

Case Study

African Swine Fever Virus (ASFV)

Phylo-Genomic & Pop-Genomic Analysis

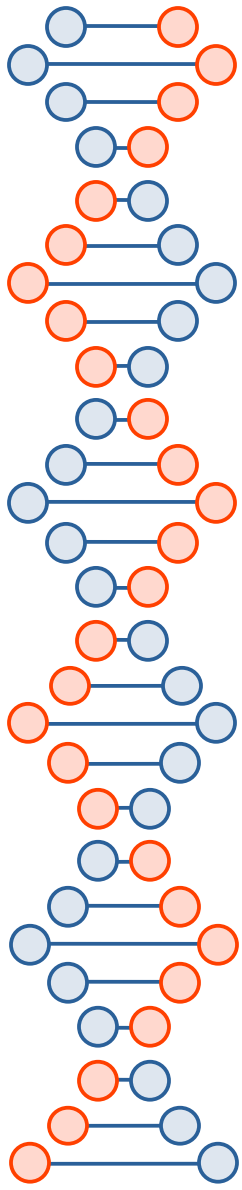
Project Coordinators:

Dr Adeniyi Charles Adeola (KIZ) & Dr Weigang Qiu (Hunter/CUNY)

System Administrator:

Ir. Liu Bei (刘贝) ; Mr Liu Lisheng (刘利生)

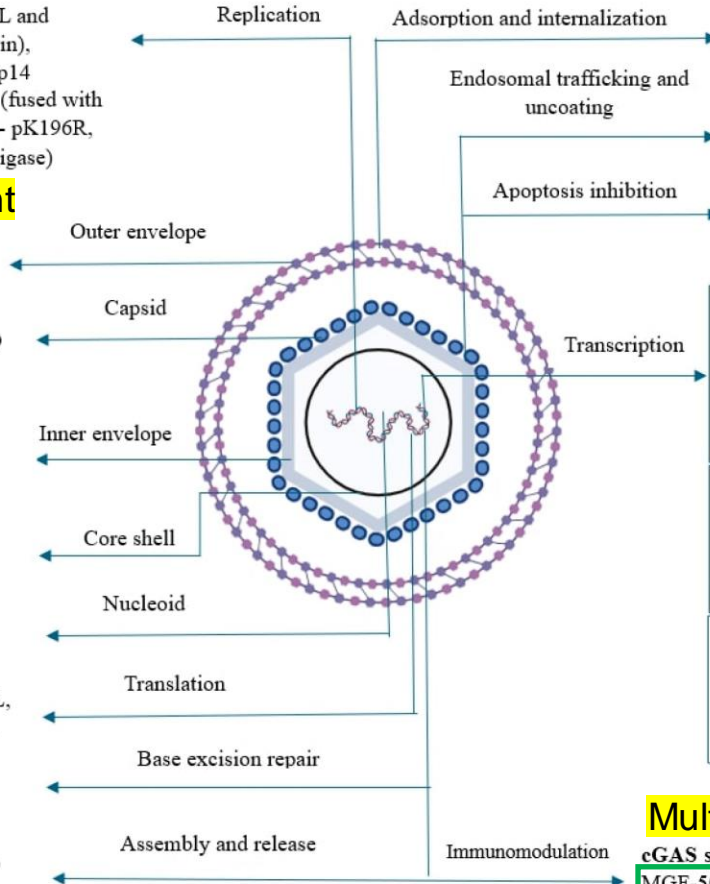
Computational Genomics Course
KIZ, Fall 2024



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, p34, p35, p15, p8, D117L (p17), D345L, S273R, H240R, E138L, E120R



p10 (K78R), **CD2v (EP402R)**, **p72 (B646L)**, p30(CP204L), p12 (O61R), p54 (E183L), P14.5 (E120R)
 p34, PE199L (E199L), pE248R (E248R), MGF-360_15R
 pA224L, **pEP153R**, pA179L

