ecdc
EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL

MONITORING

European SARS-CoV-2 and influenza Bioinformatics External Quality Assessment (ESIB-EQA), 2023

ECDC MONITORING

European SARS-CoV-2 and influenza Bioinformatics External Quality Assessment (ESIB-EQA), 2023

This report was commissioned by the European Centre for Disease Prevention and Control (ECDC), under the framework contract ECDC/2022/001 and specific contract No.2 (request for services RS/2022/DPR/25222), coordinated by Olov Svartström, Annette Kraus and Eeva Broberg, and produced by AURORAE Consortium.

Authors
Ivo van Walle, Centre for Infection Research, Diagnostics and Laboratory Surveillance, Dutch National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands

Etienne Simon-Loriere, G5 Evolutionary Genomics of RNA Viruses, Virology Department, Institut Pasteur, Paris, France

Suggested citation: European Centre for Disease Prevention and Control. European SARS-CoV-2 and influenza Bioinformatics External Quality Assessment (ESIB-EQA), 2023. Stockholm: ECDC; 2024.

Stockholm, November 2024

ISBN 978-92-9498-753-2
doi: 10.2900/6481347
Catalogue number TQ-01-24-009-EN-N

Contents

Figures

Tables

Abbreviations

CDS	Coding Sequence.
COVID-19	COronaVIrus Disease of 2019
CS	comparator sequence
EQA	External Quality Assessment
GR	Wuhan-1 global reference
I	Intermediate reduction in drug susceptibility
Indel	A mutation with the mixture of insertion and deletion
NA	Influenza neuraminidase protein
ORF	Open Reading Frame
PA	One of three influenza virus polymerase proteins
PB1	One of three influenza virus polymerase proteins
PB2	One of three influenza virus polymerase proteins
R	High-level resistance to antiviral drugs
S	Sensitive to antiviral drugs
S-gene	The spike glycoprotein gene of SARS-CoV-2 virus
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SNP	Single nucleotide polymorphism.
VIC	Influenza type B Victoria lineage
YAM	Influenza type B Yamagata lineage

Executive summary

SARS-CoV-2 and influenza viruses are pathogens of major public health impact. Surveillance has played a vital role in monitoring the COVID-19 pandemic as well as seasonal influenza, including public health actions such as updates on vaccines and antiviral drugs.

The objective of this External Quality Assessment (EQA) is to strengthen the capacity for genomic epidemiology and public health bioinformatics, which is crucial for response during a pandemic or unexpected major public health event.

The EQA was divided into SARS-CoV-2 and influenza and was composed of different components that could be performed individually, together representing both short read and long read sequencing technology, and included the bioinformatics processes relevant for public health activities, i.e. consensus sequence generation and quality control, clustering and classification of genomes and mutation analysis for prediction of phenotypic properties. Laboratories from the 27 European Union (EU) and the three European Economic Area (EEA) countries as well as the six Western Balkan countries and Türkiye were invited to participate. The EQA was open between 1 March and 1 April 2023, and the correctness of the responses and throughput time was assessed.

The EQA had participation from 25 of 30 EU/EEA countries and two of six Western Balkan countries. The components they participated in varied. The quartile of the adjusted throughput time for these four components, indicative of the more performant laboratories, was in the range of one to two working days. A large majority of the laboratories correctly classified the samples according to their expected quality status. For the consensus generation of SARS-CoV-2 genomes, issues were mostly related to insertions or deletions ('indels') in comparison to the comparator sequence and for influenza, the issues were mostly related to that consensus sequences were not generated in their full length and primer sequences were not considered artificial sequences. For the genome clustering component, a process important for outbreak investigations and for following virus evolution, laboratories performed well. For both SARS-CoV-2 and influenza, the most common discrepancies were due to using different thresholds of cluster definitions (number of mutations). The mutation analysis component for SARS-CoV-2 was in general accurate, however for influenza there was significant fraction of mutation-based misclassifications of resistance to antiviral drugs oseltamivir, zanamivir and baloxavir marboxil.

The issues identified in this EQA may facilitate several improvements in the laboratories' bioinformatic processes. The example of indels that are either falsely assigned or missed may lead to incorrect variant classification or delayed detection of an emerging variant or that cases are not assigned to an outbreak which they belong to. Furthermore, for influenza, standardisation of parameters such as flanking regions, primer sequence masking and antiviral susceptibility predictions can facilitate inter-laboratory comparisons. Novel viruses provided as educational samples that does not affect the score may improve detecting reassortment and determination of subtype and origin.

This EQA establishes a baseline that could be used for future comparisons. The participants received individual feedback with suggestions for improvement and laboratories from three countries were invited to and completed a five-day twinning training activity. Furthermore, general recommendations are compiled in this report in order to improve the laboratory practices and performance.

1 Introduction

SARS-CoV-2 and influenza viruses are pathogens of major public health impact [1-4]. These viruses pose a significant threat to global health, but their impact varies in terms of transmission, severity, and response measures. SARS-CoV-2, the virus responsible for the COVID-19 pandemic, has demonstrated high transmissibility, leading to widespread outbreaks and a significant burden on healthcare systems worldwide. As of yet, there is no evidence of a seasonal pattern for SARS-CoV-2, instead recurring waves such as those caused by emergence of variants with major impact such as Delta (https://www.ecdc.europa.eu/en/publications-data/threat-assessment-emergence-and-impact-sars-cov-2-delta-variant) and Omicron (https://www.ecdc.europa.eu/en/publications-data/covid-19-assessment-further-emergence-omicron-18th-risk-assessment). In contrast, influenza is responsible for seasonal increased burdens and, in some cases, severe epidemics. While the impact of influenza and SARS-CoV-2 on populations are mitigated by vaccine development efforts, the continuous evolution of both viruses poses challenges for effective prevention and control efforts. Understanding the public health impact of both SARS-CoV-2 and influenza is crucial for informing policy decisions, resource allocation, and the development of effective strategies to mitigate their respective impacts on population health.

In this area, surveillance plays a vital role to monitor the spread and impact of SARS-CoV-2 and influenza [5]. It involves the collection, analysis, and interpretation of data to inform public health actions. ECDC recognises the importance of disease surveillance and has set specific surveillance objectives to address these pathogens [6]:

1. Monitor the intensity, geographical spread and temporal patterns of influenza, COVID-19, and other respiratory virus infections to inform mitigation measures.
2. Monitor severity, risk factors for severe disease, and assess the impact on healthcare systems of influenza, COVID-19, and other respiratory virus infections to inform mitigation measures.
3. Monitor changes and characteristics of circulating and emerging respiratory viruses, particularly virological changes of influenza viruses, SARS-CoV-2, and other respiratory viruses to inform treatment, drug, and vaccine development.
4. Describe the burden of disease associated with influenza, COVID-19, and other respiratory virus infections.
5. Assess vaccine effectiveness against influenza, COVID-19, and other respiratory virus infections where locally feasible.

EQA programs are designed to assess and improve the performance of laboratories involved in disease surveillance. These programs have specific objectives that align with and contribute to ECDC's surveillance objectives. The primary objectives of EQAs include evaluating the accuracy and reliability of laboratory data analysis methods and identifying areas for improvement. By ensuring standardised and high-quality laboratory testing, EQA programs aim to optimise laboratory performance and ultimately enhance the overall surveillance efforts. To do so, one objective of this Framework Contract is to provide laboratory support for surveillance, preparedness and response to COVID-19 and influenza, and to specifically strengthen the capacity for genomic epidemiology and public health bioinformatics through interdisciplinary trainings, which is crucial for response during a pandemic or unexpected major public health events. The geographic scope of this Specific Contract covers European Union/European Economic Area (EU/EEA) countries, the Western Balkans (Albania, Kosovo*, Montenegro, Serbia, North Macedonia, Bosnia and Herzegovina), and Türkiye.

In order to effectively monitor and respond to the public health impact of SARS-CoV-2 and influenza, a robust laboratory network is crucial. ECDC coordinates the European COVID-19 Reference Laboratory Network (ECOVID-LabNet) and the European Reference Laboratory Network for Human Influenza (ERLI-Net), both with the aim to carry out virological surveillance and strengthen laboratory capacity and foster data-sharing, for example through EQAs (https://www.ecdc.europa.eu/en/about-ecdc/what-we-do/partners-and-networks/disease-and-laboratory-networks/european-covid-19; https://www.ecdc.europa.eu/en/about-us/networks/disease-and-laboratory-networks/erlinet-about). These networks comprise laboratories across Europe, which play a pivotal role in supporting disease surveillance by contributing to the detection, confirmation, and characterisation of cases. Through their coordinated efforts, this laboratory networks strengthen the surveillance infrastructure and enables the generation of accurate and reliable data for effective public health response.

The main objectives of this bioinformatics EQA are (i) to assess the capability of participating laboratories for public health bioinformatics for SARS-CoV-2 and influenza; and (ii) to provide troubleshooting advice where needed and possible.

* This designation is without prejudice to positions on status, and is in line with UNSCR 1244/1999 and the ICJ Opinion on the Kosovo declaration of independence.

This bioinformatics EQA consists of eight different components designed to each independently test one particular type of analysis for both SARS-CoV-2 and influenza virus, as described in Table 1. The influenza virus components were managed by Institut Pasteur and the SARS-CoV-2 components by the Dutch National Institute for Public Health and the Environment (RIVM).

Table 1. Independent EQA components to which the laboratories could choose to participate

Name	Organism	Description
INFL1	Influenza virus	Consensus sequence generation from complete amplicons based on Nanopore reads.
SARS1	SARS-CoV-2	Consensus sequence generation from complete amplicons based on Nanopore reads.
INFL2	Influenza virus	Consensus sequence generation from fragmented amplicons based on Illumina reads.
SARS2	SARS-CoV-2	Consensus sequence generation from fragmented amplicons based on Illumina reads.
INFL3	Influenza virus	Clustering and classification of full-length genomes.
SARS3	SARS-CoV-2	Clustering and classification of full-length genomes.
INFL4	Influenza virus	Prediction of reduced susceptibility to antivirals.
SARS4	SARS-CoV-2	Detection of particular amino-acid substitutions.

2 Study design and methods

Participants and process

The invitation for the EQA was sent to ECDC's Operational Contact Points for COVID-19, for influenza and for bioinformatics, with ECDC's National Focal Points for Viral Respiratory Diseases and National Microbiology Focal Points in copy. These contact points cover the 27 European Union countries as well as the three European Economic Area countries. In addition, the Western Balkan countries and Türkiye were contacted by ECDC's Country Cooperation Office.

