

Benchmarking AmpliDiff for monkeypox, HIV-1 and Influenza-A

1. AmpliDiff

AmpliDiff takes an input of MSA sequences and their lineages. AmpliDiff then uses this input to find amplicons and their primers such that we can differentiate as many sequences as possible based on these amplicons. It does this by:

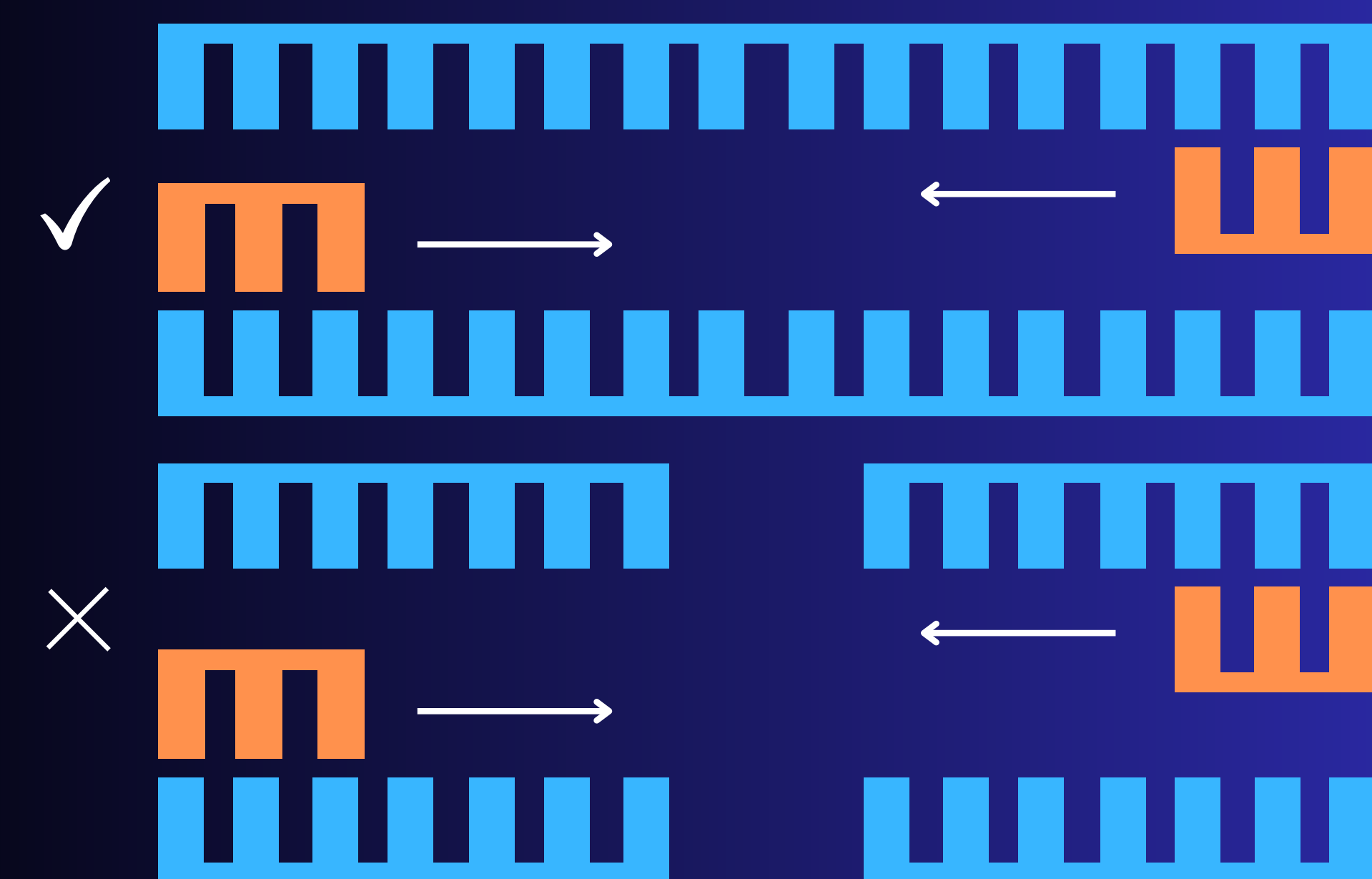
- Generating a list of all possible primers
- Generating a list of all possible amplicons
- Sorting the amplicons on differentiability
- Greedily taking the most differentiable amplicon and checking for feasibility

It does this until we have either found a specified amount of amplicons, or until every single sequence can be differentiated.