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

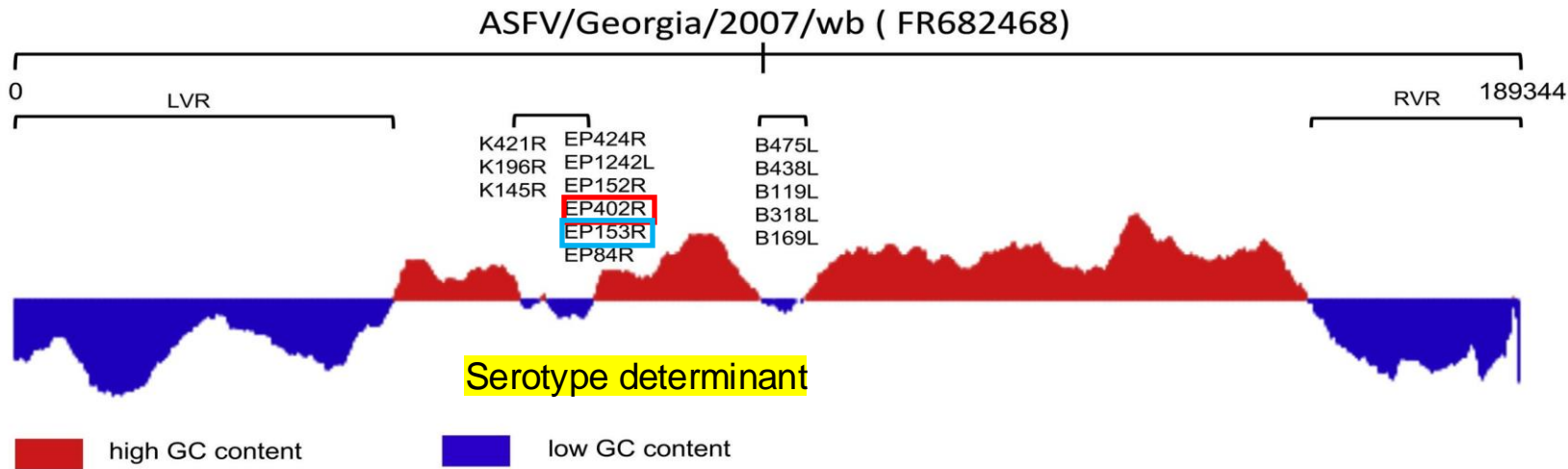
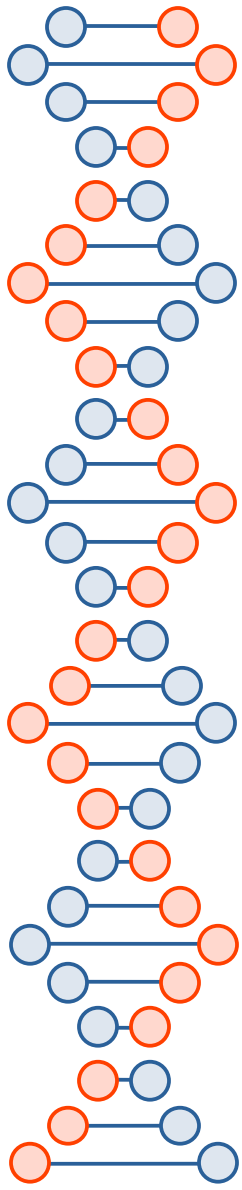
Other transcription proteins:

pA859L, pM448R, pB125R, p15, pQP509L (VV 18-like)

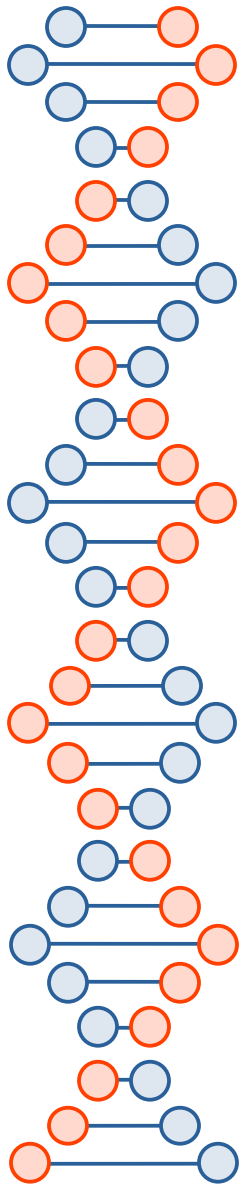
Multigene Families

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.](#)

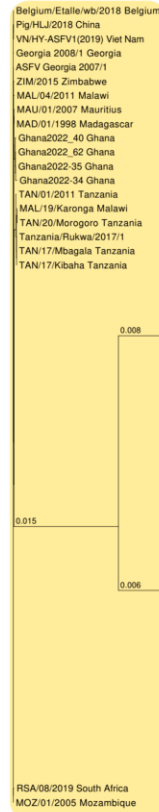


[Malogolovkin & Kolbasov \(2019\). "Genetic and antigenic diversity of African swine fever virus". *Virus Research* 271:197673.](#)



