Genomic surveillance of SARS-CoV-2 in Belgium

Report of the National Reference Laboratory (UZ Leuven & KU Leuven)

Situation update – 26 of October 2021
(report 2021_51)

Executive summary

58,561 Belgian sequences of SARS-CoV-2 are now publicly available on GISAID; compared to last week's report 6,335 sequences have been added.

813 sequences of positive SARS-CoV-2 samples collected between 11/10/2021 and 24/10/2021 have at this stage been analysed in the context of baseline surveillance. Among these, B.1.617.2 (*Delta*) and its sublineages represented 100% of the circulating strains. In the context of active surveillance, also 100% of the 133 genomes obtained so far, were classified as Delta.

The genomic diversity of SARS-CoV-2 in Belgium is comparable with the situation described over the last 12 weeks.

Authors (National Reference Laboratory – UZ Leuven and KU Leuven):
Lize Cuypers, Piet Maes, Els Keyaerts, Guy Baele, Simon Dellicour, Tom Wenseleers, Barney Potter, Sunita Janssenswillen, Elke Wollants, Marc Van Ranst, Emmanuel André.

With the collaboration of the laboratories of UCL, ULB, UMons, UNamur, ULiège, UGent, UZA/UAntwerpen, Jessa ZH, AZ Delta, AZ Klina, IPG, AZ St Lucas Gent, OLVZ Aalst, Briant network, ZNA, AZ St Jan Brugge, UZ Brussel, LHUB-ULB, and UZ Leuven/KU Leuven; and Sciensano HealthData.

Previous reports can be downloaded using the following link:
https://www.uzleuven.be/nl/laboratoriumgeneeskunde/genomic-surveillance-sars-cov-2-belgium

Table of contents

1. Monitoring of VOCs in Belgium

While first identified on 6 April 2021 in Belgium, the B.1.617.2 Variant of Concern (Delta) is now the dominant lineage in the country, representing 100% of the baseline surveillance samples sequenced (see Figure 1).

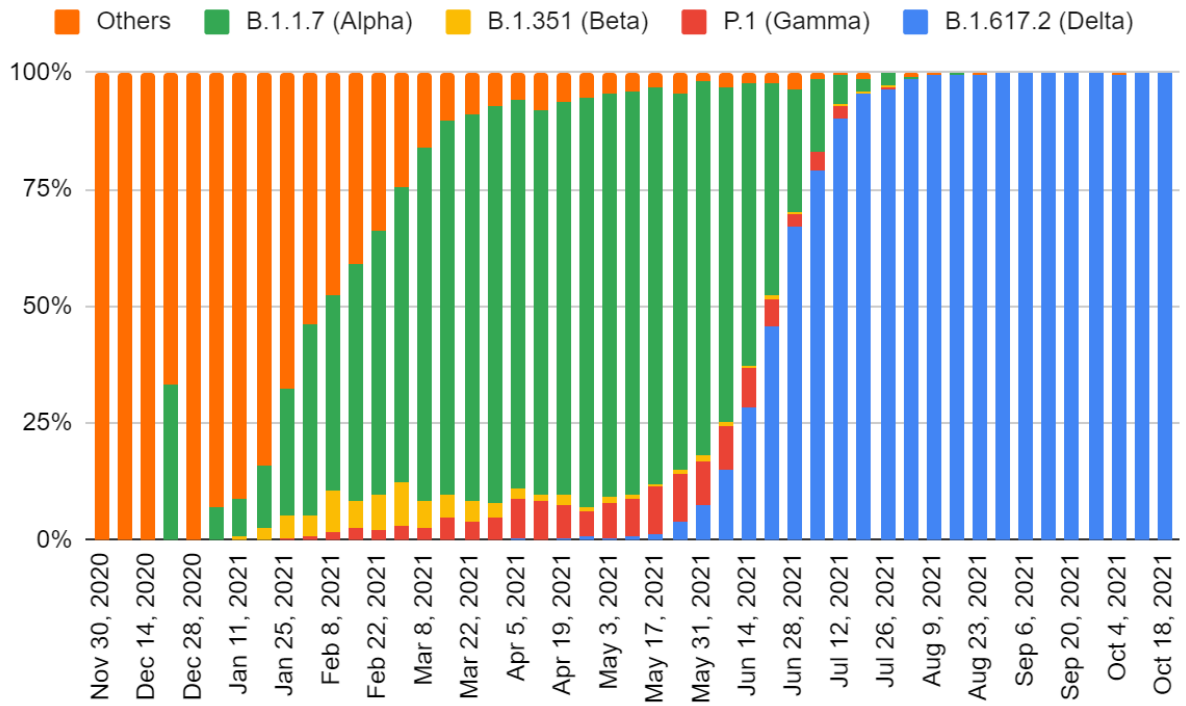

Figure 1: *Weekly evolution of the frequency of variants of concern reported by the baseline surveillance network using a whole genome sequencing (WGS) approach.*

Belgium is currently observing a significant acceleration in the number of infections (Re=1.55) and hospitalisations (Re=1.39), as can be seen in Figure 2.

