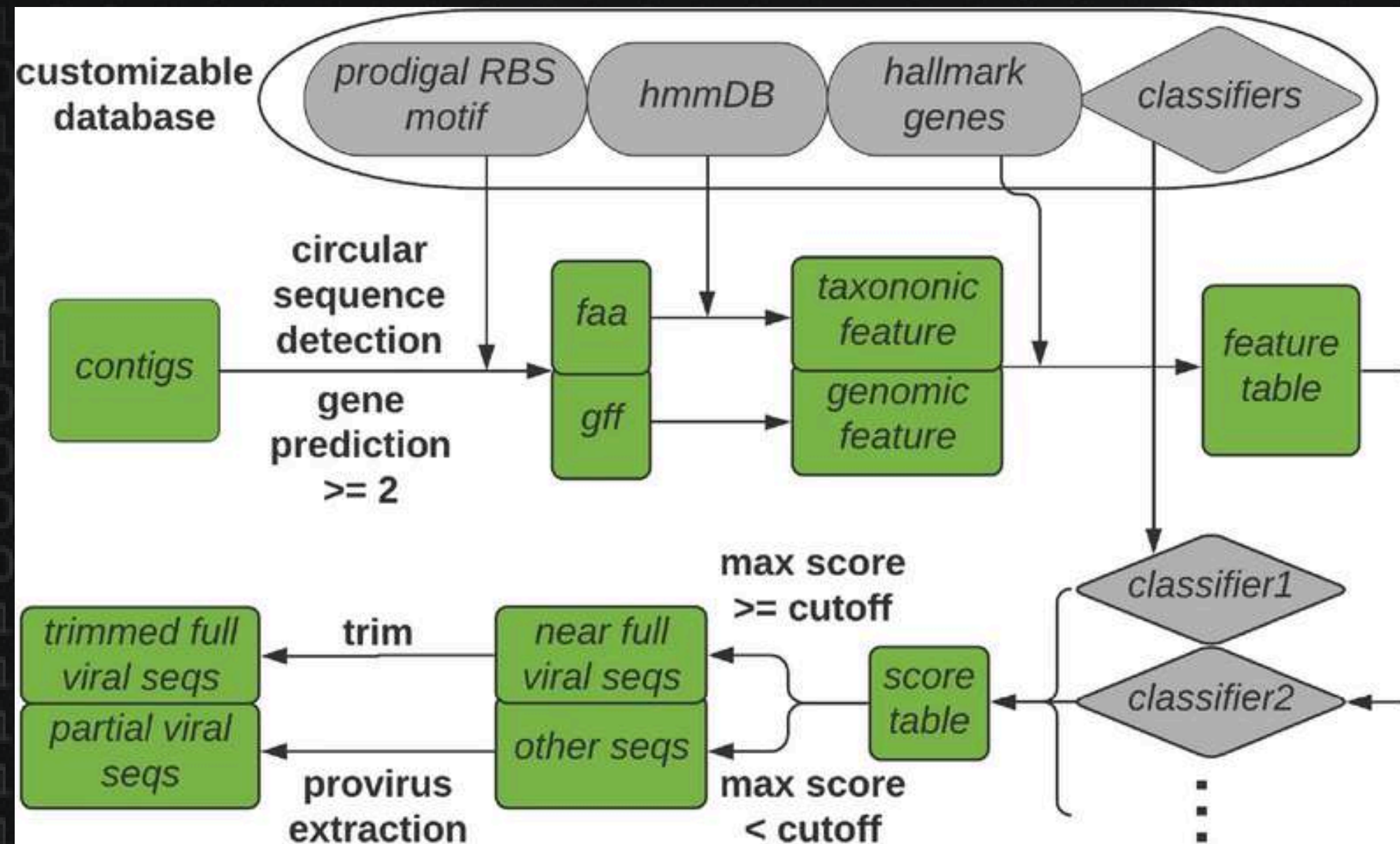
Li et al. 2015



# Virus discovery

## VirSorter2

Guo et al. 2021

