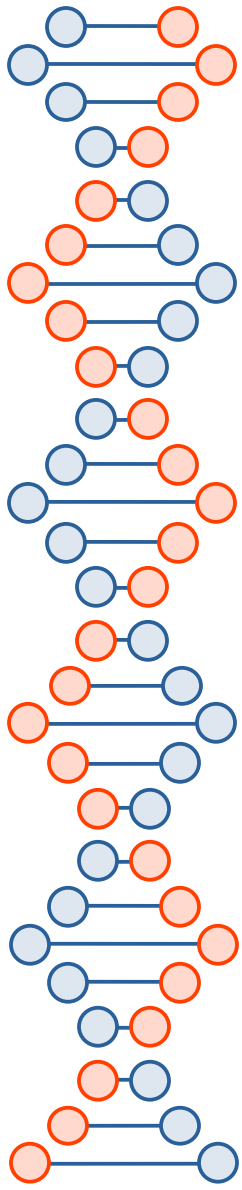# Case Study
African Swine Fever Virus (ASFV)
Phylo-Genomic & Pop-Genomic Analysis

Project Coordinators:
Dr Charles Adeola (KIZ) & Dr Weigang Qiu (Hunter/CUNY)
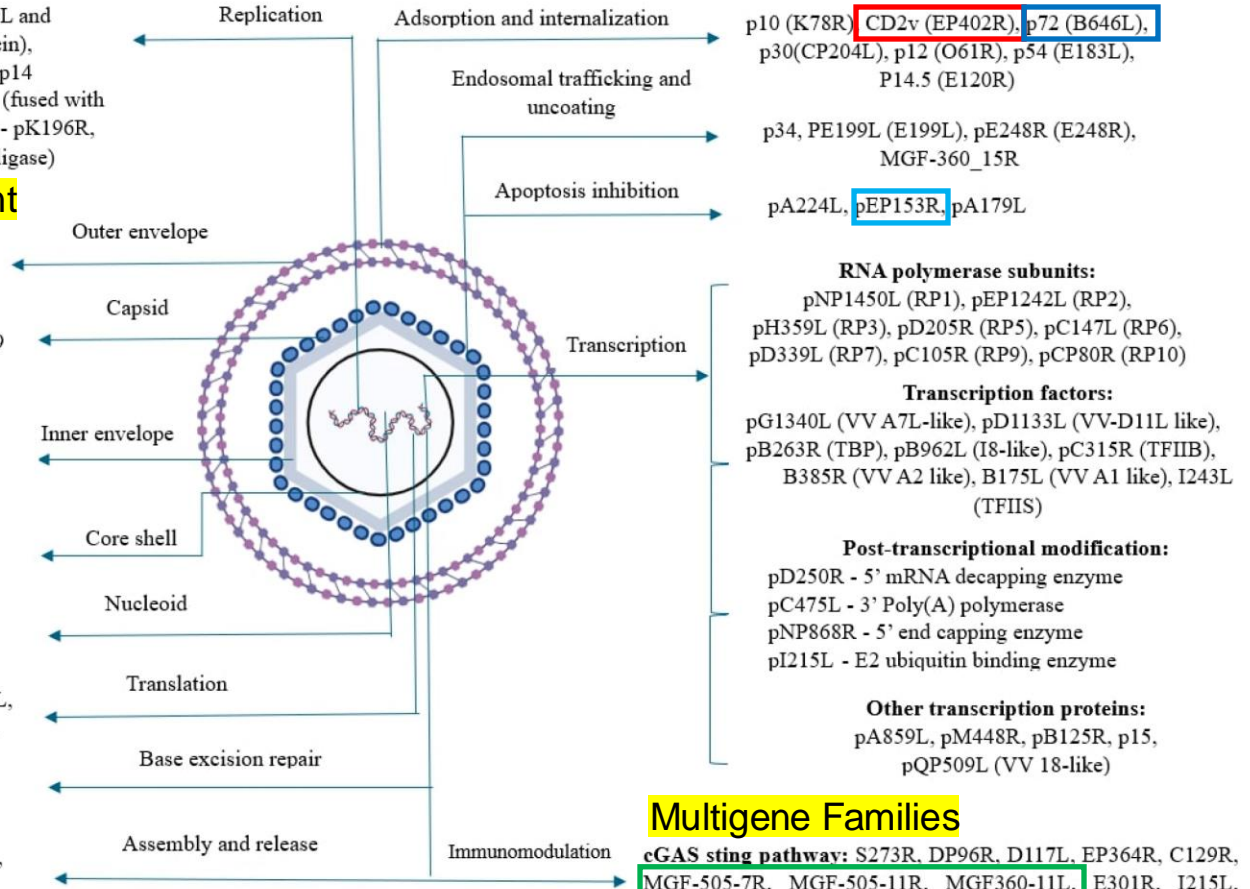
System Administrator:
Ir. Liu Bei (刘贝)

