

YAVSAP

Versatile viral quasispecies analysis for veterinary samples

Thomas Christensen II ¹ Steven Stanic ¹ Andrea Lu ¹ Dana
Mitzel ² William Wilson ² Rachel Palinski ¹

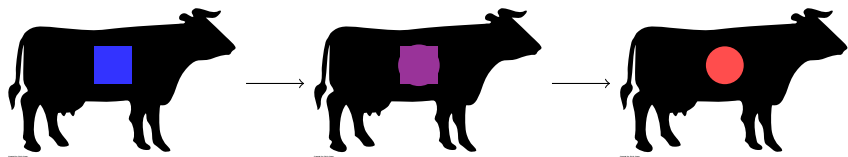
Kansas State University Veterinary Diagnostic Laboratory, Kansas State University,
Manhattan, KS

Foreign Arthropod Borne Animal Disease Research, National Bio- and Agro-Defence Facility,
United States Department of Agriculture Agricultural Research Service, Manhattan, KS

September 15, 2024

What is a quasispecies?

Consensus level



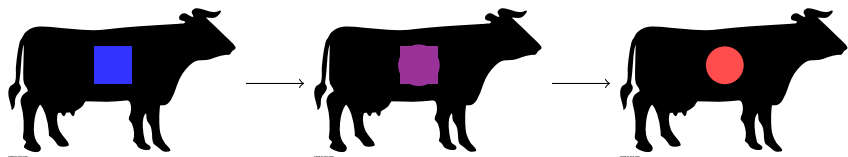
¹Domingo, Sheldon, and Perales, "Viral Quasispecies Evolution".

²Andino and Domingo, "Viral Quasispecies".

³Vignuzzi et al., "Quasispecies Diversity Determines Pathogenesis through Cooperative Interactions in a Viral Population".

What is a quasispecies?

Consensus level



Quasispecies model



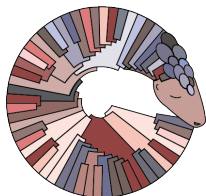
¹Domingo, Sheldon, and Perales, "Viral Quasispecies Evolution".

²Andino and Domingo, "Viral Quasispecies".

³Vignuzzi et al., "Quasispecies Diversity Determines Pathogenesis through Cooperative Interactions in a Viral Population".

Existing quasispecies pipelines

nf-core/
viralrecon



⁴Harshil Patel et al., *Nf-Core/Viralrecon*.

⁵O'Toole et al., "Assignment of Epidemiological Lineages in an Emerging Pandemic Using the Pangolin Tool".

⁶Hadfield et al., "Nextstrain".

⁷Wang et al., "An Integrated Software for Virus Community Sequencing Data Analysis".

Gap in veterinary coverage: viralrecon

```
N E X T F L O W ~ version 23.04.2
Pulling nf-core/viralrecon ...
Already-up-to-date
Launching 'https://github.com/nf-core/viralrecon' [nostalgic_curran] DSL2 - revision: 3731dd3a32 [master]

ERROR ~ ~~~~~
Genome 'genome.fasta' not found in any config files provided to the pipeline.
Currently, the available genome keys are:
NC_045512.2, MN908947.3, NC_063383.1, ON563414.3, MT903344.1
~~~~~