[Forth et al \(2020\) PLoS Pathog 16\(8\): e1008779.](#)

[Entfellner et al \(2024\) Viruses 16\(9\), 1466](#)



II (2) 2

I (1) 1

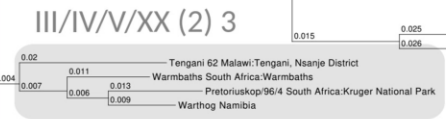


IX (9) 5



X (9) 4

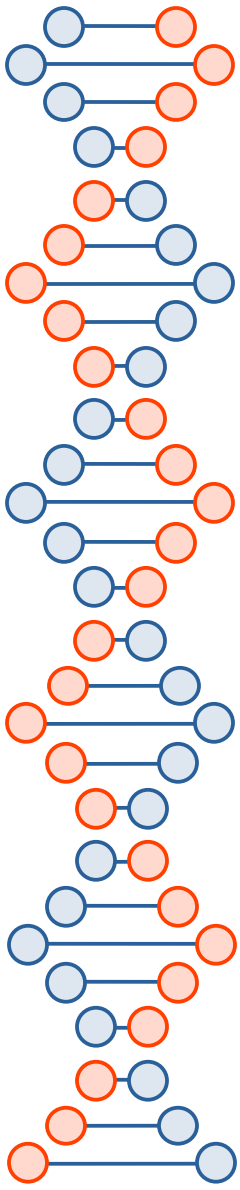
VIII/XV (8/15) 6



0.03

0.01

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

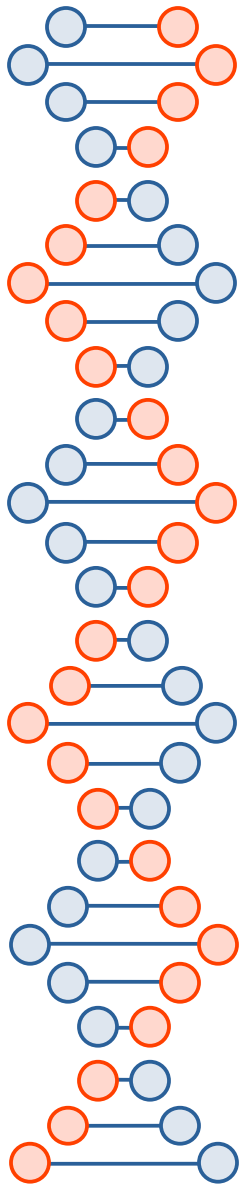
Identify gene gains and losses

Genome-Wide Association Study (GWAS)

Host specificity

Virulence genes

Project Overview: Software Tools



Genome Phylogeny

Minimap2 & Samtools

VCFTools & BCFTools

IQ-Tree

iToL

Pop Genome Analysis

VCFTools & BCFTools

ClonalFrameML

Pan-Genome Analysis

Ppangolin

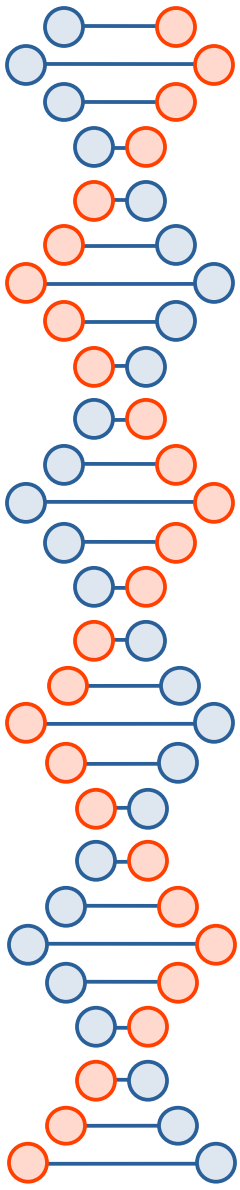
APE package (R)

Genome-Wide Association Study (GWAS)

BinaryPGLMM (R)

GEMMA

Project Overview: Data-Science Skills



Domain Knowledge

Mol Biology

Genomics

Evolution: pop genetics & phylogenetics

Linux & Git

Batch-processing

Source control & distributed computing

Commands (CLI)