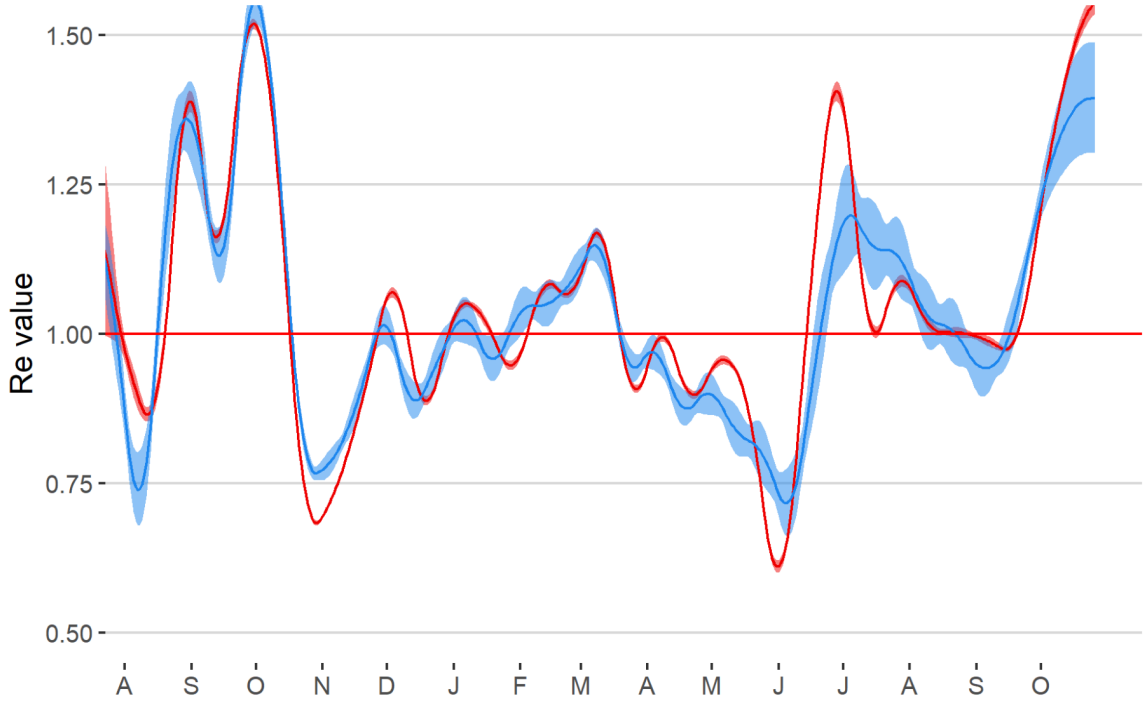

Figure 2: *Re value at the moment of infection in Belgium calculated from case (red : 1.55 [1.54-1.57] 95% CL) and hospitalisation data (blue: 1.39 [1.30-1.49] 95% CL). Corrected for variable testing intensity.*

2. Sequencing coverage and flow of sequencing process in detail

WGS: from a positive RT-PCR result to detailed typing

Within the SARS-CoV-2 genome surveillance initiative, two main pillars are being addressed: **baseline and active surveillance**. In the context of baseline surveillance, the purpose is to sequence a random and unbiased set of positive SARS-CoV-2 samples for the entire country. The aim has been set to minimally sequence 5% of all positive cases that are identified on a daily basis. To achieve this objective and to sequence a representative set of samples, a network of sentinel labs has been established throughout the country. The choice of this network was based on the geographical location of the laboratory, the number of SARS-CoV-2 tests being performed on a daily basis, the nature of samples (mix of hospital, private, federal testing labs) and the possibility to use an already existing transport flow or easily setup a new flow from the sentinel to the sequencing lab. By using this approach, more than 40 different labs agreed to send samples to the geographically closest sequencing facility (at least one sequencing lab per province was set up) in the context of baseline surveillance. Either a predefined percentage or fixed number of positive samples with a high viral load (corresponding to a Ct value <25) were requested to be randomly selected and sent to a sequencing facility, with a minimum frequency of once per week to not largely impact the time from a positive RT-PCR result to WGS.

