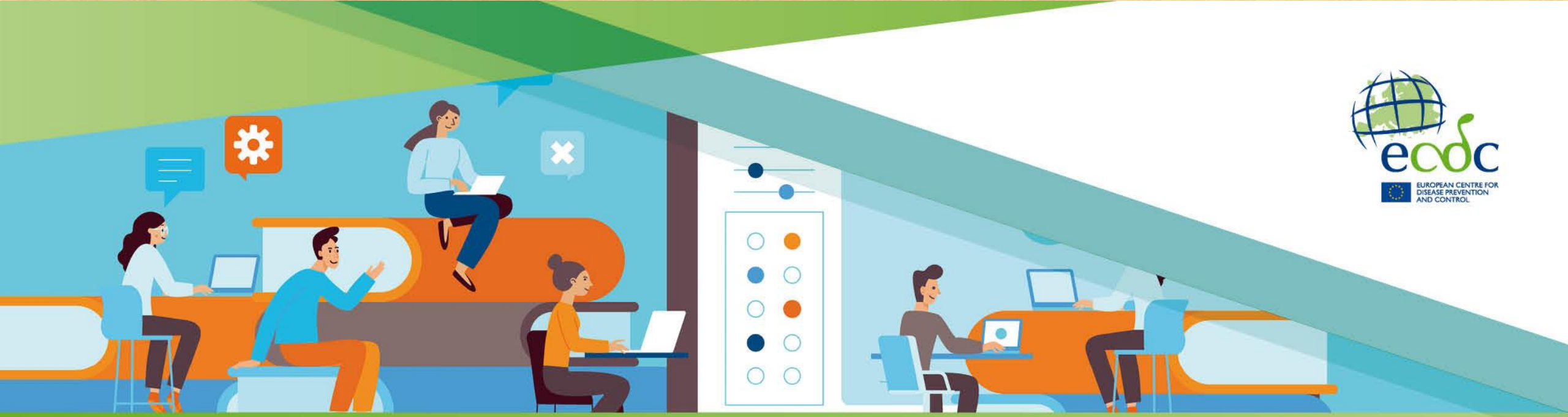


GenEpi-BioTrain Virtual Training 29 - Genomic epidemiology of mpox

Guided discussion

12-10-2026

David Nieuwenhuijse and Bas Oude Munnink

Aim and learning objectives

Aim

Recap of the exercise and phylogenetic placement on the sequences

Specific learning objectives

At the end of this session, you will be able to:

1. Interpreted the results
2. Identify potential clusters
3. Learn how to discriminate between ongoing human-to-human transmission of mpox or if there are repeated zoonotic transmission events

Quiz!

Do not edit
How to change the design



What are drawbacks of Nanopore based amplicon sequencing relative to other approaches?

ⓘ The Slido app must be installed on every computer you're presenting from

slido

Do not edit
How to change the design



What are bioinformatics steps needed to generate a consensus genome sequence from Nanopore amplicon data?

ⓘ The Slido app must be installed on every computer you're presenting from

slido

Do not edit
How to change the design



What sequence errors can occur when generating a consensus?

ⓘ The Slido app must be installed on every computer you're presenting from

slido

Do not edit
How to change the design



What databases are there to find virus reference sequence?

ⓘ The Slido app must be installed on every computer you're presenting from

slido

Do not edit
How to change the design



What is important when generating a multiple sequence alignment?

ⓘ The Slido app must be installed on every computer you're presenting from

slido

Do not edit
How to change the design



Which often used genetic phylogeny method did we discuss last time?

ⓘ The Slido app must be installed on every computer you're presenting from

slido

Do not edit
How to change the design



What do APOBEC3 mutations signify in the analysis of mpox sequence data?

ⓘ The Slido app must be installed on every computer you're presenting from

slido

Recap of the exercise

- Genomic surveillance of hospitalized patients with suspected Mpox in South-Kivu – Democratic Republic of Congo (DRC)
- Lineage differentiating PCR detected common *Clade Ib* cases but also previously undetected *Clade Ia* cases
- Three Clade Ia samples are fit for sequencing
- Genomic epidemiology was deployed to investigate
- After successful sequencing you are tasked with the analysis of the data

Steps

1. Download the course material
2. Acquire mpox reference data
3. Align sequences using squirrel
4. Use provided clade Ia masking and compare the difference
5. Manually curate the alignment (if necessary)
6. Run IQ-TREE on the curated alignment using squirrel
7. Determine if there is one or multiple clusters
8. Use squirrel to determine the number of APOBEC3 mutations for the new strains

Do not edit
How to change the design



Did you manage to complete the exercise?

ⓘ The Slido app must be installed on every computer you're presenting from

slido

Do not edit
How to change the design



What do you think of the level of the exercise on scale from 1-10 with 1 being too easy and 10 being too difficult?

ⓘ The Slido app must be installed on every computer you're presenting from

slido

Do not edit
How to change the design



What challenges did you encounter during the execution of the exercise?

ⓘ The Slido app must be installed on every computer you're presenting from

slido

Do not edit
How to change the design



Where did you acquire your reference sequences to make your tree?

ⓘ The Slido app must be installed on every computer you're presenting from

slido

Do not edit
How to change the design



How many clusters did you identify?

ⓘ The Slido app must be installed on every computer you're presenting from

slido

Do not edit
How to change the design



Based on your analysis, do you think the Clade Ia cases represent multiple spillover events or ongoing human-to-human transmission?

