MPXV Phylogenetics

—

**Wytamma Wirth (@wytamma)**

Objectives

- Understand complexities of MPXV phylo reconstruction
- Learn best practices for MPXV phylogenetics
- Run Galaxy workflow – MPXV Phylogenetic Workflow
- Understand outputs and visualise using Nextstrain

Galaxy
PROJECT

Features of MPXV Genome

Unique Tandem Repeats (UTR)

- 16-nt tandem repeats (AACTAACTTATGACTT)
- located in the inverted terminal repeats (ITR)
- Copy number different among clade I, clade IIa, and clade IIb viruses.
- Only present in MPXV, not other Pox viruses
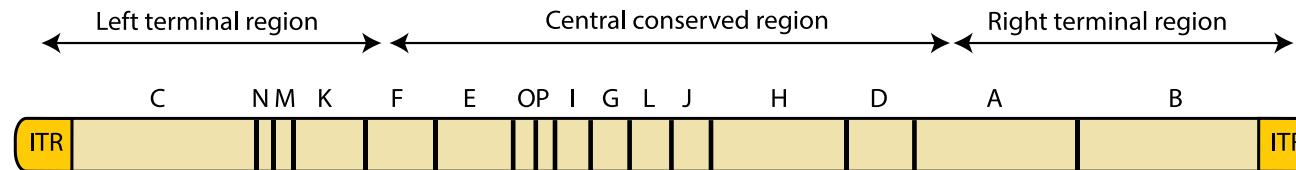
ITR= Inverted terminal repeats

References:
1. https://pmc.ncbi.nlm.nih.gov/articles/PMC10101126/

Features of MPXV Genome

<u>Unique Tandem Repeats (UTR)[1]</u>

- 16-nt tandem repeats, located in the inverted terminal repeats (ITR), Copy number different among clade I, clade IIa, and clade IIb viruses.
- Only present in MPXV, not other Pox viruses

<u>Short Tandem Repeats (STR) across the genome</u>

- Areas with more variation, possible role in MXPV biology and evolution.

<u>Deletions, frameshifts</u>
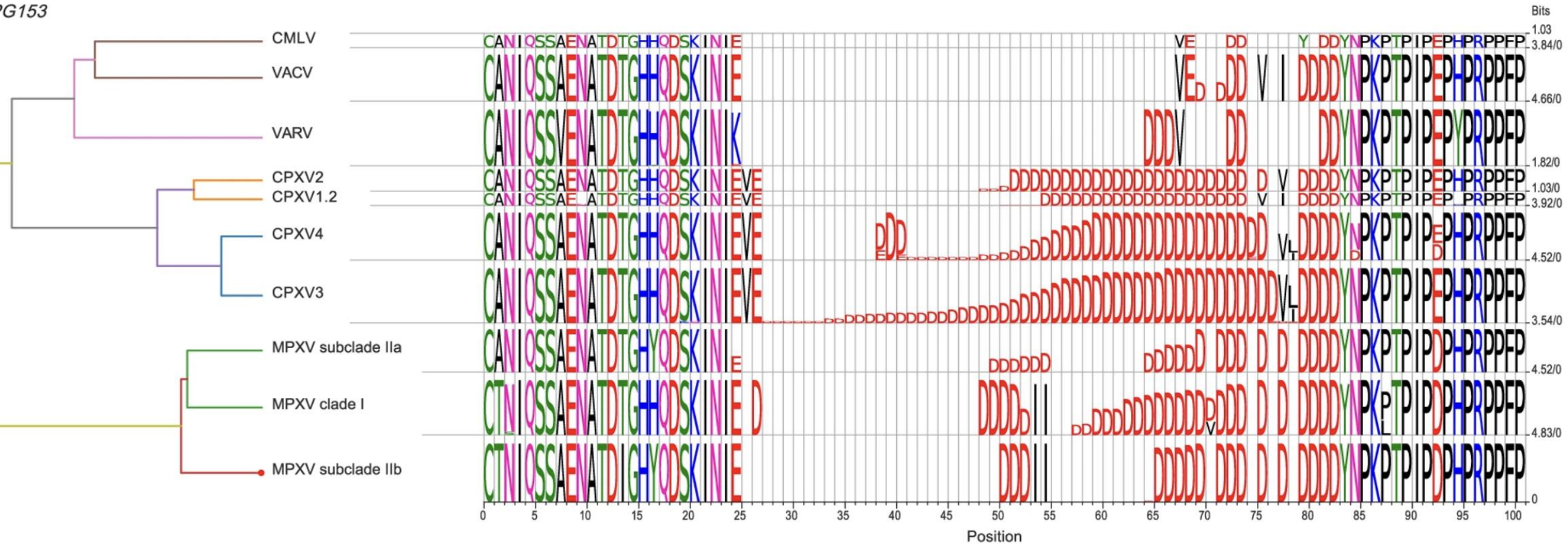
<u>Homopolymers across the genome</u>