The second pillar is active surveillance, including specific indications such as **outbreak settings, returning travellers and reinfections**. In this context, samples can be sent from any lab in Belgium that is performing SARS-CoV-2 diagnostics to one of the sequencing facilities or to the NRC UZ/KU Leuven for particular indications as defined in the RIZIV/INAMI convention.

Due to the nature of the analytical process of WGS, samples are processed in a batch, largely driven by efficiency of the process but also by cost, to make sure that the nomenclature price of €75 is not exceeded. This price is not billed to the patient, but facturation proceeds through the mutuality of the patient for which a sample was sequenced. When a WGS result is obtained, it will be communicated to the laboratory that has sent the sample to the sequencing facility as well as to Sciensano HealthData through the use of the 'Laboratory test results variants' message. On top of that, all sequencing facilities are asked to submit the sequencing data together with a minimum of metadata (i.e. age, gender, location and context of surveillance) to the international database GISAID. Through this approach, **all Belgian SARS-CoV-2 genome data are publicly available together with an enormous amount of genomic data from various countries across the world**. Having this central database for sequencing information allows us to perform follow-up analyses on Belgian data as well as international data, for various purposes, such as description of a new lineage, characterizing the importance of influx and export of lineages to and from Belgium, study of trends, follow-up of sequencing coverage, …

Although the upload of sequencing data to GISAID is a crucial and highly valuable step in the process, since we depend on the availability of the data for numerous downstream analyses, it can also be one of the delaying factors in submitting sequences go GISAID (see Figure 3). Since the sequencing result is already communicated to the requesting lab and Sciensano, the submission to GISAID is no longer a form of clinical reporting. Therefore, some laboratories are also batching GISAID submissions, which can explain a delay in sequences being available through this source. Furthermore, the frequency of transport for samples from a sentinel lab to a sequencing facility is of key importance for a short turn-around-time (TAT) between obtaining a positive RT-PCR result and having a WGS result available. Sentinel labs are asked to send these samples at least once a week, and every sequencing facility performs at least one sequencing run per week. However, in the worst-case scenario, when a sentinel lab ships samples in the context of baseline surveillance every Monday, and a positive RT-PCR result is only

obtained on Tuesday, that respective sample will only be sent to the sequencing facility the week after, already having a TAT of six days when starting the count at the time of the positive SARS-CoV-2 test result. The technical process of WGS is organised in batches, requiring a few days of wet-lab work and actual sequencing as well as analyzing the sequencing data through a bioinformatics pipeline. A third potential delaying factor can be a technical error or challenging factor (i.e. too low viral concentration after amplification step despite a Ct value <25) during the WGS analytical process, which demands the repetition of the sequencing process for a particular sample. Overall, a sequencing result can be expected approximately 10 days after obtaining the positive PCR result, at least in case of absence of the mentioned delaying factors.

Sequencing coverage

The start of the Belgian SARS-CoV-2 whole-genome analysis consortium was in February 2021, although prior to this date already positive samples were sequenced by a limited number of labs (approximately 1-2% of the positive cases). The first SARS-CoV-2 positive cases were confirmed by WGS at the NRC UZ/KU Leuven. From February 2021 on, the proportion of sequenced samples over all identified positive cases ('sequencing coverage') significantly increased over time. From February to April the aim of 5% coverage was reached, while from May 2021 on, over 10% of all positive samples have been sequenced. Sequencing coverage is calculated as the number of sequences (source: GISAID) compared to the overall positive cases (source: dashboard Sciensano). However by using this approach the coverage is even underestimated since an important share of the total positive cases were classified as 'weakly positive', hence due to a low viral load simply not feasible to be processed by whole-genome sequencing.