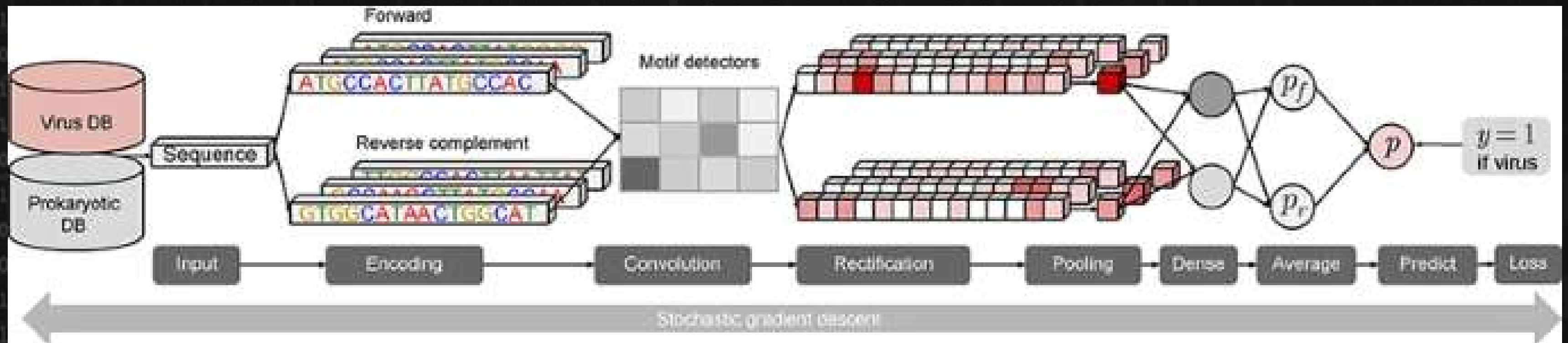




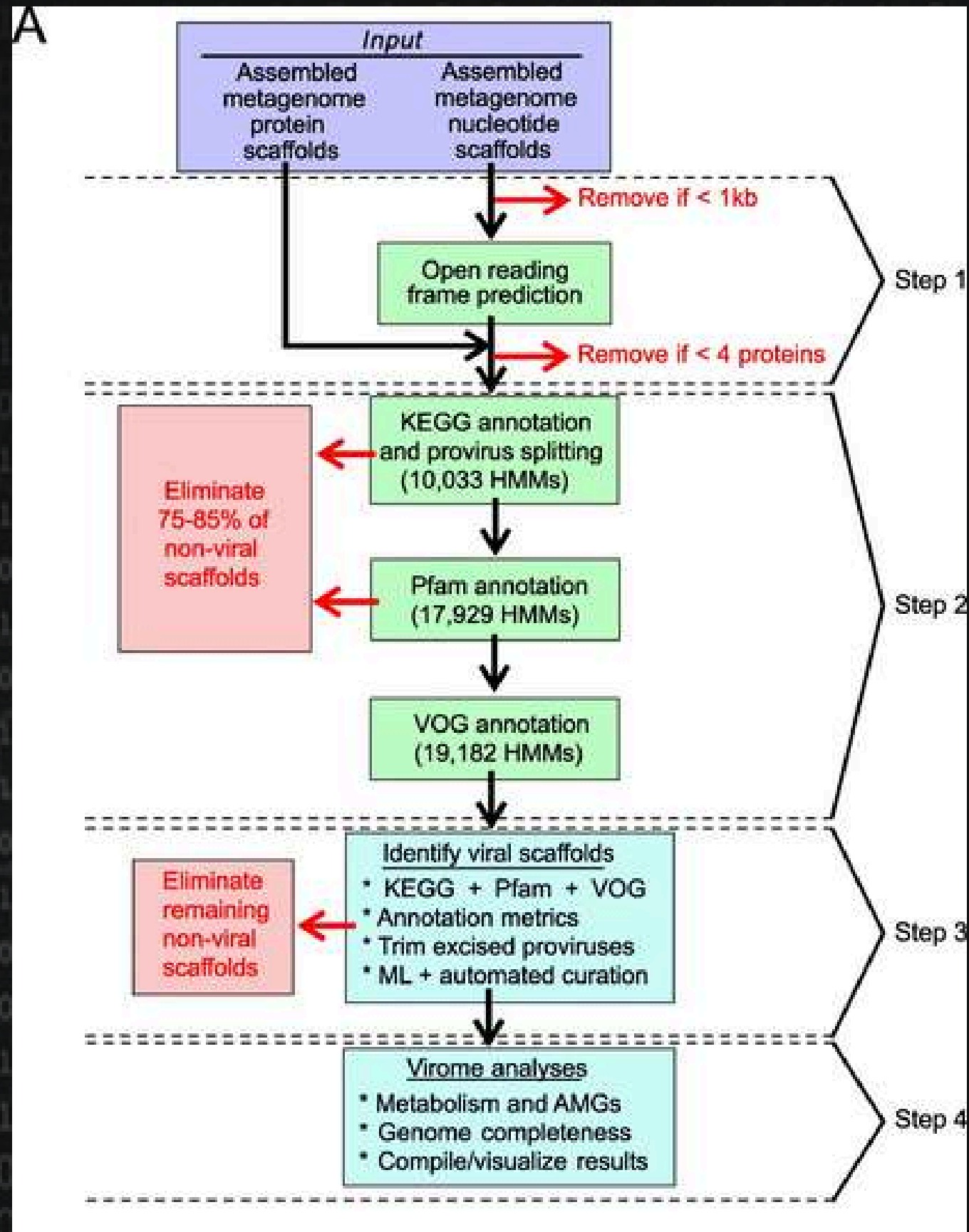