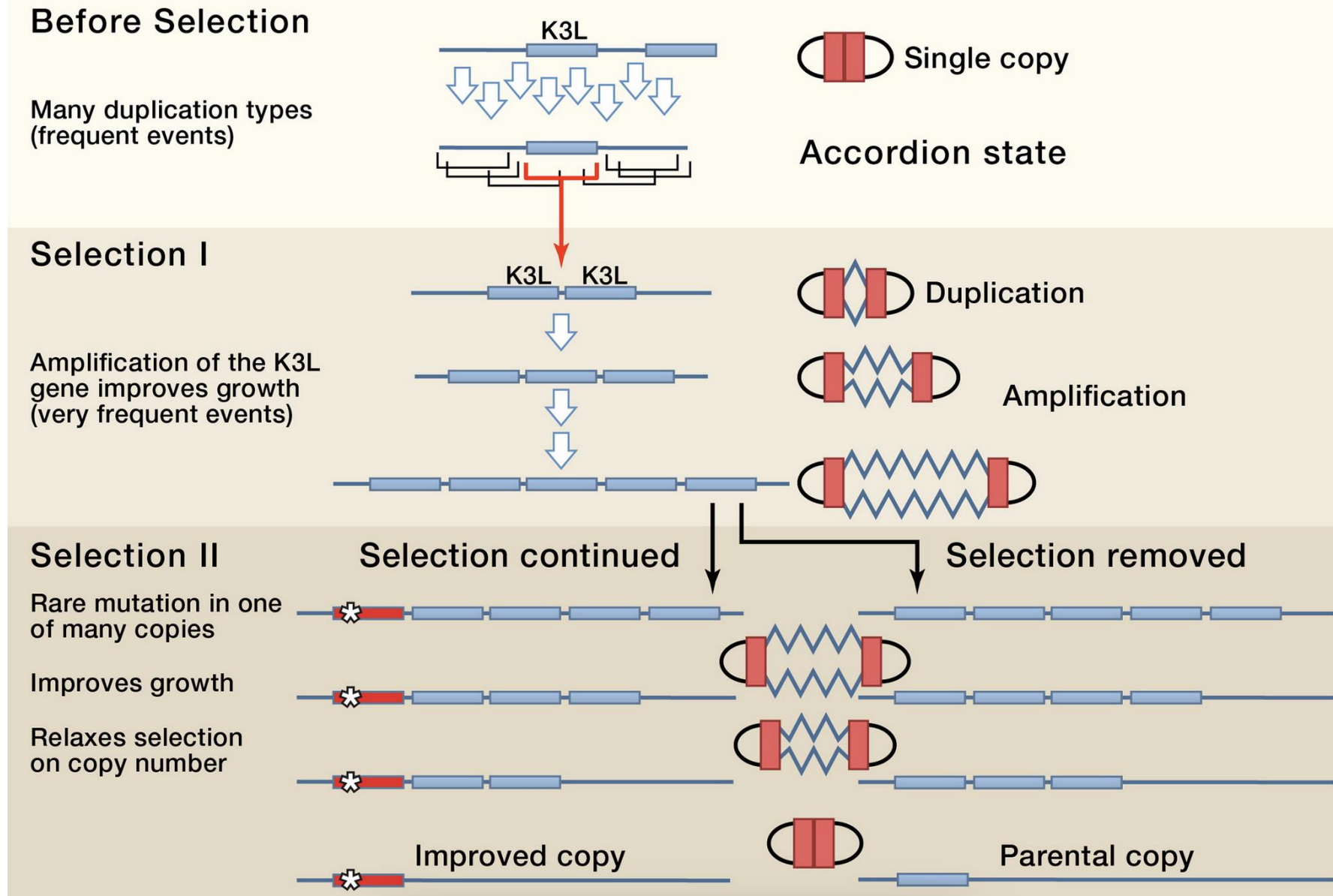<u>Slow mutation rate (1–2 SNPs per year)</u>