Computational Genomics Course
KIZ, Fall 2024

1

pC962R (Primase), pG1211R (DNA pol), QP509L and Q706L (Helicases), pA104R (DNA binding protein), pP1192R (DNA topoisomerase II), p10, p37 and p14 (nucleocytoplasmic shuttling proteins), pF1055L (fused with DNA primase), pE301R (Sliding clamp), dNTPs - pK196R, pA240L, pF134L, pF778R, pE165R, pNP419L (ligase)
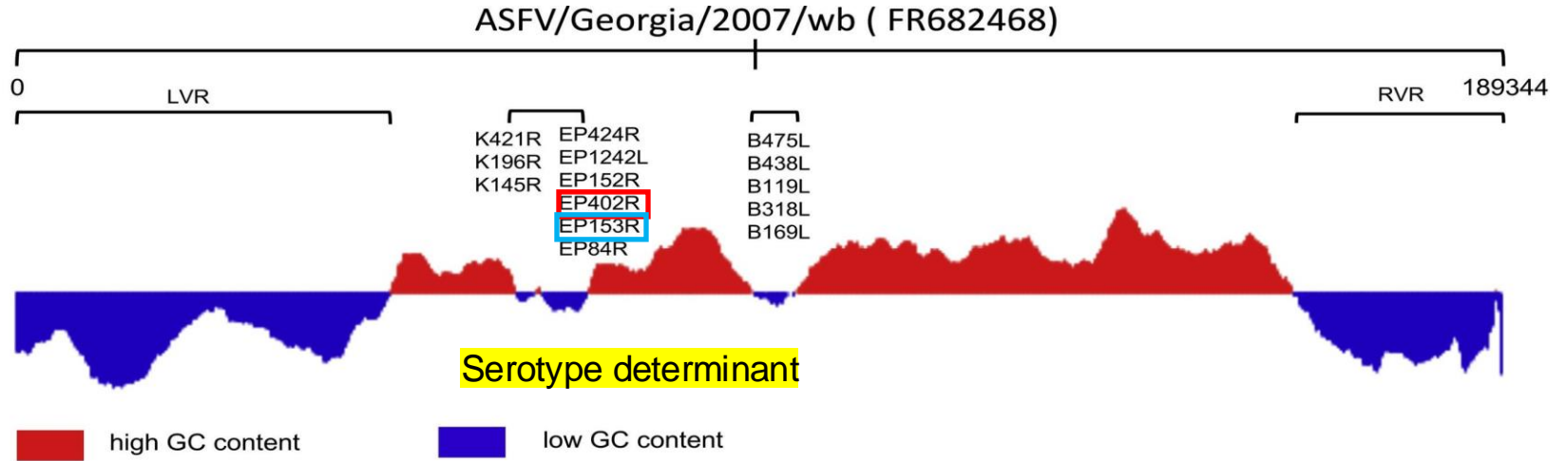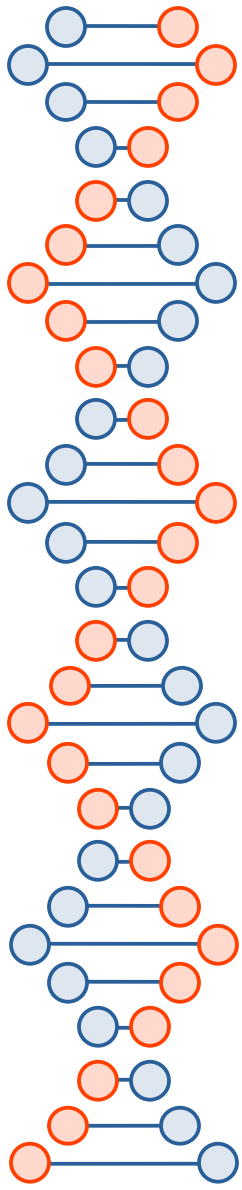
Serotype determinant

p12 (O61R), CD2V/pEP402R (E402R), p22 (KP177R)

p72 (B646L), pH240R (H240R), pM1249L (M1249L), p17 (D117L), p14.5 (E120R), p49 (B438L)

p12 (O61R), p22 (KP177R), CD2V/pEP402R (E402R)

pS273R (S273R)
pp220 (CP2475L) - p150, p37, p34, p14, p5
pp62 (CP530R) - p35, p15, p8

pA104R (A104R), p10 (K78R)

pI215L, pE66L, MGF 110-7L, pA224L, pDP71L, pEP424R, pD250R, pI73R

pO174L (Pol X), pD345L (5'/3' exonuclease), pE296R (AP endonuclease)