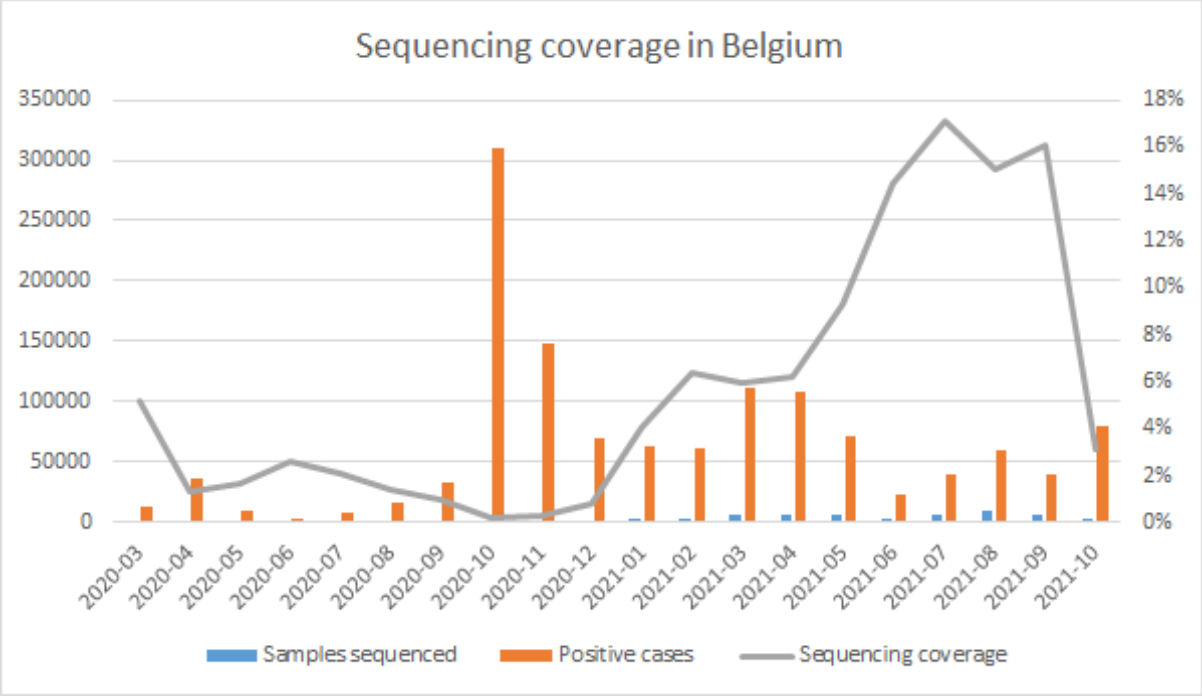

Figure 3: *Monthly evolution of the sequencing coverage in Belgium, as calculated by the ratio of samples sequenced (in blue, source: GISAID) over the overall number of positive cases (in orange, source:*

dashboard Sciensano), since the start of the pandemic. The sequencing coverage of the last month is underestimated due to analytical delay as mentioned above.

Since for the majority of sequences submitted to GISAID, location information is annotated to the sequence, sequencing coverage can be followed over time for each of the 11 provinces/regions. While small differences are observed, all provinces have a coverage above the minimum threshold of 5% for the last half year (Figure 4). Similar to the plot for Belgium (Figure 3), **most regions even exceed a coverage of 10%, which is a good marker for the representative and stable genomic surveillance that is in place in Belgium**. Of note, every month almost 3% of the sequences that are submitted to GISAID are annotated without location information or with incomplete information which cannot be mapped to province level. While currently the positivity rate and hence the number of positive cases and potential candidate samples for sequencing is rapidly increasing, additional efforts will be put in place to guarantee the high sequence coverage over the entire country.