Each component had its own dataset consisting of a number of samples that could be downloaded by participants, processed, and the results provided back in an Excel template. In addition to the assessment of the particular type of analysis for which the component was designed, two general aspects were assessed as well for each component: quality control of the input data for the analysis, and the throughput time of the analysis. For the SARS1, SARS2 and SARS4, as well as for the INFL1 and INFL2 components, some educational samples were included with issues that were not considered part of the primary objective but nonetheless relevant. The methodology used for all these aspects is described in the sections below.

For each dataset, a comparator sequence was prepared. All reported differences versus the comparator sequence of good quality samples were manually investigated to determine the reason. Where the difference was found to be due to an issue in the comparator sequence, the latter was adjusted, and the determination of differences was re-calculated. In practice, this occurred in a very small number of cases. The discrepancies were aggregated and compared to the methodology provided by the laboratory, to try to identify concrete methodological issues such as inappropriate parameter settings. The educational samples were, where relevant, assessed separately. No criteria were applied to classify the laboratory as having passed or failed each EQA component. Instead, the focus was on providing useful information for improvement, by highlighting potential issues and potential causes.

After the period allotted to laboratories to provide their results, laboratories received an individual report with their results assessed and compared to those of other laboratories in an anonymised form. Laboratories were subsequently given the opportunity to discuss the report with the organisers, and finally feedback could be given through an optionally anonymous questionnaire.

Throughput time

The total throughput time needed to process the component was measured for educational purposes only, from the moment the password was provided to decrypt the dataset for a particular component until the moment the results for that component were submitted. An adjusted version of the throughput time was calculated to better reflect actual throughput time by subtracting non-working hours from the total throughput time. The latter included weekend days as a whole and outside 8 am to 6 pm on working days. In addition, when passwords were provided and/or results submitted for two or more components at the same time, these were generally excluded from statistics on adjusted throughput time.

Quality control

Quality control is an essential element of each analysis. Performing the analysis in question on sequence reads with bad quality can often lead to misleading results, such as missing mutations or unjustified clusters. In addition, the results from processed sequence reads can also be of bad quality, either due to the input data or the algorithm used and would as such affect any downstream analyses. Each dataset therefore contained some clearly bad quality samples, and participants needed to indicate for each sample whether the quality was either OK or BAD.

A distinction was made between overall and specific quality. Overall quality considered the entire genome and therefore the usability for analyses that take the entire genome as input, such as clustering or classification. Specific quality only considered a specific part of the genome that is used for a more targeted analysis, such as detection of mutations in a particular gene. The assessment of the differences in quality control versus the comparator sequence was performed identically for both types.

For the SARS1/2 components, the overall quality control needed to be carried out on the output genome sequences. For INFL1/2 components, the overall quality control mainly needed to be carried out on HA and NA segments, although the rejection of a sample by a laboratory for the poor quality of other segments was allowed. For SARS3 and INFL3, the overall quality control needed to be carried out on the input genome sequences. For SARS4 and INFL4, the specific quality control needed to be carried out on, respectively, the S-gene sequence, and the HA, NA and PA input gene sequences.

Consensus sequence generation

The purpose of the SARS1/2 and INFL1/2 components is to assess the generation of full-length SARS-CoV-2 and influenza virus genomes from amplicons sequenced by respectively the Nanopore platform (SARS1 and INFL1 components) and the Illumina platform (SARS2 and INFL2 components), by comparing them with a comparator sequence. To this end, a strict and a less strict consensus genome was generated from the reads for each sample in the respective datasets, corresponding to a high and a low coverage threshold.

For SARS-CoV-2, this was carried out using the ViroConstrictor pipeline v1.1.0 (https://github.com/RIVM-bioinformatics/ViroConstrictor) with a modified version of TrueConsense specific for this EQA. The Nanopore reads were created with R9.4.1 flowcells and base calling was done with the High-Accuracy preset of Guppy, the high and low coverage thresholds were set to respectively 50x and 20x, and for Illumina reads to 20x and 10x. Ambiguity symbols representing two nucleotides were put when coverage was sufficiently high and the difference between the most frequent and the second most frequent nucleotide was at most 10% or 20% for respectively the strict and the less strict consensus genome. Similarly, for influenza virus a dedicated pipeline was used with coverage thresholds set to respectively 50x and 20x for both types of reads. Ambiguity symbols were assigned when the difference between the most frequent and the second most frequent nucleotide was at most 10% for the strict and the less strict consensus genomes.

Low coverage regions were assigned Ns for both types of genomes and subsequently manually curated by a person doing this curation routinely for respectively SARS-CoV-2 and influenza national surveillance, and a comparator sequence was created from that which allows for each sequence position both the strict and the less strict version. As such, an acceptable range of variation in the consensus sequence was created to reflect the fact that there are no exact criteria for defining correctness of a sequence in routine application. In addition, the sequence regions upstream and downstream the first/last primer were set to allow both N and gaps, and positions where an ambiguity symbol (not N) was allowed, were also set to allow N.

For influenza, this was carried out with the internal routine workflow for Illumina, and a combination of IRMA (https://wonder.cdc.gov/amd/flu/irma) and wf-flu (https://github.com/epi2me-labs/wf-flu) for Nanopore. A manual curation was also performed on all consensus sequences.

The genome sequences returned by the participants were compared with the comparator sequence and differences found were classified into eight types of differences listed in 2. The SARS1, SARS2, INFL1 and INFL2 datasets each consisted of 32 samples that were selected as described in Tables A2, A3, A6, and A7.

Table 2. Types of differences versus the comparator sequence consensus sequence that were individually assessed

The types of differences are sorted in decreasing order of overall impact on the quality.

Type of difference	Description
Insertion	An insertion was introduced versus the comparator sequence. The entire insertion is counted as a single difference.
Deletion	A deletion was introduced versus the comparator sequence. The entire deletion is counted as a single difference.
Wrong nucleotide	A different nucleotide or ambiguity symbol (other than N) than the one allowed was provided, or, in case an ambiguity symbol (not N) was allowed, the provided nucleotide is not represented by it.
Nucleotide instead of ambiguity	An actual nucleotide was provided where only an ambiguity symbol (other than N) was allowed.
Ambiguity instead of nucleotide	An ambiguity symbol (other than N) was provided where only an actual nucleotide was allowed.
Nucleotide stretch instead of stretch of Ns	A stretch of actual nucleotides or ambiguity (other than N) symbols was provided for a particular region (including a region of length 1), where only Ns were allowed. The entire stretch is counted as a single difference.
Stretch of Ns instead of nucleotide stretch	A stretch of Ns was provided for a particular region (including a region of length 1), where only actual nucleotides or ambiguity (other than N) symbols were allowed. The entire stretch is counted as a single difference.
Nucleotide stretch outside amplicon range	A stretch of actual nucleotides or ambiguity (other than N) symbols was provided before the first amplicon or after the last amplicon, instead of Ns or gaps. The entire stretch is counted as a single difference.

Clustering and classification

The purpose of the two components, SARS3 and INFL3 on clustering and classification, was to assess the ability to identify clusters of samples that are phylogenetically so similar that they are likely part of a very short transmission chain – less than two weeks – and as such could be used as a basis for initiating contact tracing. In addition, the ability to perform classification at sub-species level was assessed. In the absence of official recommendation for such investigation, comparator sequence clusters were defined for SARS-CoV-2 using a cut-off of a single mutation (substitution or indel).

For influenza virus, clusters were here defined using a cut-off of one to three mutations (substitution or indel) for each influenza segment. Participant-provided cluster codes were mapped in all possible combinations to the comparator sequence cluster codes, first removing singleton clusters. The combination that scored best, i.e. had the least number of cluster assignment mismatches, was used for further assessment.

The ability to perform classification at sub-species level, i.e. subtyping, was assessed as good. For SARS-CoV-2, the assignment of Pango lineages (https://cov-lineages.org/resources/pangolin.html) was assessed and for influenza virus the assignment of clades. Comparator sequence Pango lineages were assigned with Pangolin version of 4.1.3 of 2022-10-11 (https://github.com/cov-lineages/pangolin) and Pangolin data versions 1.15.1 of 2022-10-10 and 1.17.1 of 2022-12-12. For influenza virus, clades were assigned using NextClade version 2.13.0 (https://clades.nextstrain.org). In both cases, provided values were compared with comparator sequence values on the basis of identical matches.

The SARS3 and INFL3 datasets each consisted of 20 samples that were selected, as described in Tables A4 and A8. The annotated phylogenetic tree for the SARS3 component, limited to OK quality sequences, is shown in Figure 1. This tree was made using IQ_TREE COVID-19 release v2.2.0 (http://www.iqtree.org). The annotated phylogenetic trees for the INFL3 component, limited to good quality sequences, are shown in Figures 2, 3, and 4. These trees were made using the Nextstrain pipeline (https://github.com/nextstrain).

Figure 1. Phylogenetic tree of SARS3 samples, excluding bad quality samples

Samples are annotated with their Pango lineage and the cluster to which they belong, if any. Borderline cases are highlighted in green (included in a cluster) and red (not included in a cluster).

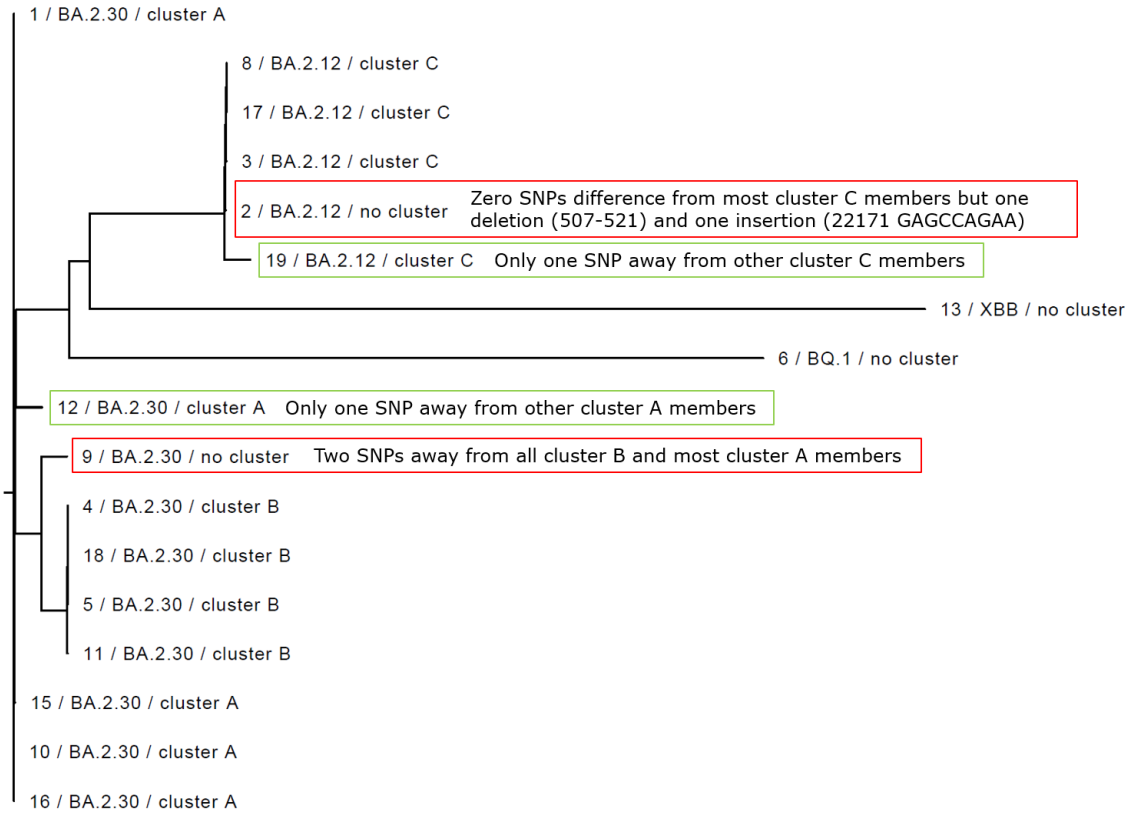


Figure 2. Phylogenetic tree of HA segment for INFL3 A(H1N1)pdm09 samples

Green coloured sample names are those belonging to cluster A, red indicates BAD quality and black the others. Clades are indicated on the branches of the tree.

