

**Background:** According to CDC, H5N1 bird flu is causing outbreaks in poultry and U.S. dairy cows with several cases of H5N1 in U.S. dairy workers. Tiled amplicon sequencing is a fast and economic way to detect and track multiple influenza A strains in wastewater.

**Goals:** design a tiled amplicon scheme targeting latest HA and NA segments of influenza A strains, including H5N1. The design should be compatible with wastewater monitoring (short amplicons).

**Summary of results:** A total of 45 tiled amplicons are designed with Olivar (beta version), covering latest HA segment of H5N1, H1N1 and H3N2, and NA segment of H5N1 and H3N2. This beta version of Olivar can take sequences as input and support multiple targets.

# Multiple sequence alignment (MSA) and variant calling

1. Latest HA and NA segments were downloaded from GISAID, including three strains: H5N1, H1N1 and H3N2.
2. MSA was made for each segment of each strain, with `mafft --auto` (version 7.525). All sequences were included for MSA.
3. A consensus sequence was made for each MSA (gaps removed).
4. Variant calling was done with a customized module of Olivar, using the consensus sequence as reference.
5. Lists of variant location and frequency, as well as consensus sequences were used for downstream Olivar design.

---

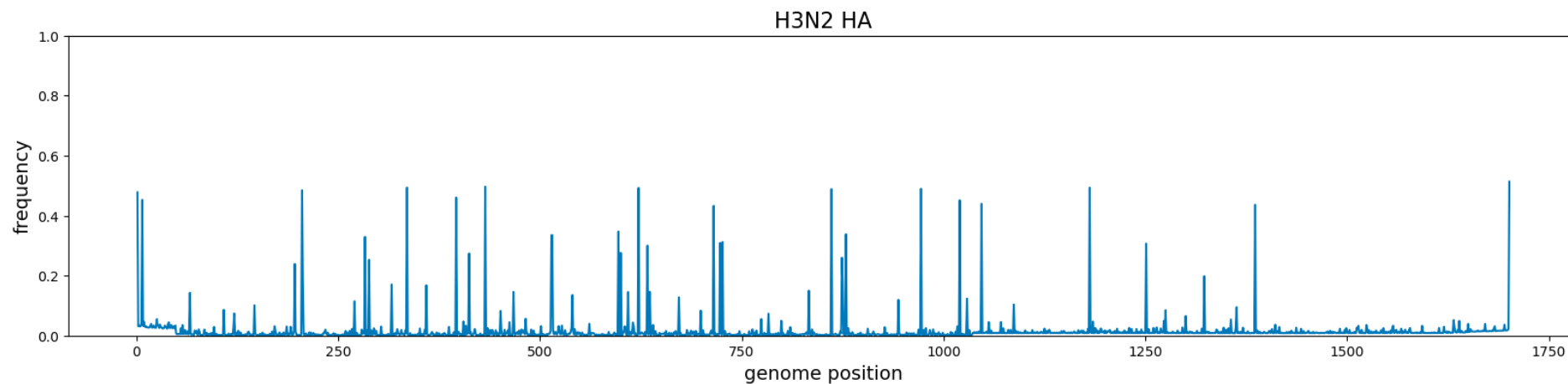
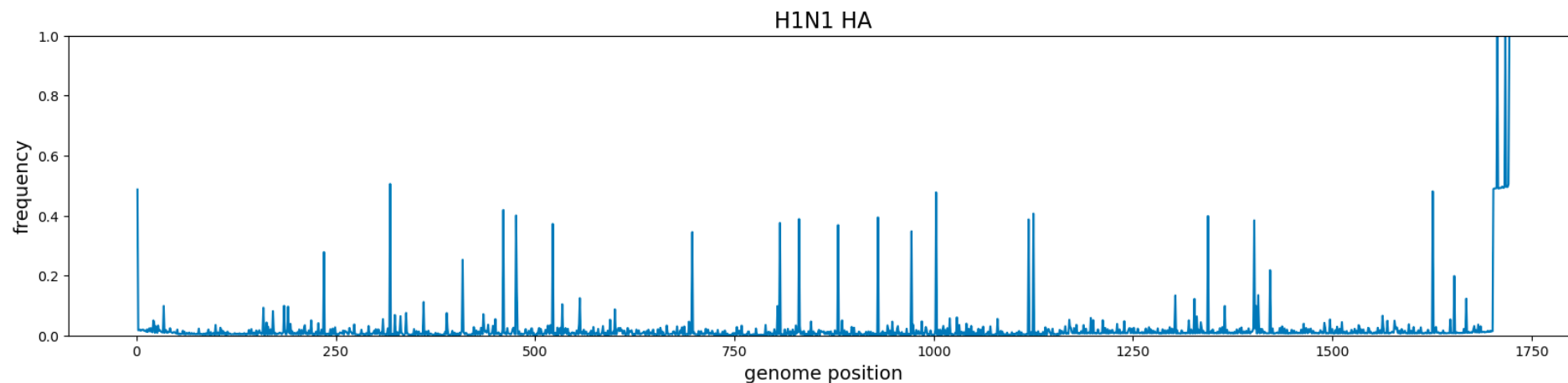
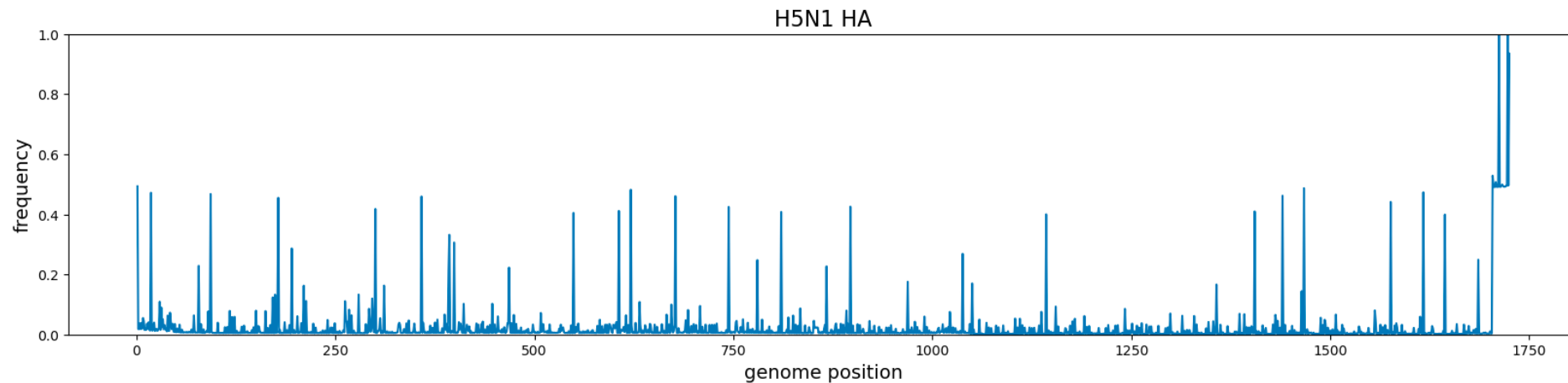
# sequences downloaded from GISAID  
collection date 05-25-2022 to 05-25-2024