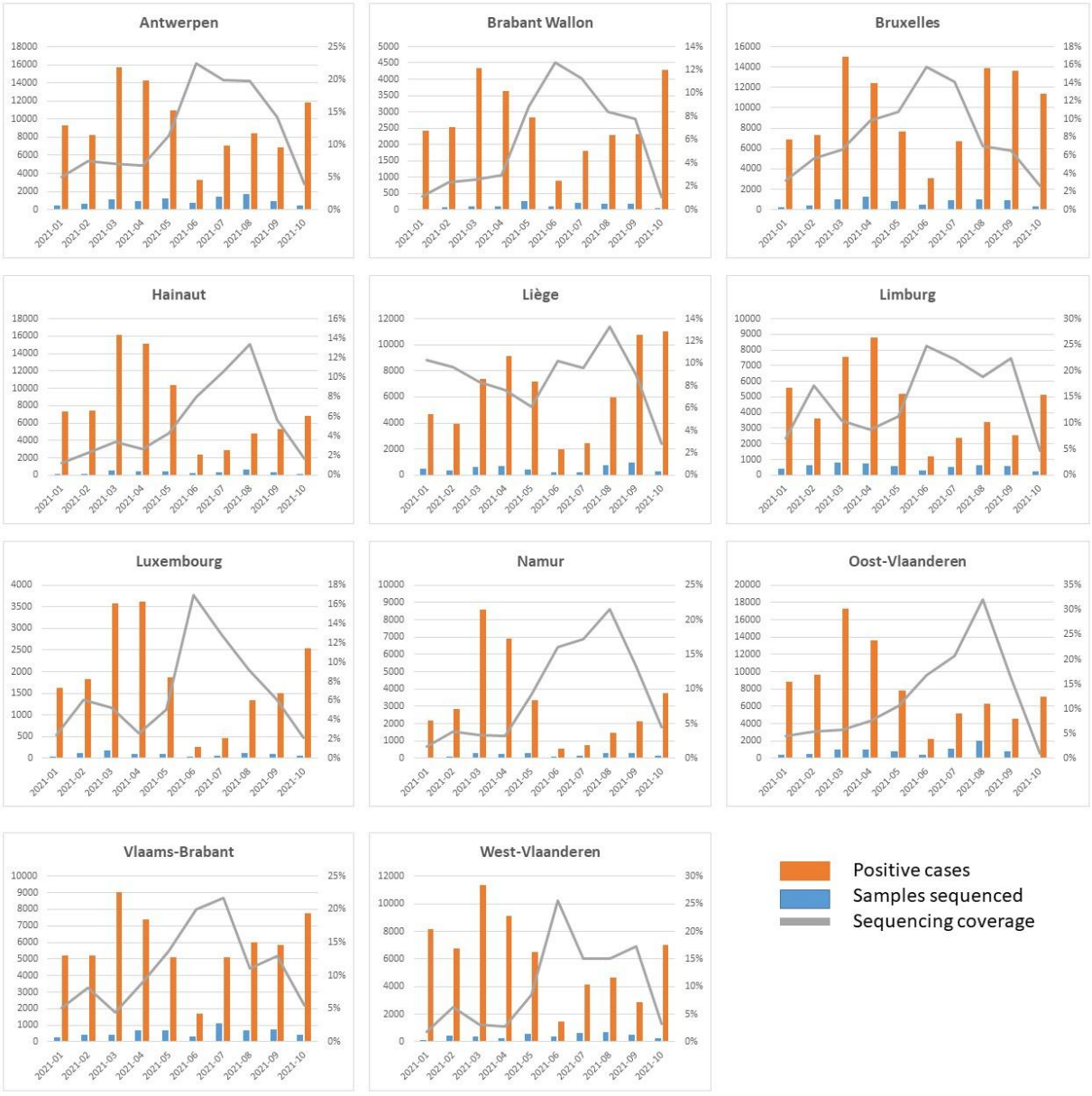

Figure 4: *Monthly evolution of the sequencing coverage per province. The sequencing coverage of the last month is underestimated due to analytical delay as mentioned above.*

3. Situation with regard to AY.4.2 and AY.5.2

As previously described, different sublineages have been delineated within the large group of Delta B.1.617.2 genomes. Currently, sublineages AY.1 up to AY.41 have been defined (for all details please see https://cov-lineages.org), with for some AY sublineages even additional subdivisions (e.g. AY.4.2). Over time, more and more sublineages have been defined due to the predominance of the Delta variant worldwide and the increased genomic surveillance on an international level.

The sub-lineage **AY.4.2** was first detected in Belgium on 25/08/2021 and currently represents less than 2% of the circulating strains (Figure 5). Overall, 74 (0.3%) out of the 23.407 sequences currently available worldwide originate from Belgium. The current estimates obtained using the international (United Kingdom) surveillance data show a limited competitive advantage, which is currently not observed in Belgium.

The sub-lineage **AY.5.2** was first detected in Belgium on 18/6/2021 and currently represents less than 2% of the circulating strains. Overall, 97 (24%) out of the 407 sequences currently available worldwide originate from Belgium. Of note, this sublineage does not originate from Belgium, as it has been identified at the international level 2 months before it was observed for the first time in Belgium. Due to this recent evolution, the very limited numbers and the scarcity of international data, the foreseen evolution of this sub-lineage remains uncertain, but we cannot exclude at this stage a significant growth advantage (>10%), as highlighted in Figure 6.