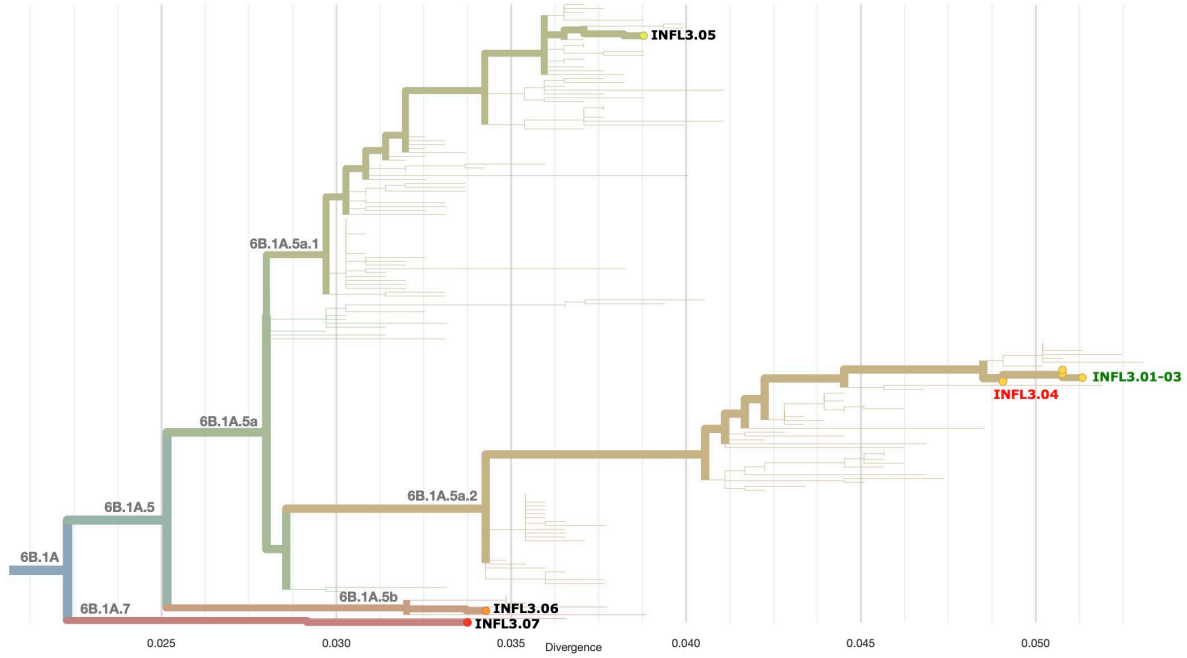


Figure 3. Phylogenetic tree of HA segment for INFL3 B/Victoria-lineage samples

Green coloured sample names are those belonging to cluster B, red indicates BAD quality and black the others. Clades are indicated on the branches of the tree.

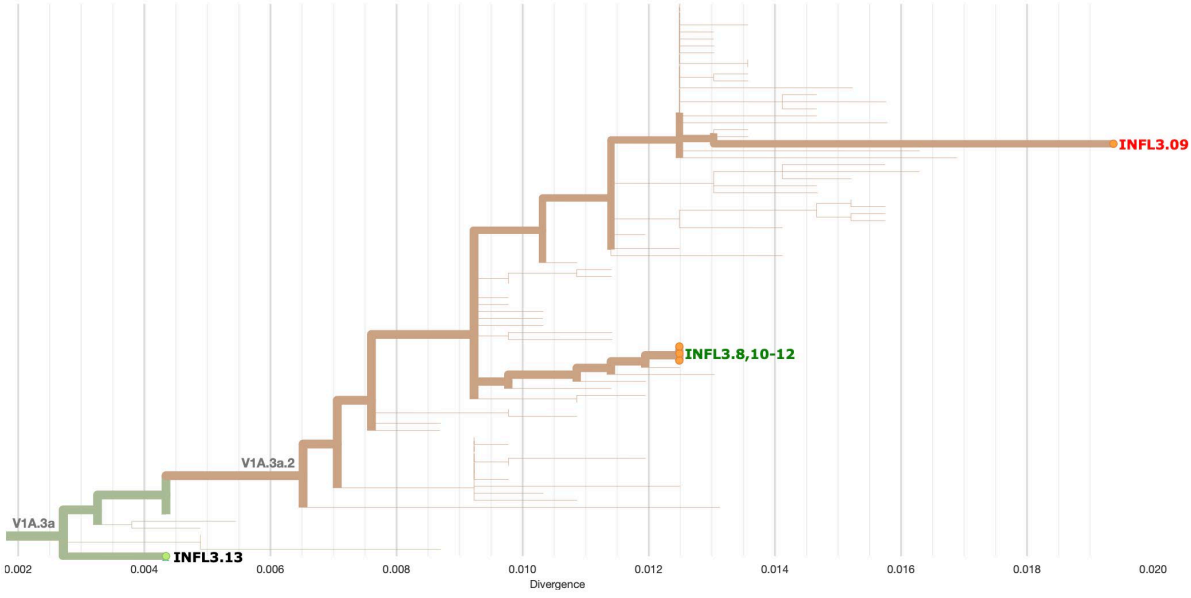
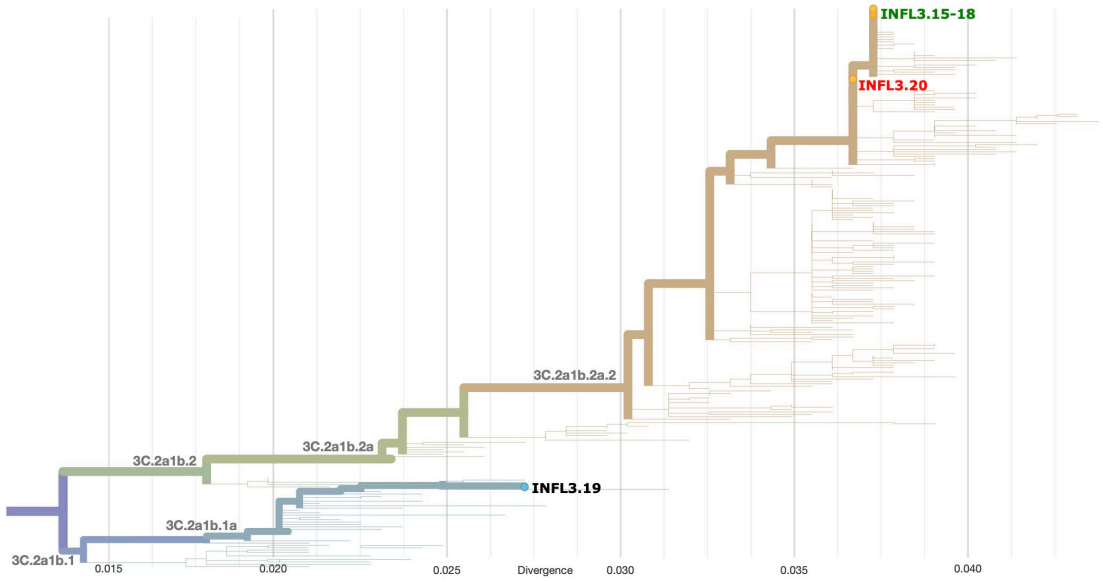


Figure 4. Phylogenetic tree of HA segment for INFL3 A(H3N2) samples

Green coloured sample names are those belonging to cluster C, red indicates BAD quality and black the others. Clades are indicated on the branches of the tree.



SARS-CoV-2 mutation detection

The purpose of this component, SARS4, was to assess the enumeration of amino acid substitutions present in a sample that has known phenotypic associations. The list of mutations to be detected for different Pango lineages was provided as part of the dataset for the component and was based on ECDC's list of variants of concern (including de-escalated ones) (https://www.ecdc.europa.eu/en/covid-19/variants-concern). Additional mutations present in the sample were also allowed, while any non-existing mutations or missed mutations compared to the list were considered a difference.

The S-gene mutations to be detected for different lineages are given in Table 3. The SARS4 dataset consisted of 10 samples that were selected as described in Table A5.

Table 3. SARS-CoV-2 amino acid substitutions in the S-gene (S) that should be detected for the SARS4 component, by lineage

Lineage	Amino acid substitutions
BA.4	S:L452R, S:F486V, S:R493Q
BA.5	S:L452R, S:F486V, S:R493Q
BA.2.75	S:W152R, S:F157L, S:I210V, S:G257S, S:D339H, S:G446S, S:N460K
BQ.1	S:K444T, S:N460K
XBB	S:N460K, S:F490S
XBB.1.5	S:N460K, S:S486P, S:F490S
BA.2.3.20	S:K444R, S:L452M, S:N460K
BF.7	S:R346T, S:F486V
XBC	S:N440K, S:F486P
BN.1	S:R346T, S:K356T, S:F490S
CH.1.1	S:K444T, S:L452R

Prediction of genotypic influenza virus susceptibility to selected antivirals

The purpose of this component, INFL4, was to assess the identification of amino acid substitutions present in a sample and that have known phenotypic associations with Oseltamivir, Zanamivir (NA segment) or Baloxavir (PA segment) susceptibility. The list of mutations to be detected for different clades is shown in Table A9 and was provided as part of the dataset for the component, and was based on the WHO lists of drug resistance mutations for NA (https://www.who.int/publications/m/item/summary-of-neuraminidase-(na)-amino-acid-substitutions-associated-with-reduced-inhibition-by-neuraminidase-inhibitors-(nais) and PA (https://www.who.int/publications/m/item/summary-of-polymerase-acidic-(pa)-protein-amino-acid-substitutions-analysed-for-their-effects-on-baloxavir-susceptibility) segments. Participants were asked to register R for high-level resistance, I for intermediate reduction in drug susceptibility or S for sensitive to antiviral drugs. R and I were accepted as resistance which, in WHO's list of drug resistance mutations, are equivalent to RI or HRI. This latter international terminology will be used in further exercises, and the mutations identified by the participants will be asked.