ⓘ The Slido app must be installed on every computer you're presenting from

slido

Do not edit
How to change the design



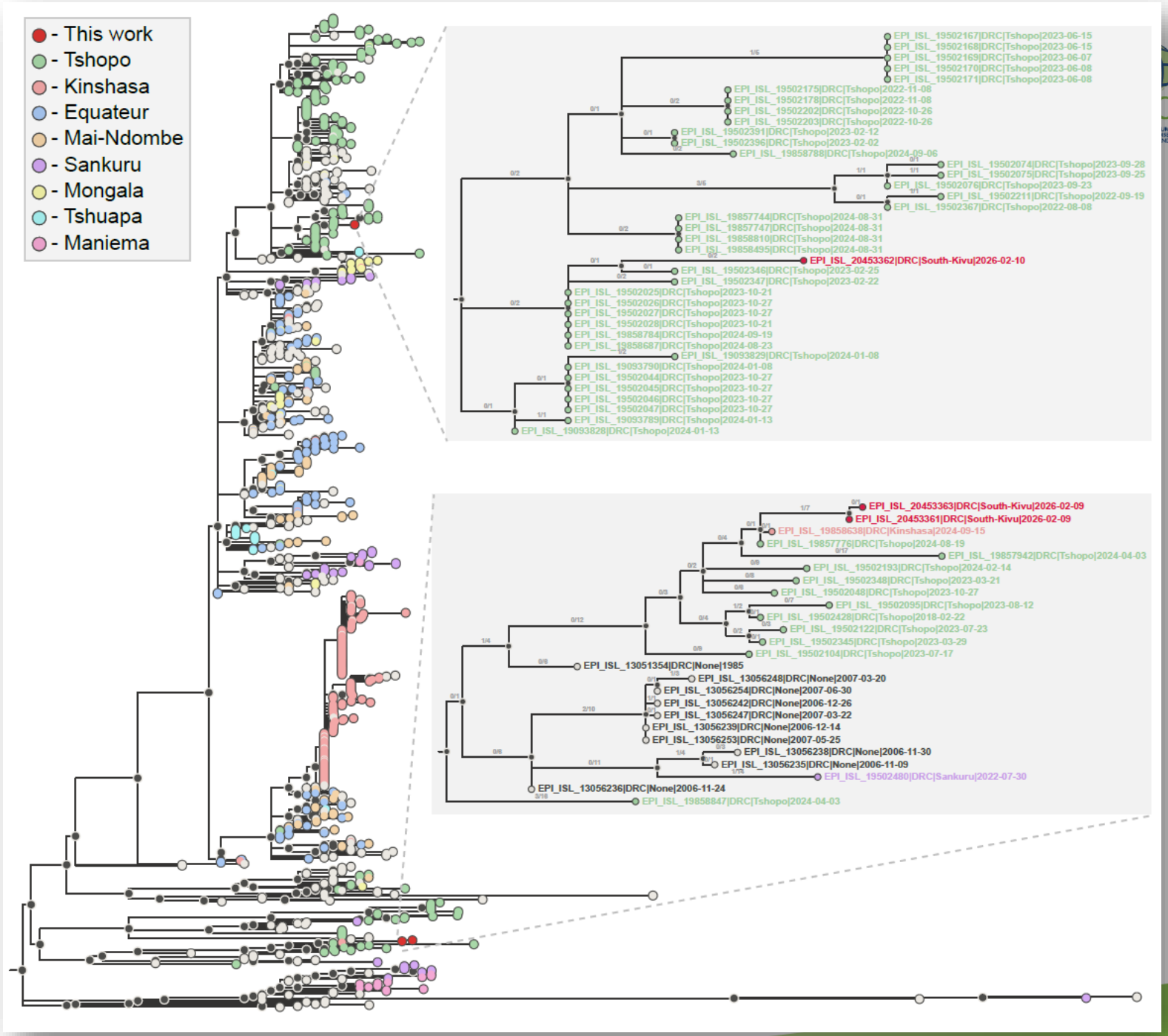
Why do you think so?

ⓘ The Slido app must be installed on every computer you're presenting from

slido

Results

- Phylogeny shows two separate introductions
- One in a cluster of sequences from Tshopo (with known Clade Ia circulation)
- One in a cluster with multiple origins



Do not edit
How to change the design



How many APOBEC-3 style mutations did you find

ⓘ The Slido app must be installed on every computer you're presenting from

slido

Results

APOBEC-3 style mutations

No strong signature caused by sustained human-to-human transmission



Do not edit
How to change the design



What are reasons to mask certain parts of the genome and why shouldn’t you just delete the entire sequence?

ⓘ The Slido app must be installed on every computer you’re presenting from

slido

Masking

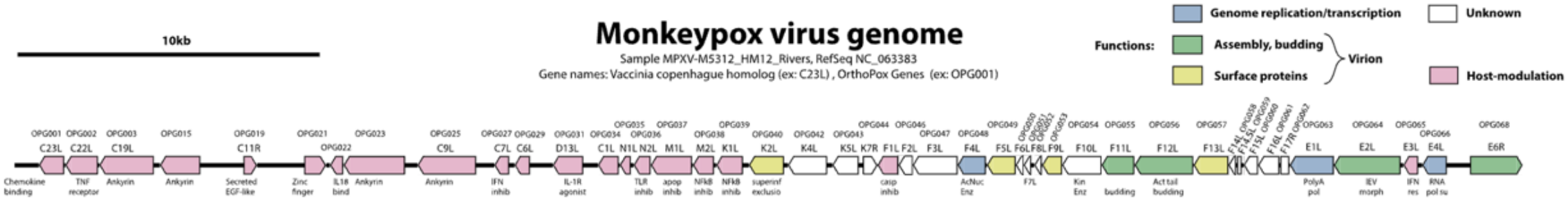
- Sometimes there is only limited genomic information from a certain region/continent

- Sometimes there is only limited genomic information from a certain time period (e.g. old sequences)

- It might help retain valuable epidemiological information

Mpox genomics

- Large genome of around 200kb encoding for around 200 different proteins
- Function of these proteins remain largely unknown
- Contains inverted terminal repeats
- Initial data suggests that the virus only changes 1-2 nucleotides per year
- Deletions and insertion are frequently observed



Manually inspect your alignment (200kb… x a lot of viruses)



Masking certain parts of the genome

Reference	Name	Type	Minimum	Maximum	Length	Direction
NC_003310	noisy region	misc_feature	1727	1735	9	none
NC_003310	repetitive region	misc_feature	8566	8593	28	none
NC_003310	repetitive region	misc_feature	10403	10556	154	none
NC_003310	homopolymeric runs	misc_feature	20446	20470	25	none
NC_003310	homopolymeric runs	misc_feature	21344	21366	23	none
NC_003310	repetitive region	misc_feature	31764	31784	21	none
NC_003310	noisy region	misc_feature	34219	34228	10	none
NC_003310	noisy region	misc_feature	41092	41097	6	none
NC_003310	noisy region	misc_feature	84242	84253	12	none
NC_003310	repetitive region	misc_feature	138325	138413	89	none
NC_003310	noisy region	misc_feature	145244	145268	25	none
NC_003310	noisy region	misc_feature	146186	146188	3	none
NC_003310	repetitive region	misc_feature	148716	148767	52	none
NC_003310	repetitive region	misc_feature	150192	150219	28	none
NC_003310	repetitive region	misc_feature	152389	152600	212	none
NC_003310	noisy region	misc_feature	158508	158514	7	none
NC_003310	messy region	misc_feature	162917	162948	32	none
NC_003310	noisy region	misc_feature	169251	169253	3	none
NC_003310	noisy region	misc_feature	169348	169354	7	none
NC_003310	repetitive region	misc_feature	169482	169502	21	none
NC_003310	noisy region	misc_feature	169715	169720	6	none
NC_003310	repetitive region	misc_feature	170635	170660	26	none
NC_003310	repetitive region	misc_feature	177521	177554	34	none
NC_003310	homopolymeric runs	misc_feature	177733	177743	11	none
NC_003310	homopolymeric runs	misc_feature	178772	178947	176	none