## Virus discovery

# DeepVirFinder

Ren et al. 2020



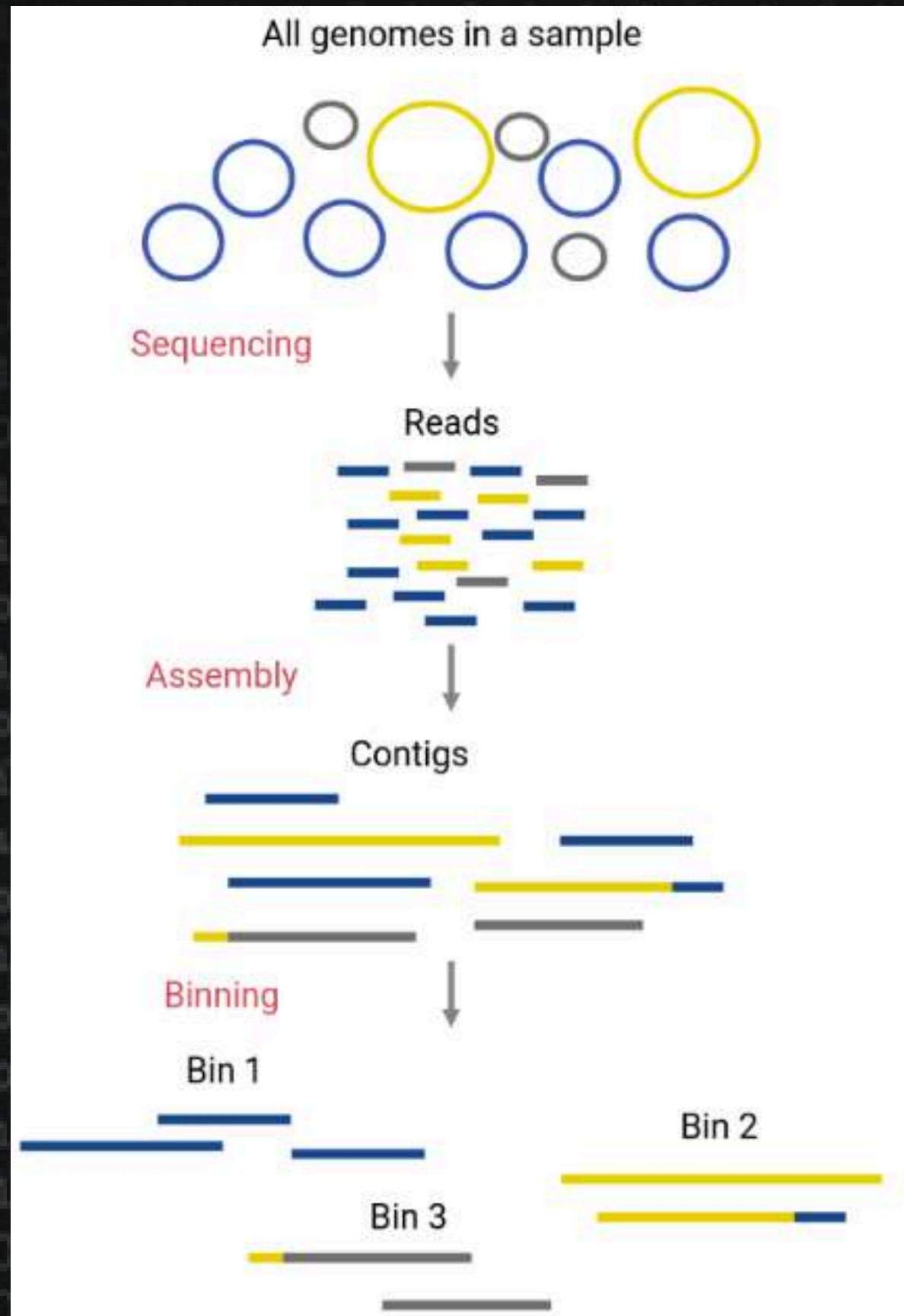