3 Results

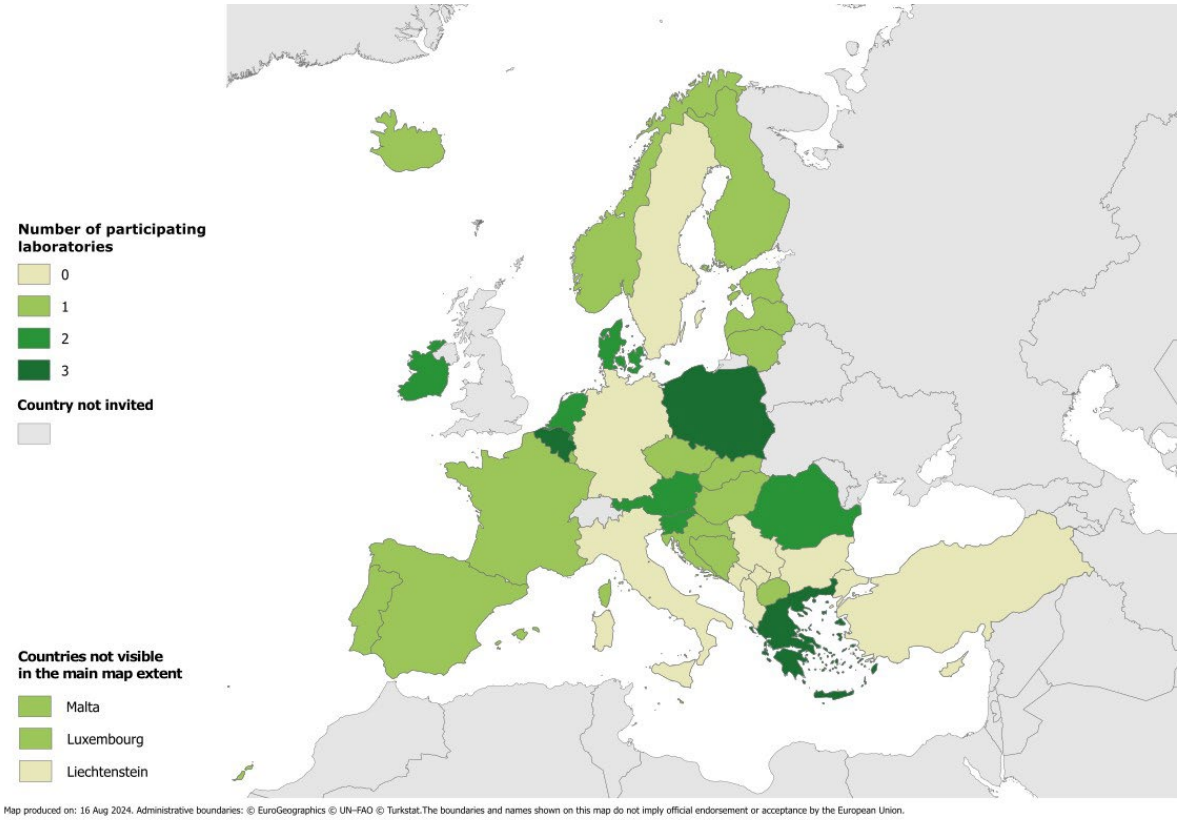
Participation

The EQA was open between 1 March and 1 April 2023 and had participation from 25 of 30 EU/EEA member countries and two of six Western Balkan countries. A total of 38 laboratories from 26 countries confirmed their participation in one or more of the eight components (Table 4, Figure 5). A total of 20 laboratories participated in both SARS1 and SARS2 (i.e. performing the same task from different sequencing platforms), whereas 16 participated in only one of these two components. For INFL1 and INFL2, these numbers were, respectively, 13 and 16 laboratories. The list of participating laboratories is provided in Table A1.

Table 4. Participation and completion of individual components by all laboratories that participated in at least one component

Status	SARS1	SARS2	SARS3	SARS4	INFL1	INFL2	INFL3	INFL4
Completed	25	26	33	32	18	22	29	26
Participated but not completed	1	4	2	2	1	2	0	0
Not participating	14	8	3	4	21	16	10	13

Figure 5. Number of participating laboratories in the EU/EEA, the Western Balkans and Türkiye

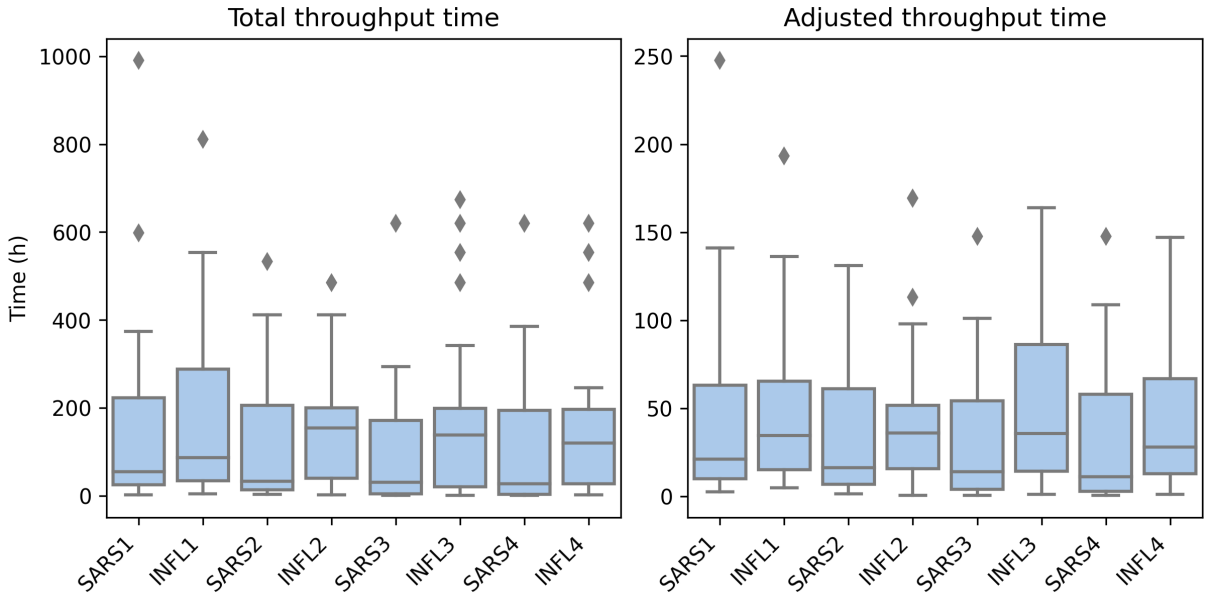


Throughput time

The total and adjusted throughput times are shown in Figure 6. The median adjusted throughput time for the INFL components was higher for all components compared to those of SARS2 despite equal number of samples, while the inter-quartile range shows less variation. For the consensus sequence generation components, the median ratio between Nanopore reads (SARS1, INFL1) and Illumina reads (SARS2, INFL2) adjusted throughput time was 1.03 (IQR 0.69-1.65), indicating very limited impact of the platform technology on throughput time. The quartile of the adjusted throughput time for these four components, indicative of the more performant laboratories, ranged from 7.0 to 15.8 hours or one to two working days.

Figure 6. Total and adjusted throughput time for the different components

Distributions are represented as box plots with whiskers at max. 1.5x the inter-quartile range.



Quality control

A large majority of the laboratories correctly classified the samples according to their quality, including a small fraction with 100% correct classification for a specific component. The number of samples for which the quality assessment was identical or different between comparator sequence and provided value is given in Table 5. For the BAD/OK combination, i.e. where the sample was classified as bad quality in the comparator sequence but not by the laboratory, respectively 37/51 (SARS1) and 78/81 (SARS2) occurred for samples where a co-infection with distinct variants was observed, and which was most likely either not detected by the laboratory or not considered as bad quality. For SARS1, another 12/51 were most likely due to having too high a threshold on the maximum number of missing nucleotides (N's). The remaining five BAD/OK samples for SARS1 and SARS2 were due to a misunderstanding by the laboratory on how to assign the quality control result. For the OK/BAD combination, i.e. where the sample was classified as good quality in the comparator sequence but not by the laboratory, 5/12 (8 SARS1 and 4 SARS2) occurred for samples where most likely the presence of a true frameshift, a shifted start codon or a true large deletion led to a classification as bad quality. Another 3/12 occurred for samples where most likely a too low threshold on the maximum number of missing nucleotides (N's) was used, with the remainder unclear.

For INFL1 and INFL2 the differences in terms of quality assessment were generally due to distinct criteria on minimal segment coverage or read quality. For INFL1/2, six samples had low quality PA, PB1 and PB2 data (detected by a drop of coverage in the middle of segments), leading to rejection of the whole sample by some laboratories. More generally, several samples were rejected by laboratories because of low coverage on segments other than HA or NA. INFL1.09, a 'non-pandemic' A(H1N1) close to the A/Puerto Rico/38 strain was often rejected because too distant from the A(H1N1)pdm09 clade. Some laboratories did not analyse B samples due to a lack of solution to compute consensus sequences for it. Finally, a few laboratories did not give any quality control values or provided an 'OK' value for all samples.

Table 5. Number of samples for which the quality assessment was identical (OK/OK or BAD/BAD) between comparator sequence/laboratory provided value, and number of samples where the assessment was different (OK/BAD or BAD/OK)

The latter two should ideally have zero samples. Educational samples are excluded.

Component	Correct classification		Incorrect classification	
	OK/OK	BAD/BAD	OK/BAD	BAD/OK
SARS1	592	49	8	51
SARS2	620	49	4	81
SARS3	561	85	0	14
SARS4	252	31	4	1
INFL1	451	66	6	46
INFL2	646	49	23	50
INFL3	509	87	1	3
INFL4	234	27	0	0

Consensus sequence generation

SARS1 and SARS2 components

Of 26 laboratories that participated in SARS1, 25 completed the component and for SARS2, 26 of 30 completed the component.

