

1 **Title:**

2 National-scale surveillance of emerging SARS-CoV-2 variants in wastewater

3 **Authors:**

4 Fabian Amman^{1*}, Rudolf Markt^{2*}, Lukas Endler¹, Sebastian Hupfauft², Benedikt Agerer¹, Anna Schedl¹,
5 Lukas Richter³, Melanie Zechmeister⁴, Martin Bicher^{4,5}, Georg Heiler^{6,7}, Petr Triska¹, Matthew Thornton¹,
6 Thomas Penz¹, Martin Senekowitsch¹, Jan Laine¹, Zsófia Keszei¹, Beatrice Daleiden⁸, Martin
7 Steinlechner⁸, Harald Niederstätter⁸, Christoph Scheffknecht⁹, Gunther Vogl¹⁰, Günther Weichlinger¹¹,
8 Andreas Wagner², Katarzyna Slipko¹², Amandine Masseron¹², Elena Radu^{12,13}, Franz Allerberger³, Niki
9 Popper^{4,5}, Christoph Bock¹, Daniela Schmid³, Herbert Oberacher⁸, Norbert Kreuzinger¹², Heribert Insam²,
0 Andreas Bergthaler^{1#}

1
2 **Affiliations:**

3 ¹ CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Lazarettgasse 14
4 AKH BT 25.3, 1090, Vienna, Austria

5 ² Department of Microbiology, University of Innsbruck, Technikerstr. 25d, 6020 Innsbruck, Austria

6 ³ Austrian Agency for Health and Food Safety (AGES), Währinger Str. 25A, 1090, Vienna, Austria

7 ⁴ dwh GmbH, Neustiftgasse 57-59, 1070 Vienna, Austria

8 ⁵ Institute for Information Systems Engineering, TU Wien, Favoritenstr. 11, 1040 Vienna, Austria

9 ⁶ Complexity Science Hub, Josefstädter Str. 39, 1080 Vienna, Austria

0 ⁷ Institute of Information Systems Engineering, TU Wien, 1040 Vienna, Austria

1 ⁸ Institute of Legal Medicine and Core Facility Metabolomics, Medical University of Innsbruck, Müllerstr.
2 44, 6020 Innsbruck

3 ⁹ Institut für Umwelt und Lebensmittelsicherheit des Landes Vorarlberg, Montfortstr. 4, Bregenz, Austria

4 ¹⁰ Institut für Lebensmittelsicherheit, Veterinärmedizin und Umwelt des Landes Kärnten, Kirchengasse 43,
5 9020 Klagenfurt.

6 ¹¹ Abteilung 12 – Wasserwirtschaft, Amt der Kärntner Landesregierung, Flatschacher Straße 70, 9021
7 Klagenfurt am Wörthersee

8 ¹² Institute for Water Quality and Resource Management, TU Wien, Karlsplatz 13/226, 1040 Vienna,
9 Austria

0 ¹³ Ștefan S. Nicolau Institute of Virology, Mihai Bravu Avenue 285, 030304, Bucharest, Romania

1
2 * These authors contributed equally

3 # Corresponding author: bergthaler@cemm.oeaw.tu.ac.at

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

1 Main text

2 Introduction

3 The recent introduction of SARS-CoV-2 into the human population led to a pandemic with immense health
4 and socioeconomic impact worldwide ^{1,2}. The pace of the successful development and use of protective
5 vaccines has been unprecedented and is widely considered to be instrumental to end the pandemic ^{3,4}. Yet,
6 the emergence of new variants with higher transmissibility and/or immune escape properties may pose
7 future challenges for containing the circulation of the virus ^{5,6}. The evolution of SARS-CoV-2 is widely
8 studied but remains incompletely understood ^{7,8}. Importantly, potential contributing factors like antiviral
9 immunity by natural infections or vaccines as well as social behaviour, reservoirs such as long-term infected
0 immunosuppressed patients and animals, testing strategies and global transmission make it difficult to
1 predict the dynamic nature of the future evolution of SARS-CoV-2 ⁹⁻¹².

2 More than twenty months into the pandemic, the importance of virus surveillance for public health
3 continues to increase worldwide. Typically, this requires the identification of SARS-CoV-2 positive cases
4 and the underlying viral genome signature by either mutation-specific polymerase chain reaction (PCR) or
5 sequencing-based methods. Comprehensive SARS-CoV-2 surveillance programs allow to rapidly monitor
6 the epidemiological situation and link it to viral variants and clinical outcomes ^{13,14}. Results thereof are of
7 pivotal importance for decision makers to assess the current situations and prepare for imminent
8 developments within but also beyond the respective countries ¹³. Nonetheless, such programs do come with
9 limitations in their applicability on a global scale. Certain population groups including asymptomatic and
0 underprivileged individuals may be inherently underrepresented depending on the given national testing
1 and surveillance strategy in place ¹⁵. Furthermore, few countries can muster enough resources for such
2 comprehensive monitoring, which includes adequately funded organizational structures, cross-disciplinary
3 scientific expertise, and the readiness to integrate these resources into public health operations ¹⁶⁻¹⁸.

4 Wastewater-based epidemiology (WBE) promises to overcome sampling bias and economic constraints of
5 epidemiological case-based monitoring programs ¹⁹. SARS-CoV-2 is excreted by more than half of infected
6 patients via faeces ²⁰⁻²². Additionally, sputum, saliva and urine are also discharged into the sewage system
7 and contribute to the virion load in the wastewater (WW). Hence, municipal WW drainage systems can
8 serve as point of use for representative sampling of circulating SARS-CoV-2 variants in a qualitative and
9 quantitative manner ^{23,24}. Initially, refined PCR-based approaches were applied to detect, at times even
0 retrospectively, viral RNA in the sewage ²⁵⁻³⁰. Deduced prevalence was shown to empower forecasts of
1 infection dynamics for the near future ³¹. The successful sequencing of SARS-CoV-2 genomes from WW
2 was reported ^{32,33} and used to detect regional occurrence of selected virus variants based on the presence of
3 selected characteristic mutations. ^{26,34-48}. Later, WW was used to construct and quantify haplotype
4 signatures of variants of concern and to deduce the reproductive number ^{49,50}. Yet, the suitability of WBE

5 to survey the spatiotemporal viral dynamics and integrate variant-specific epidemiological parameters on a
6 national-scale was not assessed so far.

7

8 In Austria, a Central European country of around 9 million inhabitants surrounded by eight countries and
9 with a high degree of international mobility due to tourism and external economic ties, the first pandemic
0 wave from February to April 2020 was curbed by strict lockdown measures (**Extended Data Fig. 1**)^{51–53}.

1 This coincided with early efforts to establish a viral surveillance system in WW in Austria. In March 2020,
2 a broader surveillance approach was successively implemented. Since December 2020, close to the peak of
3 the second wave, up to 95 wastewater treatment plants (WWTP) were sampled on a weekly basis, covering
4 >57% of the Austrian population (**Extended Data Fig. 2**). Like in other countries in Europe and beyond,
5 the Alpha variant (pangolin nomenclature⁵⁴: B.1.1.7) was introduced in December 2020, which was
6 followed by a near complete take-over by the Delta variant (pangolin nomenclature: B.1.617.2 and AY.*)
7 by August 2021⁵⁵. Notably, the Austrian province of Tyrol harboured one of the largest clusters of the Beta
8 (B.1.351) variant outside of South Africa from January to March 2021⁵⁶. Such spatiotemporal dynamics
9 of emerging viral variants coupled with a national epidemiological surveillance system⁵⁷, including
0 comprehensive and routine genotyping of positive cases, provided the stage to develop and validate the
1 necessary methodology for WW-based variant monitoring by sequencing.

2 In this study, we developed and validated a robust framework to deduce public health-relevant
3 epidemiological indicators such as relative variant abundance and variant specific reproduction numbers
4 from WW-derived deep sequenced SARS-CoV-2 genomes within the context of a national-scale WBE
5 program.

6 **Results**

7 **Data collection.** For the presented study, samples from selected WWTPs across Austria, collected between
8 December 1, 2020, and September 15, 2021, were considered. All samples were subjected to RT-qPCR
9 based measurements of SARS-CoV-2 genome quantity, using internal spike-in controls to deduce absolute
0 copy number per volume. Wastewater collection, pre-processing and quantitative screening was performed
1 at three different laboratories. Thereof, a total of 2,093 representative samples from 95 different WWTP
2 across Austria were further analysed by amplicon-based whole genome deep sequencing (**Fig. 1a**). The
3 selected WWTP varied in size between 1,490 to 1,900,000 connected persons (median size: 22,370).
4 Associated catchment areas included 57.3% of the total Austrian population. Obtained sequencing data
5 were analysed with a PCR amplicon-based deep sequencing pipeline⁵². Of all samples 14.7% failed during
6 sequencing, most of the time due to low viral load in the WW as indicated by C_T values >35. The remaining
7 85.3% samples exhibited a convincing detection of SARS-CoV-2, but only 70.7% passed our quality
8 criteria of at least 40% genome coverage, to be considered for follow up analysis (**Extended Data Fig. 2**).