# Virus discovery



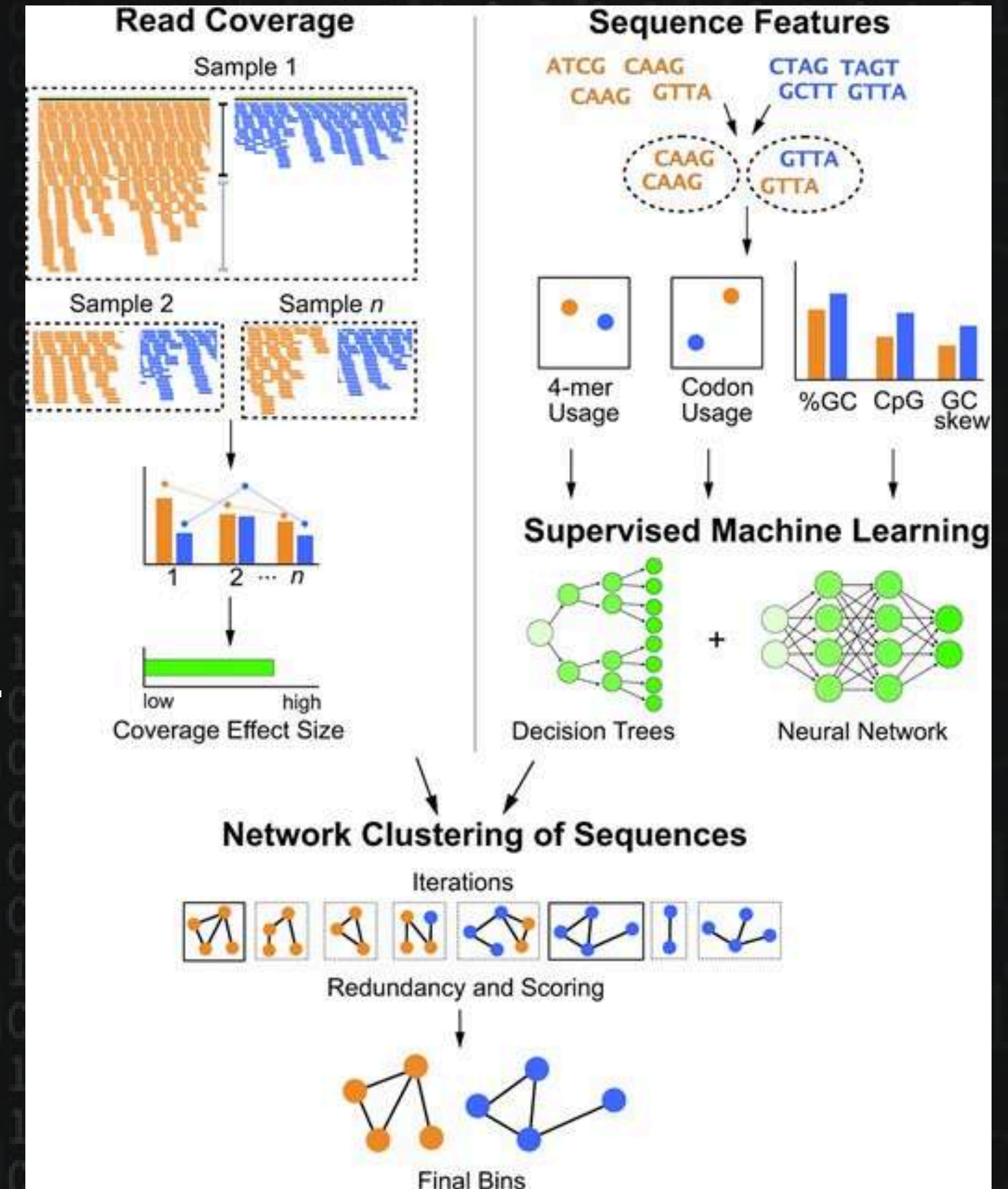
**VIBRANT**  
Kieft, Zhou & Anantharaman 2022