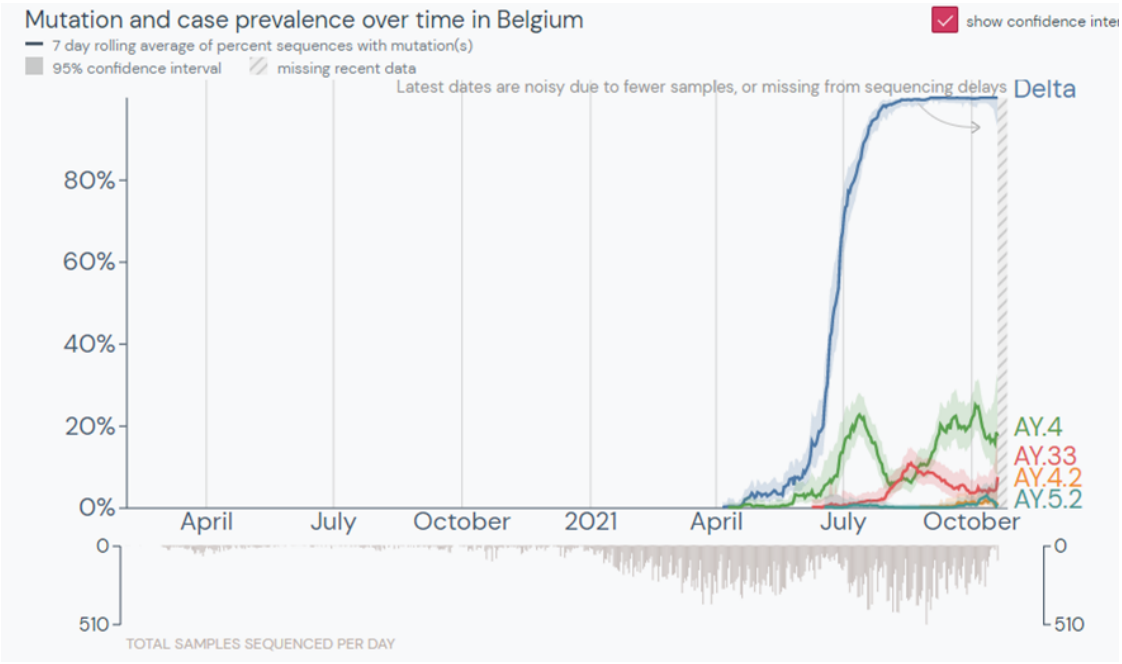

Figure 5: *Evolution of the prevalence of the Delta sublineage AY.4, AY.4.2, AY.5.2 and AY.33 over time in Belgium.*