The average number of sequence differences per good-quality non-educational sample, stratified by the type of difference, is shown in Figure 7. See Table 2 for a description of the types of differences that were distinguished between. Across all laboratories and all good quality samples, 2 990 (SARS1) and 3 699 (SARS2) differences were found. These differences are summarised by type in Table 6, and subdivided by potential cause. It should be noted that these differences, including insertions and deletions (indels), are with respect to the comparator sequence (CS) rather than the Wuhan-1 global reference (GR). An insertion in a provided sequence compared to the comparator sequence can therefore be at the same locus as where the GR has an insertion compared to the comparator sequence, or at a different locus. The same applies to deletions compared to the comparator sequence.

Mistakes in the calling of insertions and deletions are sometimes reported as root causes of other differences. These types of mistakes frequently lead to regions being called as N, missing regions being called as nucleotides and wrongly called nucleotides.

Table 6. Number and type of sequence differences versus the comparator sequence for SARS1 and SARS2, excluding bad quality sequences and including educational samples

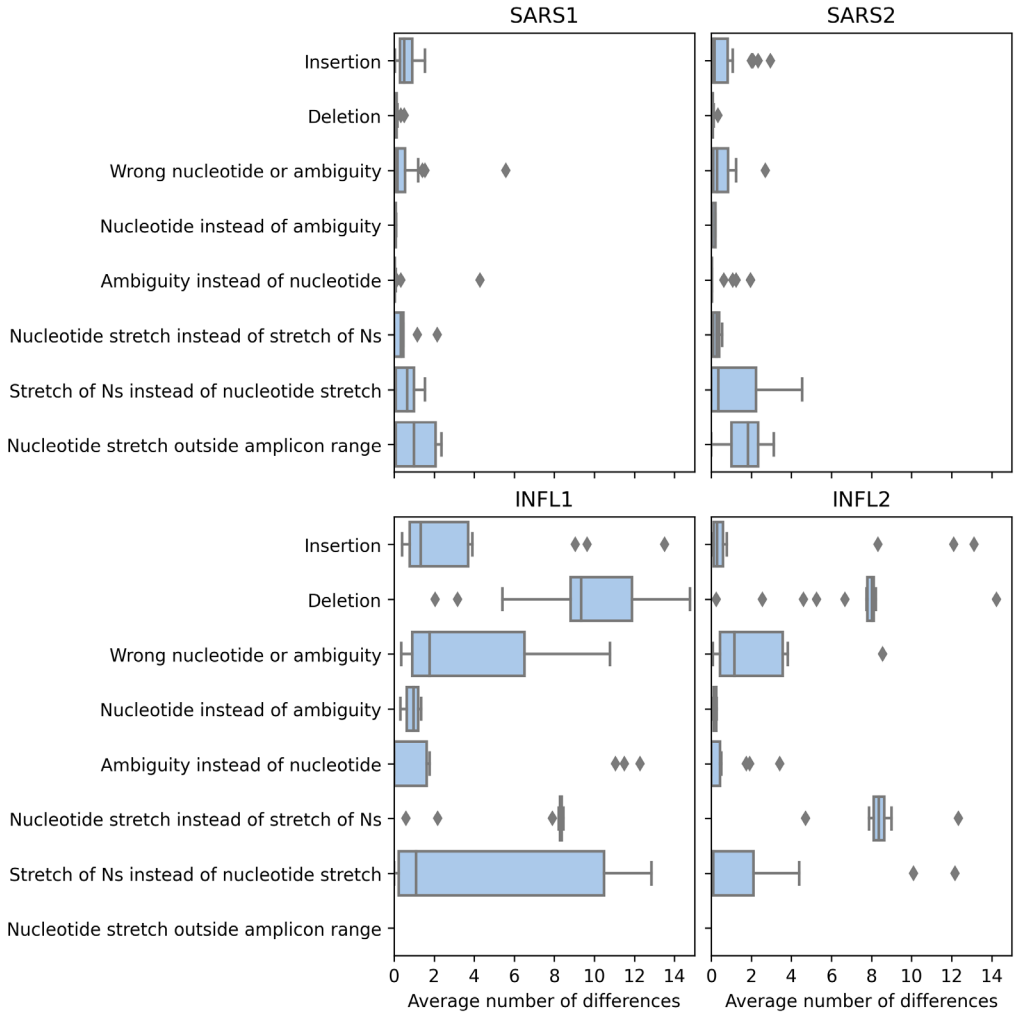
The main types are those described in Table 2. Subtypes are given when relevant, or when the cause of the difference is evident. Values for each subtype are expressed as the number and proportion (%) of the differences of that subtype within that type. The totals per type are shown in the header row for that type.

Type and subtype of sequence difference versus comparator sequence	SARS1	SARS2
Insertion	n=360	n=410
Same locus as GR insertion vs CS, same length, and put as stretch of N's	233 (64.7)	112 (27.3)
Same locus as GR insertion vs CS, same length, and put as GR nucleotides	87 (24.2)	227 (55.4)
Same locus as GR insertion vs CS, but shorter length	14 (3.9)	17 (4.1)
Other locus than GR insertion vs CS, but not outside amplicon range	18 (5.0)	44 (10.7)
Extra nucleotides or N's outside amplicon range compared to GR[1]	8 (2.2)	10 (2.4)
Deletion	n=52	n=27
Homopolymer-related	22 (42.3)	0 (0.0)
Single-nucleotide in structurally critical Open Reading Frame (ORF)	8 (15.4)	0 (0.0)
Same locus as GR deletion vs CS, but longer length	7 (13.5)	0 (0.0)
Insertion not called (missing calls compared to CS counted as deletion)	3 (5.8)	7 (25.9)
Missing nucleotides outside amplicon range compared to GR[1]	3 (5.8)	9 (33.3)
Missing deletions in low coverage regions	5 (9.6)	0 (0.0)
Missing deletions in or near primer regions	4 (7.7)	0 (0.0)
No distinguishable pattern	0 (0.0)	11 (40.7)
Wrong nucleotide	n=401	n=340
In or near primer regions	175 (43.6)	114 (33.5)
In or near a differentiating primer scheme region[2]	46 (11.5)	0 (0.0)
In difficult to call region[4]	14 (3.5)	25 (7.4)
Near an incorrect insertion	7 (1.7)	0 (0.0)
Mutation of CS vs GR at last nucleotide of last amplicon	10 (2.5)	0 (0.0)
Called minority nucleotide on ambiguous position	0 (0.0)	34 (10.0)
Called minority or reference nucleotide on non-ambiguous position	0 (0.0)	104 (30.6)
No distinguishable pattern	149 (37.2)	63 (18.5)
Nucleotide instead of ambiguity	n=37	n=88
In or near primer regions	22 (59.5)	35 (39.8)
No distinguishable pattern	15 (40.5)	53 (60.2)
Ambiguity instead of nucleotide	n=127	n=798
In or near primer regions	107 (84.3)	81 (10.2)
Very low threshold for calling ambiguity	0 (0.0)	133 (16.7)
No distinguishable pattern	20 (15.7)	584 (73.2)
Nucleotide stretch instead of stretch of Ns	n=304	n=148
In or near primer regions	11 (3.6)	0 (0.0)
Nucleotides in no-coverage or low-coverage regions	28 (9.2)	0 (0.0)
Nucleotides called in region with ambiguous deletion[5]	20 (6.6)	0 (0.0)
Nucleotides called instead of N	245 (80.6)	148 (100.0)
Stretch of Ns instead of nucleotide stretch	n=1088	n=732
At position with low frequency minority variant[3]	540 (49.6)	88 (12.0)
Differentiating primer scheme[2]	197 (18.1)	18 (2.5)
In or near primer regions	0 (0.0)	115 (15.7)
Coverage threshold	0 (0.0)	304 (41.5)
No distinguishable pattern	351 (32.3)	207 (28.3)
Nucleotide stretch outside amplicon range	621	1 156
One or more stretches of nucleotides at start or end of genome	621 (100.0)	1 156 (100.0)

CS: comparator sequence; GR: global reference (Wuhan-Hu-1). [1]Missing and extra nucleotides compared to the GR sequence length outside the amplicon range. [2]Differentiating primer scheme refers to the educational samples that were processed with a different primer scheme than the other samples. This corresponds to samples 7, 13 in SARS1 and sample 13 in SARS2 (Table A2 and A3). [3]For SARS1, a single laboratory accounts for 509/540 cases. [4]These regions can probably only be resolved by manual curation. [5]These regions had a deletion in approximately 50% of the reads, so an N was expected.

Figure 7. Average number of sequence differences per sample for all good quality, non-educational samples as compared to the comparator sequence

Results are stratified by the type of difference. Distributions are represented as box plots with whiskers at max. 1.5x the inter-quartile range.



INFL1 and INFL2 components

Of 19 laboratories that participated in INFL1, 18 completed the component and for INFL2, 22 of 24 completed the component.

The main sources of differences between consensus sequences and comparator sequences were found to be the followings:

1. Lab consensuses were truncated, e.g. starting at the initial start codon (ATG) instead of the beginning of segment (due to usage of references consisting only of the coding sequence part). In that case, nucleotides before ATG could not be called and were counted as deletion compared to comparator sequence.
2. In other cases, primer matching sequences were not masked (in 5' and 3'), and thus counted as nucleotides instead of Ns when compared to comparator sequences.
3. In the manual curation phase, ambiguity in nucleotide calling not resolved the same way in comparator sequences and in the consensus. This case ended up with wrong nucleotides, ambiguity instead of nucleotide, or nucleotide instead of ambiguity.
4. Diversity of reference sequences and pipelines used for influenza components may explain the number of wrong nucleotides called. Some labs incorrectly defined PB1 as PB2 (and vice versa) for VIC and YAM samples.

- No intensive curation of viruses with evidence of defective genomes (educational samples). These cases generated many wrong nucleotides, because of the difficulty of calling nucleotides in the concerned regions. These cases are not shown in Figure 7.

Clustering and classification

For the SARS3 component, and across 33 laboratories that completed the component (of 35 participants), a total of 442/561 (78.8%) good quality sequences were correctly assigned, either to the correct cluster or as not belonging to any cluster. The 119 different assignments were inspected, and the likely cause determined (Table 7). Two thirds of these assignments were due to laboratories using a cluster cut-off that was one higher, i.e. two, or one lower, i.e. 0, than the cut-off of 1 used for the comparator sequence. In addition, two thirds of the laboratories likely did not take into account indels when performing the clustering, i.e. only considered substitutions. Finally, a small number of laboratories (n=5) included bad quality sequences in their analyses. With respect to the assignment of the Pango lineage, 8 (24.2%) laboratories likely used an outdated version of the software, as a result of which an ancestral, i.e. less high resolution, lineage was assigned. The remainder of the Pango lineage assignments were identical.