	HA segment	NA segment
H5N1	6,680	6,578
H1N1	42,042	40,846
H3N2	44,431	42,639

---

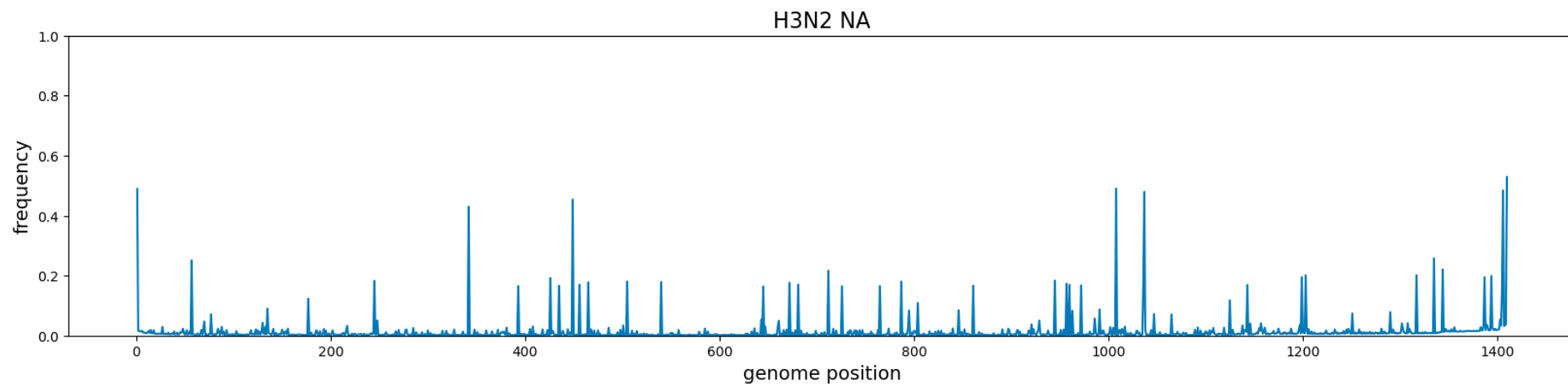
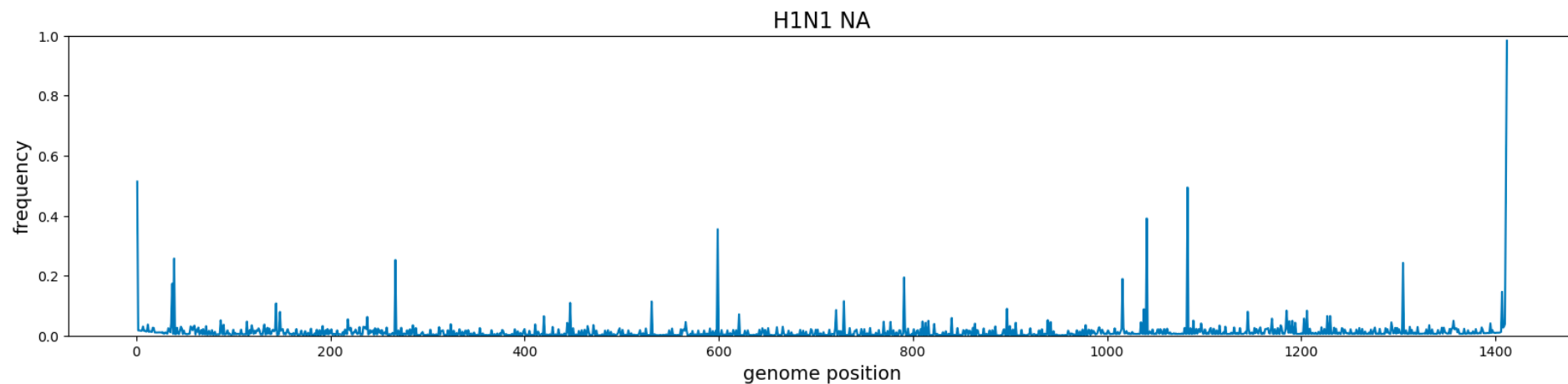
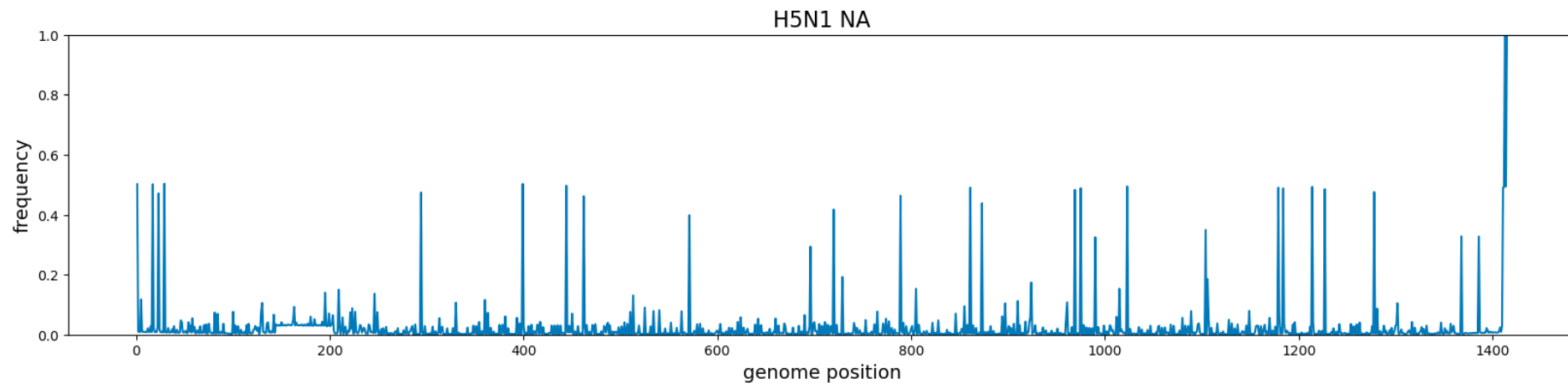
HA segment variant frequency by position on the reference (consensus).