9 The reproducibility of our sequencing pipeline to call low-frequency mutations down to 1% in clinical
0 samples was shown previously ⁵². To validate mutation calling in wastewater samples, we sequenced
1 triplicate samples from two representative locations, WWTP Hofsteig with a high industrial share and more
2 rural WWTP Kuchl. Comparing the observed mutations qualitatively between the triplicates revealed that
3 only 29% and 21% of the mutations in the WWTPs Hofsteig and Kuchl, respectively, were observed in all
4 three replicates (**Extended Data Fig. 3a**). For mutations called by at least two of the three replicates this
5 number rose to 47% and 40%, respectively. Despite the high dropout rate, the overall correlation of
6 observed mutation frequencies was encouraging with correlation factor in pairwise comparisons between
7 0.69 and 0.87 (**Extended Data Fig. 3b**). Notably, the read coverage along the genome deviated from a
8 uniform distribution, leaving substantial regions with only a few informative reads. The positioning of
9 lowly covered regions did not show a strong reproducibility between the replicas, indicating that variable
0 sample quality and stochastic reason have a more pronounced effect, than systematic technical causes, such
1 as amplicon primer affinities (**Extended Data Fig. 3c**). Consequently, mutations which were expected to
2 be present, e.g. defining mutations for a viral strain dominating at a certain time, often failed to manifest in
3 the data (**Extended Data Fig. 3d**). Comparing the deduced mutation frequencies from the single samples
4 to the median value across the respective triplicate also revealed substantial deviation from the expected
5 mutation frequency, especially in the range up to 50% relative abundance (**Extended Data Fig. 3e**). Most
6 missed mutations were in regions of low read coverage which explains their dropout. However, some of
7 the missed mutations would have been covered by a substantial number of reads. This lack of detection
8 may be owed to sampling errors during the amplification step due to the overall low amounts of viral RNA
9 in the complex WW matrix (**Extended Data Fig. 3d,e**).

0 **Variant characterization and quantification.** Our technical assessments of WW sequencing determined
1 the overall reliability and reproducibility of the approach, but also showed inherent shortcomings related to
2 dropouts at individual sites. Consequently, we aimed to develop a robust and error-tolerant method for the
3 detection and quantification of different viral variants, which we named *VaQuERo* (Variant Quantification
4 in Sewage designed for Robustness). To increase the positive predictive value, a reduced set of relevant
5 variants for targeted quantification is defined first. Thereby, detectable variants are restricted to all variants
6 defined as variants of concern, of interest or under monitoring, as defined by the European Centre for
7 Disease Prevention and Control (ECDC), and variants circulating in Austria according to GISAID database
8 records ^{58,59}. In total 36 variants of relevance were included (**Extended Data Fig. 4**). Each mutation which
9 occurred in more than 85% of all genomes of a particular variant deposited in the GISAID database, was
0 considered a marker mutation for the respective variant. Mutations found in only one of the variants of
1 relevance were denominated as unique markers (**Extended Data Fig. 4**). Based on this comprehensive set
2 of marker mutations, we developed a method to detect and quantify variants. To gain robustness against
3 the anticipated dropouts and hence inflated number of mutations with an observed frequency of zero, we

4 implemented a hurdle model inspired scheme⁶⁰. In a first step, we removed all mutations not found in the
5 set of the marker mutations or that were observed at a mutation frequency below 2%⁶¹. Variants of
6 relevance, with at least three unique mutations remaining, were designated as detected and subjected to the
7 subsequent quantification step. For quantification both unique and non-unique marker mutations were used
8 in a SIMPLEX regression, to deduce the overall variant frequencies of the detected variants of relevance.
9 Applying *VaQuERo* to our national-scale WW data enabled us to quantify the temporal development of the
0 variant composition for each WWTP, and hence for its contributing population in the sewer-catchment area
1 (**Fig. 1b**). Thereby, the *VaQuERo* approach converted the detailed but complex mutation patterns as
2 observed in the WW into a format that was reminiscent of infection case-based epidemiological reports and
3 made it more accessible for public health stakeholders in Austria. Moreover, as these data provides
4 information on variant-specific relative abundances over time, it can also serve as the basis for further in-
5 depth analysis, e.g. analysis of differential fitness of distinct variants (**Extended Data Fig. 5**).

6 To visualize the regional expansion and decline of single variants, we reduced the Austrian geographical
7 drawn-out shape (West-East: 575 km [7.6°], North-South: 294 km [2.6°]) to its longitudinal dimension and
8 supplemented it by the factor time as an additional dimension (**Fig. 1c**). Interpolation in this reduced space-
9 time coordinate system allowed to illustrate the spatiotemporal development of variant incidence. This was
0 demonstrated by the steady retreat of B.1.160 and its replacement by Alpha, ushering in the eastern
1 provinces of Austria during January and being completed in the western provinces only around three
2 months later. In contrast, the exchange of Alpha by the variant B.1.617.2 started almost concomitantly
3 throughout Austria and was already completed (>98%) a month later during July 2021 (**Fig. 1c**).

4 **Validation of the results.** In a next step, we sought to validate our WBE analysis results through a
5 comprehensive integration with available epidemiological data from individual cases in the catchment areas
6 (> 130,000) collected by the Austrian Agency of Health and Food Safety (AGES). Noteworthy, during the
7 period from Jan 1st and Sept 15th, 2021, 70.8% of all positive cases in Austria were examined if they
8 constitute a variant of concern. Together with the overall high testing rate and the low positive rate⁵⁷, the
9 number of unreported cases is assumed to be comparatively low. The absolute case counts per detected
0 variants from communities within the catchment areas, were aggregated to proportions per total genotyped
1 cases per week per catchment area. Contrasting the two independent data sources demonstrated a high
2 concordance of the WW data analysis with the epidemiological trends and the time of introduction of to-
3 be dominant variants (**Fig. 2a**). WBE also reliably captured the detection of smaller epidemiologically
4 confirmed outbreaks as shown, for example, for Beta (pangolin nomenclatur: B.1.351) and Mu (pangolin
5 nomenclature: B.1.621) in the municipal area discharging to the WWTP Bregenz (**Fig. 2a**, top left).
6 Occasional flare-ups were detected in the WW even with only a marginal number of positively tested
7 individuals residing in the respective catchment area. In some instances, such as the Mu variant in WWTP
8 Hohenems (population: 43,722) or B.1.318 in WWTP Rosegg (population: 23,600), even timepoints with

9 a single confirmed infection case in the catchment area were reflected by a proportionate signal in the WW
0 analysis (**Fig. 2a**). Collectively, the signal from the WW variant surveillance and the epidemiological case
1 monitoring correlated with an R^2 of 0.73 (**Fig. 2b**). To estimate the sensitivity of the WW based approach,
2 the signal from the epidemiological case monitoring program was examined for regions and timepoints
3 where no corresponding variant-specific signal in the WW surveillance was detected. These missed variant
4 occurrences constituted between 1 and 46 epidemiologically assigned cases, with a median of just 2 cases.
5 In terms of relative frequencies, missed variant signals range from 0.02% to 100% with a median value of
6 2.38%, which is close to the expected detection limit of our method, given that mutations with frequencies
7 below 2% in wastewater samples are dismissed as not reliable in the first place (**Fig. 2c**). Considering the
8 median missed case count and the median missed variant frequency altogether, 81% of all missed datapoints
9 were below either of them, indicating an estimated effective detection limit of the applied method.
0 Conversely, a variant linked to more than 2 cases and more than 2.38% of all cases in the catchment area
1 can be expected to be reliably detected in the WW. Overall, the comparative analyses revealed, that WBE
2 provides a detailed reflection of the dynamic epidemiological situation down to the regional level.

3 **De-novo recognition of undescribed mutation constellation.** Beside the reliable detection and relative
4 quantification of known variants, detecting novel emerging variants is the second prospect a variant
5 surveillance program should serve. Effectively, this is equivalent to a deconvolution of single haplotypes
6 in the amalgamation of genotypes found in the sewage. Our amplicon-based sequencing approach relies on
7 a modified version of the widely used ARTIC primer set and amplicons of around 400 bases ⁶², which
8 renders direct linkages of mutations across amplicons largely unfeasible. To overcome this obstacle, we
9 conceived to conflate associated mutations by their corresponding frequency pattern. Thereby, it was
0 possible to designate different mutation constellations based on their observed frequencies across different
1 samples in time and space by a hierarchical, unsupervised two-step clustering approach (see Methods). The
2 produced constellations of correlated mutations were characterized by incorporating them into the SARS-
3 CoV-2 reference genome with subsequent variant typing with the software tool pangolin ⁵⁴. Evaluation of
4 this approach showed that single samples typically feature not enough supportive information to deduce
5 telling mutation constellations. Apparently, a sensible composition of the sample set is pivotal to produce
6 reliable results. In this line, our de novo mutation constellation detection approach was applied to all
7 samples from the same calendar week and the same federal state. To demonstrate at what point in time the
8 emergence of an Alpha typic mutation constellation could have been identified, we exemplarily
9 investigated WWTP from the province of Carinthia, for which dense sequencing data well before the arrival
0 of Alpha is available. With our approach, Alpha-related mutations were observed in a dedicated
1 constellation deduced from six samples from February 1st to February 6th, 2021 (calendar week 5). The
2 respective constellation contained a total of 30 mutations, 19 of which are defining mutations for Alpha
3 (**Fig. 3a**). To evaluate the relevance of the obtained mutation constellations, this way determined mutation

4 can again be used as input for *VaQuERo* to examine the timely trends and geographic spread of the de novo
5 identified mutation constellations. Such an analysis immediately displays that the observed constellation is
6 indeed novel and sharply rising in three out of four WWTP (**Extended Data Fig. 6**).