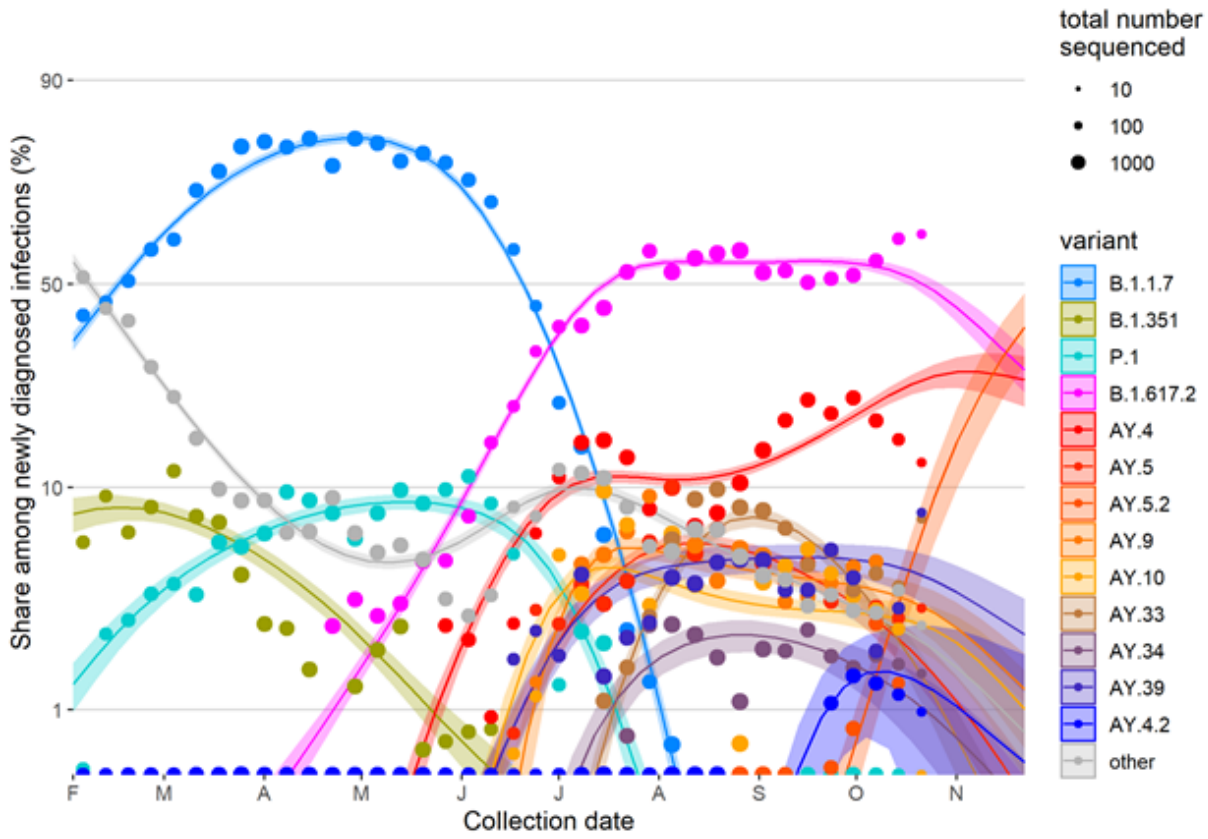

Figure 6. *Evolution of SARS-CoV-2 pangolin lineages in Belgium (GISAID data), including Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2) and Delta descendant lineages (AY.X). A multinomial spline fit (LINEAGE ~ PROVINCE + ns(DATE, df=3), with ns=3 df natural cubic spline) demonstrates that currently, AY.4.2 does not have a significant growth rate advantage relative to either B.1.617.2 (original Delta lineage) (growth rate difference per day = -0.8% [-5.3 – 3.7%] 95% CLs) or AY.4 (the direct ancestor lineage) (growth rate difference per day = -2.8% [-7.3 – 1.7%] 95% CLs). This conclusion, however, remains uncertain, given the small number of AY.4.2 infections occurring in Belgium (n=74 genomes sequenced out of 58,321 since February 1, 2021). Lineage AY.5.2, however, currently does have a significant growth rate advantage relative to both B.1.617.2 (growth rate difference per day = 13% [11 – 15%] 95% CLs) and AY.4 (growth rate difference per day = 11% [9 – 13%] 95% CLs). This conclusion, however, again remains highly uncertain, given the small number of AY.5.2 infections occurring in Belgium (n=97 genomes sequenced out of 58,321 since February 1, 2021), and that no systematic advantage is seen for AY.5.2 based on data from other countries (cfr. GISAID data outbreak.info). Hence, given these low counts, the estimated growth rate advantage could also represent chance events linked to localised infection clusters.*

4. Post-vaccination breakthrough cases: trends in lineage, age and brand

A breakthrough infection is defined as a positive SARS-CoV-2 test at least 7 days after the full completion of a vaccination scheme. According to a recent RAG advice, there is no longer a need to systematically sequence all breakthrough cases. It was agreed upon that only samples of infections of fully vaccinated persons with a severe disease course (hospitalisation) should be systematically sequenced, as well as samples of fully vaccinated residents of nursing homes.

According to data provided by Sciensano, the weekly evolution of the frequency of variants of concern is summarized in Figure 7 for the post-vaccination breakthrough infections, as well as the age of the person and brand of the administered vaccine related to the post-vaccination breakthrough case for which variant information was available. Of note, for the figure the raw data is visualized, not corrected for the age distribution of the Belgian population nor for the overall administered vaccine doses per brand, reflecting the vaccine roll-out in Belgium for which the elderly population was prioritized and for the brand Pfizer being the largest number of vaccine doses being purchased and administered. Furthermore, these data are purely based on post-vaccination cases for which variant information is available. Similar to baseline surveillance, it is to be expected that only a subset of samples (estimation of 10%) is sequenced and therefore only those data are considered for this analysis.

What can be concluded from this graph, is only that the frequency of variants of concern that are detected within breakthrough cases follows the same trends as what is observed in the context of baseline surveillance (see Figure 1). Furthermore, for both age categories and vaccine brand, the observed numbers were compared to the expected numbers, respectively based on the number of vaccines distributed (with one vaccine being two doses, except for J&J) and the total number of people vaccinated in each age category, using a Chi-Squared test. Based on these data, the number of post-vaccination infections is observed to be higher than expected for persons that were vaccinated by AstraZeneca and J&J, while in general higher than expected in persons with an age over 54 years. Based on these preliminary data, where no additional confounders were considered, this trend is believed to be rather linked to age than waning immunity since we don’t observe a clear and recent increase in the absolute number of post-vaccination breakthrough cases for the older age categories.