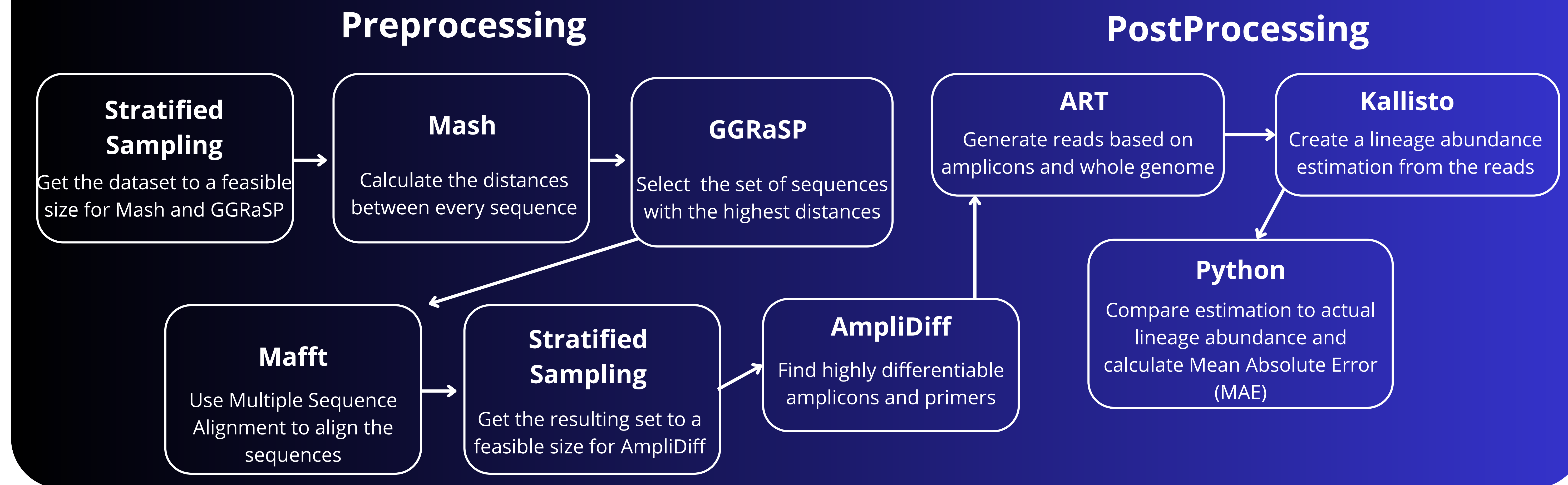
2. Segmented genome

Influenza has a genome split into 8 segments. To run Influenza on AmpliDiff:

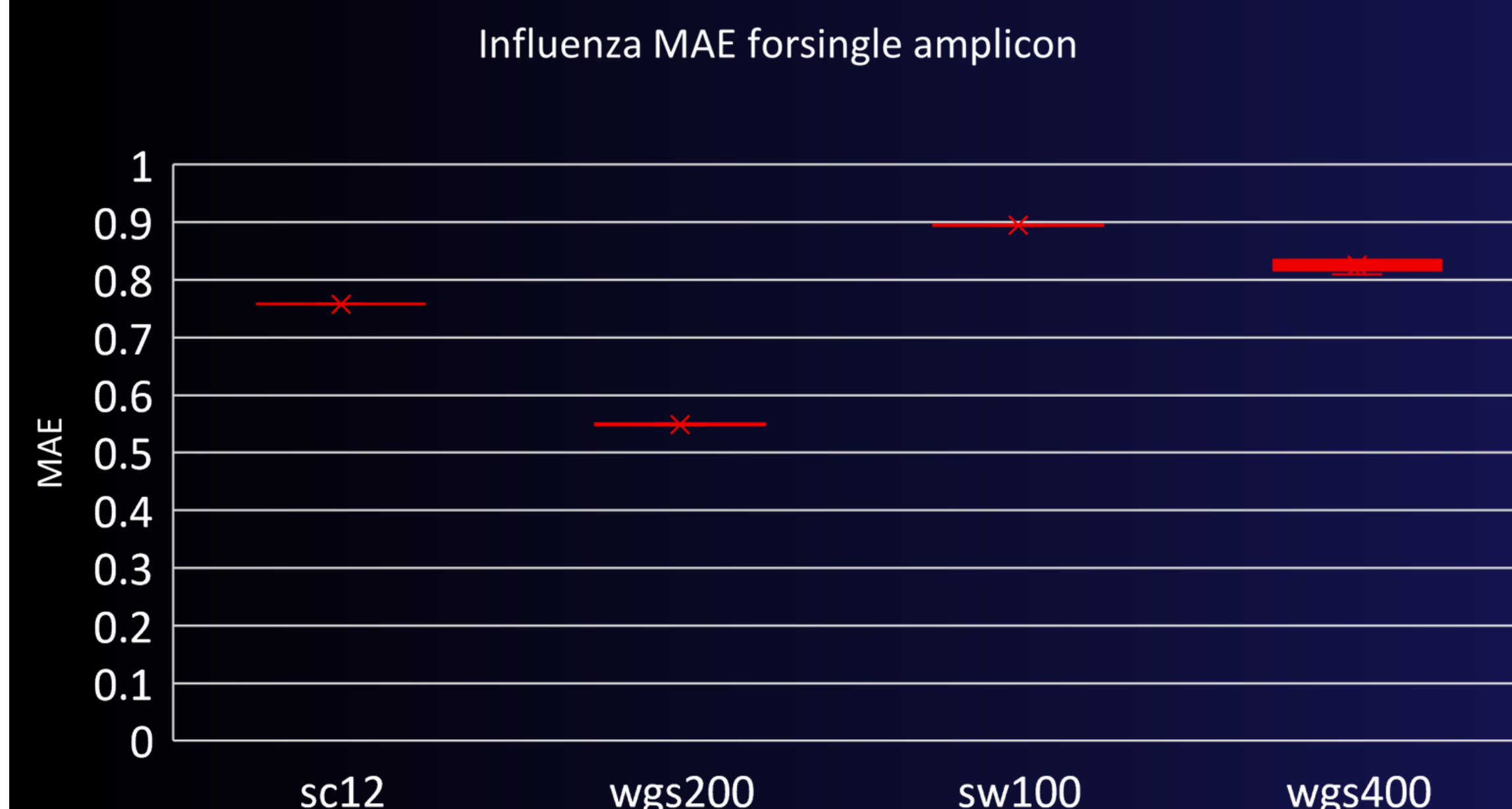
- Paste together the segments with '8'
- Modify AmpliDiff to not generate amplicons or primers that contain '8'
- Run as if a single segment genome