7 **Nucleotide Diversity.** One important concept in molecular population biology is the nucleotide diversity
8 π , expressing the mean number of nucleotide differences across all loci between two genome sequences
9 from all possible pairs ⁶³. A small π indicates a very homogeneous population. Emerging variants with a
0 higher reproductive fitness, gradually outdistancing prevailing variants, are expected to concomitantly
1 reduce observed π -values. In contrast, repeated introduction of virus through infected individuals should
2 manifest in an elevated π -value. To test these hypotheses, we first calculated the nucleotide diversity π for
3 each sequenced sample and compared its progression with the relative amount of the Alpha and Delta
4 variants, respectively (**Fig. 3b**). The recurring pattern seen, is an increasing or stable nucleotide diversity π
5 prior to the introduction of Alpha. After Alpha had arrived and became the dominant variant, the level of π
6 was significantly reduced in the majority of WWTPs (**Fig. 3b**). This reduction, which even coincides with
7 an increase of prevalence, demonstrates the ongoing selective sweep in the genomic diversity due to Alpha
8 replacing previously prevalent variants (**Fig. 3b**). In contrast, upon the abrupt takeover of the Delta variant
9 no further reduction of π can be observed (**Fig. 3b**). Since the effect of the selective sweep during the Alpha
0 transition dominates the nucleotide diversity in the measured time period, further hypotheses were
1 subsequently evaluated only on samples with at least 95% relative frequency of the Alpha variant.
2 Population size and number of active infections in the catchment area positively correlated with observed
3 nucleotide diversity (**Extended Data Fig. 7a,b**). Yet, this explained only a small fraction of the observed
4 variability (R^2 of 0.03). To test the hypothesis that regional migration of persons contributed to elevated π -
5 values, we resorted to time resolved mobile communication network records and official numbers of
6 Austria's Federal Statistical Office on registered commuters and registered touristic overnight stays. None
7 of these mobility-based analyses could confirm a correlation between people's regional movements and
8 diversity of the virus population measured in the sewage (**Extended Data Fig. 7c-f**).

9 **Deduction of a variant specific reproduction number from wastewater.** Sequencing-based
0 quantification typically produces only relative, and not absolute quantification of single variants. Yet, it
1 can be combined with ex-ante measured RT-qPCR based quantification of the total viral load of the same
2 sample ⁵⁰. Thereby, it can be assumed that the SARS-CoV-2 load normalized per population equivalents
3 λ reflects a proxy of the prevalence of COVID-19 within the population connected to a sewage-system ⁶⁴.
4 Accordingly, its first derivative with respect to time $d\lambda_v/dt$ is a surrogate for the effective reproduction
5 number R_{eff} ⁶⁵, which we denoted as reproduction number from wastewater R_{ww} . Hence, our data allowed
6 a time-, region-, and variant-resolved tracing of the effective reproduction number (R_{ww}), an important
7 epidemiological parameter (**Extended Data Fig. 8**). Variant specific R_{ww} values were calculated

8 considering the total viral load according to RT-qPCR normalized to the ammonium-nitrogen load as
9 population size marker and the variant specific relative abundances derived by *VaQuERo* from the same
0 WW sample. R_{ww} of Delta could not be deduced this way since the rapid spread of Delta during a low
1 incidence phase did not produce enough timepoints with a signal of Delta and a concomitant signal of a
2 competing variant. For the variant of concern Alpha, the variant specific R_{ww} was increased by a factor 1.34
3 compared to concomitantly circulating variants. This result was robustly confirmed in all examined WWTP
4 (**Fig. 4**) and is supported by epidemiological studies ⁶⁶⁻⁶⁸.

7 **Discussion**

8 Foremost, we aimed with this study to convert insights deduced from WBE into classical, established
9 concepts of case-based SARS-CoV-2 epidemiology. This includes the detection and quantification of a
0 broad range of variants as defined by pangolin and the calculation of WW derived reproduction number
1 (R_{ww}). Thereby, we facilitate its direct utilization in decision making processes of public health authorities.
2 We demonstrated that our newly developed *VaQuERo* approach allows to robustly deduce relative virus
3 variant frequencies from WW. For the first time it could be shown in large-scale that variant surveillance
4 from WBE and case-based epidemiology agree qualitatively and quantitatively. Onset, duration and scale
5 of variant prevalence are in good agreement. The robustness of this finding is warranted by the
6 comprehensive testing strategy implemented in Austria, and the longitudinal and transversal breadth of the
7 here presented and publicly shared WW sequencing data. With respect to detection sensitivity of single
8 cases, we note that the accuracy of sequencing based WBE depends on prevalence and population number
9 in the catchment area. In a catchment area with too many positive cases, the signal of a particular case
0 becomes undetectable in the overall signal. Too low of a prevalence rate leads to a high dilution of the
1 virions in the WW, thus impedes the detection of the overall signal. On average, we observed with the given
2 relation of (time variable) prevalence and catchment size, that an absolute signal >2 cases or a relative
3 signal >2.38% could be reliably identified. Thus, the presented method may prove helpful to survey and
4 identify new viral variants in an early stage after regional introduction.

5
6 Methods for the detection of de novo variants from the complex genotype mixtures in the WW are still in
7 their infancy. Our approach to assort mutations based on their individual mutation frequencies across
8 several samples, serves as a first proof of principle and can be integrated directly with the variant definition
9 guided *VaQuERo* approach to identify emerging mutation constellations, and investigate their temporal and
0 spatial development patterns. A robust confirmation that the deduced constellations indeed are novel
1 haplotypes remains to be shown based on individual patient samples.

2 In contrast to clinical samples, virus sequencing from WW remains a technically challenging undertaking
3 due to the overall low concentration of target molecules as well as numerous interfering factors in the
4 complex WW matrix. The former can be dealt with using appropriate enrichment strategies while the latter
5 may be more difficult to control for. The manifold possible combinations between the WW matrix
6 influencing factors (e.g. seasonal changes, municipal versus industrial discharger), sewage system
7 architecture (e.g. stormwater implementation), sample pre-processing and sequencing techniques, still
8 await a thorough investigation for a detailed understanding of interferences and protocol adaptations. It is
9 important to highlight that the here presented results were accomplished despite considerable deficiencies in
0 the raw sequencing data quality caused by the complex wastewater matrix, making it applicable to versatile
1 operating conditions. However, fine-grained sample designs⁶⁹ and/or more customized extraction methods
2⁷⁰ may promise an even higher resolution of variant designation, detection and quantification for appropriate
3 settings.

4 Our results indicate that WBE recapitulates epidemiological screening programs at a high spatiotemporal
5 resolution, at reduced sample number and logistical effort. Thereby, WBE proved to complement classical
6 surveillance programs based on individual epidemiological cases. In the long run, comprehensive WBE
7 might be an economic way to inform traditional case-based epidemiology to adaptively adjust testing
8 strategy and sampling density on a regional level. Furthermore, once in place, WBE provides synergies to
9 survey prevalence and serotypes of a wider range of public health relevant pathogens, such as influenza
0 and enterovirus^{71,72}. This will be especially relevant for countries with limited resources. In this regard, it
1 is critical to iterate the here presented methodology to surface waters, cesspits and other modes of WW
2 disposal in absence of a centralised sewer⁷³.