-- Check '.nextflow.log' file for details
```

Gap in veterinary coverage: Nextstrain

Dataset

nCoV

dengue

ebola

enterovirus

flu

lassa

measles

mpox

mumps

nCoV

nextclade


rsv

tb

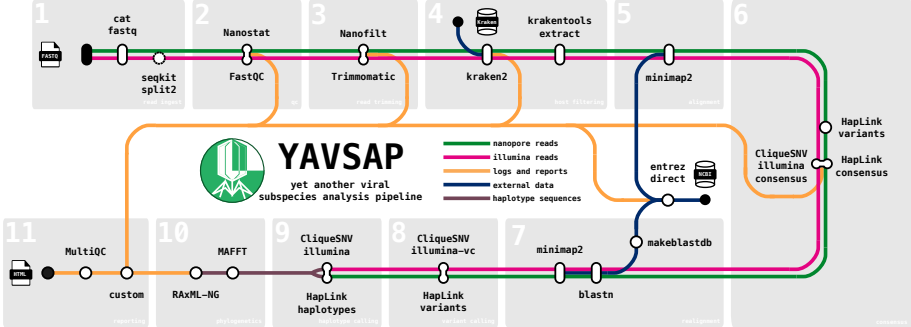
WNV

zika

Clade

Filter Data 

Introducing YAVSAP



Sequencing technologies

Illumina



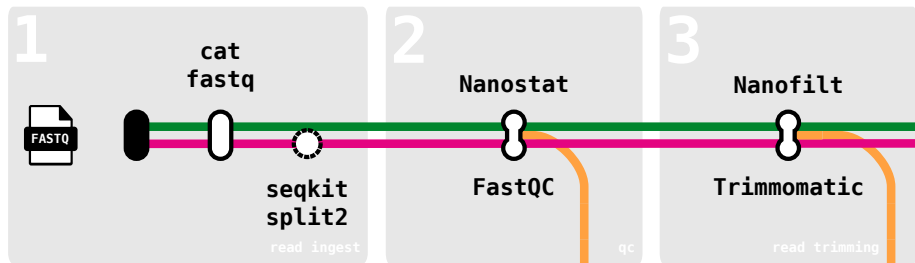
- short reads
- high quality
- reads come in pairs

Oxford Nanopore



- long reads
- low quality
- single reads

Ingestion, quality control, and filtering



Host filtering

4

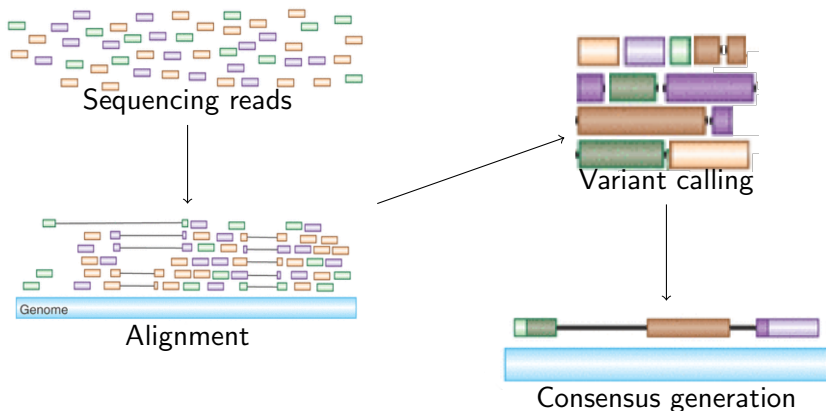


**krakentools
extract**

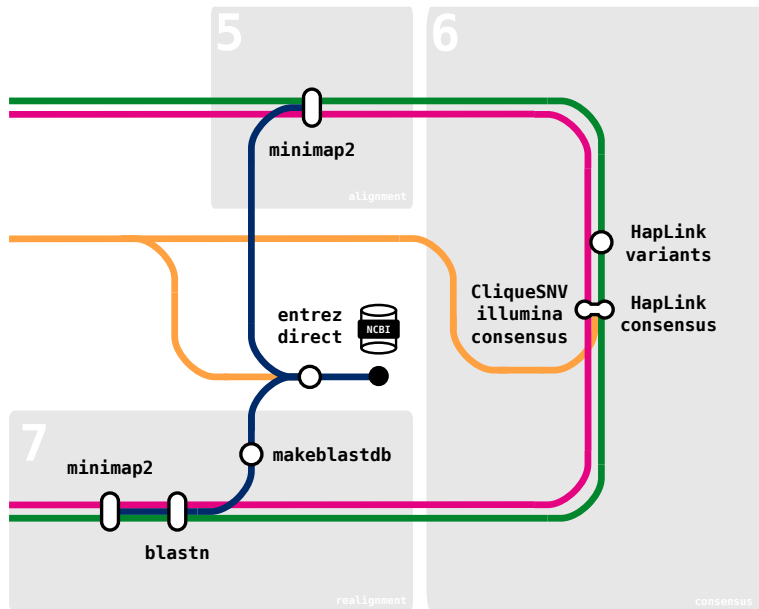
kraken2

host filtering

Alignment: process overview

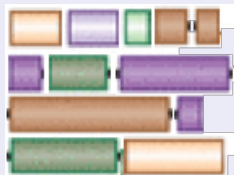


Alignment, consensus generation, and realignment



Variant and haplotype calling: process overview

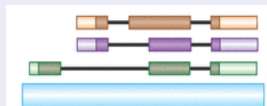
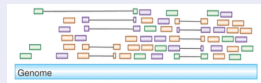
Variant calling



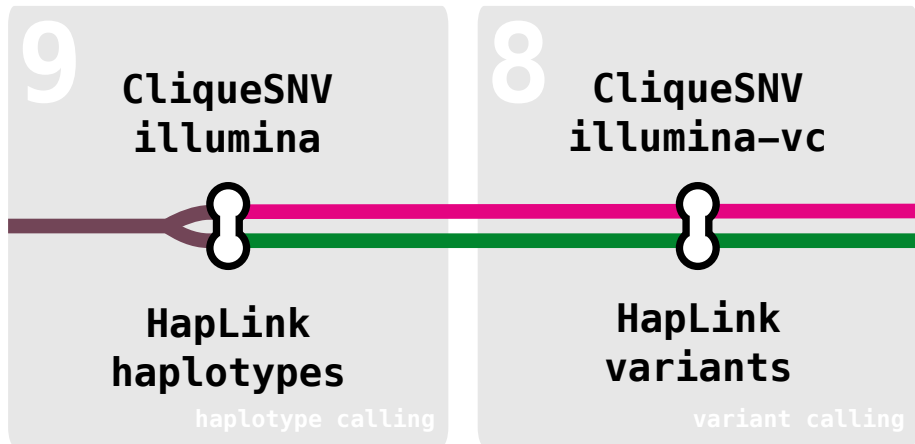
Consensus calling



Haplotype calling




Variant and haplotype calling



Phylogenetics

10

MAFFT



RAXML-NG

phylogenetics

>ROOT

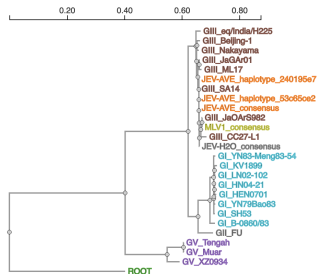
```
ggcataagcagaggatgtcactactattaac
```

>pool01_consensus