3. Full pipeline



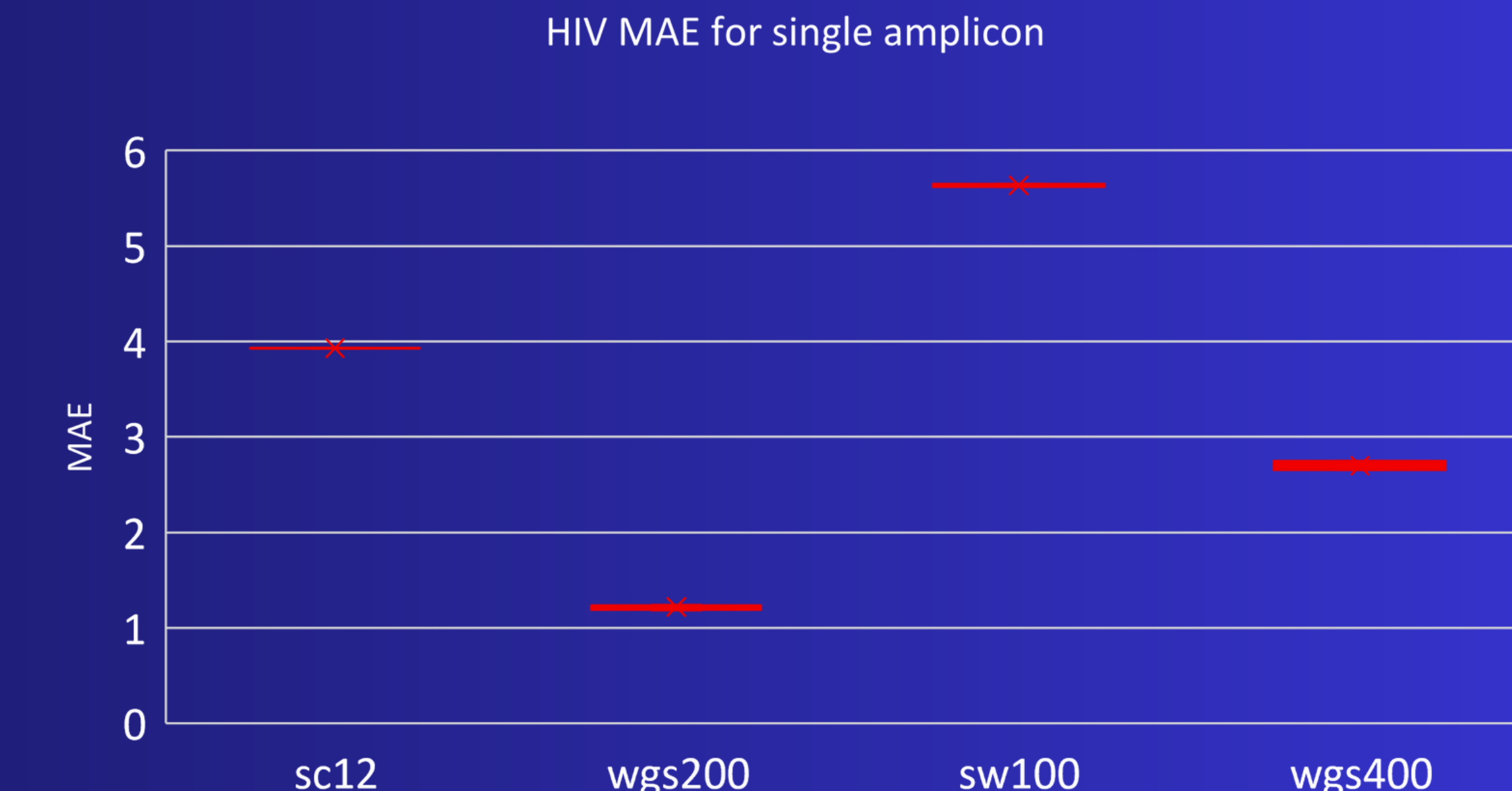
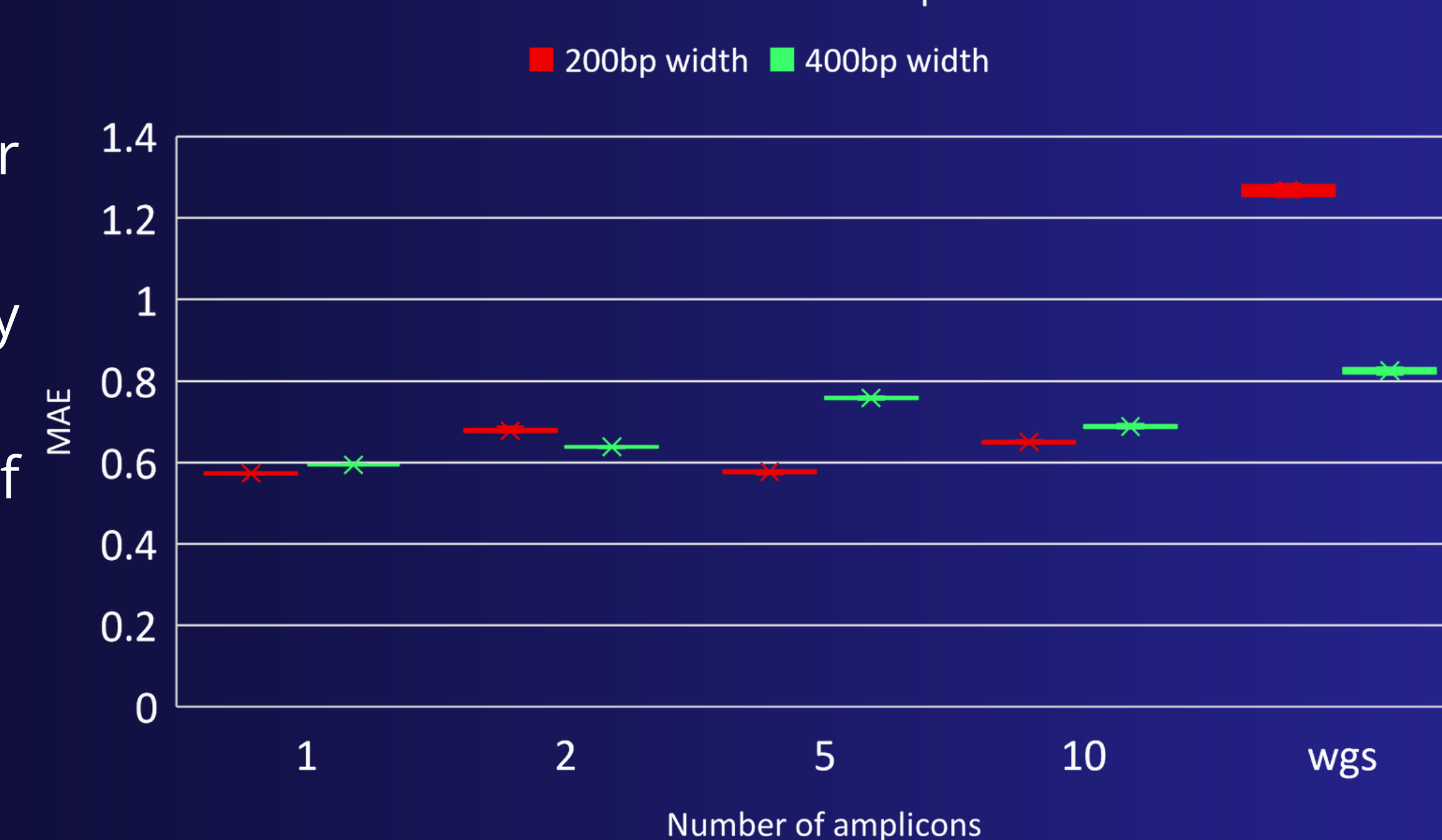
Results



Influenza-A

- Length of ~15.000bp
- Single amplicon found, results for runs with specific settings:
 - Increased self complementary threshold of primers (sc12)
 - Increased primer search width of 100 (sw100)
- amplicon performs similar to wgs
- sc12 runs outperforms sw100 run

Pox MAE vs number of amplicons



HIV-1

- Length of ~9.000bp
- Single amplicon found, runs with same settings as for Influenza-A
- wgs vastly outperforms the amplicon found
- Had trouble with Multiple Sequence Alignment, increasing size of ~9.000 to ~15.000.