Frequency includes substitution, deletion and insertion.



NA segment variant frequency by position on the reference (consensus).

Frequency includes substitution, deletion and insertion.

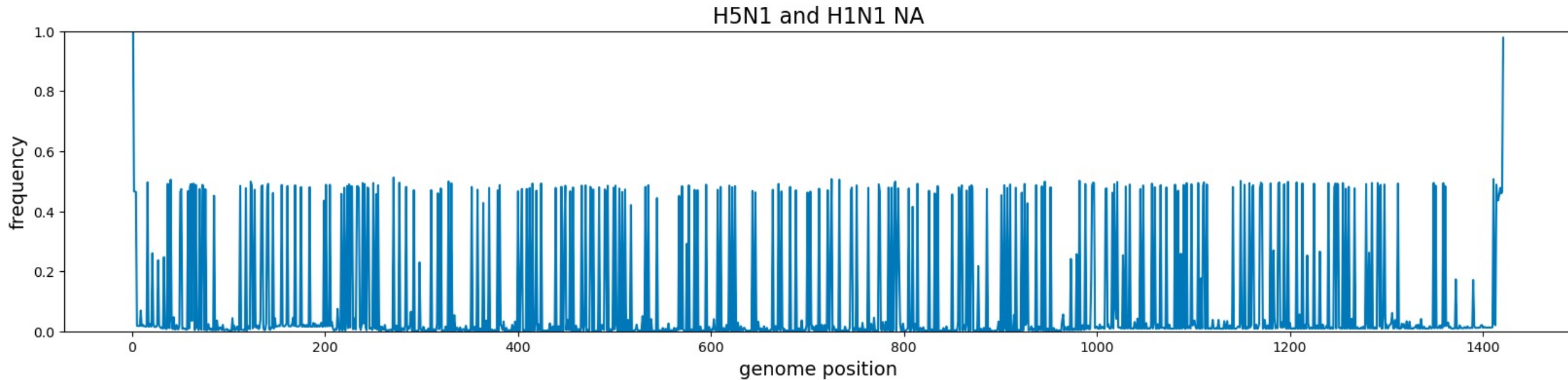


- To test the feasibility of covering the NA segment of both H5N1 and H1N1 with the same tiled amplicon set, an additional MSA was made combining the two groups of NA segments below.