References:
1. https://pmc.ncbi.nlm.nih.gov/articles/PMC10101126/

Features of MPXV Genome

https://www.nature.com/articles/s41467-024-46949-7#Sec1

Features of MPXV Genome – Genomic Accordion

Features result in downstream complexities

MPXV genome has:
- Tightly packed gene contents
- Evolutionary strategies such as 'genomic accordion', homopolymer, STR, UTR regions result in complicated
- Large genome (197 kbp)

Result in complicated phylogenetic analysis

https://www.nature.com/articles/s41467-024-46949-7#Sec1

Solution: Squirrel

- **S**ome **QUI**ck **R**econstruction to **R**esolve **E**volutionary **L**inks

- https://github.com/aineniamh/squirrel

- Integrated into Galaxy

Squirrel Phylo Phylogenetic and APOBEC3 analysis of MPXV (Mpox virus)

Squirrel QC QC of MPXV (Mpox virus) sequences

- Dedicated pipeline:

🌐 edited less than a minute ago   ☰ workflow runs: 3   ☆ ▾

MPXV (Mpox) Phylogenetic Analysis with Squirrel

Run Squirrel tool to perform MPXV phylogenetic reconstruction. Pipeline will perform masking (using squirrel tool) of SNPs.

Add Tags 🏷   🔗 Link to Workflow   📋 Copy   ⬇ Download   ◁ Share   ✎ Edit   ▶

Running Squirrel

Running Squirrel

🔀 **Workflow: MPXV (Mpox) Phylogenetic Analysis with Squirrel** (Version: 13)   ✎ ⚙ ▶ Run Workflow

🕐 *edited 1 minute ago*   🌐 👤 **aaziz**

                                                          ☰ **workflow runs: 3**

Run Squirrel tool to perform MPXV phylogenetic reconstruction. Pipeline will perform masking (using squirrel tool) of SNPs.

Consensus Sequences *

| 📄 | 📋 | ... | 139: all.consensus.fasta ⬅ **Output of MPXV Pipeline** ▾ |

accepted formats ▾

Input sequences for phylogenetic reconstruction. Must be a single clade (CladeI or CladeII). Do not mix clades.

MPXV Clade *

cladeii   ▾

Select MPXV Clade.

Run APOBEC Analsis

🔘 Yes

Perform APOBEC3-reconstruction and map mutations onto the phylogeny.

Expand to full workflow form.

Running Squirrel

Workflow: MPXV (Mpox) Phylogenetic Analysis with Squirrel (Version: 13) ▶ **Run Workflow**

🕐 *edited 1 minute ago*

🌐 👤 **aaziz**

≡ **workflow runs: 3**

Run Squirrel tool to perform MPXV phylogenetic reconstruction. Pipeline will perform masking (using squirrel tool) of SNPs.