Table 7. Likely causes of different cluster assignments for the SARS3 component

Likely cause	Different assignments	Laboratories
Cluster cut-off of 2 instead of 1	66 (53.2%)	19 (57.6%)
Cluster cut-off of 0 instead of 1	17 (13.7%)	11 (33.3%)
Not taking into account indels	22 (17.7%)	22 (66.7%)
Too large cluster cut-off	10 (8.1%)	1 (3%)
Incorrect interpretation of tree	1 (0.8%)	1 (3%)
Use of bad quality sequence	5 (4%)	5 (15.2%)
Unknown	3 (2.4%)	1 (3%)

For the INFL3 component, and across 29 laboratories that completed the component (of 29 participating), a total of 95/600 (15.8%) good quality sequences were correctly assigned. Discrepancies between comparator sequence and laboratories can be explained by: (i) different definition of the cut-off between the comparator sequence and what laboratories consider as a cluster; and (ii) different ways of counting the number of mutations. As an example, some laboratories counted mutations in a concatenation of all segments, while the comparator sequence counted mutations segment by segment.

SARS-CoV-2 mutation detection

There were 32 of 34 participants that completed the SARS4 component where only 2/288 (0.7%) incorrectly detected substitutions in the S-gene, each by a different laboratory. In both cases, the amino acid position was correct but not the actual amino acid. The reason for these incorrect results is not clear.

Influenza virus susceptibility to antivirals prediction

With INFL4 (completed by all 26 participants), a total of 35 resistance assignments (13% of 270 total assignments, ranging from 0 to 3 depending on the labs) were assigned a different resistance to Oseltamivir than identified in the comparator sequence. 66 samples (24% of total assignments, ranging from 0 to 5 depending on the labs) were assigned a different resistance to Zanamivir than identified in the comparator sequence, and 59 (22% of total assignments, ranging from 0 to 4 depending on the labs) were assigned a different resistance to Baloxavir than identified in the comparator sequence.

4 Discussion

Throughput time

For consensus sequence generation (SARS1/2 and INFL1/2 components), very limited or no impact on the throughput time was observed of the type of reads, i.e. Illumina short reads or Nanopore long reads. This indicates that inherently there are no substantial technical differences in difficulty in processing between these types of reads. Of note, very few laboratories appear to have performed manual curation on samples with specific quality issues, and in our experience such time-consuming manual curation is required more for Nanopore reads. Nonetheless, the proportion of samples with specific quality issues is usually not as high as in this EQA, so the differential impact on throughput time of improving the quality through manual curation is likely not that high.

The time of one to two working days that the second quartile of the laboratories – i.e. not the fastest 25% but the subsequent 25% of laboratories – needed to complete one of these components would translate to four to eight working days for all four components. This can be substantial, although it should be kept in mind that these throughput times are upper estimates, measured from provision of the dataset until return of the results. The actual throughput time will be less than that. For the SARS3/4 and INFL3/4 components, it is less clear what might have driven the throughput time.

Quality control

For SARS-CoV-2 consensus sequences, most sequences incorrectly classified as OK quality were likely due to not properly taking into account the possibility of samples corresponding to co-infections with distinct variants. Such co-infections should lead to ambiguity symbols in the consensus sequence, and should subsequently be taken into account for the assessment of the quality. At the same time, co-infections are rare events and this dataset was enriched in them, so the overall impact of this issue is likely not very high.

For influenza sequences, the majority of sequences incorrectly classified as BAD were likely due to either (i) a misclassification of a 'non-pandemic' H1N1 sample; (ii) a rejection of samples because of low coverage in non-HA and non-NA segments (especially for the educational samples with evidence of defective genomes) or, less frequently; (iii) an impossibility to analyse samples belonging to B lineages. On the other hand, some samples were misclassified by some labs as OK while having a low coverage on HA and / or NA. For latter assignments, we could ask, in addition to the OverallQuality score, a SegmentScore (i.e. a quality control for each segment) for INFL datasets. This would better describe the reasons why some laboratories accepted or rejected a sample differently.

Consensus sequence generation

A total of eight types of differences were distinguished as part of the assessment (Table 2). The clearly wrong – and therefore critical – types of difference are insertions, deletions or a wrong nucleotide compared to the comparator sequence. The other types have less of an impact as they are not necessarily wrong or merely reduce the available information, and are therefore considered non-critical differences: nucleotide instead of ambiguity, ambiguity instead of nucleotide, nucleotide stretch instead of stretch of Ns, stretch of Ns instead of nucleotide stretch, nucleotide stretch outside amplicon range.

SARS1 and SARS2 components

Among the critical differences, considering only the OK quality non-educational samples, the pattern was very similar between Nanopore (SARS1) and Illumina (SARS2) reads. Insertions with respect to the comparator sequence occurred the most, and the large majority of them was likely due to over-reliance on the (global) reference sequence in reference-sequence based pipelines. Instead of correctly introducing a deletion versus the reference sequence, the reference sequence was either kept or multiple N characters were introduced, thereby introducing an insertion versus the actual comparator sequence sequence. Large deletions compared to the global reference, which can occur for example in the ORF7/8 region, are particularly hard to address with such pipelines because of the large gap that needs to be introduced in reads that span this locus when aligning them to the reference.

The same problem of over-reliance on the reference sequence occurred with deletions, i.e. where an insertion was not introduced although it was present in the comparator sequence. This problem of 'filling in' the global reference is one of the main issues that should clearly be improved about consensus sequence generation for SARS-CoV-2. Apart from this, wrong nucleotides occurred in about half of the cases in or around primer regions, with most of the other cases not having a distinguishable pattern.

The root cause of this problem is likely the absence of or incomplete trimming of primers from reads. As a result, when the primer has a difference compared to the actual sequence, the reference (primer) nucleotide is called, or an ambiguous nucleotide is called. When this happens, it is impossible for downstream processes, e.g. typing tools, to distinguish sequences on mutations in primer regions. This is a second main issue that should be improved about consensus generation for SARS-CoV-2.

For the non-critical differences, the pattern is also quite similar between Nanopore and Illumina reads. The most frequently occurring one is one or more nucleotide stretches outside the amplicon range, instead of putting Ns or gaps. Here as well, it is possible that when a global reference is used as part of the algorithm for consensus sequence generation, the global reference is filled in outside the amplicon range. Alternatively, it is also possible that the presence of reads from the original material, i.e. not from amplicons, leads to still being able to resolve (parts of) the sequence outside the amplicon range. The 'filling in' of the global reference nucleotides outside the reference range should in any case be avoided, since the actual sequence may have differences, and can likely be addressed in the pipelines used by the laboratories at the same time as the analogous issue with the critical differences.

The second most frequent non-critical difference was putting a stretch of Ns instead of nucleotides. For Nanopore reads, and excluding a single laboratory that accounted for half of the cases there, no likely cause could be identified. For Illumina reads on the other hand, about half of the cases could be attributed to the application of a too high a coverage threshold. Correcting this is likely only a question of parameterisation, by lowering the minimum coverage threshold to e.g. 10–20x. The opposite issue, where a nucleotide stretch was put instead of a stretch of Ns, occurred substantially less frequently. Finally, putting an ambiguity rather than a nucleotide while there is sufficient evidence for only one particular nucleotide, occurred more frequently for Illumina reads, however with no likely cause distinguishable.

INFL1 and INFL2 components

One of the main issues found in these components relate to missing nucleotides at the beginning and at the end of segments which, in most cases, corresponded to flanking and non-coding stretches in 5' or 3' regions; some laboratories produced segments so they began at the first start codon and ended at the last stop codon of the related genes. In addition, some other laboratories did not mask the primers at the beginning and the end of the segments. Both particularities explain the numbers of deletions and nucleotides instead of Ns.

Another main issue appeared with B and non-pandemic A(H1N1) lineages that were simply not analysed by some laboratories. As an example, one laboratory indicated that they do not process B lineage data and thus rejected Victoria and Yamagata samples from their analysis. Possibly because of a different nomenclature for PB1 and PB2 genes for A subtypes and B lineages, several laboratories swapped the two segments. While this is not a major issue, this specific example, together with the previous issues, showed a possible lack in the analysis of non-seasonal and less prevalent influenza lineages.

One last point is related to defective genomes, that are particularly difficult to analyse, and require time-consuming manual curation when detected. These samples are the source of many differences between comparator sequences and lab consensus.

Clustering and classification

For SARS-CoV-2 clustering, most of the discrepancies came from the use of a cluster cut-off that was one higher or one lower than the one used for the comparator sequence. The choice of cluster cut-off is debatable since no guidance on this exists, and as such we do not consider this as an issue that would need to be improved.

The main other issue was not taking into account the presence of indels in the clustering. This is an issue that needs to be improved since these evolutionary events are at least equally important to take into account, in particular when deciding on public health actions. For the SARS-CoV-2 Pango lineage classification, the main issue was that several laboratories did not use sufficiently up-to-date software or rather the continuously updated lineages definitions (pangolin-data), resulting in less precise lineage assignments. This issue should also be addressed.

For influenza virus, most of the discrepancies came from different thresholds of cluster definitions (number of mutations) between comparator sequence and laboratories. Another major issue, related to the previous one, concerns the genetic material used for the analysis. To build the comparator sequence, we performed the cluster analysis separately on each segment, in order to check for intra-segment evolution, but also to compare the phylogenetic placement of individuals on the segment basis and thus possible reassortments. Some laboratories likely preferred a more global approach by concatenating all segments. Both issues could be addressed, mainly by more specific guidelines on how to perform the exercise through a clearer definition of the cut-off and strategy of analysis.

SARS-CoV-2 mutation detection

This component did not reveal any particular issues that would need to be addressed.

Influenza virus susceptibility to antivirals prediction

One major issue was the definition of resistance with either I (RI) and R (HRI) levels. The component statement should have used directly the standard nomenclature, or, in a simpler manner, a binary nomenclature where samples are either resistant (RI / HRI) or susceptible (NI) to drugs. For this reason, samples where I was given instead of R (or vice-versa) were considered as correct, as if the reference level of resistance was given.

Feedback from laboratories

A total of ten laboratories provided feedback through the questionnaire, of which one did so anonymously. In nine out of 10 cases, the time period to complete all the components was deemed sufficiently long, with the one laboratory stating that they had other EQAs in parallel and that the number of samples in the components was too high for them. With respect to how satisfied laboratories were with the practical organisation of the EQA in general, regardless of the content of the different components, four answered Very satisfied, six Somewhat satisfied, and 0 answered Neither satisfied nor dissatisfied, Somewhat dissatisfied or Very dissatisfied. With respect to the usefulness of the individual EQA report, six answered Very satisfied, three Somewhat satisfied, one Neither satisfied nor dissatisfied, and 0 Somewhat dissatisfied or Very dissatisfied.