```
ggcataagcagaggatgtcactactattaac
```

>sample03_haplotype_8637fdf8

```
ggcataagcagaggatgtcactactattaac
```



MultiQC

v1.12

General Stats

NanoStat

Fastq stats

Reads by quality

Kraken

HapLink.jl (Haplotypes)

RAxML-NG

yavsap Software Versions

yavsap Workflow Summary

MultiQC

A modular tool to aggregate results from bioinformatics analyses across many samples into a single report.

This report has been generated by the [yavsap](#) analysis pipeline. For information about how to interpret these results, please see the [documentation](#).

Report generated on 2022-09-17, 22:32

Welcome! Not sure where to start?

[Watch a tutorial video](#) (6:06)

[don't show again](#) ✕

General Statistics

[Copy table](#)

[Configure Columns](#)

[Plot](#)

Showing $\frac{3}{3}$ rows and $\frac{3}{3}$ columns.

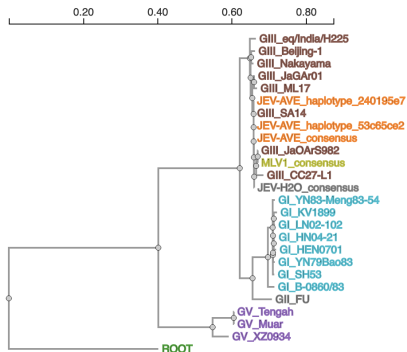
Sample Name	% <i>Mesocricetus auratus</i>	% Top 5 Species	% Unclassified
JEV-AVE	18.2%	49.2%	4.7%
JEV-H2O	17.3%	44.0%	6.1%
MLV1	0.7%	26.3%	17.5%

Toolbox



Viral analyses

- Japanese Encephalitis Virus
- Influenza D Virus
- Bovine Coronavirus
- SARS CoV2
- Rift Valley Fever Virus
- Porcine Reproductive and Respiratory Syndrome Virus



Next steps

Internal

- Novel virus discovery
- Multi-strand virus improvements
- Protein transcript prediction

External

- Identify JEV “bottleneck” genomes
- Model transmission of JEV and RVFV
- Surveillance and monitoring of enteric and respiratory viruses
- Vaccine development

Acknowledgments

Molecular NGS Department

- Rachel Palinski
- Steven Stanic
- Andrea Lu
- Tyler Doerksen
- Jayme Souza-Neto

KSVDL

- Nanyan Lu
- Gregg Hanzlicek

USDA

- William Wilson
- Dana Mitzel
- Steven Lakin
- SCINet High-performance computing center

The nf-core community

- Harshil Patel
- Philip Ewels
- James Fellows Yates
- Edmund Miller

The BioJulia community

- Jakob Nissen
- Ciaran O'Mara
- Kevin Bonham

Funding

This material is based upon work supported by the U.S. Department of Agriculture, Agricultural Research Service, under agreement No. 58-3022-0-001.