Consensus Sequences *

139: all.consensus.fasta ⬅ **Output of MPXV Pipeline**

accepted formats ▾

Input sequences for phylogenetic reconstruction. Must be a single clade (CladeI or CladeII). Do not mix clades.

MPXV Clade *

cladeii ⬅ **Select Clade**

Select MPXV Clade.

Run APOBEC Analsis

🔵 Yes

Perform APOBEC3-reconstruction and map mutations onto the phylogeny.

Expand to full workflow form.

Running Squirrel

Workflow: MPXV (Mpox) Phylogenetic Analysis with Squirrel (Version: 13)   ▶ Run Workflow

🕐 *edited 1 minute ago*   🌐 👤 aaziz
                                            ≡ workflow runs: 3

Run Squirrel tool to perform MPXV phylogenetic reconstruction. Pipeline will perform masking (using squirrel tool) of SNPs.

Consensus Sequences *

139: all.consensus.fasta   ⬅ **Output of MPXV Pipeline**

accepted formats ▾

Input sequences for phylogenetic reconstruction. Must be a single clade (CladeI or CladeII). Do not mix clades.

MPXV Clade *

cladeii   ⬅ **Select Clade**

Select MPXV Clade.

Run APOBEC Analsis

🔵 Yes   ⬅ **Optional – Do you want to perform APOBEC3 Analysis?**

Perform APOBEC3-reconstruction and map mutations onto the phylogeny.

Expand to full workflow form.

Running Squirrel

Workflow: MPXV (Mpox) Phylogenetic Analysis with Squirrel (Version: 13)

▶ Run Workflow

🌐 👤 aaziz

⏱ *edited 1 minute ago*

≡ workflow runs: 3

Run Squirrel tool to perform MPXV phylogenetic reconstruction. Pipeline will perform masking (using squirrel tool) of SNPs.

Consensus Sequences *

139: all.consensus.fasta ← **Output of MPXV Pipeline**

accepted formats ▾

Input sequences for phylogenetic reconstruction. Must be a single clade (CladeI or CladeII). Do not mix clades.

MPXV Clade *

cladeii ← **Select Clade**

Select MPXV Clade.

Run APOBEC Analsis

🔵 Yes ← **Optional – Do you want to perform APOBEC3 Analysis?**

Perform APOBEC3-reconstruction and map mutations onto the phylogeny.

Expand to full workflow form.

Squirrel Outputs

1. Branch Reconstruction (SNP and AA)
2. Visualised Tree in PNG/SVG
3. Tree in Newick format
4. MSA – Multiple Sequence Alignment

166: **Branch SNPS Reconstruction**	👁 ✏ 🗑
165: **Branch AA Reconstruction**	👁 ✏ 🗑
164: **Tree (PNG)**	👁 ✏ 🗑
163: **Tree (SVG)**	👁 ✏ 🗑
162: **MSA**	👁 ✏ 🗑
161: **Tree (Newick)**	👁 ✏ 🗑

Squirrel Outputs

1. Branch Reconstruction (SNP and AA)
2. **Visualised Tree in PNG/SVG**
3. Tree in Newick format
4. MSA – Multiple Sequence Alignment

Inspect the APOBEC3 annotated tree

166: **Branch SNPS Reconstruction** 👁 ✏ 🗑

165: **Branch AA Reconstruction** 👁 ✏ 🗑

164: **Tree (PNG)** 👁 ✏ 🗑

163: **Tree (SVG)** 👁 ✏ 🗑

162: **MSA** 👁 ✏ 🗑

161: **Tree (Newick)** 👁 ✏ 🗑

Squirrel Outputs

1. Branch Reconstruction (SNP and AA)
2. **Visualised Tree in PNG/SVG**
3. Tree in Newick format
4. MSA – Multiple Sequence Alignment

Inspect 👁 the APOBEC3 annotated tree

PNG/SVG are the same – but only PNG is viewable on Galaxy.