Apart from these fixed-choice answers, several laboratories provided more in-depth feedback. These were mainly about practical improvements to the instructions and reporting format on the one hand, and to the individual report on the other hand. In particular, instructions should be fully self-contained per pathogen, be clearer on how the sequence quality control should be carried out, and where applicable should clearly state the reference sequence used. Reporting of sequence data could be easier in Fasta format than in Excel, that can be more easily uploaded and analysed if the appropriate naming convention is adopted. For the individual report, consensus generation components, a distinction should be made between critical sequence differences and non-critical ones. The clustering methodology could be described more clearly, and the colour red should be reserved for clearly wrong issues.

5 Conclusions

The EQA had participation from 25 of 30 EU/EEA countries and two of six Western Balkan countries. Overall, this participation was good in EU/EEA for the first iteration of this EQA, while for the Western Balkan countries, and Türkiye, there may be room for improvement. There was a slightly higher participation rate for SARS-CoV-2 than for influenza virus, which may have been due to the longer than usual influenza season in some countries.

For the consensus sequence generation, about half of the laboratories signed up for both Nanopore and Illumina reads based components and the other half only for one, indicating that both technologies are often available in these laboratories. No substantial difference in throughput time for both technologies could be observed, indicating a similar workload for the consensus generation process. Several recurring issues in the generation of the consensus sequence were observed and comparatively more issues were observed for influenza virus than for SARS-CoV-2. This may mainly be due to (i) usage of references with only coding sequences; (ii) non-masking of primer sequences; and iii) difficulty to analyse (either automatically or manually) data corresponding to defective genomes. Altogether, consensus sequence generation should be improved for several laboratories and in general for influenza virus. The most important of these improvements are described further in the Recommendations section.

For clustering, the main issue that was observed was not taking into account indels when determining phylogenetic relatedness. Apart from that, for SARS-CoV-2, the choice of cluster cut-off varied with plus minus one mutation from the cut-off used for the comparator sequence. However, in the absence of guidance on which cut-off or cut-off range to use for which purpose, such differences cannot be considered as critical issues. For influenza virus, the strategy used to count the number of mutations in a sample and the choice of cut-off explained most of the issues. Thus, a better description of the assignment should be given, and the global result of the INFL3 component should not be considered as critical. For classification, in some cases the data definitions used were outdated for SARS-CoV-2, which should be avoided. For influenza virus, no classification-related issues were documented.

The SARS-CoV-2 mutation detection component did not reveal significant issues. It should be noted that this component is intended to be converted into prediction of antiviral resistance, as antivirals and corresponding resistance data become available. The influenza virus resistance prediction showed, in some cases, a misdetection of Drug-Resistance Mutations, leading to a misclassification of some samples as resistant/susceptible to Oseltamivir, Zanamivir and/or Baloxavir. Clearer instructions should help identify the issues that led to these misclassifications.

6 Recommendations

With the first iteration of this EQA, a baseline was established for the quality that the processes of consensus sequence generation, clustering, classification, mutation detection and antiviral resistance prediction at EU/EEA level. Subsequent iterations may aim to assess the evolution of the quality of these processes and apply more stringent quality criteria where necessary.

All the recommendations derived from the previous sections have been compiled in Table 8. In order to improve quality, laboratories that have quality issues and that are at present not well-placed to address them, should be given the opportunity to participate in relevant training. This training can cover consensus sequence generation e.g. understanding of the differences between and limitations of existing tools, appropriate parameterisation of tools, local installation of tools, quality control and manual curation. For clustering, the usage and limitations of tools, as well as guidance on appropriate cluster cut-offs may also be helpful. Over time, results from this EQA can then also be used to monitor the impact of training and support.

Finally, based on training needs identified in this trial, selected staff members from laboratories from three countries were invited to and completed five-day twinning training activities on topic related to the topics covered in this EQA.

Table 8. Recommendations based on the first ESIB-EQA (2023)

Id	Target audience	Process	Description
1	Individual laboratories	Consensus sequence generation	The introduction of indels in the consensus sequence that are not present in the raw data but are instead due to alignment to a reference sequence, should be avoided. This can be done through e.g. manual curation, less stringent reads mapping with respect to gaps, and/or an additional steps of aligning reads to the generated (intermediate) consensus sequence.
2	Individual laboratories	Consensus sequence generation	The introduction of wrong nucleotides or wrong non-N ambiguities due to absence of or incomplete removal of primers should be avoided. The primer removal step should be well-validated.
3	Individual laboratories	Consensus sequence generation	Clearly ambiguous positions should be assigned a non-N ambiguity symbol, so that subsequent analyses, including quality controls, can take this information into account. When e.g. the two most frequent nucleotides at a certain position have less than 10% difference in frequency, the corresponding ambiguity symbol should be put instead of that of the most occurring nucleotide.
4	Individual laboratories	Consensus sequence generation	The minimum coverage threshold that is normally applied should be neither too high nor too low, to avoid stretches of Ns where actual nucleotides can be put and vice versa. For Nanopore reads, a threshold in the range between e.g. 20-50x could be applied whereas for Illumina reads, a threshold in the range between e.g. 10-20x could be applied. It should be kept in mind these figures are also dependent the exact protocol applied and that as technology evolves, increased accuracy may lower these thresholds.
5	Individual laboratories	Consensus sequence generation	The completion of the genome in 5' of the first amplicon and in 3' of the last amplicon with the corresponding part of a reference sequence need to be avoided, since this wrongly assumes that no mutations may be present in these regions. Instead, an equivalent number of Ns can be put or no nucleotides at all.
6	Individual laboratories	Consensus sequences generation influenza	Up-to-date and complete reference influenza genomes should be used. These should include including H1N1, H1N1pdm and B reference genomes and should include sequences outside CDS, as well as properly named segments.
7	Individual laboratories	Consensus sequences generation influenza	Primer regions should be masked in all segments.
8	Individual laboratories	Consensus sequences generation influenza	Samples indicative of defective viral genomes, which may lead to many wrong nucleotides reporting, should be detected and manual curation performed on them.

Id	Target audience	Process	Description
9	Individual laboratories	Sequence quality control	The presence of non-N ambiguity symbols in the sequence should be part of quality control since this can be indicative of co-infection with distinct variants or contamination with material from other samples. Such sequences should not be considered to be of ok quality, and only used in specific analyses such as on co-infections.
10	Individual laboratories	Clustering	When determining the phylogenetic distance between sequences, indels should also be taken into account rather than only substitutions.
11	Individual laboratories	Classification	Sufficiently recent versions should be used for classification algorithms such as Pangolin for SARS-CoV-2. This also applies to any (reference) dataset that they use. If these algorithms are run as part of a broader pipeline, the versions should be checked as well.
12	ECDC	Clustering	Guidance in terms of which cut-offs should be used for clustering and for which purpose could be helpful.
13	EQA organisers	n.a.	Make instructions fully self-contained per pathogen, and clearer on how the sequence quality control should be carried out. Reporting of sequence data in Fasta format rather than in Excel could be allowed. For the individual report, consensus generation components, a distinction should be made between critical sequence differences and non-critical ones. Make INFL3 and INFL4 instructions more specific.

References

1. Grubaugh ND, Hodcroft EB, Fauver JR, Phelan AL, Cevik M. Public health actions to control new SARS-CoV-2 variants. Cell. 2021 Mar 4;184(5):1127-32. Available at: https://www.ncbi.nlm.nih.gov/pubmed/33581746
2. Abdool Karim SS, de Oliveira T. New SARS-CoV-2 Variants - Clinical, Public Health, and Vaccine Implications. N Engl J Med. 2021 May 13;384(19):1866-8. Available at: https://www.ncbi.nlm.nih.gov/pubmed/33761203
3. World Health Organization (WHO). Factsheet - Influenza (Seasonal) 2023. Geneva: WHO; 2023. Available at: https://www.who.int/en/news-room/fact-sheets/detail/influenza-(seasonal)
4. European Centre for Disease Prevention and Control (ECDC). Public health impact of SARS-CoV-2 variants of concern: scoping review protocol. Stockholm: ECDC; 2021. Available at: https://www.ecdc.europa.eu/sites/default/files/documents/Public-health-impact-of-SARS-CoV-2-variants-of-concern-scoping-review-protocol.pdf
5. World Health Organization (WHO). Interim guidance - Maintaining surveillance of influenza and monitoring SARS-CoV-2 – adapting Global Influenza surveillance and Response System (GISRS) and sentinel systems during the COVID-19 pandemic. Geneva: WHO; 2020 Available at: https://iris.who.int/bitstream/handle/10665/336689/WHO-2019-nCoV-Adapting_GISRS-2020.1-eng.pdf?sequence=1
6. European Centre for Disease Prevention and Control (ECDC). Objectives for integrated surveillance of respiratory viruses at EU level. Stockholm: ECDC; 2022. Available at: https://www.ecdc.europa.eu/en/seasonal-influenza/surveillance-and-disease-data/facts

Annex 1. Participating laboratories

Table A1. Participating laboratories, sorted by country and name

Country	Name
Austria	Austrian Agency for Health and Food Safety (AGES), Vienna
Austria	Center for Virology, Medical University of Vienna, Vienna
Belgium	Sciensano (Laboratory 1), Brussels
Belgium	Sciensano (Laboratory 2), Brussels
Belgium	University Hospitals UZ Leuven – Laboratory Medicine – National reference centre for respiratory pathogens, Leuven
Bosnia and Herzegovina	OU Clinical microbiology, Clinical Center of the University of Sarajevo, Sarajevo
Croatia	Croatian Institute of Public Health, Zagreb
Czechia	Centrum Epidemiology and Microbiology, NIPH, Prague
Denmark	Statens Serum Institut (laboratory 1), Copenhagen
Denmark	Statens Serum Institut (laboratory 2), Copenhagen
Estonia	Estonian Health Board, Tallinn
Finland	Finnish institute for health and welfare, THL, Helsinki
France	CNR des virus à transmission respiratoire (dont la grippe), Groupement Hospitalier Nord, Lyon
Greece	Hellenic Pasteur Institute, Athens
Greece	National Influenza Centre for N. Greece, Microbiology Laboratory of Medical School, Aristotle University of Thessaloniki, Thessaloniki
Greece	National Public Health Organisation of Greece, Marousi
Hungary	National Public Health Center, Budapest
Iceland	Landspitali University Hospital, Reykjavik
Ireland	Cork University Hospital, Cork
Ireland	UCD National Virus Reference Laboratory, Dublin
Latvia	Riga East University Hospital, National Microbiology Reference Laboratory, Riga
Lithuania	National public health surveillance laboratory, Vilnius
Luxembourg	Laboratoire national de santé, Dudelange
Malta	Mater Dei Hospital, Msida
Netherlands	Erasmus MC, dept. Viroscience, Rotterdam
Netherlands	National Institute for Public Health and the Environment, Bilthoven
Norway	Norwegian Institute of Public Health, Oslo
Poland	National Institute of Public Health NIH - NRI, Warsaw
Poland	Wojewódzka Stacja Sanitarno-Epidemiologiczna w Olsztynie (The Regional Sanitary Inspectorate, Olsztyn), Olsztyn
Poland	Wojewódzka Stacja Sanitarno-Epidemiologiczna w Rzeszowie (Regional Sanitary Inspectorate, Rzeszów), Rzeszów
Portugal	National Institute of Health Dr. Ricardo Jorge, National reference Laboratory for Influenza and Respiratory Viruses /Bioinformatics Unit, Lisboa