Masking adviced by squirrel

Name	Minimum	Maximum	Length	note
2426	2426	2426	1	convergent_snp
2626	2626	2626	1	reversion
6948	6948	6948	1	clustered_snps
6949	6949	6949	1	clustered_snps
13744	13744	13744	1	reversion
19128	19128	19128	1	clustered_snps
19130	19130	19130	1	clustered_snps
19131	19131	19131	1	clustered_snps
20250	20250	20250	1	N_adjacent
20269	20269	20269	1	N_adjacent
24871	24871	24871	1	clustered_snps
24874	24874	24874	1	N_adjacent;clustered_snps
24876	24876	24876	1	N_adjacent;clustered_snps
26654	26654	26654	1	N_adjacent
31230	31230	31230	1	clustered_snps
31232	31232	31232	1	clustered_snps
41097	41097	41097	1	N_adjacent
48612	48612	48612	1	N_adjacent
52227	52227	52227	1	N_adjacent
61479	61479	61479	1	N_adjacent
84244	84244	84244	1	N_adjacent
84253	84253	84253	1	N_adjacent
89000	89000	89000	1	N_adjacent
96895	96895	96895	1	N_adjacent
96897	96897	96897	1	N_adjacent
115484	115484	115484	1	convergent_snp
138322	138322	138322	1	N_adjacent
145002	145002	145002	1	N_adjacent
145010	145010	145010	1	N_adjacent
145011	145011	145011	1	N_adjacent
145012	145012	145012	1	N_adjacent
145013	145013	145013	1	N_adjacent
145015	145015	145015	1	N_adjacent;clustered_snps
145019	145019	145019	1	clustered_snps
145020	145020	145020	1	clustered_snps
145021	145021	145021	1	clustered_snps
145022	145022	145022	1	clustered_snps
145023	145023	145023	1	clustered_snps
145024	145024	145024	1	clustered_snps
145030	145030	145030	1	clustered_snps
145031	145031	145031	1	N_adjacent
145033	145033	145033	1	N_adjacent
145035	145035	145035	1	clustered_snps
145037	145037	145037	1	clustered_snps
145038	145038	145038	1	clustered_snps
145040	145040	145040	1	clustered_snps
159742	159742	159742	1	clustered_snps
159744	159744	159744	1	clustered_snps
159745	159745	159745	1	N_adjacent
162916	162916	162916	1	N_adjacent
162924	162924	162924	1	N_adjacent
162942	162942	162942	1	N_adjacent
169253	169253	169253	1	clustered_snps
169254	169254	169254	1	clustered_snps
169305	169305	169305	1	N_adjacent
169452	169452	169452	1	N_adjacent
169483	169483	169483	1	N_adjacent
169500	169500	169500	1	convergent_snp;N_adjacent
169501	169501	169501	1	convergent_snp;N_adjacent
171805	171805	171805	1	N_adjacent

Name	Minimum	Maximum	Length	note
178343	178343	178343	1	N_adjacent
178348	178348	178348	1	N_adjacent
180878	180878	180878	1	clustered_snps
180879	180879	180879	1	clustered_snps
180880	180880	180880	1	clustered_snps
180881	180881	180881	1	clustered_snps
180883	180883	180883	1	N_adjacent
19127	19127	20281	1155	Gappy
31227	31227	31232	6	Gappy
41092	41092	41099	8	Gappy
52206	52206	52206	1	multiple_mutations
52220	52220	52220	1	multiple_mutations
52227	52227	52227	1	multiple_mutations
84244	84244	84253	10	Gappy
144998	144998	145046	49	Gappy
145244	145244	145255	12	Gappy
157454	157454	157458	5	Gappy
162893	162893	162947	55	Gappy
169251	169251	169254	4	Gappy
180872	180872	180891	20	Gappy
165597	165597	165597	1	Reversion
1	1	6378	6378	ITR
101965	101965	101965	1	near_N
178235	178235	178235	1	near_N
135721	135721	135721	1	near_N
183181	183181	183181	1	near_N
135375	135375	135375	1	convergent_IIb
99854	99854	99854	1	convergent_other_clades
169452	169452	169501	50	repetitive
172844	172844	172844	1	near_N
27697	27697	27697	1	convergent_some_IIb
86815	86815	86815	1	near_N
35666	35666	35666	1	differe
71952	71952	71952	1	convergent_some_IIb
64117	64117	64117	1	convergent_some_IIb
repetitive region	10403	10556	154	squirrel_default_mask
homopolymeric runs	20449	20470	22	squirrel_default_mask
homopolymeric runs	21344	21366	23	squirrel_default_mask
repetitive region	31764	31784	21	squirrel_default_mask
repetitive region	138325	138413	89	squirrel_default_mask
repetitive region	148716	148767	52	squirrel_default_mask
repetitive region	150192	150219	28	squirrel_default_mask
repetitive region	152389	152600	212	squirrel_default_mask
repetitive region	169482	169502	21	squirrel_default_mask
repetitive region	170635	170660	26	squirrel_default_mask
homopolymeric runs	177733	177743	11	squirrel_default_mask
homopolymeric runs	178772	178947	176	squirrel_default_mask
8293	8293	8293	1	near_N
18157	18157	18157	1	Reversion
87834	87834	87834	1	N_region
123603	123603	123603	1	near_N
135322	135322	135322	1	near_N
135779	135779	135779	1	Reversion
148582	148582	148582	1	unresolved_bases
152301	152301	152301	1	Reversion_clustered_snps
152345	152345	152345	1	Reversion_clustered_snps
152370	152370	152370	1	Reversion_clustered_snps
159115	159115	159115	1	Gappy
168857	168857	168857	1	near_N
178397	178397	178397	1	Gappy