3 The virus conglomerates in the WW reflects the whole virus population, which differs substantially from
4 the canonical approach of summing consensus sequences of virus isolates from individual cases. This
5 holistic perspective on the population provides opportunities and is expected to offer additional value for
6 WBE. It is well established in population genomics that certain applications benefit from a high sampling
7 rate⁷⁴. Here, we aimed to synergise the two fields by connecting the concept of nucleotide diversity and
8 important epidemiological parameters, such as prevalence and mobility, as a surrogate for introductions.
9 We observed that the absolute case number is indeed imprinted into the observed nucleotide diversity. In
0 contrast, no effect of mobility was found. Further studies are needed to clarify if this is due to a limited role
1 of regional mobility during viral spread, or if the used data does not provide enough sensitivity for its
2 detection. A cause for such a reduced sensitivity might be found in the pronounced effect of the selective
3 sweep caused by the displacement by the Alpha variant, which led to a reduction of the overall nucleotide
4 diversity. Remarkably, no similar reduction could be observed after the takeover of the Delta variant. It is
5 important to highlight that both displacements, by Alpha and Delta, respectively, as the dominant
6 circulating variants differed in many aspects. For example, Alpha was introduced during winter amidst a

7 fading seasonal wave while Delta emerged during a low-incidence period in early summer, after widespread
8 rollout of vaccines. Both replacements were accompanied by an increase in case prevalence. Interestingly,
9 Alpha took a month to be spread across Austria and even three months to gain dominance, while Delta
0 replaced Alpha within a month. Interestingly, during the dominance of Delta, the nucleotide diversity rose
1 in most of the WWTP. In contrast, the nucleotide diversity remained stable during the dominance of Alpha.
2 It is tempting to account this, beside increased population size, to a higher plasticity of Delta in contrast to
3 Alpha. Distinctions between the phylogeny of Delta and other variants of concern are already described ⁷⁵.
4 The dynamic differences of Alpha and Delta emergence may hold the key for a better understanding and
5 anticipation of future shifts in the circulating virus population.

6 The field of WBE, which has come a long way since its beginnings using monkeys as biosensors for
7 poliovirus detection ⁷⁶, received a strong impetus driven by the needs of the current pandemic management
8 ^{23,70}. Data analysis, interpretation and integration with classical epidemiological data of infectious diseases
9 remains a formidable challenge that will require concerted interdisciplinary efforts of the scientific
0 community, supported by water engineers, mathematicians, chemists, molecular biologists, virologists,
1 medical scientist and computational biologists. This national-scale study demonstrates the usefulness of
2 sequencing-based WW surveillance for the current SARS-CoV-2 pandemic and its future impact on
3 improved global surveillance of other infectious diseases.

4
5 **Data and software availability.** All raw sequencing data used during this study, with additional
6 sequencing data produced for samples drawn from Austrian WWTP before December 2020, are available
7 under the following ENA accession number: PRJEB48985. The source code for the developed and applied
8 software for variant quantification (*VaQuERo*) and mutation deconvolution (*DeViVa*) is available on
9 GitHub (<https://github.com/fabou-uobaf/VaQuERo> and <https://github.com/SebH87/DeViVa>).

1 **Acknowledgement**

2 This project was funded in part by the Austrian Agency for Health and Food Safety (AGES), the Vienna
3 science and Technology Fund (WWTF) as part of the WWTF COVID-19 Rapid Response Funding 2020
4 (A.B.), the Austrian Science Fund (FWF1212P) and the Austrian Academy of Sciences. Access to
5 WWTP and sample logistics was possible through the Coron-A Project, which was supported and funded
6 by the Austrian Federal Ministry of Education, Science and Research, the Austrian Federal Ministry of
7 Agriculture, Regions and Tourism, the Provincial Governments of Lower Austria, Upper Austria,
8 Salzburg, Carinthia, Tyrol, Styria, Burgenland, Vorarlberg and the Austrian Association of Cities and
9 Towns. Additional samples and background information were provided by the covid-19 wastewater
0 surveillance programs of the federal states Carinthia, Vorarlberg and Vienna. Sequencing was performed
1 at the biomedical sequencing facility (<https://www.biomedical-sequencing.org/>).

2

3

4 **Author contributions**

5 F.Am., R.M., L.E., S.H., L.R., M.Z., P.T., N.P., D.S., C.B. analysed data. R.M., B.A., A.S., M.Z., M.B.,
6 G.H., P.T., M.T., T.P., M.Se., J.L., Z.K., B.D., M.St., H.N., C.S., G.V., G.W., A.W., K.S., A.M., E.R.,
7 F.Al., H.O., N.K., H.I. provided processed wastewater samples, generated SARS-CoV-2 genome sequences
8 and/or provided data. F.Am., R.M. and A.B. wrote the manuscript.

9 **Methods**

0 **Analysis of WW parameters.** COD, total nitrogen (tN) and ammonium nitrogen (NH₄-N) were analysed
1 on site by respective WWTP operators using cuvette-tests. Deposition and daily WW flow were also
2 recorded on site.

3 **Sample processing.** Sample drawing, RNA extraction and RT PCR-based virus quantification was
4 performed by the members of the Coron-A consortium (www.coron-A.at), at three different laboratories
5 (TU Wien, Medical University Innsbruck, University Innsbruck). The protocols for WW sample pre-
6 treatment and PEG-NaCl-based precipitation of the virus fraction were based on the methods described by
7 Ye et al. and Wu et al.^{77,78}. In details, the implemented protocols varied in each of the executing laboratory.
8 Generally, all workflows followed a common procedure. 24 h volume equivalent mixed samples were
9 stored at 4°C until processing. Prior to PEG-precipitation based concentration, large particles were
0 removed. Nucleic acid extracted from pellet after concentration were analysed via one-step RT-qPCR for
1 presence of the N-gen of SARS-CoV-2 and passed on for sequencing library preparation. Details per
2 laboratory as follows. *University Innsbruck:* Larger particles were removed to decrease the amount of non-
3 viral RNA and PCR-inhibitors. For this purpose, 40–70 ml of WW were transferred to centrifugation tubes
4 and centrifuged at 4,500 g for 30 min at 4°C. To precipitate viral fragments, the resulting supernatant was
5 immediately transferred into a fresh tube containing 10 % w/v polyethylene glycol (PEG) 8000 (CarlRoth,
6 Karlsruhe, Germany) and 2.25 % w/v NaCl. The Reax2™ head over-head shaker (Heidolph, Schwabach,
7 Germany) was used until both additives were dissolved. Subsequently, the samples were centrifuged at
8 12,000 g for 99 min at 4°C to obtain a pellet containing the viral fragments. The supernatant was removed
9 by decantation and an additional centrifugation at 12,000 g for 5 min to remove remaining fluids with a
0 pipette. Following the precipitation of the viral fragments, samples were resuspended in 800 µl lysis-buffer
1 (Monarch™ total RNA Miniprep Kit, NEB, Ipswich, USA) and purified according to manufacturer protocol
2 of the Monarch™ total RNA Miniprep Kit with non-enzymatic gDNA-removal. RNA was eluted in 40 µl
3 RNase free water. RNA concentrations of the templates were quantified via Nanodrop and extracts with
4 RNA concentrations above 200 ng/µl were diluted as needed. RNA copy numbers were determined using
5 the N1 primers/probe according to the CDC-protocol⁷⁹ targeting the nucleocapsid-gene of SARS-CoV-2.
6 RT-qPCR reactions contained per 20 µl: 10 µl Luna Universal Probe One-Step Reaction Mix (2X) from
7 NEB, 1 µl Luna WarmStart® RT Enzyme Mix (20X) from NEB, 0.8 µl primer (final concentration 0.4
8 µM), 0.4 µl probe (final concentration 0.2 µM), 2 µl PCR grade water, and 5 µl template. Analyses were
9 conducted on a RotorGene cycler (Qiagen, Hilden, Germany). After an initial reverse transcription at 55°C
0 for 10 min, followed by 95°C for 1 min of denaturation, 45 cycles of 95°C for 10 sec and 60°C for 40 sec
1 were performed. To calculate copy numbers, a plasmid standard containing the N-gene of SARS-CoV-2
2 (2019-nCoV_N_Positive Control, IDT, Leuven, Belgium) was used.⁸⁰

3 *Medical University Innsbruck*: Untreated sewage was collected with a 24 h volume-equivalent sampler,
4 cooled to 4°C and shipped under cooled conditions to the processing laboratory. Wastewater samples were
5 processed within 24 hours of arrival. They were stored at 4°C until analysis. In a first step, 250 ml aliquots
6 of the WW samples were cleared by centrifugation at 4,500 x g for 30 min. In 225 ml of the resulting
7 supernatant, 22.5 g of PEG 8000 and 5.1 g of NaCl (both Sigma-Aldrich, St. Louis, MO) were dissolved
8 by careful swirling. These solutions were allowed to sit at 4°C for 15 min before they were spun at 12,000
9 g for 50 min to pellet the precipitated virus fraction. After carefully discarding the supernatant, the pellets
0 were re-centrifuged at 12,000 g for 5 min and the residual liquid was removed with a pipette. Finally, the
1 pellets were resuspended in 700 µl of buffer MTL (Qiagen, Hilden, Germany).