Country	Name
North Macedonia	Institute of Public Health, Skopje
Romania	Cantacuzino Institute, Bucharest
Romania	National Institute of Public Health, Bucharest
Slovakia	Comenius University Science Park, Bratislava
Slovenia	National Laboratory for Health, Environment and Food Slovenia, Laboratory for Public Health Virology, Ljubljana
Slovenia	University of Ljubljana, Faculty of Medicine, Institute for Microbiology and Immunology, Ljubljana
Spain	Institute of Health Carlos III. Bioinformatics Unit, Madrid

Annex 2. Composition of the datasets

Table A2. Composition of the SARS1 dataset

Samples	Description
1, 2, 3, 7, 14, 15, 16, 25, 27	No specific challenges. Consists of 7 Omicron, 1 Alpha and 1 Delta variant.
6, 10, 13, 17, 18, 28	Challenges in the S gene.
29	Challenges in the N gene.
19, 24, 30, 31	Challenges in the ORF1ab region.
8, 9, 20	Challenges in the ORF7 gene.
21, 32	Challenges in the ORF8 gene.
4, 5	Coinfected samples, of which 1 Alpha/Delta and 2 Omicron/Omicron. These are considered as bad overall quality, since they are not usable for routine analyses that do not specifically take into account co-infection.
12, 26	Clearly bad overall quality.
11	Omicron/Omicron recombination.
22, 23	Hepatitis E virus, i.e. different species, samples. These are educational samples intended to indicate the importance of detecting potential miscommunications between wetlab and drylab/bioinformatics processes, i.e. in this case a sequencing run with several species where only one species was expected. Some SARS-CoV-2 reads are still present in each sample, likely due to the adapter ligation reaction continuing a little bit after pooling of the samples, complicating the detection of the issue.
13	V3 rather than V4.1 primers used. This is an educational sample intended to indicate the importance of detecting potential miscommunications between wetlab and drylab/bioinformatics processes, i.e. in this case a change in the primers used.

Table A3. Composition of the SARS2 dataset

Samples	Description
1, 2, 4, 8, 10, 11, 12, 19, 28, 29	No specific challenges. Consists of 8 Omicron, 1 Delta and 1 Alpha variant.
5, 21, 31	Challenges in the S gene.
24	Challenges in the N gene.
6, 14, 16	Challenges in the ORF1ab region.
3, 15, 25	Challenges in the ORF7 gene.
26	Challenges in the ORF8 gene.
9, 17, 32	Coinfected samples, all Omicron/Omicron. These are considered as bad overall quality, since they are not usable for routine analyses that do not specifically take into account co-infection.
20, 27	Recombination, of which one Delta/Omicron and one Omicron/Omicron.
18, 30	Clearly bad overall quality.
22, 23	Influenza virus, i.e. different species, samples. These are educational samples intended to indicate the importance of detecting potential miscommunications between wetlab and drylab/bioinformatics processes, i.e. in this case a sequencing run with several species where only one species was expected.
7, 13	V4 rather than V3 primers used. These are educational samples intended to indicate the importance of detecting potential miscommunications between wetlab and drylab/bioinformatics processes, i.e. in this case a change in the primers used.

Table A4. Composition of the SARS3 dataset

Samples	Pangolin lineage	Description
1, 10, 15, 16	BA.2.30	Cluster A, zero SNPs and indels difference.
12	BA.2.30	Cluster A, one SNP difference from the rest of the cluster.
4, 5, 11, 18	BA.2.30	Cluster B, zero SNPs and indels difference.
9	BA.2.30	Not in a cluster; two SNPs difference from both cluster A and B.
3, 8, 17	BA.2.12	Cluster C, zero SNPs and indels difference.
19	BA.2.12	Cluster C, one SNP difference from the rest of the cluster.
2	BA.2.12	Not in a cluster, zero SNPs, one deletion and one insertion from cluster C.
6,	BQ.1	Not in a cluster.
13	XBB	Not in a cluster.
7	Unassignable	Clearly bad overall quality.
14	BA.5	Clearly bad overall quality.
20	BA.2.30	Clearly bad overall quality.

Table A5. Composition of the SARS4 dataset

The S gene substitutions column only contains those substitutions that were included in the list of to be detected mutations.

Samples	S gene substitutions	Description
1	S:L452M, S:N440K, S:K444R, S:N460K	BA.2.3.20
2	S:W152R, S:G257S, S:I210V, S:F157L, S:N460K, S:N440K, S:G446S	BA.2.75 with good overall quality but with a portion of the S gene not resolved and thus considered as bad specific quality since the presence of specific mutations is important for this analysis and they may be masked by ambiguities. This sample was added for educational purposes, to indicate that care should be taken as to which results can be considered reliable.
3	S:F486V, S:R493Q	Clearly bad overall quality sample with substantial parts of the S gene not resolved and thus also bad specific quality.
4	S:K444T, S:F486V, S:L452R, S:N440K, S:R493Q	BA.5. Contains a reversion, S:R493Q, that should be detected.
5	S:F486V, S:L452R, S:N460K, S:N440K, S:R346T	BF.7
6	S:W152R, S:F490S, S:R346T, S:G257S, S:I210V, S:F157L, S:N460K, S:K356T, S:N440K, S:G446S	BN.1
7	S:K444T, S:F486V, S:L452R, S:N460K, S:N440K	BQ.1
8	S:W152R, S:K444T, S:L452R, S:R346T, S:G257S, S:I210V, S:F157L, S:N440K, S:G446S	CH.1.1
9	S:F490S, S:R346T, S:N460K, S:N440K, S:G446S	XBB
10	S:F490S, S:R346T, S:F486P, S:N460K, S:N440K, S:G446S	XBB.1.5

Table A6. Composition of the INFL1 dataset

Samples	Description
1, 2, 6, 8, 10, 11, 12, 13, 14, 15, 16, 18, 21, 24, 25, 27, 29, 30, 31, 32	specific challenge. Consist of 6 H1N1pdm, 8 H3N2, 4 Victoria and 2 Yamagata samples. Some samples may have occasional low coverage sections in some segments or few mixtures at random positions.
3,4,7	Consist of 2 H1N1pdm and a H3N2 with PA, PB1 and/or PB2 that are defective to the point where the analysis can be difficult.
5	Consist of a H3N2 with drops of coverage in random positions for most segments. Because the majority of segments can still be analysed, we have decided to accept this sample. It is understandable to reject it.
9	Consist of a H1N1 not pandemic, but rather close to the A/Puerto-Rico/1938 strain. The main challenge remains the capacity to analyse such H1N1 sample, especially since it has lots of mixtures along multiple segments. Rejecting such sample is acceptable.
17	Consist of a H3N2 with a low coverage on all segments except MP. MP, NS and NP are the only exploitable segments.
19	Consist of a H1N1pdm with low coverage on all segments except MP which is the only exploitable segment.
20	Consist of a H3N2 with a very low coverage on NA, PA, PB1 and PB2. Although it may be possible to accept this individual based on the information from the HA segment, it is acceptable to reject it.
22, 23	Consist of two SARS-CoV-2.
26	Consist of a H3N2 with a low coverage on almost all segments. Some segments are not retrievable by our standards, but this sample can be accepted as OK for HA and NA for the comparator sequence with 20 as minimal coverage. It is acceptable to reject it.
28	Consist of a Victoria with a bad coverage on all segments except MP. Some segments are not retrievable by our standards, but this sample can be accepted as OK for HA and NA for the comparator sequence with 20 as minimal coverage. It is acceptable to reject it.

Table A7. Composition of the INFL2 dataset

Samples	Description
1,2,3,4,5,6,8,9,11,12,13,14,15,16,18, 19,20,21,24,25,26,27,28,29,31,32	No specific challenge. Consist of 9 H1N1pdm, 11 H3N2, 4 Victoria and 2 Yamagata individuals. Some samples may have occasional low coverage sections in some segments or few mixtures at random positions.
7, 10, 17	Consist of a H1N1pdm and 2 H3N2 with PA, PB1 and/or PB2 that are defective to the point where the analysis can be difficult or lead to a rejection.
22, 23	Consist of two SARS-CoV-2
30	Consist of a Victoria with NP and MP segments missing.

Table A8. Composition of the INFL3 dataset

Samples	Description
1,2,3	H1N1pdm viruses belonging to the 6B.1A.5a.2 clade and to a cluster (termed A for example)
4	BAD sample, should not be analysed
5	H1N1pdm virus belonging to the 6B.1A.5a.1 clade
6	H1N1pdm virus belonging to the 6B.1A.5b clade.
7	H1N1pdm virus belonging to the 6B.1A.7 clade
8, 10, 11, 12	Victoria viruses belonging to the V1A.3a.2 clade and to a cluster (termed B for example)
9	BAD sample, should not be analysed
13	Victoria virus belonging to the V1A.3a clade
14	Yamagata virus belonging to the Y3 clade
15, 16, 17, 18	H3N2 viruses belonging to the 3C.2a1b.2a.2 clade and to a cluster (termed C for example)
19	H3N2 virus belonging to the 3C.2a1b.1a clade
20	BAD sample, should not be analysed.

Table A9. Composition of the INFL4 dataset with the overview of samples of high-level resistance (R), intermediate reduction in drug susceptibility (I) and sensitive to antiviral drugs (S) and their drug-associated mutations (DRM)

Sample	Oseltamivir	Zanamivir	Baloxavir	NA DRM	PA DRM
INFL4.01	R	S	S	H275Y	-
INFL4.02	S	S	R	-	I38M
INFL4.03	R	S	R	H275Y	I38M
INFL4.04	R	R	S	E119V;T148I	-
INFL4.05	S	R	S	D151G	-
INFL4.06	I	I	S	S331R	-
INFL4.07	S	S	I	-	E23G
INFL4.08 (BAD)	-	-	-	-	-
INFL4.09	S	S	I	-	E23K
INFL4.10	I	I	S	D197N	-

European Centre for Disease Prevention and Control (ECDC)

Gustav III:s Boulevard 40
16973 Solna, Sweden

Tel. +46 858601000
ECDC.info@ecdc.europa.eu

www.ecdc.europa.eu

Follow ECDC on social media
Twitter: @ECDC_EU
Facebook: www.facebook.com/ECDC.EU
Linkedin: www.linkedin.com/company/ecdc/

Publications Office
of the European Union