Masking adviced by WHO ISPN

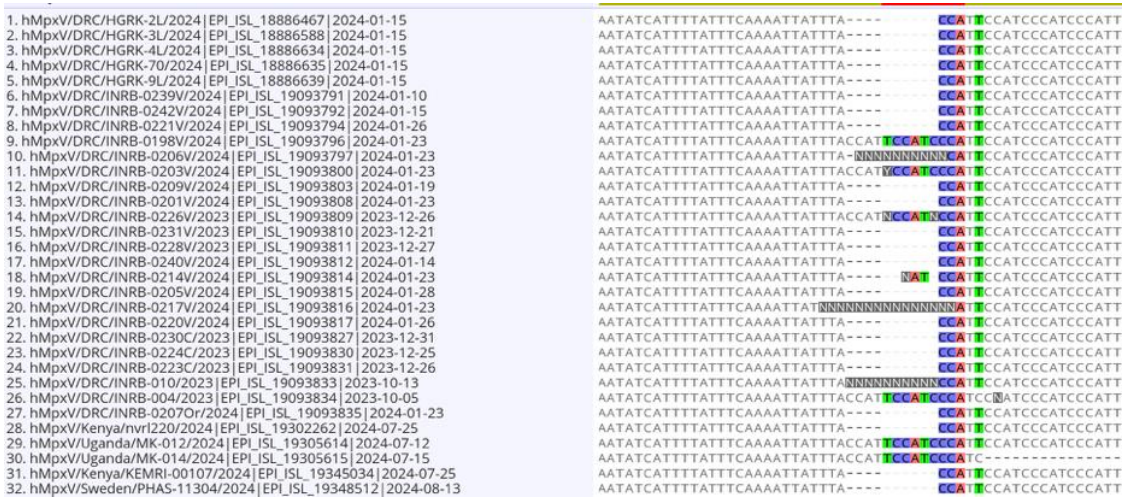
Reason for masking

- Repetative regions
- Homopolymeric regions
- Noisy regions

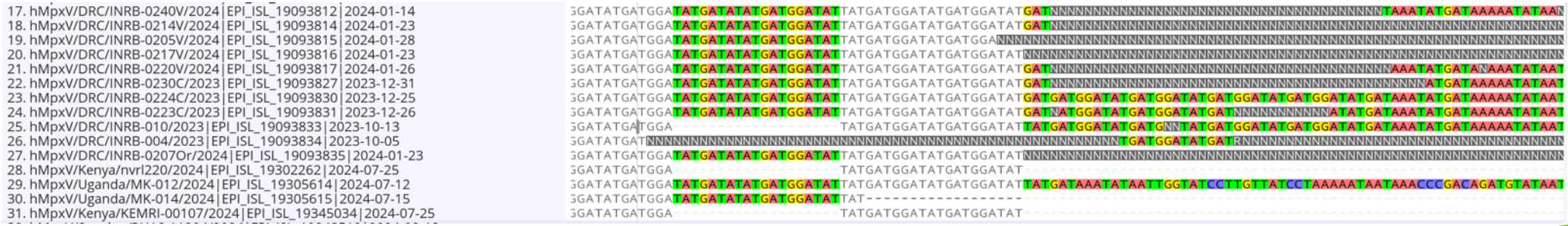
- Near regions containing a lot of “NNN”
- Gappy regions
- Clustered SNPS