6,680 H5N1 NA sequences (05-25-2022 to 05-25-2024)

7,104 H1N1 NA sequences (01-20-2024 to 05-25-2024)

- A consensus was made, and variant calling was done in the same way.
- **While impractical to cover both segments at the same time, we decided to cover H5N1 only to avoid formation of primer dimers or other complexes.**

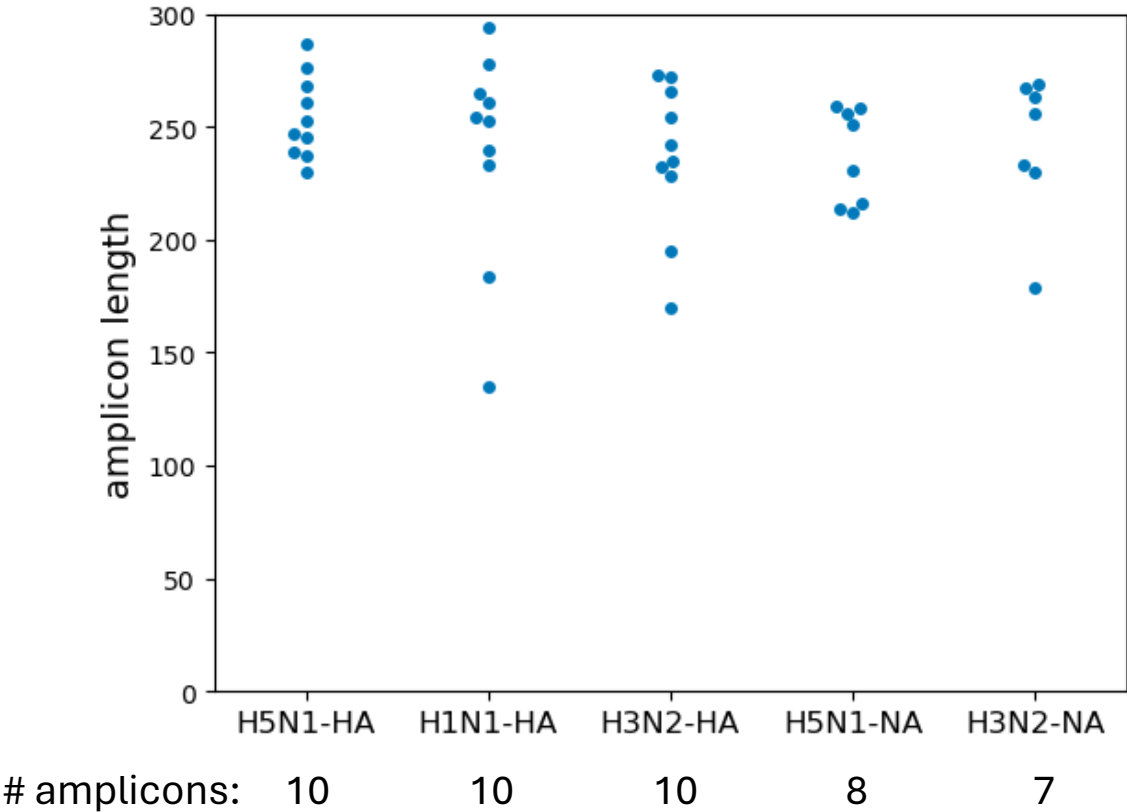


## Design of tiled amplicons

- Amplicon length set as 150 to 300, since shorter amplicons have better sensitivity for fragmented wastewater DNA/RNA. Annealing temperature set as 60°C.
- BLAST database: `ref_viruses_rep_genomes` (downloaded from NCBI FTP on May. 18, 2024, BLAST v2.15.0). Influenza A is excluded from BLAST search with `-negative_taxids 11320`
- NA segment of H1N1 is not included in the design
- Design outputs are uploaded to GitHub (<https://github.com/treangenlab/InfA-amplicon>)
  - Output primers found in `InfA-tiling.csv` and `InfA-tiling.scheme.bed` (ARTIC format)
  - Interactive html figures found in the Figures folder. Download html files and view in browser. Refer to Fig1 of Olivar manuscript for details ([doi.org/10.1101/2023.02.11.528155](https://doi.org/10.1101/2023.02.11.528155)).

# Summary of output primers

**Total # amplicons: 45**  
**Amplicon length range: 135 to 294**



## Sensitivity of primer pairs

**y-axis:** percentage of perfectly matched segment for each pair of fP and rP