Python & Notebook

Pre-processing

BioPython

BioPerl / BpWrapper

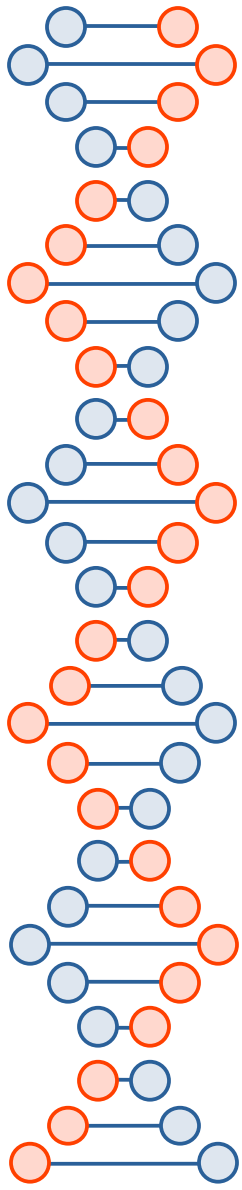
Machine learning

R/RStudio/R Markdown

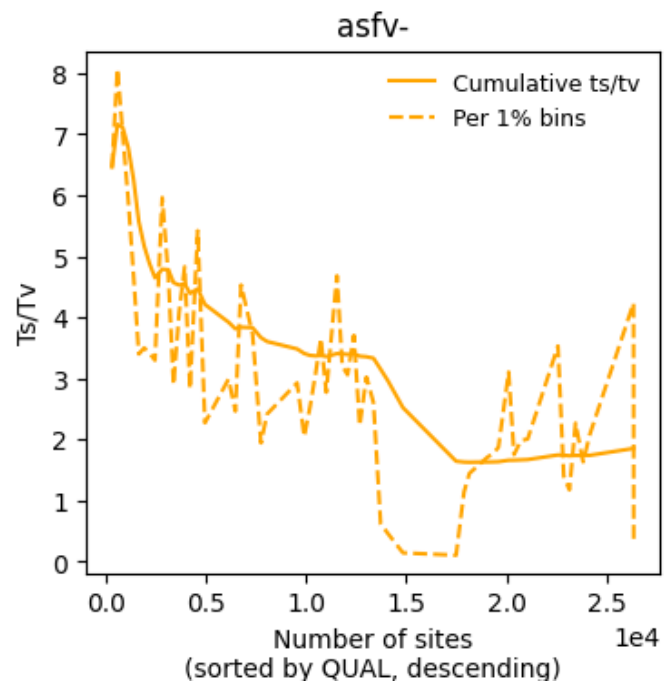
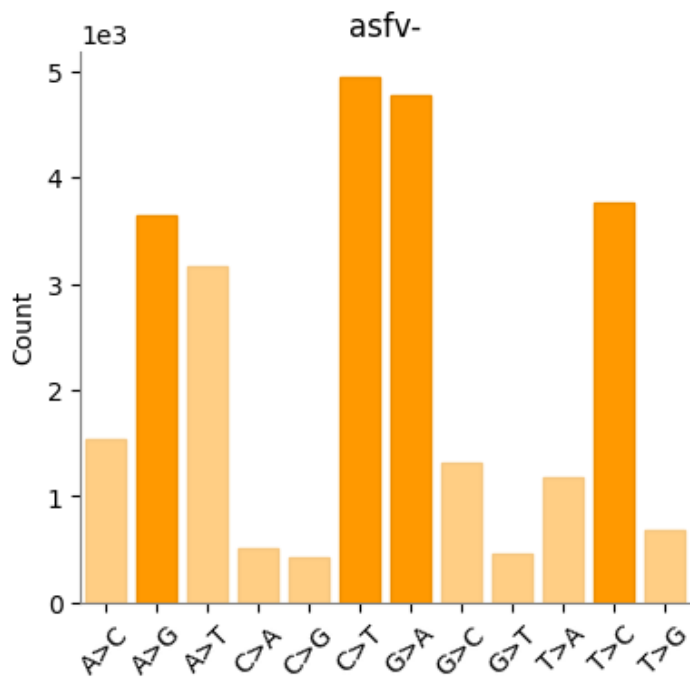
Visualization