Reason for masking

- **Repetative regions**
- Homopolymeric regions
- Noisy regions

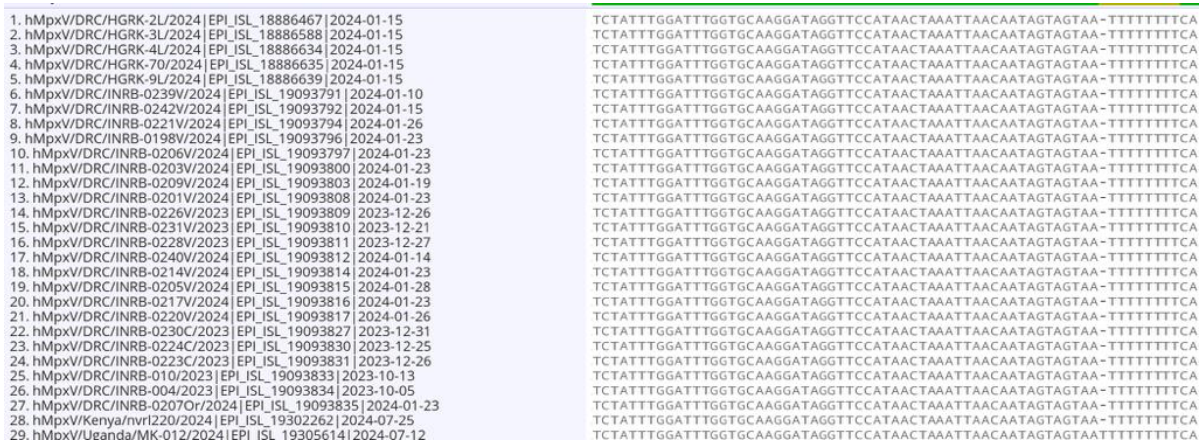


- Near regions containing a lot of “NNN”
- Gappy regions
- Clustered SNPS

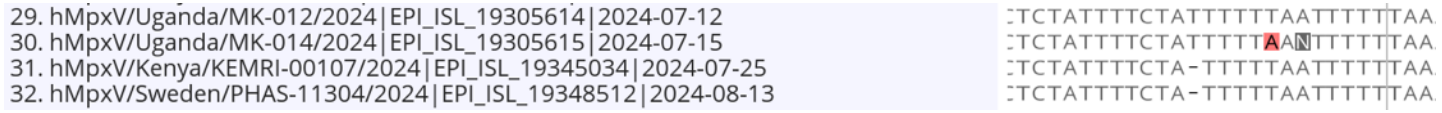


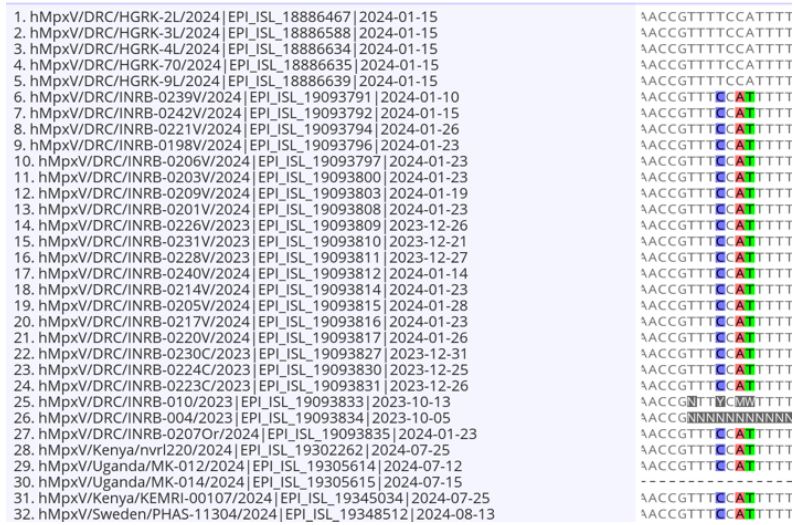
Reason for masking

- Repetative regions
- **Homopolymeric regions**
- Noisy regions



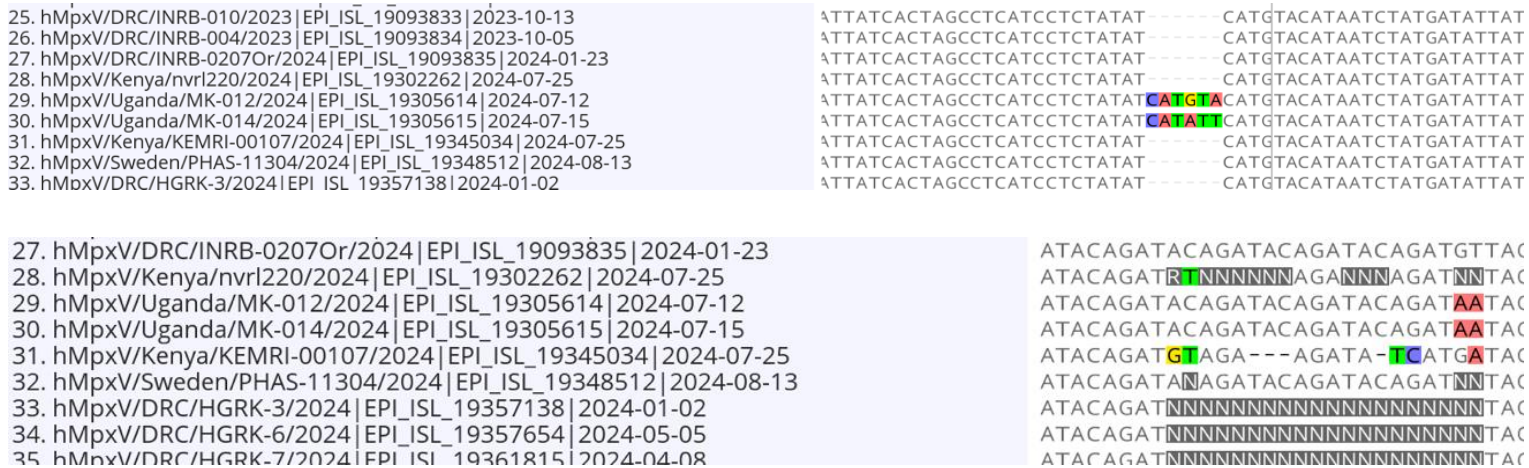
- Near regions containing a lot of “NNN”
- Gappy regions
- Clustered SNPS