# Binning



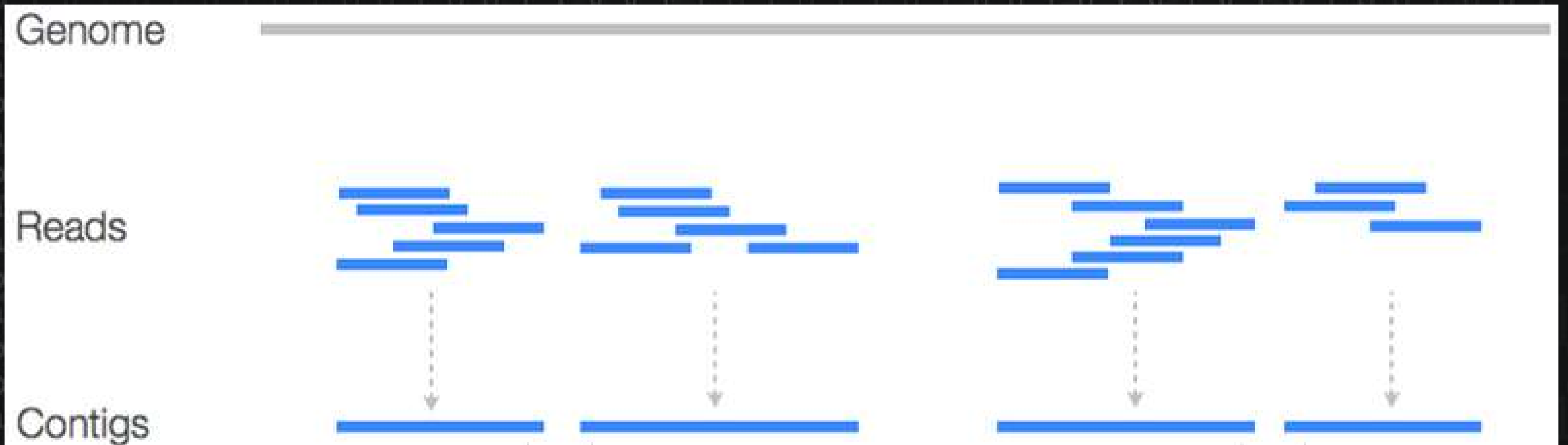
VRhyme  
Kieft et al. 2022



Quantification

CoverM

<https://github.com/wwood/CoverM>