2 Four-hundred microliters of pellet suspension were supplemented with 60 µl of a solution comprising 60
3 ng carrier RNA per microliter AVL buffer (both Qiagen). Subsequently, samples were subjected to
4 automated RNA extraction on a Biorobot EZ1 Advanced XL instrument (Qiagen) by using the EZ1 Virus
5 Mini kit (v2.0, Qiagen) according to the manufacturer's recommendations. Total RNA was eluted in 60 µl
6 AVE buffer (Qiagen) and immediately used in quantitative reverse transcription real-time polymerase chain
7 reaction (RT-qPCR) experiments targeting the N1 gene. For subsequent sequencing, 45 µl of the extracts
8 were further purified with the Monarch™ total RNA Miniprep Kit (NEB, Ipswich, USA) according to the
9 manufacturer's protocol. Total RNA was eluted in 45 µl water.

0 *TU Wien*: Thoroughly mixed 45 ml aliquots from each sample were centrifuged at 4,500 g for 30 min at
1 4°C (the centrifuge was previously cooled down to 4°C). 40 ml of the supernatant were further added to
2 fresh 50 ml falcon tubes containing 100 g/L PEG (Sigma Aldrich) 8000 and 22.5 g/L NaCl (Sigma Aldrich),
3 followed by centrifugation at 12,000 g for 1h 30 min at 4°C. The supernatant was carefully removed, and
4 an additional centrifugation step was carried out in order to remove the remaining liquid (12,000 g, 5 min,
5 4°C). The pellet was resuspended in 400 µl molecular biology grade water (Sigma Aldrich) and 400 µl
6 CTAB (Hexadecyltrimethylammonium bromide; Promega) buffer. Prior to the RNA extraction, 40 µl of
7 Proteinase K (Promega) were added to each tube and inverted carefully in order to obtain a homogenous
8 suspension. The suspension was further transferred to a tube containing disruptor beads and run on
9 FastPrep24 (MP Biomedicals™) machine for 40 s at 6 m/s. Subsequently, 400 µl of each sample were used
0 for further RNA extraction. The RNA extraction was performed on Maxwell® RSC AS4500 (Promega)
1 device using Viral RNA/DNA Concentration and Extraction Kit from Wastewater (Promega). The SARS-
2 CoV-2 viral RNA was eluted in 100 µl TE buffer.

3 **RNA library prep and sequencing.** For amplicon cleanup AMPure XP beads (Beckman Coulter) were
4 used at a 1:1 ratio and amplicon concentrations and size distributions were checked with the Qubit
5 Fluorometric Quantitation (Life Technologies) and the 2100 Bioanalyzer system (Agilent), respectively.
6 Amplicon concentrations were normalized, and sequencing libraries created with the NEBNext Ultra II
7 DNA Library Prep Kit for Illumina (NEB) according to the manufacturer's instructions. As before library
8 concentrations and size distribution were assessed again, the pools mixed at equimolar concentrations and
9 sequencing carried out on the NovaSeq 6000 platform (Illumina) using an SP flow cell with a read length
0 of 2x250 bp in paired-end mode.

1 **Mutation calling.** After demultiplexing, adapter sequences were trimmed with BBDUK, and the
2 overlapping regions of paired reads were corrected with BBMERGE from the BBTools suite ⁸¹. BWA-
3 MEM ⁸² (v 0.7.17) was then used to map read pairs to the combined GRCh38 and SARS-CoV-2 genomes
4 (RefSeq: NC_045512.2) with a minimal seed length of 17. Only reads mapping uniquely and properly
5 paired to the SARS-CoV-2 genome were kept and primer sequences were masked with iVar ⁸³. LoFreq ⁸⁴
6 (v 2.1.2) was used for low frequency variant calling, after first using its Viterbi method to realign reads
7 around insertions and deletions (InDels) and adding InDel qualities. Variants were then filtered with LoFreq
8 and Bcftools ⁸⁵ (v 1.9) only considering variants with a minimum coverage of 75 reads, a minimum Phred
9 scaled calling quality value of 90 and InDels with a HRUN value of at less than 4. Variant annotation was
0 performed with SnpEff ⁸⁶ (v 4.3) and SnpSift ⁸⁷ (v 4.3).

1 **Virus variant quantification, including marker mutation definition.** Multiple genome alignments of
2 1.9 million global SARS-CoV-2 genomes provided by the global initiative on sharing avian influenza data
3 (GISAID database, retrieved July 5th, 2021) ^{58,59}, and the associated pangolin lineage assignments, were
4 used to extract all nucleotide variants compared to the reference genome. For each pangolin lineage, all
5 mutations that could be observed in at least 85% of the analysed genomes were defined as sensitive marker
6 mutations for the respective lineage. To reduce complexity, the scope was focused on all lineages which
7 were detected in Austria (at least ten sequenced genomes according to GISAID) and all variants of
8 concern/interest and variants under monitoring as defined by the European Centre for Disease Prevention
9 and Control (ECDC; <https://www.ecdc.europa.eu/en/covid-19/variants-concern>; accessed on July 7th,
0 2021). A total of 36 different regionally relevant variants were included (**Extended Data Fig. 4**). Deduced
1 marker mutation, together with inferred mutation frequencies, were further utilized to quantify all
2 respective variants of relevance per WWTP in the available time course data according to the following
3 scheme: (i) remove all samples with more than 60% of the genomic position covered with less than 10
4 reads; (ii) filter marker mutations with observed frequency >0.02 and >75
5 supportive reads; (iii) foreach timepoint, denote all variants with at least three unique mutations as detected;
6 (iv) add mutation frequencies from samples from the preceding 8 days; (v) filter frequencies of marker
7 mutation of detected variants; (vi) transform ⁸⁸ remaining mutation frequencies to avoid zeros and ones by

8 $f' = (f(d - 1) + 0.5)/d$, whereas d denotes the sequencing depth at the respective locus, f and f' denote
9 the observed and the transformed mutation frequency, respectively; (vii) infer expected frequency per
0 variant applying a regression model $y = X\beta$, whereas y represents the inferred variant frequencies, b the
1 mutation frequencies and X represents the $n \times m$ design matrix assigning the considered n mutations to one
2 or more of the m variants. To this end the package `gamlss`⁸⁹ (generalized additive model for location scale
3 and shape) is used with a SIMPLEX⁹⁰ linker function and weights w per mutation i $w_i = \left(\frac{1}{\Delta t + 1}\right) \log_{10}(d_i)$,
4 with the time difference to the current sample in focus in days Δt and the sequencing depth d_i for at mutation
5 position i . If the SIMPLEX function does not converge, resort to the Beta function. If sum of expected
6 values exceeds 1, scale down proportionally. An implementation of the described software is available on
7 GitHub (<https://github.com/fabou-uobaf/VaQuERo>). Spatiotemporal reconstruction was performed from
8 the deduced variant frequencies across Austria using the Multilevel B-Spline Approximation⁹¹ using the
9 `mba.surf` function implemented in the R package MBA with the following non-default parameters: `no.X =`
0 `no.Y = 30`, `m = range of longitude [°] divided by range of date [d]`, and `extend set to TRUE`.

1 **Deduction of a variant specific reproduction number from WW.** SARS-CoV-2 load was normalized to
2 the ammonium nitrogen (NH₄-N) load as population size marker, assuming a NH₄-N load per capita of 8 g
3 per day⁶⁴. The normalization results in the viral load parameter λ , representing viral copies per population
4 equivalent per day, which is proportional to the absolute number of infected individuals⁹². In combination
5 with the relative proportions from the sequencing based variant quantification (derived from the *VaQuERo*),
6 the absolute quantity λ_v from a certain variant v was estimated. Time series at a daily base, necessary for
7 the following R_{ww} calculation, were interpolated by LOESS smoothing (smoothing factor 0.25).
8 Subsequently, R_{ww} was calculate using the EpiEstim Package⁶⁵ (<https://github.com/mrc-ide/EpiEstim>)
9 based on λ_v , applying a fixed serial interval of $3.37 \text{ d} \pm 1.83$ (see ref.⁹³). The R_{ww} was separately calculated
0 for the Alpha variant (R_{ww}^{Alpha}) and the sum of all other pre-Alpha variants (R_{ww}^{Others}). The ratio
1 $R_{ww}^{Alpha} / R_{ww}^{Others}$, as a measure for relative fitness disparity, was analysed for the transition period January
2 to June 2021 during which Alpha gained dominance in Austria.

3 **Evaluation by epidemiological records:** The national Covid-19 surveillance system, maintained by
4 AGES, consist of all confirmed SARS-CoV-2 cases since the beginning of the pandemic. In the period of
5 examination between Jan 1st and Sep 15th, 2021, on average 70.8% of all identified SARS-CoV-2 cases in
6 all towns connected to any examined WWT were tested for SARS-CoV-2 mutations and variants.
7 Sequencing based tests were performed by partial sequencing of the spike gene⁹⁴, Sanger sequencing or
8 WGS. However, most cases were analysed using PCRs for specific mutations, performed in the responsible
9 diagnosis laboratory and reported to AGES. Based on the set of reported mutations a case was classified as
0 a particular variant of concern or as the wild type, accordingly. Definition of indicator mutations were
1 adapted during the pandemic. Individual cases are associated to their place of residence. Data were

2 extracted from the surveillance system for all towns connected to any examined WWTP and aggregated for
3 all towns within the same catchment area and from the same calendar week, to deduce absolute case counts
4 and relative frequency per variant per catchment area per week. For plotting purpose, and for direct
5 comparison with WW derived data, both epidemiological and WW data was shifted to Wednesday of that
6 respective week.