166: **Branch SNPS Reconstruction**	👁 ✏ 🗑	
165: **Branch AA Reconstruction**	👁 ✏ 🗑	
164: **Tree (PNG)**	👁 ✏ 🗑	
163: **Tree (SVG)**	👁 ✏ 🗑	
162: **MSA**	👁 ✏ 🗑	
161: **Tree (Newick)**	👁 ✏ 🗑	

Squirrel Outputs

1. Branch Reconstruction (SNP and AA)
2. Visualised Tree in PNG/SVG
3. **Tree in Newick format**
4. MSA – Multiple Sequence Alignment

166: **Branch SNPS Reconstruction** 👁 ✏ 🗑

165: **Branch AA Reconstruction** 👁 ✏ 🗑

164: **Tree (PNG)** 👁 ✏ 🗑

163: **Tree (SVG)** 👁 ✏ 🗑

162: **MSA** 👁 ✏ 🗑

161: **Tree (Newick)** 👁 ✏ 🗑

Tree (Newick) can be used for custom visualisation eg Nextstrain (Auspice)

Squirrel Outputs - Nextstrain

1. Download newick file:
   - Click on "Tree (newick)"
   - Click on the floppy disk icon 💾
   - This will download the .newick to your local computer
2. Rename the file to something sensible, eg: mpxv-{date}.newick
3. Go to auspice.us (nextstrain for your trees)
4. Drag and drop the .newick file into the browser

Squirrel Outputs - Nextstrain

We need to add meta data to the tree.

Created in excel, column required:
- 'strain'

Then any optional data, eg:
- location
- country
- **date** <- NOTE!
- lineage
- etc

	A	B	C	D	E
1	strain	continent	country	date	lineage
2	B1.fastq.g	Australia	Australia	2025-01-02	B.1.20
3	B2.fastq.g	Australia	Australia	2025-01-02	B.1.20.1
4					

'date' must be lower case
'date' must be in format: YYYY-MM-DD eg 2024-03-01

Save as tsv or csv!

Explore your phylogenetic tree.

mpxv-2025-06-01.newick

Showing 17 of 17 genomes.

auspice.us

Color By ⓘ

country ▼

Filter Data ⓘ

Type filter query here...

⌄ **Tree** ⓘ ⬤

Layout

▥ RECTANGULAR

⋔ RADIAL

⋉ UNROOTED

⣿ SCATTER

◯ Focus on selected ⓘ

Branch Labels

none ▼

Tip Labels

Sample Name ▼

Explode Tree By ⓘ

Select... ▼

Phylogeny

🔍 | ZOOM TO SELECTED | ZOOM TO ROOT

country ⌃

🟦 Australia

- B2.fastq.gz
- B1.fastq.gz
- OP415257|human|United_Kingdom|ex_Nigeria|2022-08
- ON563414|human|USA||2022-05-19
- MN648051|human|Israel|ex_Nigeria|2018-10-04
- OP413718|human|United_Kingdom||2022-08
- MK783033|human|Nigeria|Rivers|2017-10-09
- MK783027|human|Nigeria|Rivers|2017-11-09
- KJ642617|human|Nigeria|Abia|1971-04-14
- KJ642615|human|Nigeria||1978
- MT903346.1|pouched_rat|USA||2003
- MT724769.1|swamp_rat|DRC||2012
- AY741551.1|human|Sierra_Leone||
- KP849470.1|human|Cote_dIvoire||1971
- DQ011156.1|monkey|Liberia||1970
- KJ136820.1|chimp|Cote_dIvoire||2012
- MN346698.1|chimp|Cote_dIvoire||2017

0.0000   0.0002   0.0004   0.0006   0.0008   0.0010   0.0012   0.0014   0.0016   0.0018