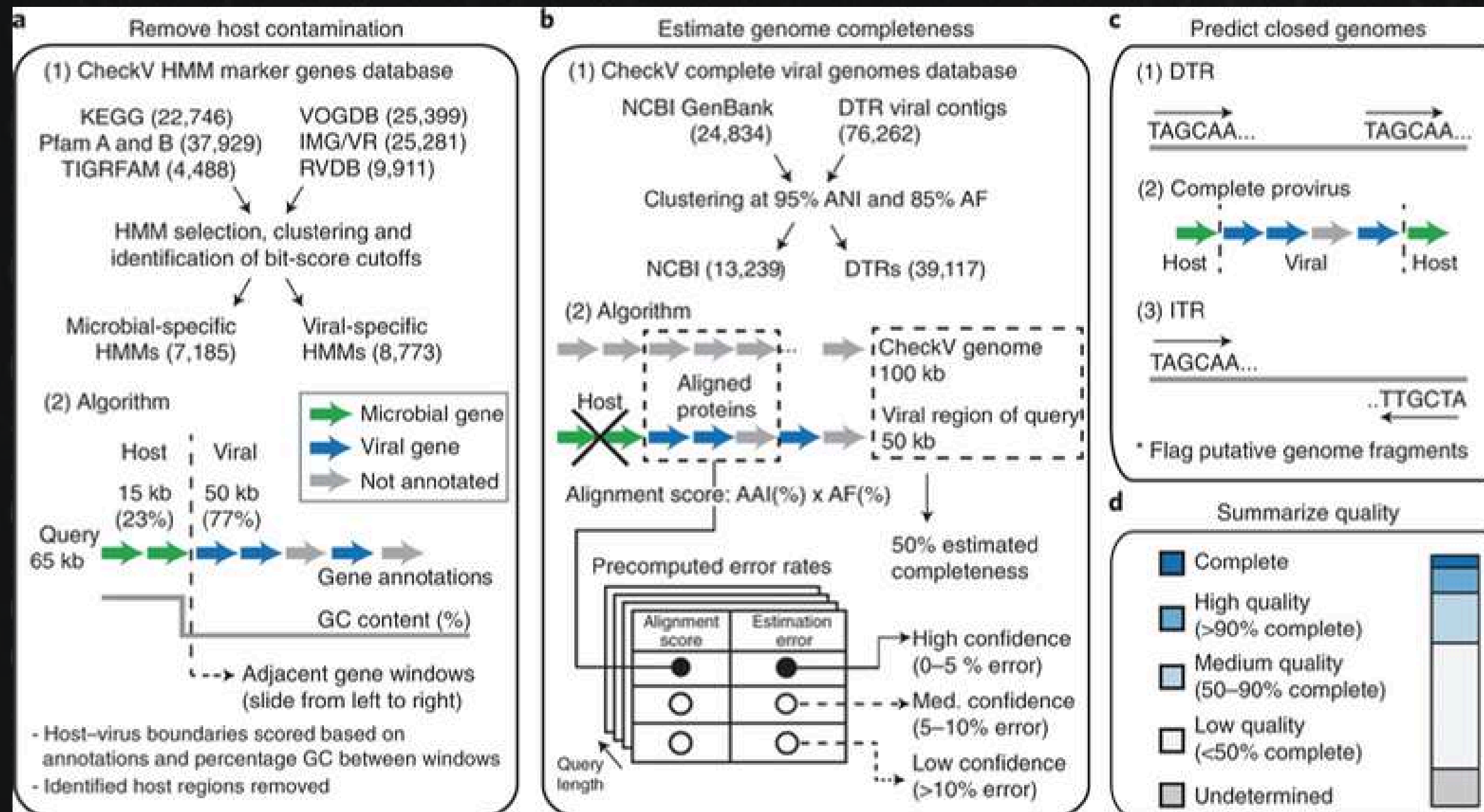


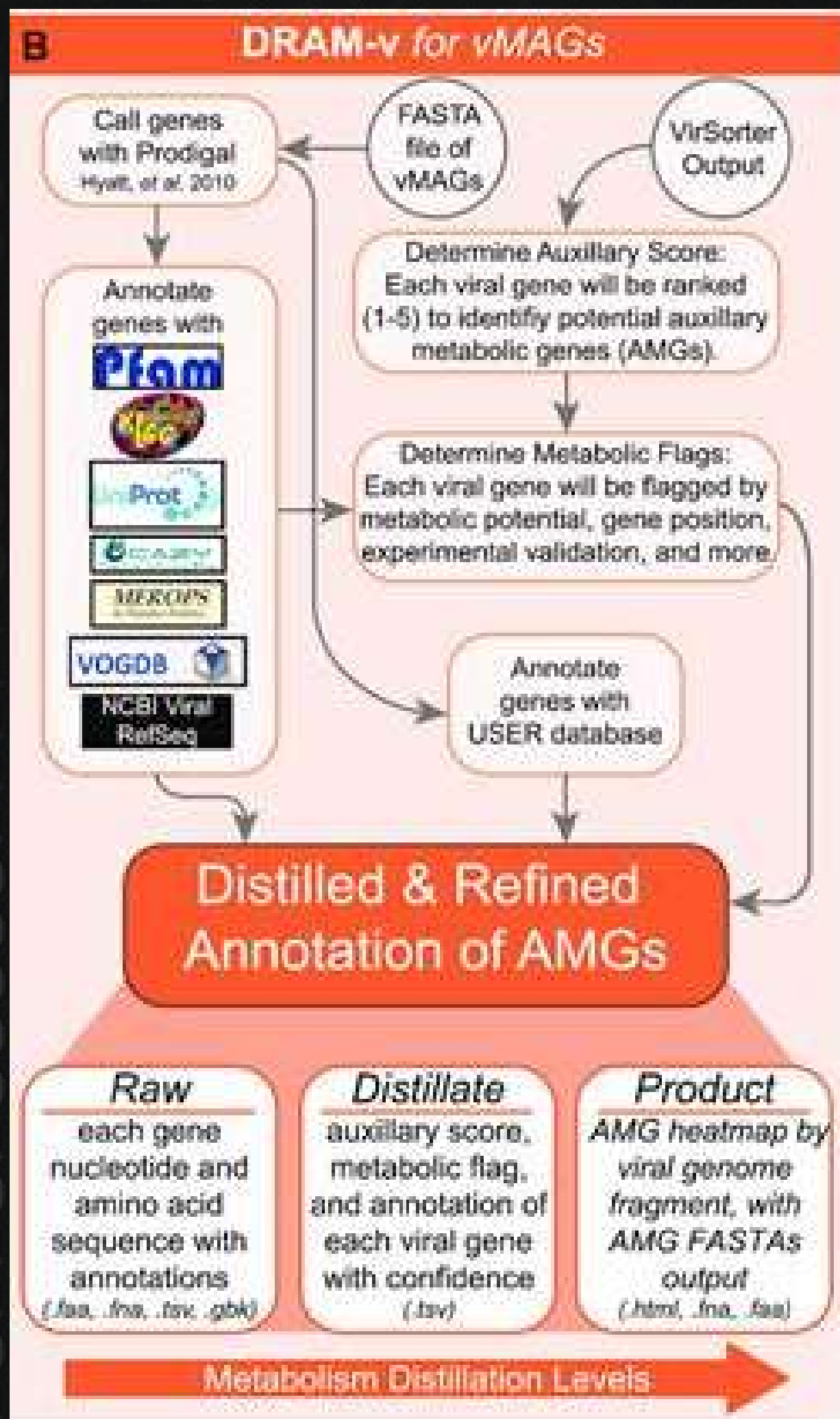


# Checking contigs quality

## CheckV

<https://www.nature.com/articles/s41587-020-00774-7>