7 **Mutation Deconvolution.** For mutation deconvolution, a hierarchical, unsupervised two-step clustering
8 approach was used. First, a silhouette-analysis collected mutations with similar pattern of relative
9 abundances within a specific set of samples. The method is implemented in a software tool named *DeViVa*
0 (Deconvolution of Virus Variants), which was written in python (version 3.8) and which can be obtained
1 from GitHub (<https://github.com/SebH87/DeViVa>). The first clustering step used Ward's method based on
2 Euclidian distance in order to shed spurious mutations. To guarantee consistent results, only mutations with
3 a relative abundance >5% were accepted. The relative abundance of the remaining mutations was
4 transformed (centered Log-ratio transformation; 0 was replaced by 10^{-7}) and subjected to a second
5 hierarchical clustering step. Therefore, distances between mutations were calculated by applying the
6 complete-linkage algorithm based on square Euclidian distance. To identify distinct mutation constellations
7 a silhouette analysis was performed. Thereby, a silhouette coefficient is calculated for each clustering
8 produced by a critical distance threshold between 2 and 20. The clustering exhibiting the highest silhouette
9 coefficient is considered optimal and is further used. Finally, each cluster is variant-typed by introducing
0 its mutations into the reference-genome of SARS-CoV-2 and analysed by the Pangolin-lineage tool ⁵⁴.

1 **Nucleotide diversity.** The loFreq filtered variant call format files of all WWTP for which more than 18
2 timepoints were available with at least 2/3 of the genome covered with more than ten reads, were used to
3 extract variants with an allele frequency above 1%. Multiallelic SNPs were joined using Bcftools ⁸⁵ (v1.12).
4 SNPgenie ⁹⁵ (v2019.10.31) was used to calculate genome wide nucleotide diversity π . For samples with
5 more than 18 timepoints, the data was integrated with abundances of variant of concern as deduced by
6 *VaQuERo*. Significance of differences in π between different periods was tested using a Mann-Whitney U
7 test between π -values from ± 4 weeks around observed introduction of Alpha and from ± 4 weeks around
8 maximal observed dissemination of Alpha.

9 **Mobility data.** Statistics on registered accommodation in the hospitality industry (Beherbergungsstatistik
0 2019/20, https://www.statistik.at/atlas/?mapid=them_tourismus_winter_uebernachtungen), and the
1 statistic on registered commuting dynamics (Registerzählung 2011, Abgestimmte Erwerbsstatistik,
2 https://www.statistik.at/atlas/?mapid=them_bevoelkerung_pendler) were gathered from publicly available
3 databases. Of note, both registers do not represent the same period, and moreover, it is incalculable how
4 social adaptation (work from home and circumvention of hospitality restrictions) has affected the registered
5 behaviour. Therefore, mobility data based on the mobile communications network records from one
6 undisclosed mobile provider were furthermore used. A movement from region A to B in the origin

7 destination matrix was measured if a mobile phone was first logged in to A for at least 15 minutes and
8 afterwards logged in to B for at least 15 minutes. Detailed information about the origin destination matrix
9 calculation process have been described previously ⁹⁶. This movement data was used to deduce the
0 proportion of population who at least once per day leaves the catchment area in the direction of another
1 area or even another federal state. Therefore, the number of movements leaving the districts which are
2 assigned to the WWTP were counted for each catchment. The retrieved statistics were normalized to the
3 number of people connected to the sewage system and log₁₀ transformed.

4 References

- 5 1. Nicola, M. *et al.* The socio-economic implications of the coronavirus pandemic (COVID-19): A
6 review. *International Journal of Surgery* vol. 78 185–193 (2020).
- 7 2. Josephson, A., Kilic, T. & Michler, J. D. Socioeconomic impacts of COVID-19 in low-income
8 countries. *Nature Human Behaviour* **5**, 557–565 (2021).
- 9 3. Case, J. B., Winkler, E. S., Errico, J. M. & Diamond, M. S. On the road to ending the COVID-19
0 pandemic: Are we there yet? *Virology* **557**, 70–85 (2021).
- 1 4. Krammer, F. SARS-CoV-2 vaccines in development. *Nature* vol. 586 516–527 (2020).
- 2 5. Harvey, W. T. *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews*
3 *Microbiology* vol. 19 409–424 (2021).
- 4 6. Callaway, E. Heavily mutated coronavirus variant puts scientists on alert. *Nature* (2021).
- 5 7. Martin, D. P. *et al.* The emergence and ongoing convergent evolution of the N501Y lineages
6 coincides with a major global shift in the SARS-CoV-2 selective landscape. *medRxiv : the preprint*
7 *server for health sciences* (2021) doi:10.1101/2021.02.23.21252268.
- 8 8. Valesano, A. L. *et al.* Temporal dynamics of SARS-CoV-2 mutation accumulation within and across
9 infected hosts. *PLoS Pathogens* **17**, (2021).
- 0 9. Truong, T. T. *et al.* Increased viral variants in children and young adults with impaired humoral
1 immunity and persistent SARS-CoV-2 infection: A consecutive case series. *EBioMedicine* **67**,
2 (2021).
- 3 10. Luo, R., Delaunay-Moisan, A., Timmis, K. & Danchin, A. SARS-CoV-2 biology and variants:
4 anticipation of viral evolution and what needs to be done. *Environmental Microbiology* **23**, 2339–
5 2363 (2021).
- 6 11. Chandler, J. C. *et al.* SARS-CoV-2 exposure in wild white-tailed deer (*Odocoileus virginianus*).
7 doi:10.1101/2021.07.29.454326.
- 8 12. Lucas, C. *et al.* Impact of circulating SARS-CoV-2 variants on mRNA vaccine-induced immunity.
9 *Nature* (2021) doi:10.1038/s41586-021-04085-y.
- 0 13. Furuse, Y. Genomic sequencing effort for SARS-CoV-2 by country during the pandemic.
1 *International Journal of Infectious Diseases* **103**, 305–307 (2021).
- 2 14. The COVID-19 Genomics UK (COG-UK) consortium. An integrated national scale SARS-CoV-2
3 genomic surveillance network. *The Lancet Microbe* vol. 1 e99–e100 (2020).
- 4 15. Treibel, T. A. *et al.* COVID-19: PCR screening of asymptomatic health-care workers at London
5 hospital. *The Lancet* **395**, 1608–1610 (2020).
- 6 16. Brito, A. F. *et al.* Global disparities in SARS-CoV-2 genomic surveillance. *medRxiv : the preprint*
7 *server for health sciences* (2021) doi:10.1101/2021.08.21.21262393.
- 8 17. Belman, S., Saha, S. & Beale, M. A. SARS-CoV-2 genomics as a springboard for future disease
9 mitigation in LMICs. *Nature Reviews Microbiology* (2021) doi:10.1038/s41579-021-00664-y.
- 0 18. Majid, F., Omer, S. B. & Khwaja, A. I. Optimising SARS-CoV-2 pooled testing for low-resource
1 settings. *The Lancet Microbe* vol. 1 e101–e102 (2020).
- 2 19. Larsen, D. A., Green, H., Collins, M. B. & Kmush, B. L. Wastewater monitoring, surveillance and
3 epidemiology: a review of terminology for a common understanding. *FEMS Microbes* **2**, (2021).
- 4 20. Sharma, V. K. *et al.* COVID-19 epidemiologic surveillance using wastewater. *Environmental*
5 *Chemistry Letters* vol. 19 1911–1915 (2021).
- 6 21. Foladori, P. *et al.* SARS-CoV-2 from faeces to wastewater treatment: What do we know? A review.
7 *Science of the Total Environment* **743**, (2020).
- 8 22. Chen, Y. *et al.* The presence of SARS-CoV-2 RNA in the feces of COVID-19 patients. *Journal of*
9 *Medical Virology* **92**, 833–840 (2020).
- 0 23. Bonanno Ferraro, G. *et al.* A State-of-the-Art Scoping Review on SARS-CoV-2 in Sewage Focusing
1 on the Potential of Wastewater Surveillance for the Monitoring of the COVID-19 Pandemic. *Food*
2 *and Environmental Virology* (2021) doi:10.1007/s12560-021-09498-6.
- 3 24. Hassard, F., Lundy, L., Singer, A. C., Grimsley, J. & di Cesare, M. Innovation in wastewater near-
4 source tracking for rapid identification of COVID-19 in schools. *The Lancet Microbe* vol. 2 e4–e5
5 (2021).

