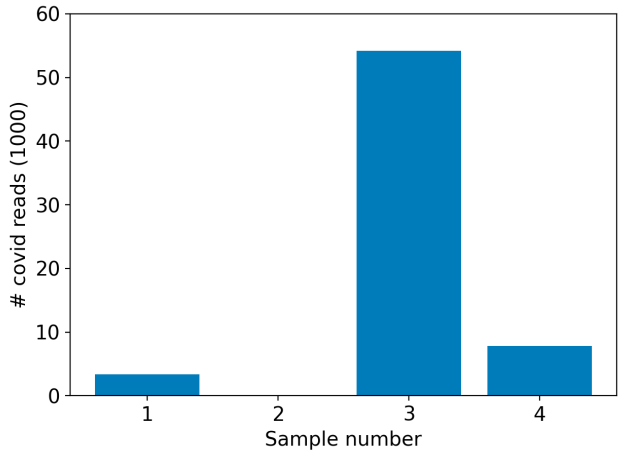
U.S. FOOD & DRUG ADMINISTRATION

CFSAN/OAO
BIOSTATISTICS AND BIOINFORMATICS STAFF

WASTEWATER SARS-COV2 ANALYSIS REPORT

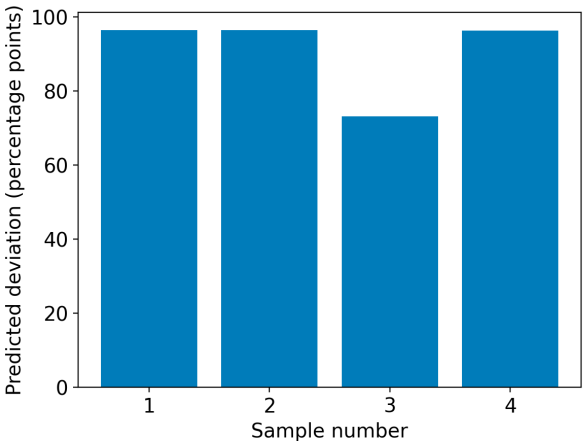
Summary

Sample#	Sample name	Total #reads	Reads aligned PF*	Genomic coordinates 0X	Genomic coordinates <10X
1	SRR16828013	906676	3399 (0%)	18832nt (62%)	28129nt (94%)
2	SRR16828014	4096	154 (3%)	22376nt (74%)	29615nt (99%)
3	SRR16828015	1366018	54184 (3%)	6793nt (22%)	12997nt (43%)
4	SRR16828016	1114642	7861 (0%)	15002nt (50%)	24582nt (82%)



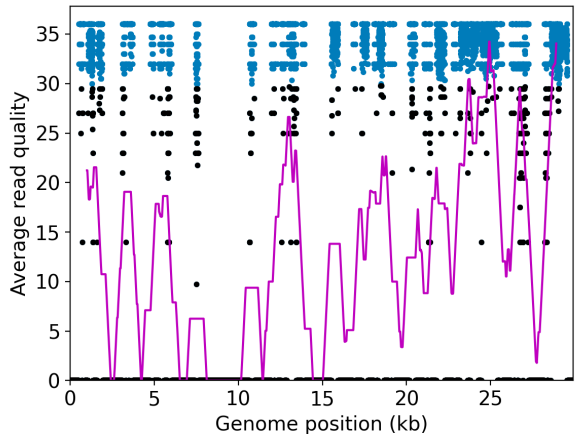
*Quantity of raw reads that align to the reference sequence and pass filter, i.e. the read length after adaptor trimming ≥30 and minimum read quality ≥20 within a sliding window of width 4. SNR refers to the ratio of SC2-mapping reads aligned that pass filter in the sample vs. that in the auto-detected negative control samples (if any). The dashed line represents the baseline level of covid reads detected from the negative control or their average if multiple negative controls we included.

QC-bot (Not yet functional)

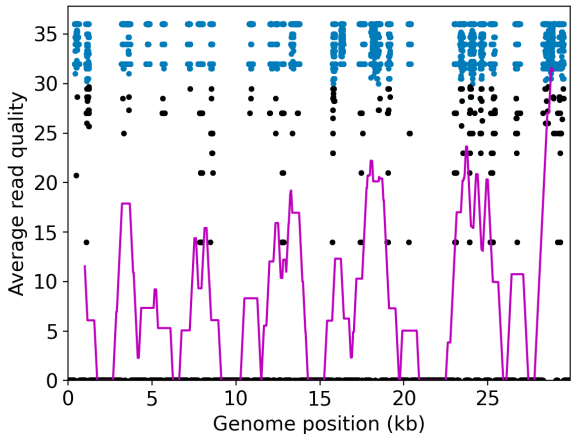


Machine-learning based prediction of the SC2 variant calling accuracy of Freyja of this dataset. The model is a random forest trained on FDA/CFSAN's experimental wastewater WGS data obtained in January 2022 and aims to assess the impact of the potential coverage gaps on the variant abundance estimates. The plotted values represent the predicted deviation of the omicron percentage points from the value that would have been obtained if the coverage was near-complete.

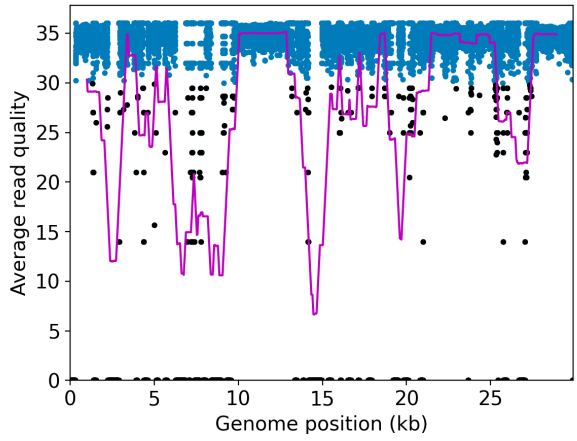
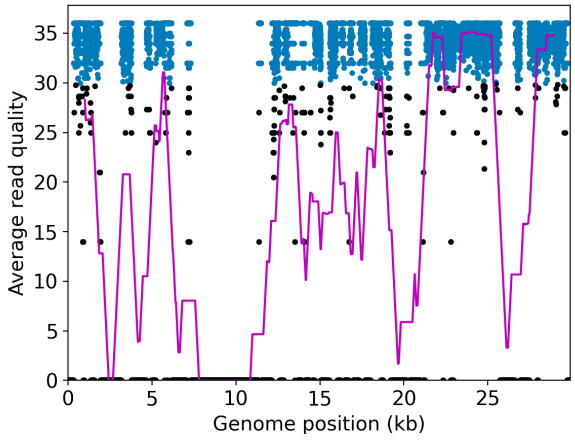
SRR16828013



SRR16828014



SRR16828015



SRR16828016