Statistical analysis

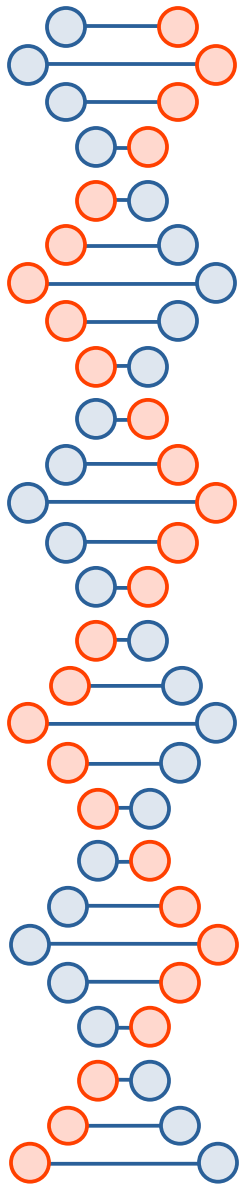
Modeling & inference



SNP stats

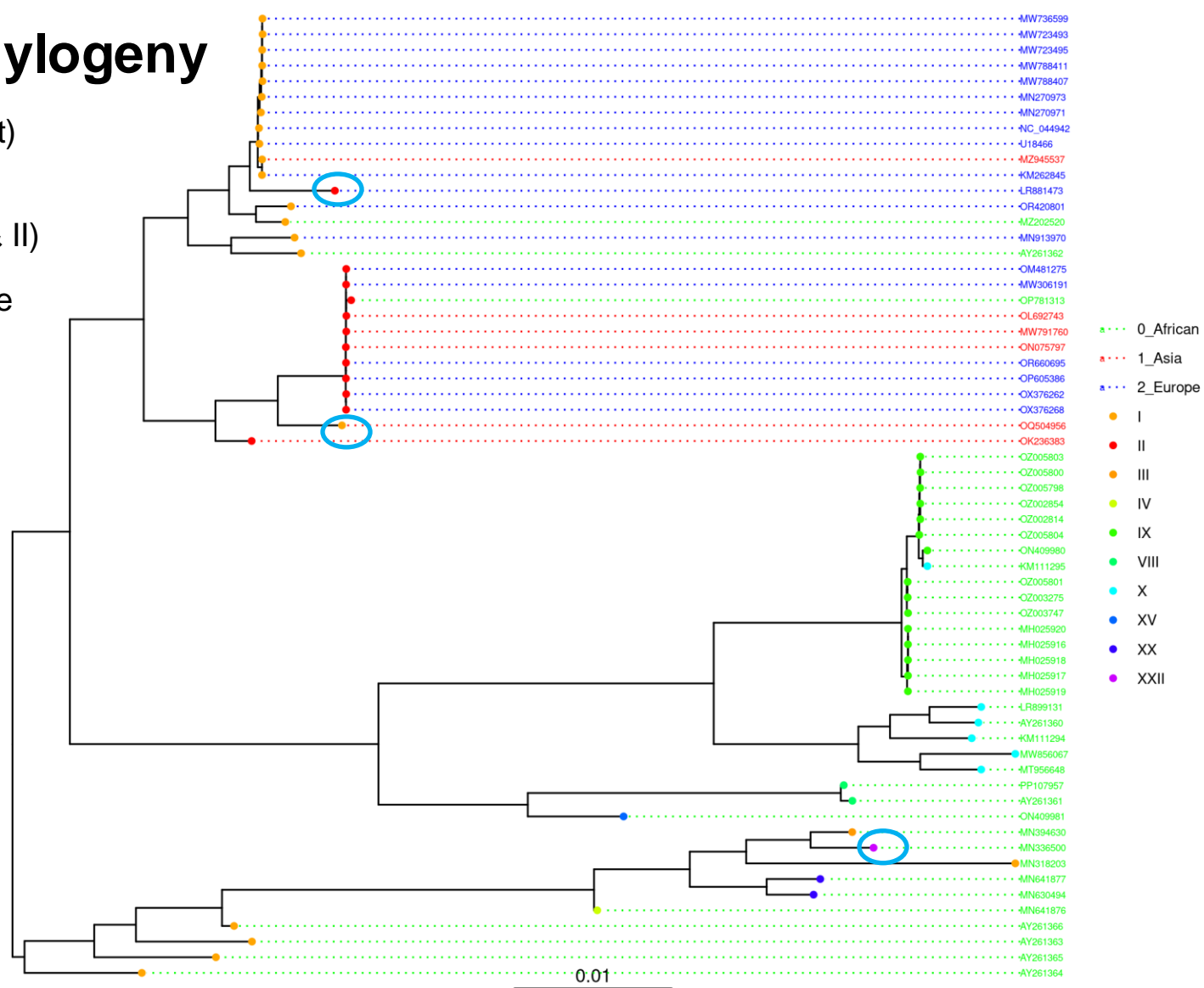


- N=316 genomes (from NCBI); Ref: BA71 (NC_044942)
- N=26,372 (L=180,365 bp) bi-allelic (-m2 -M2 -c2)