A104R, K78R(p10), EP84R, M1249L, B318L, B438L, B354L, B602L, B646L (p72), p150, p 34, p35, p15, p8, D117L (p17), D345L, S273R, H240R, E138L, E120R
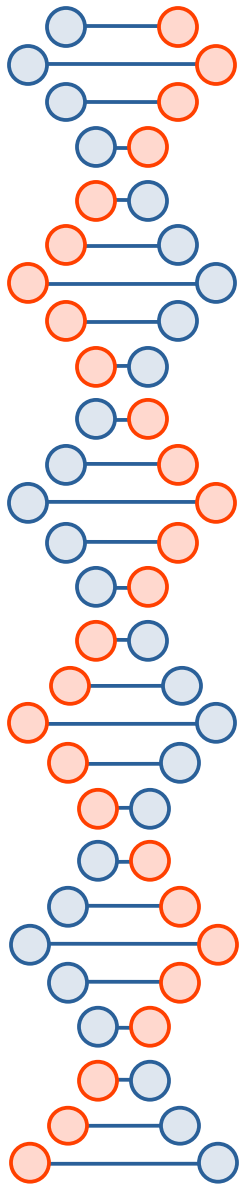
Replication
Adsorption and internalization

p10 (K78R) CD2v (EP402R), p72 (B646L), p30(CP204L), p12 (O61R), p54 (E183L), P14.5 (E120R)

Endosomal trafficking and uncoating

p34, PE199L (E199L), pE248R (E248R), MGF-360_15R

Apoptosis inhibition

pA224L, pEP153R, pA179L

Outer envelope

Capsid

Inner envelope

Core shell

Nucleoid

Translation

Base excision repair

Assembly and release

Transcription

RNA polymerase subunits:
pNP1450L (RP1), pEP1242L (RP2), pH359L (RP3), pD205R (RP5), pC147L (RP6), pD339L (RP7), pC105R (RP9), pCP80R (RP10)

Transcription factors:
pG1340L (VV A7L-like), pD1133L (VV-D11L like), pB263R (TBP), pB962L (I8-like), pC315R (TFIIB), B385R (VV A2 like), B175L (VV A1 like), I243L (TFIIS)

Post-transcriptional modification:
pD250R - 5' mRNA decapping enzyme
pC475L - 3' Poly(A) polymerase
pNP868R - 5' end capping enzyme
pI215L - E2 ubiquitin binding enzyme

Other transcription proteins:
pA859L, pM448R, pB125R, p15, pQP509L (VV 18-like)

Immunomodulation

Multigene Families

cGAS sting pathway: S273R, DP96R, D117L, EP364R, C129R, MGF-505-7R, MGF-505-11R, MGF360-11L, E301R, I215L, MGF360-14L, MGF360-13L, MGF505-3R, L83L, E120R, I226R, M1249L, A137R, E184L, CD2v, QP383R, MGF505-2R

NF-κB pathway: A238L, D345L, MGF505-7R, F317L, MGF360-12L, H240R, EP402R, DP96R, A224L, I329L, K205R, E120R, I226R

JAK-STAT pathway: MGF505-7R, MGF360-9L, MGF360-10L, I215L, H240R, EP402R, I267L, A104R

NFAT pathway: A238L

NLRP3 inflammasome activation: H240R, S273R MGF505-7R

Venkateswaran et al (2024). *Animals, 14*(15), 2187.

ASFV/Georgia/2007/wb (FR682468)

0   LVR   RVR   189344

K421R   EP424R   B475L
K196R   EP1242L   B438L
K145R   EP152R   B119L
         EP402R   B318L
         EP153R   B169L
         EP84R

Serotype determinant

■ high GC content   ■ low GC content

Malogolovkin & Kolbasov (2019)."Genetic and antigenic diversity of African swine fever virus". *Virus Research* 271:197673.

Forth et al (2020) PLoS Pathog 16(8): e1008779.