Functional annotation

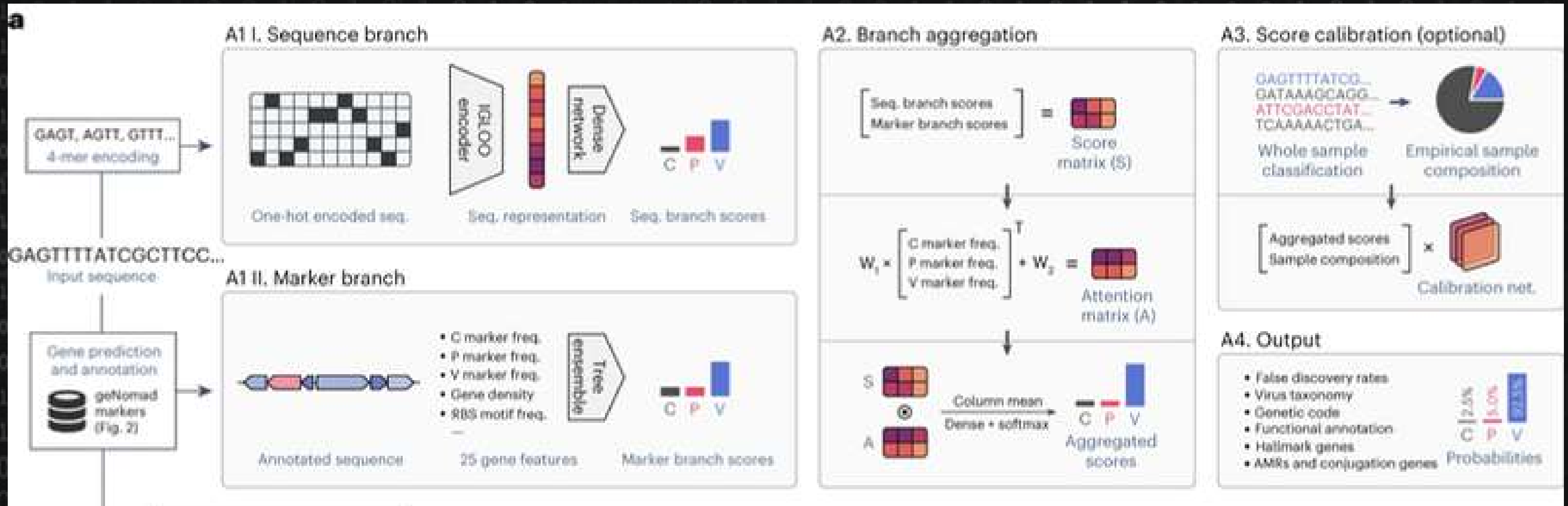
DRAM-v

Shaffer et al. 2020

Taxonomic annotation and many more....