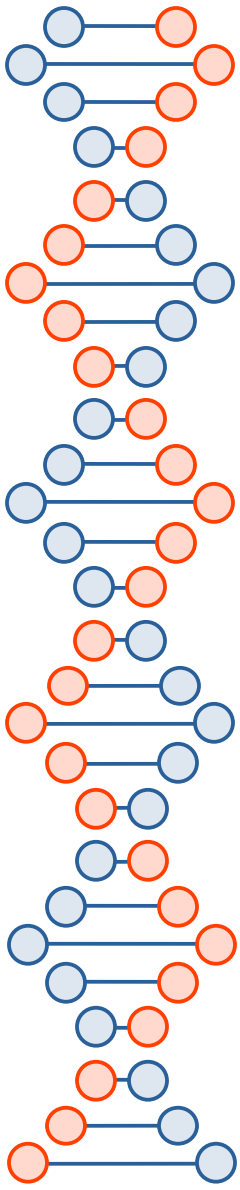


Genome phylogeny

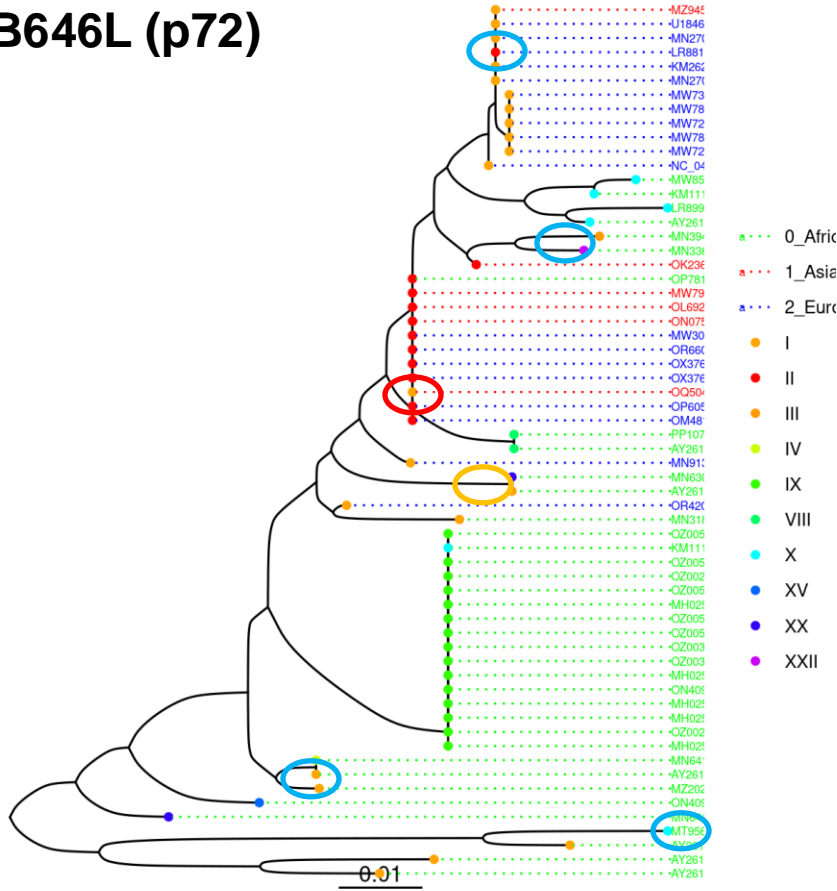
- IQ-Tree (mid-point)
- N=62 genomes (random 10 for I & II)
- Mistake or multiple serotype determinants?