- 6 25. la Rosa, G. *et al.* SARS-CoV-2 has been circulating in northern Italy since December 2019: Evidence
7 from environmental monitoring. *Science of the Total Environment* **750**, (2021).
- 8 26. Martin, J. *et al.* Tracking SARS-CoV-2 in sewage: Evidence of changes in virus variant
9 predominance during COVID-19 pandemic. *Viruses* **12**, (2020).
- 0 27. Nemudryi, A. *et al.* Temporal Detection and Phylogenetic Assessment of SARS-CoV-2 in Municipal
1 Wastewater. *Cell Reports Medicine* **1**, (2020).
- 2 28. Wurtzer, S. *et al.* Monitoring the propagation of SARS CoV2 variants by tracking identified mutation
3 in wastewater using specific RT-qPCR. *medRxiv* 2021.03.10.21253291 (2021)
4 doi:10.1101/2021.03.10.21253291.
- 5 29. Agrawal, S., Orschler, L. & Lackner, S. Long-term monitoring of SARS-CoV-2 RNA in wastewater
6 of the Frankfurt metropolitan area in Southern Germany. *Scientific reports* **11**, 5372 (2021).
- 7 30. Peccia, J. *et al.* Measurement of SARS-CoV-2 RNA in wastewater tracks community infection
8 dynamics. *Nature Biotechnology* **38**, 1164–1167 (2020).
- 9 31. Karthikeyan, S. *et al.* High-Throughput Wastewater SARS-CoV-2 Detection Enables Forecasting of
0 Community Infection Dynamics in San Diego County. *mSystems* **6**, (2021).
- 1 32. Izquierdo-Lara, R. *et al.* Monitoring SARS-CoV-2 circulation and diversity through community
2 wastewater sequencing, the netherlands and belgium. *Emerging Infectious Diseases* **27**, 1405–1415
3 (2021).
- 4 33. Crits-Christoph, A. *et al.* Genome Sequencing of Sewage Detects Regionally Prevalent SARS-CoV-
5 2 Variants. (2021) doi:10.1128/mBio.
- 6 34. Agrawal, S. *et al.* A pan-European study of SARS-CoV-2 variants in wastewater 2 under the EU
7 Sewage Sentinel System. doi:10.1101/2021.06.11.21258756.
- 8 35. Bar-Or, I. *et al.* Detection of SARS-CoV-2 variants by genomic analysis of wastewater samples in
9 Israel. *Science of the Total Environment* **789**, (2021).
- 0 36. Fontenele, S. *et al.* High-throughput sequencing of SARS-CoV-2 in wastewater provides insights
1 into 1 circulating variants 2 3 Rafaela. doi:10.1101/2021.01.22.21250320.
- 2 37. Fuqua, J. L. *et al.* A rapid assessment of wastewater for genomic surveillance of SARS-CoV-2
3 variants at sewershed scale in Louisville, KY. *medRxiv: the preprint server for health sciences*
4 (2021) doi:10.1101/2021.03.18.21253604.
- 5 38. Jahn, K. *et al.* Detection of SARS-CoV-2 variants in Switzerland by genomic analysis of wastewater
6 samples. doi:10.1101/2021.01.08.21249379.
- 7 39. Pechlivanis, N. *et al.* Detecting SARS-CoV-2 lineages and mutational load in municipal wastewater;
8 a use-case in the metropolitan area of Thessaloniki, Greece. 1–14
9 doi:10.1101/2021.03.17.21252673.
- 0 40. Smyth, D. S. *et al.* Detection of Mutations Associated with Variants of Concern Via High
1 Throughput Sequencing of SARS-CoV-2 Isolated from NYC Wastewater 2 3.
2 doi:10.1101/2021.03.21.21253978.
- 3 41. Crits-Christoph, A. *et al.* Genome Sequencing of Sewage Detects Regionally Prevalent SARS-CoV-
4 2 Variants. (2021) doi:10.1128/mBio.
- 5 42. Agrawal, S., Orschler, L. & Lackner, S. Long-term monitoring of SARS-CoV-2 RNA in wastewater
6 of the Frankfurt metropolitan area in Southern Germany. *Scientific reports* **11**, 5372 (2021).
- 7 43. Izquierdo-Lara, R. *et al.* Monitoring SARS-CoV-2 circulation and diversity through community
8 wastewater sequencing, the netherlands and belgium. *Emerging Infectious Diseases* **27**, 1405–1415
9 (2021).
- 0 44. la Rosa, G. *et al.* Rapid screening for SARS-CoV-2 variants of concern in clinical and environmental
1 samples using nested RT-PCR assays targeting key mutations of the spike protein. *Water Research*
2 **197**, (2021).
- 3 45. Nemudryi, A. *et al.* Temporal Detection and Phylogenetic Assessment of SARS-CoV-2 in Municipal
4 Wastewater. *Cell Reports Medicine* **1**, (2020).
- 5 46. Prado, T. *et al.* Wastewater-based epidemiology as a useful tool to track SARS-CoV-2 and support
6 public health policies at municipal level in Brazil. *Water Research* **191**, (2021).

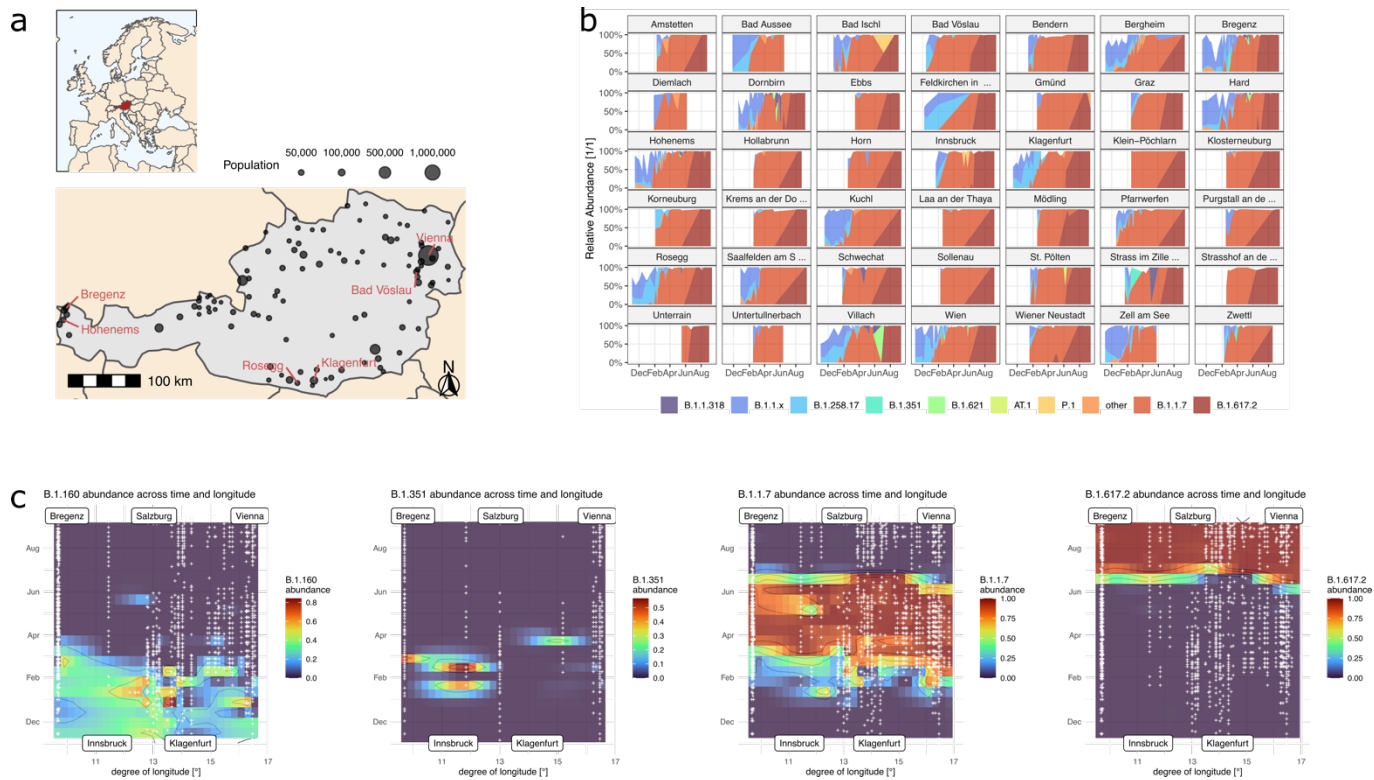
- 7 47. Rimoldi, S. G. *et al.* Presence and infectivity of SARS-CoV-2 virus in wastewaters and rivers.
8 *Science of the Total Environment* **744**, (2020).
- 9 48. Agrawal, S., Orschler, L. & Lackner, S. Metatranscriptomic Analysis Reveals SARS-CoV-2
0 Mutations in Wastewater of the Frankfurt Metropolitan Area in Southern Germany. *Microbiology*
1 *Resource Announcements* **10**, (2021).
- 2 49. Jahn, K. *et al.* Detection and surveillance of SARS-CoV-2 genomic variants in wastewater.
3 doi:10.1101/2021.01.08.21249379.
- 4 50. Huisman, J. S. *et al.* Wastewater-based estimation of the effective reproductive number of SARS-
5 CoV-2. doi:10.1101/2021.04.29.21255961.
- 6 51. Kreidl, P. *et al.* Emergence of coronavirus disease 2019 (COVID-19) in Austria. *Wiener Klinische*
7 *Wochenschrift* **132**, 645–652 (2020).
- 8 52. Popa, A. *et al.* Genomic epidemiology of superspreading events in Austria reveals mutational
9 dynamics and transmission properties of SARS-CoV-2. *Sci. Transl. Med* vol. 12
0 <https://www.science.org> (2020).
- 1 53. Haug, N. *et al.* Ranking the effectiveness of worldwide COVID-19 government interventions.
2 *Nature Human Behaviour* **4**, 1303–1312 (2020).
- 3 54. O’Toole, Á. *et al.* Assignment of epidemiological lineages in an emerging pandemic using the
4 pangolin tool. *Virus Evolution* **7**, (2021).
- 5 55. Klikovits, J. *et al.* *GISAID Österreich-Report Nr.*
6 *6* (https://www.ages.at/download/0/0/fbdce8f6e79df96cf599d90959e117034e60d123/fileadmin/AGES2015/Wissen-Aktuell/COVID19/GISAID-Report_Nummer-6.pdf). (2021).
- 7 56. Özkan, E. *et al.* High-throughput Mutational Surveillance of the SARS-CoV-2 Spike Gene.
8 doi:10.1101/2021.07.22.21259587.
- 9 57. Hasell, J. *et al.* A cross-country database of COVID-19 testing. *Scientific Data* **7**, (2020).
- 1 58. Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID’s innovative contribution to
2 global health. *Global Challenges* **1**, 33–46 (2017).
- 3 59. Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data – from vision to
4 reality. *Eurosurveillance* vol. 22 (2017).
- 5 60. Cragg, J. G. *Some Statistical Models for Limited Dependent Variables with Application to the*
6 *Demand for Durable Goods*. vol. 39 (1971).
- 7 61. van Poelvoorde, L. A. E. *et al.* Strategy and Performance Evaluation of Low-Frequency Variant
8 Calling for SARS-CoV-2 Using Targeted Deep Illumina Sequencing. *Frontiers in Microbiology* **12**,
9 (2021).
- 0 62. Itokawa, K., Sekizuka, T., Hashino, M., Tanaka, R. & Kuroda, M. Disentangling primer interactions
1 improves SARS-CoV-2 genome sequencing by multiplex tiling PCR. *PLoS ONE* **15**, (2020).
- 2 63. Nei, M. & Li, W.-H. Mathematical model for studying genetic variation in terms of restriction
3 endonucleases (molecular evolution/mitochondrial DNA/nucleotide diversity). *Genetics* **76**, 5269–
4 5273 (1979).
- 5 64. Been, F. *et al.* Population normalization with ammonium in wastewater-based epidemiology:
6 Application to illicit drug monitoring. *Environmental Science and Technology* **48**, 8162–8169
7 (2014).
- 8 65. Cori, A., Ferguson, N. M., Fraser, C. & Cauchemez, S. A new framework and software to estimate
9 time-varying reproduction numbers during epidemics. *American Journal of Epidemiology* **178**,
0 1505–1512 (2013).
- 1 66. Campbell, F. *et al.* Increased transmissibility and global spread of SARSCoV- 2 variants of concern
2 as at June 2021. *Eurosurveillance* **26**, 1–6 (2021).
- 3 67. Davies, N. G. *et al.* Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in
4 England. *Science* **372**, (2021).
- 5 68. Washington, N. L. *et al.* Emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United
6 States. *Cell* **184**, 2587-2594.e7 (2021).