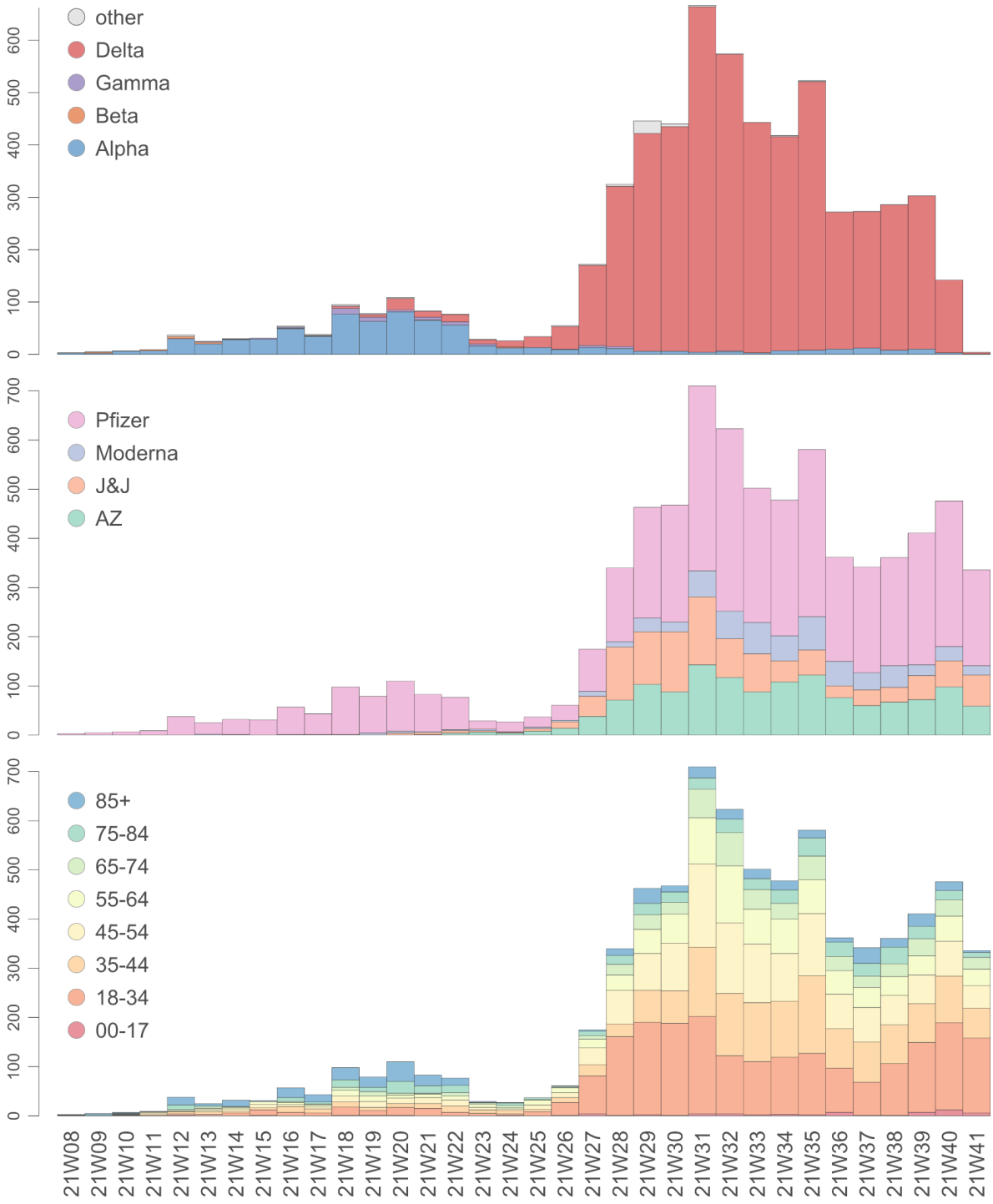

Figure 7. *Weekly evolution of the number of post-vaccination breakthrough infections for which variant information is available: 1) ranked according to WGS lineage information; 2) organized by vaccine brand that was administered; and 3) by age category. Of note, this is raw data being presented, not corrected for confounding factors such as the age distribution in Belgium or the number of vaccine doses administered overall. Furthermore, this data is purely based on post-vaccination cases for which variant information is available. Similar to baseline surveillance, it is to be expected that only a subset of samples (estimation of 10%) is sequenced and therefore only those data are considered for this analysis.*