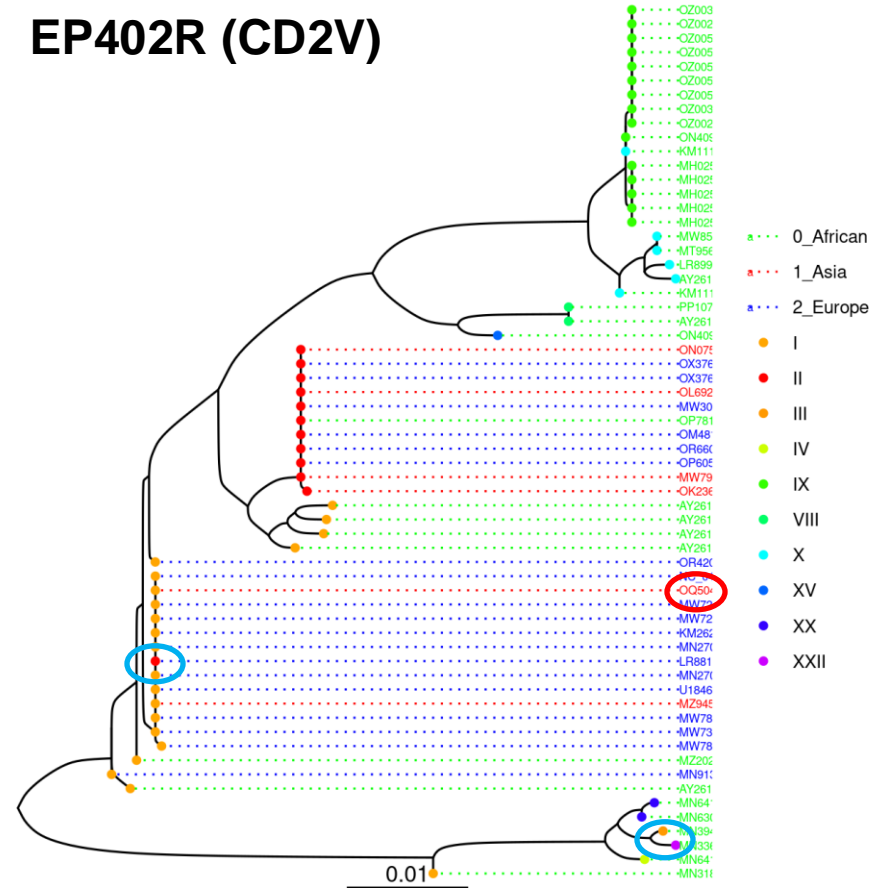
Gene tree inconsistencies: Recombination & Horizontal Gene Transfers



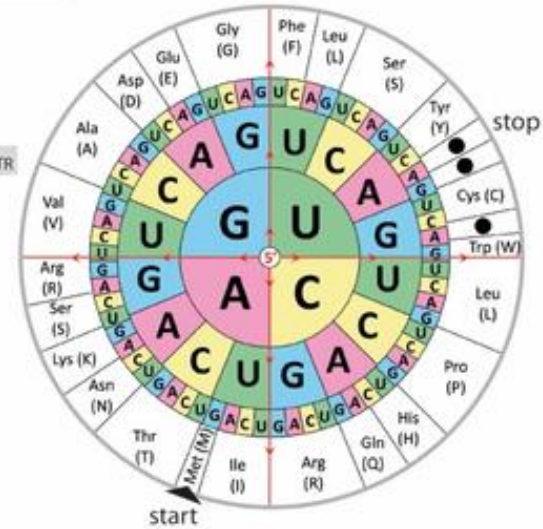
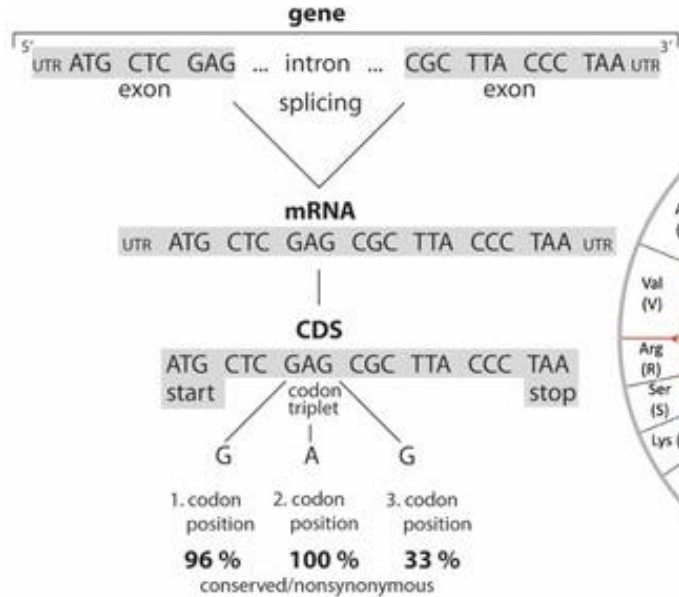
B646L (p72)