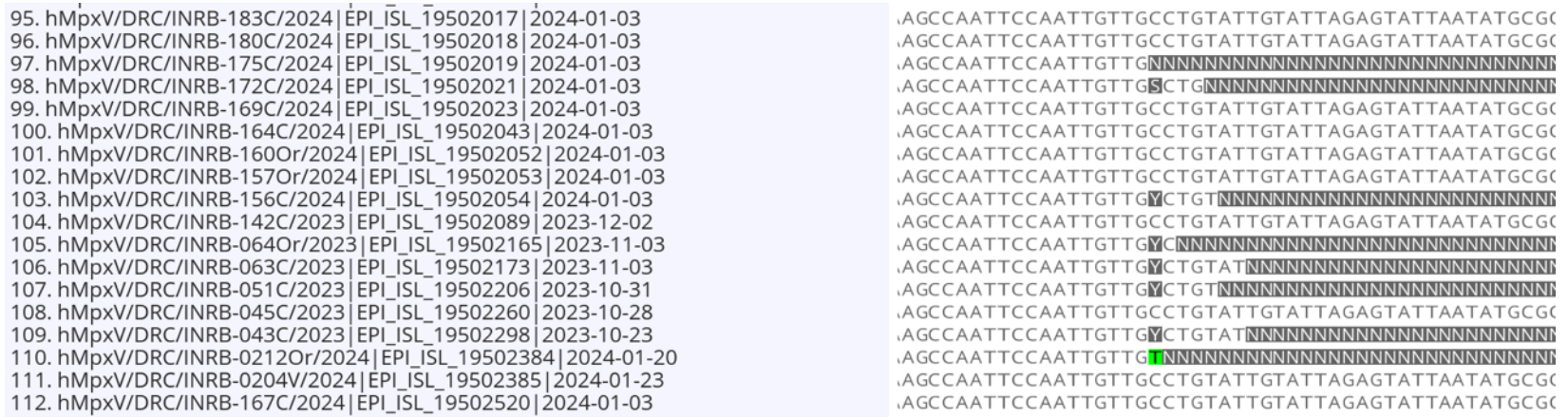


Reason for masking

- Repetative regions
- Homopolymeric regions
- **Noisy regions**

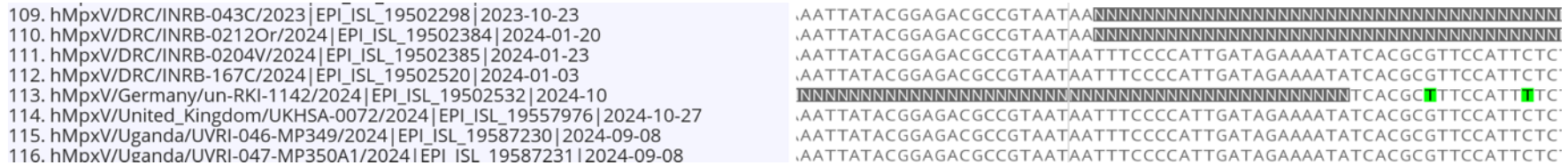


- Near regions containing a lot of “NNN”
- Gappy regions
- Clustered SNPS

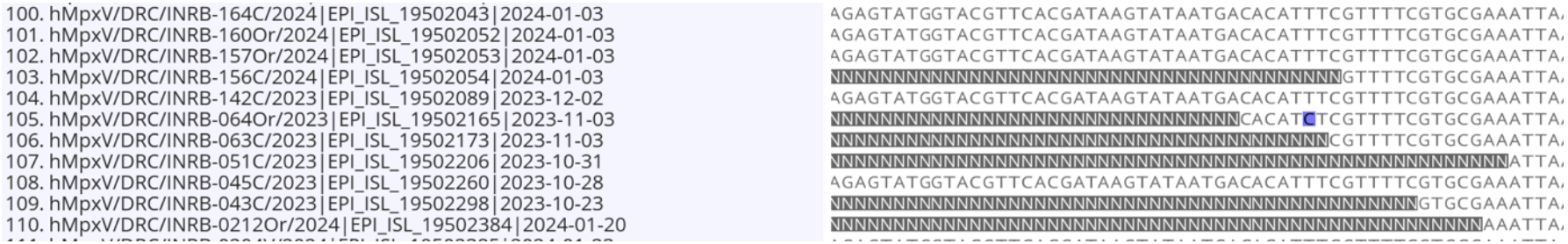


Reason for masking

- Repetative regions
- Homopolymeric regions
- Noisy regions

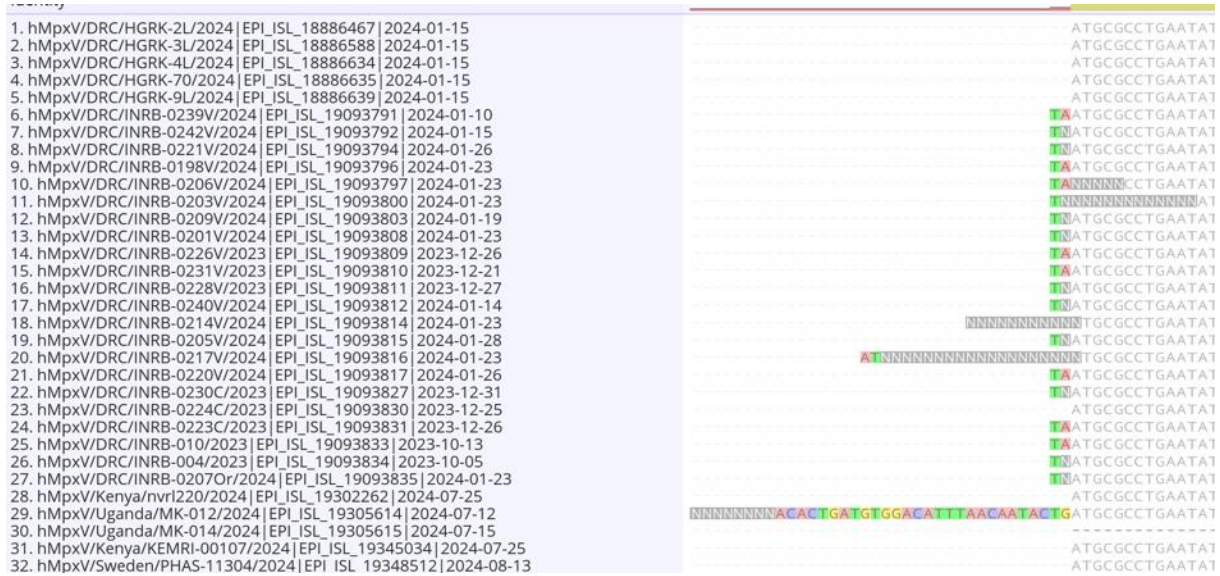


- **Near regions containing a lot of "NNN"**
- Gappy regions
- Clustered SNPS

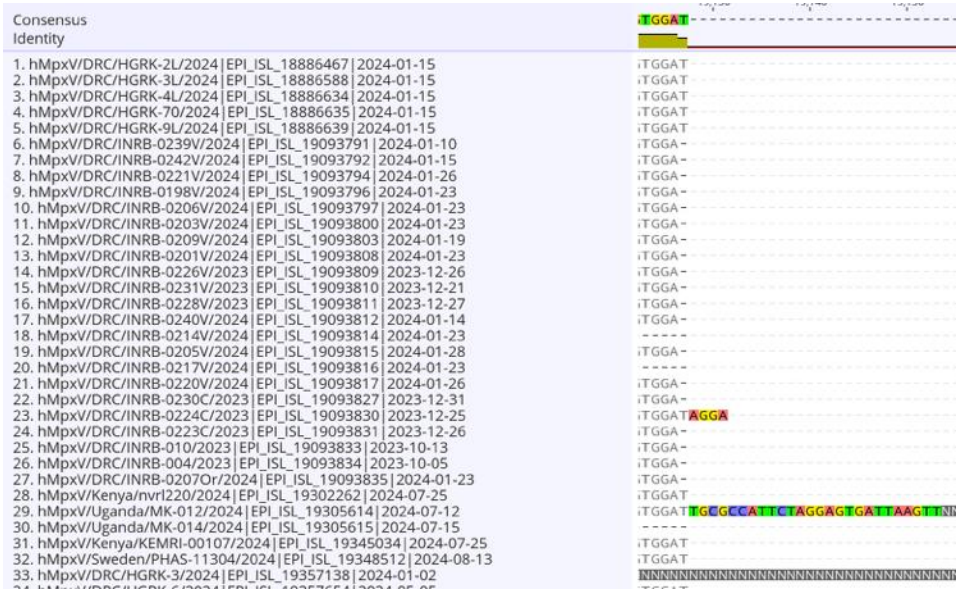


Reason for masking

- Repetative regions
- Homopolymeric regions
- Noisy regions

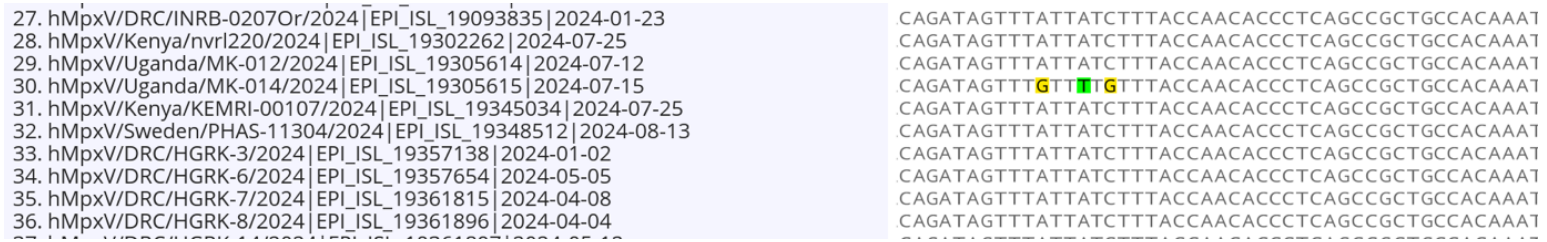


- Near regions containing a lot of “NNN”
- **Gappy regions**
- Clustered SNPS

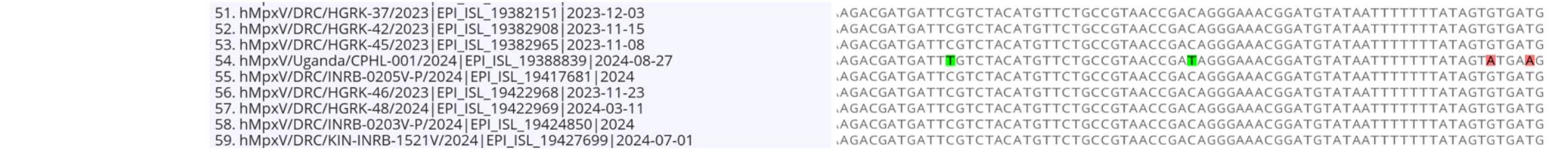