Entfellner et al (2024) *Viruses* 16(9), 1466

Project Overview: Biological Hypotheses
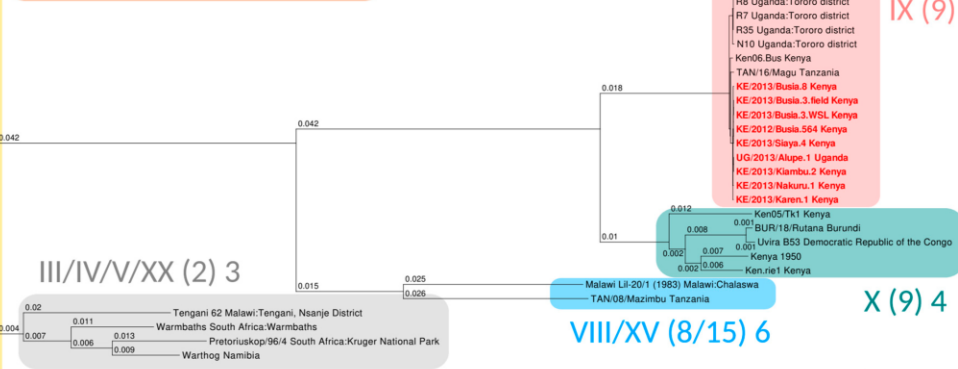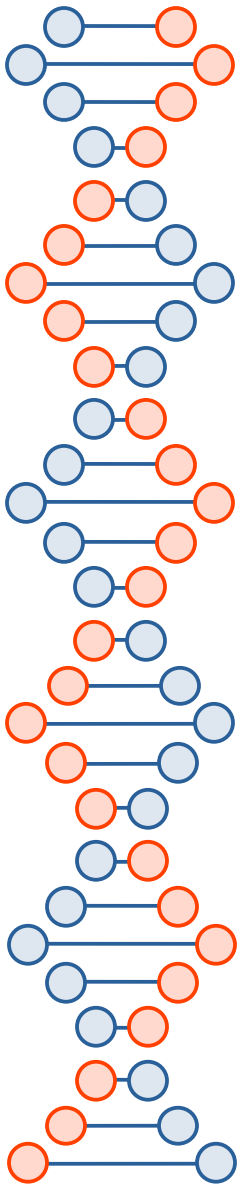
Genome Phylogeny

- Align genomes to a ref genome
- Call SNPs
- Infer phylogeny
- Annotate phylogeny

Pop Genome Analysis

- Population structure & gene flow
- Estimate recombination & mutation rates
- Identify recombination hotspots

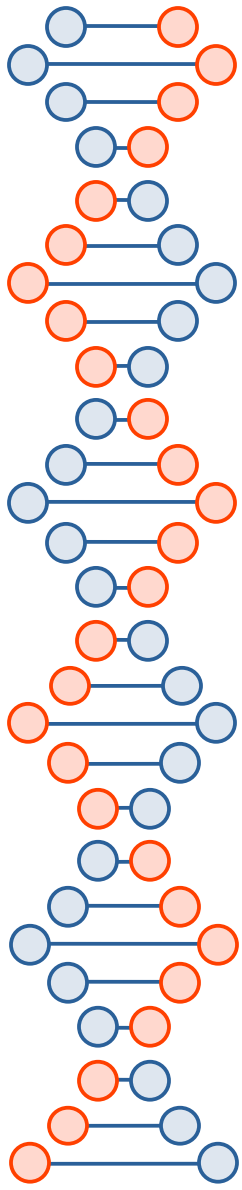## Pan-Genome Analysis

- Identify core & variable genomes
- Identify gene gains and losses

Genome-Wide Association Study (GWAS)

- Host specificity
- Virulence genes

5

Project Overview: Software Tools

Genome Phylogeny

- Minimap2 & Samtools
- VCFTools & BCFTools
- IQ-Tree
- iToL

Pop Genome Analysis

- VCFTools & BCFTools
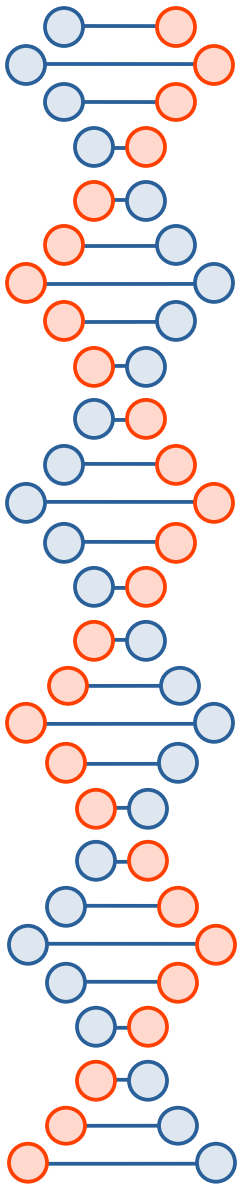- ClonalFrameML

Pan-Genome Analysis

- Ppanggolin
- APE package (R)

Genome-Wide Association Study (GWAS)

- BinaryPGLMM (R)
- GEMMA

Project Overview: Data-Science Skills

Domain Kowledge

- Mol Biology
- Genomics
- Evolution: pop genetics & phylogenetics

Linux & Git

- Batch-processing
- Source control & distributed computing
- Commands (CLI)

Python & Notebook

- Pre-processing
- BioPython
- BioPerl / BpWrapper
- Machine learning

R/RStudio/R Markdown

- Visualization
- Statistical analysis
- Modeling & inference