Divergence
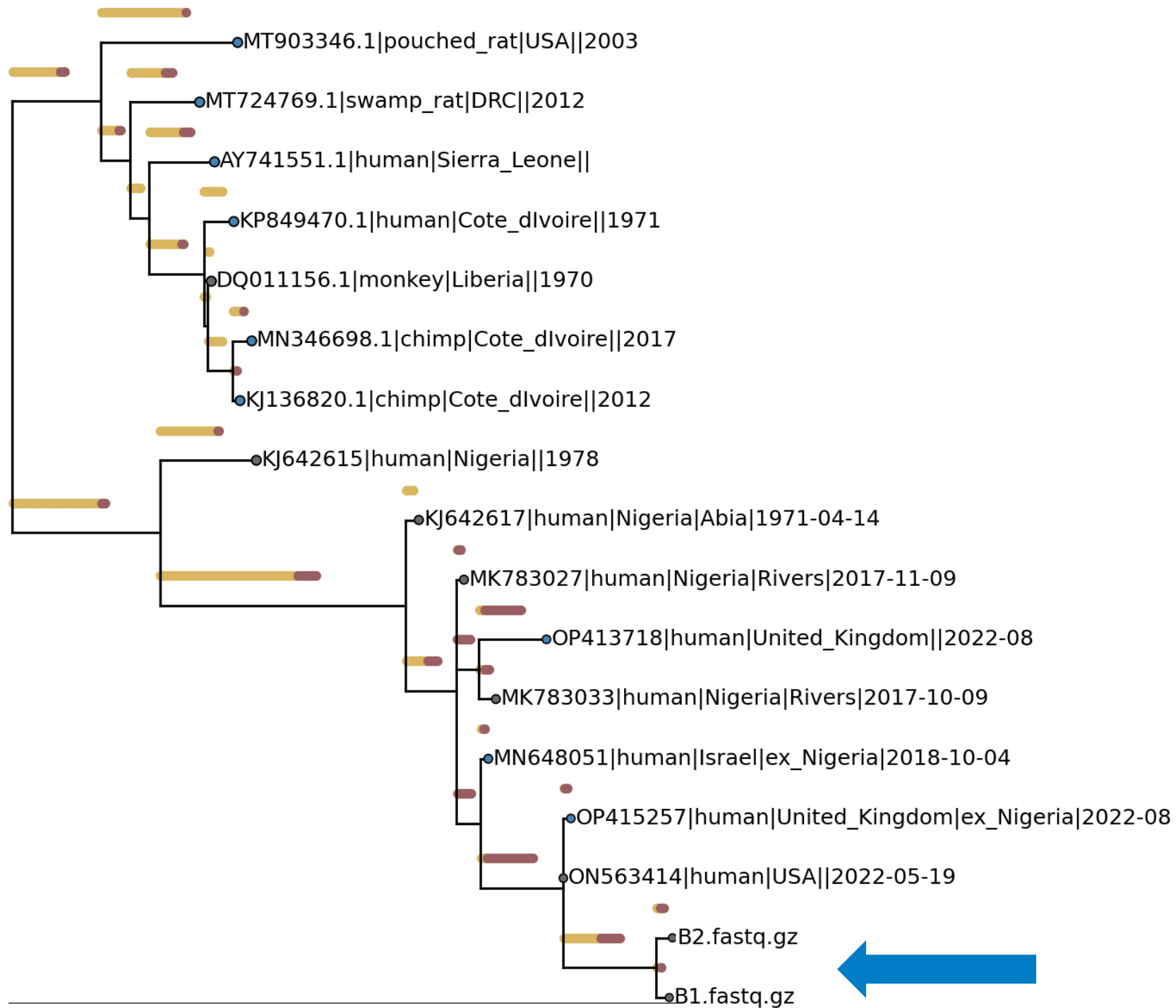
Squirrel Outputs – APOBEC Tree

Brown dots are APOBEC3 mutations.

Yellow are other SNPs (non apobec3).

Questions? + Resources

- Phylogenetic masking
  - https://github.com/WHO-Collaboratory/collaboratory-mpox-genomics-phylomasking
  - https://github.com/nextstrain/mpox/tree/master/nextclade/resources