geNomad

Camargo et al. 2024

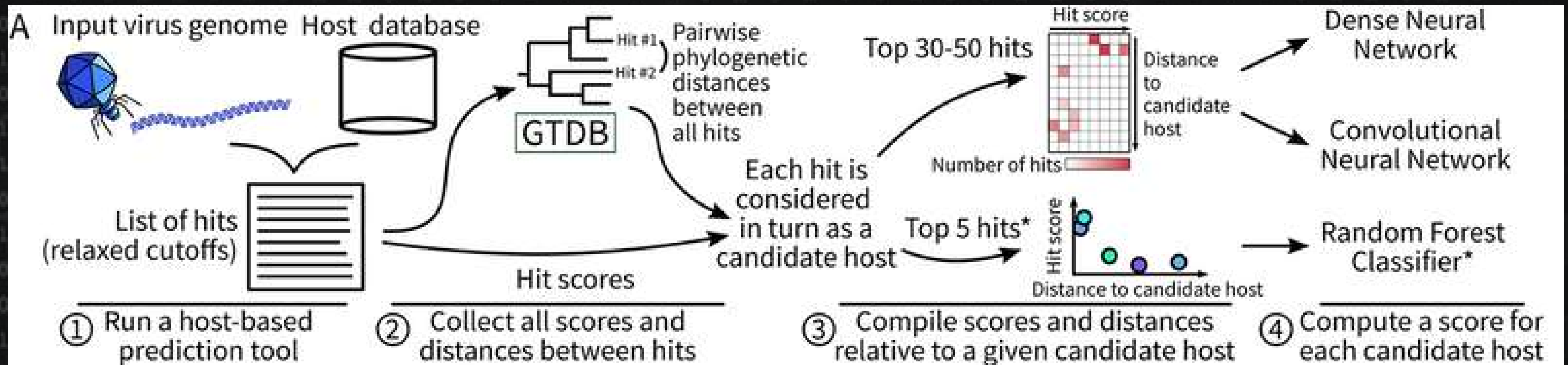


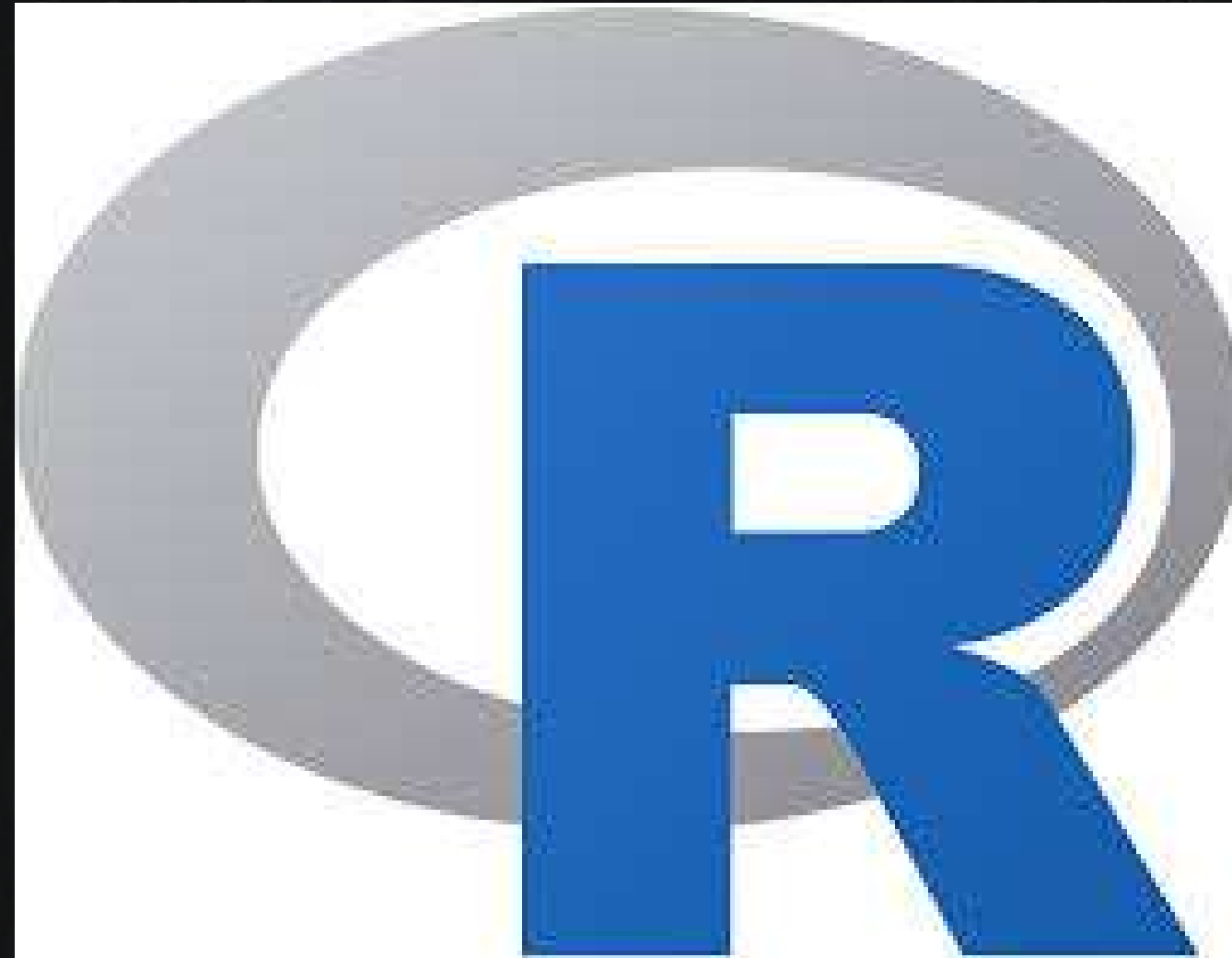


# Host prediction

## iPHoP

Roux et al. 2023





Data analysis

<https://www.r-project.org>

## Contact details:

Sergio Sánchez Carrillo

`sergio.sanchez@cbm.csic.es`

`sergio.sanchez.carrillo@csic.es`

`sergiosanchezcarrillo@gmail.com`

[www.linkedin.com/in/sergio-sánchez-carrillo-8a9121295](https://www.linkedin.com/in/sergio-sánchez-carrillo-8a9121295)

<https://orcid.org/0000-0002-9271-7484>