EP402R (CD2V)



Negative & Positive Selection in Genes (n=161 ORFs): Ka/Ks Analysis with PAML

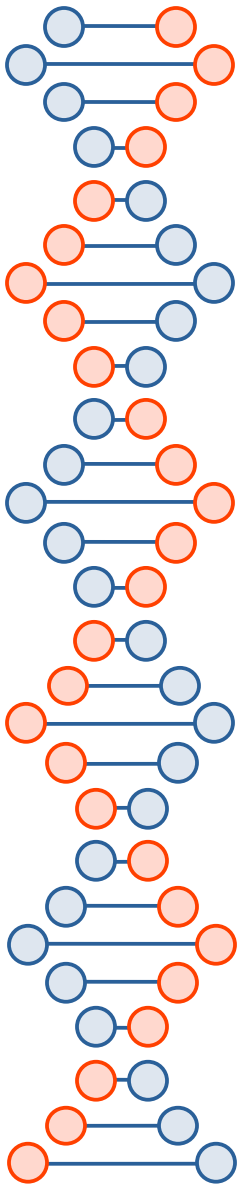


$$\omega = \frac{dN_{\text{nonsynonymous}}}{dS_{\text{synonymous}}}$$

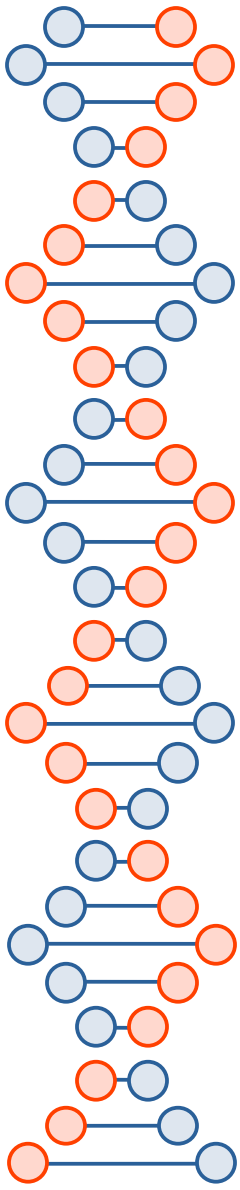
< 1 (negative - purifying selection)
= 1 (neutral)
> 1 (positive - adaptive selection)

1. Download gene alignments (git pull)
2. [Download & run IQ-Tree](#)
3. [Download & Install PAML & PAML-X](#)

Images from [Steffen et al \(2022\)](#)



Ka/Ks Analysis: PAML Site Models



Site models (M1a/M2a, M7/M8) ($\omega \leq 1/\omega > 1$) averaged over all branches	Branch-site models ($\omega = 1/\omega > 1$)	Branch models ($\omega = 1/\omega > 1$) averaged over all codons	HyPhy FEL ($H_0: \alpha = \beta$ / $H_A: \alpha \neq \beta$) averaged over all branches
<p>* Gly _ GGA _ Ala _ GCA _ Ala _ GCA _</p> <p>positive selected sites</p>	<p>* Arg _ AGA _ Thr _ ACA _ Thr _ ACA _</p> <p>positive selected sites and branches</p>	<p>1 2 3 1 2 3 1 2 3</p> <p>positive selected branches</p>	<p>* Ser _ TCA _ Ser _ TCC _ Ser _ TCG _</p> <p>negative selected sites</p>
<p>ctl default settings</p> <p>model: 0 NSsites: 1 2 7 8 fix_omega: 0 omega: 1</p>	<p>ctl default settings</p> <p>model: 2 NSsites: 2 fix_omega: 1/0 omega: 1</p>	<p>ctl default settings</p> <p>model: 2 NSsites: 0 fix_omega: 1/0 omega: 1</p>	

- Images from [Steffen et al \(2022\)](#)
- [Download & Install PAML & PAML-X](#)