Reason for masking

- Repetative regions
- Homopolymeric regions
- Noisy regions

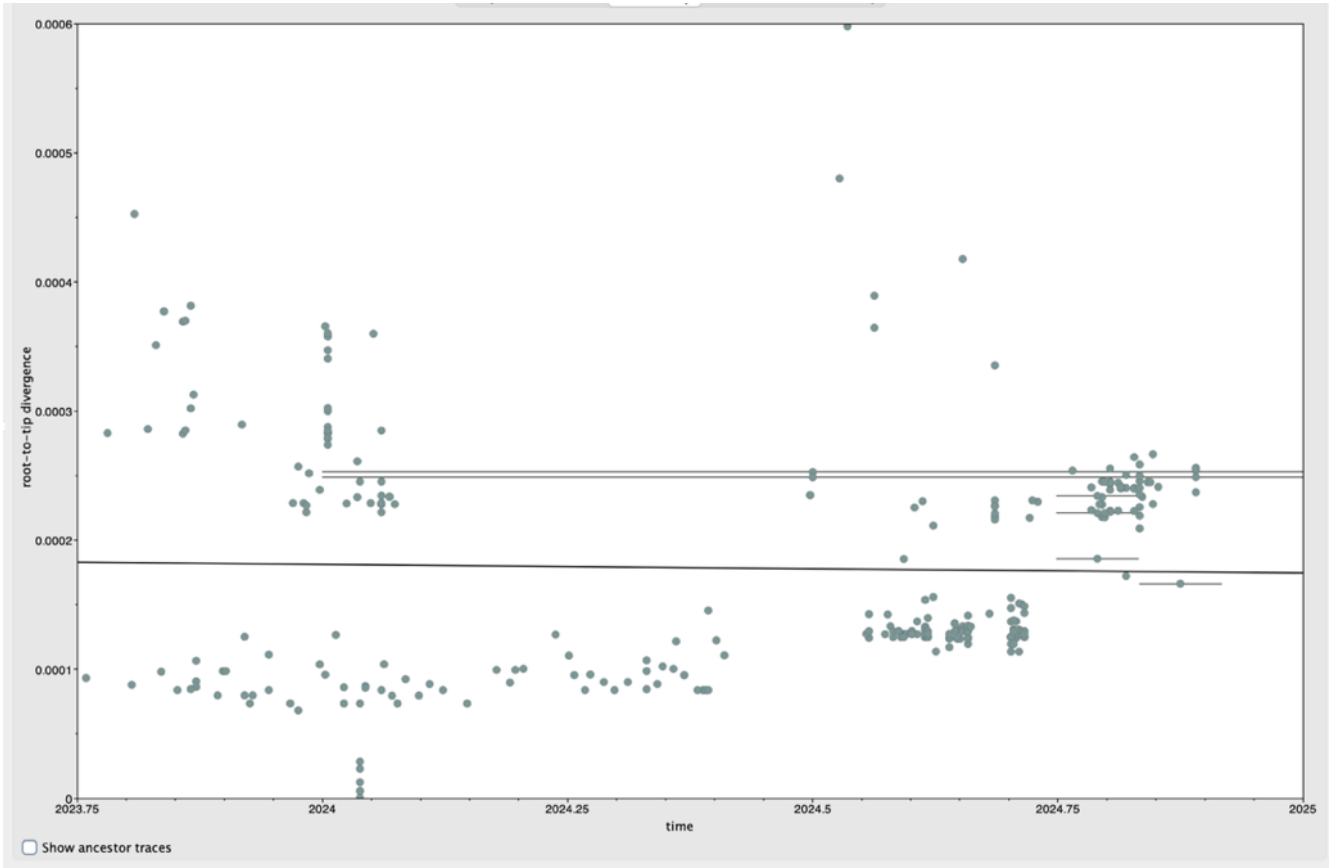


- Near regions containing a lot of “NNN”
- Gappy regions
- **Clustered SNPS**

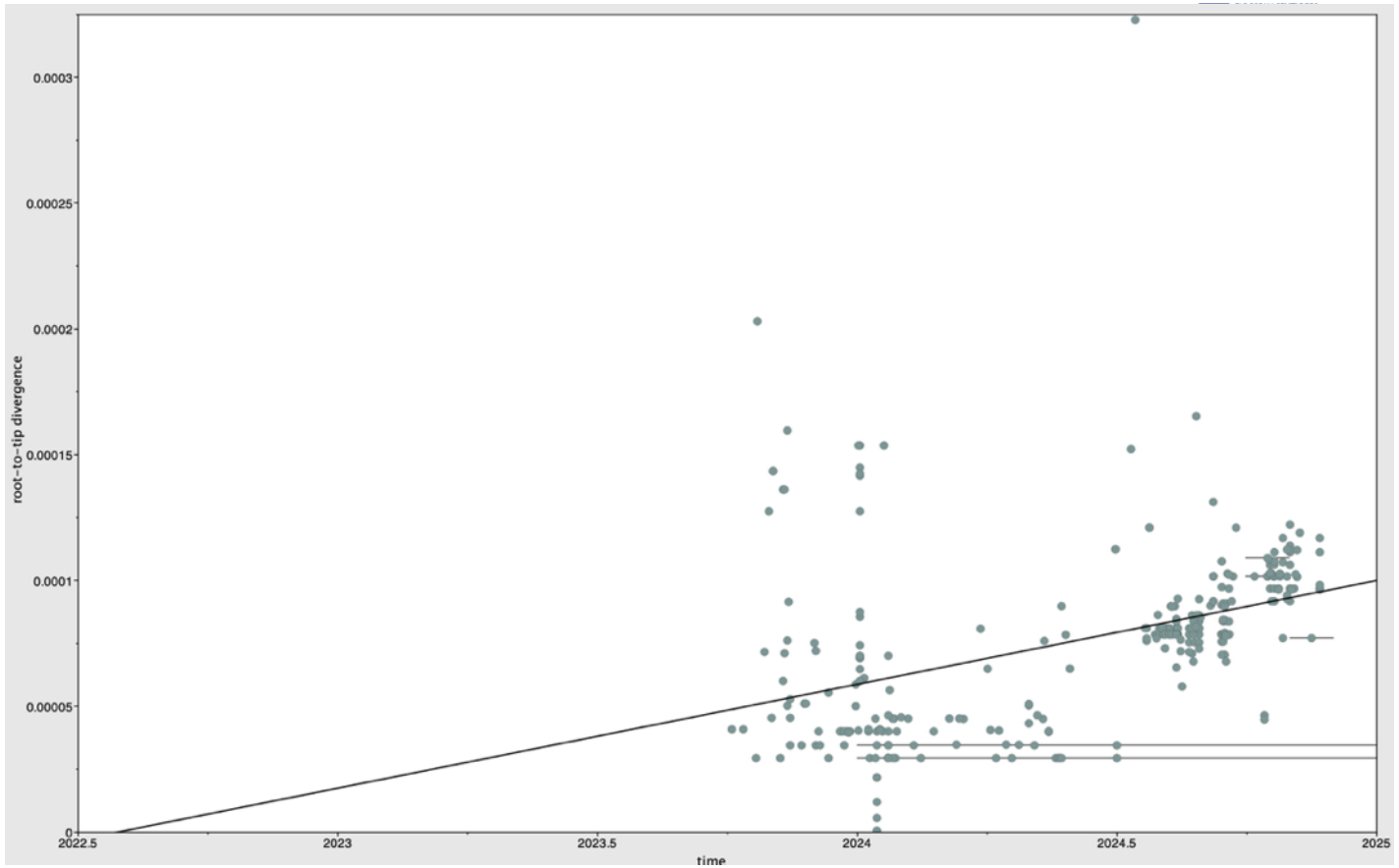


Relevance of masking



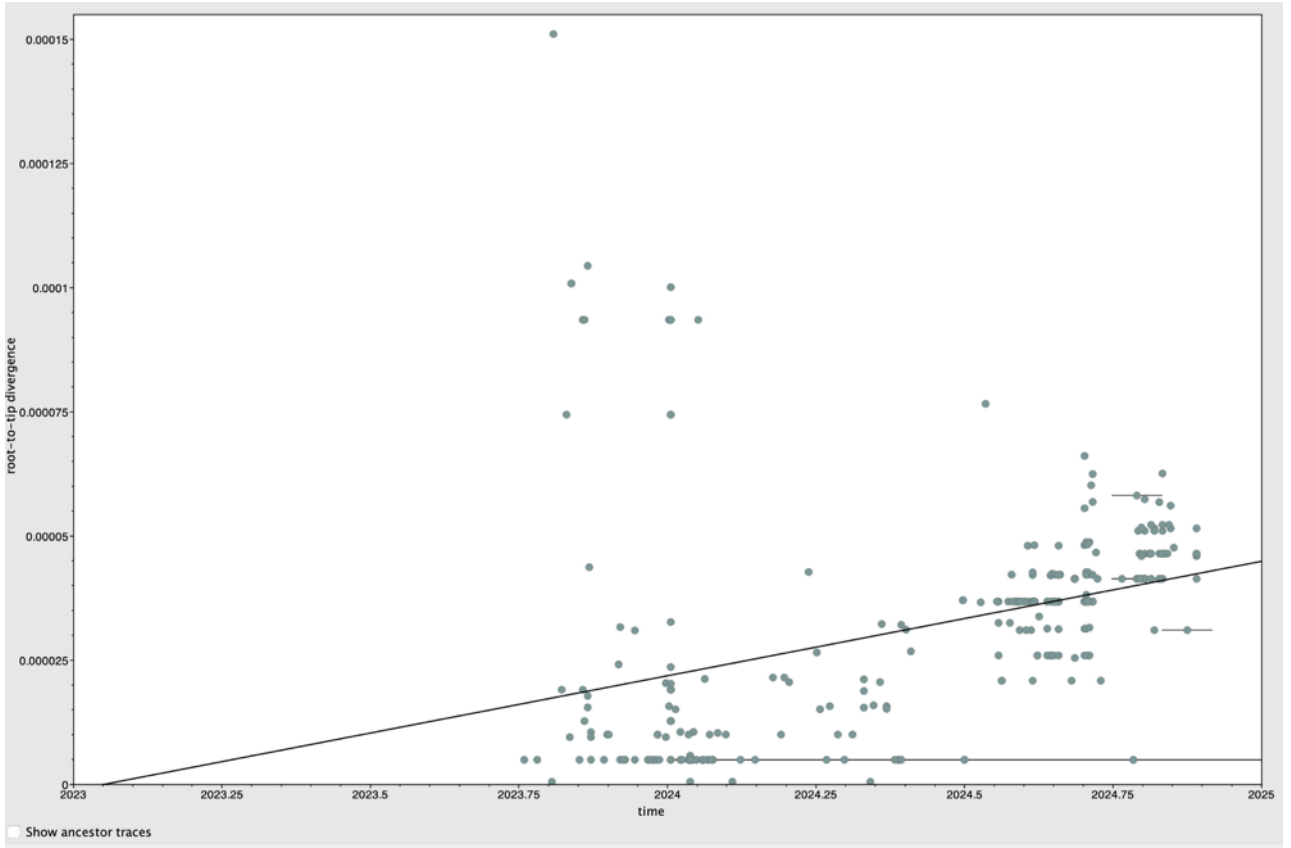


Relevance of masking (squirrel)



Relevance of masking (ISPN/WHO)





Resolution might be lost during masking

In summary

List of learning points in this session:

- Learned how to interpreted the phylogenetic results
- Learned how to identify potential mistakes in consensus sequences
- Learned how to identify potential clusters
- Learned how to discriminate between ongoing human-to-human transmission of mpox or if there are repeated zoonotic transmission events

Reusing this content

This work is licensed under the Creative Commons (http://creativecommons.org) Attribution 4.0 Generic license

You are free:
to share – to copy, distribute and transmit the work
to remix – to adapt the work

Under the following conditions; you must attribute the work in the manner specified by the author or licensor but not:
- in any way that suggests that they endorse you or your use of the work
- in any way that suggests you are author of the work.

The authors of this slide/page encourage educators, trainers, and professionals to include this slide within their documents and presentations for rightful attribution of their works and thus also allow it to be easily shared. (ECV1-29/9/11)

Terms of usage (Learning Portal)