- 7 69. Karthikeyan, S. *et al.* Rapid, Large-Scale Wastewater Surveillance and Automated Reporting System
8 Enable Early Detection of Nearly 85% of COVID-19 Cases on a University Campus. **6**, 793–814
9 (2021).
- 0 70. Calderón-Franco, D., Orschler, L., Lackner, S., Agrawal, S. & Weissbrodt, D. G. Monitoring SARS-
1 CoV-2 in sewage: Toward sentinels with analytical accuracy. *Science of the Total Environment* **804**,
2 (2022).
- 3 71. Chan, M. C. W. *et al.* Seasonal influenza a virus in feces of hospitalized adults. *Emerging Infectious*
4 *Diseases* **17**, 2038–2042 (2011).
- 5 72. Pogka, V. *et al.* Laboratory surveillance of polio and other enteroviruses in high-risk populations
6 and environmental samples. *Applied and Environmental Microbiology* **83**, (2017).
- 7 73. Hong, P. Y. *et al.* Estimating the minimum number of SARS-CoV-2 infected cases needed to detect
8 viral RNA in wastewater: To what extent of the outbreak can surveillance of wastewater tell us?
9 *Environmental Research* **195**, (2021).
- 0 74. Lynch, M., Bost, D., Wilson, S., Maruki, T. & Harrison, S. Population-genetic inference from
1 pooled-sequencing data. *Genome Biology and Evolution* **6**, 1210–1218 (2014).
- 2 75. Stern, A. *et al.* The unique evolutionary dynamics of the SARS-CoV-2 Delta variant-2 sequencing.
3 doi:10.1101/2021.08.05.21261642.
- 4 76. Melnick, J. L. Poliomyelitis virus in urban sewage in epidemic and in non-epidemic times. *American*
5 *Journal of Epidemiology* **45**, 240–253 (1947).
- 6 77. Ye, Y., Ellenberg, R. M., Graham, K. E. & Wigginton, K. R. Survivability, Partitioning, and
7 Recovery of Enveloped Viruses in Untreated Municipal Wastewater. *Environmental Science and*
8 *Technology* **50**, 5077–5085 (2016).
- 9 78. Wu, F. *et al.* SARS-CoV-2 Titers in Wastewater Are Higher than Expected from Clinically
0 Confirmed Cases. *mSystems* **5**, (2020).
- 1 79. Research Use Only 2019–Novel Coronavirus (2019-nCoV) Real-time RT-PCR Primers and Probes.
2 CDC (2020).
- 3 80. Markt, R. *et al.* Detection and stability of SARS-CoV-2 fragments in wastewater: Impact of storage
4 temperature. *Pathogens* **10**, (2021).
- 5 81. Bushnell, B., Rood, J. & Singer, E. BBMerge – Accurate paired shotgun read merging via overlap.
6 *PLoS ONE* **12**, (2017).
- 7 82. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform.
8 *Bioinformatics* **25**, 1754–1760 (2009).
- 9 83. Grubaugh, N. D. *et al.* An amplicon-based sequencing framework for accurately measuring intrahost
0 virus diversity using PrimalSeq and iVar. *Genome Biology* **20**, (2019).
- 1 84. Wilm, A. *et al.* LoFreq: A sequence-quality aware, ultra-sensitive variant caller for uncovering cell-
2 population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Research* **40**,
3 11189–11201 (2012).
- 4 85. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and
5 population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993
6 (2011).
- 7 86. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide
8 polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-
9 3. *Fly* **6**, 80–92 (2012).
- 0 87. Cingolani, P. *et al.* Using *Drosophila melanogaster* as a model for genotoxic chemical mutational
1 studies with a new program, SnpSift. *Frontiers in Genetics* **3**, (2012).
- 2 88. Supplemental Material for A Better Lemon Squeezer? Maximum-Likelihood Regression With Beta-
3 Distributed Dependent Variables. *Psychological Methods* (2006) doi:10.1037/1082-
4 989x.11.1.54.supp.
- 5 89. Rigby, R. A. & Stasinopoulos, D. M. *Generalized additive models for location, scale and shape.*
6 *Appl. Statist* vol. 54 (2005).
- 7 90. Barndorff-Nielsen, E. & Jorgensen, B. Some Parametric Models on the Simplex. *Journal of*
8 *multivariate analysis* **39**, 106–116 (1991).

- 9 91. Lee S, Wolberg G & Shin S.Y. Scattered data interpolation with multilevel B-splines. *IEEE*
0 *Transactions on Visualization and Computer Graphics* **3**, (1997).
- 1 92. Wastewater-based estimation of the effective reproductive number of SARS-CoV-2.
2 doi:10.1101/2021.04.29.21255961.
- 3 93. Richter L, Trauner F, Schmid D & Stadlober E. *Aktuelle Schätzung des seriellen Intervalles von*
4 *COVID19, 2021, Österreich.*
5 <https://www.ages.at/download/0/0/840fb2b85eb49b8f744d8d971614125ae197ab10/fileadmin/AG>
6 *ES2015/Wissen-Aktuell/COVID19/serial_interval_update2021_2021-06-14.pdf* (2021).
- 7 94. Özkan, E. *et al.* High-throughput Mutational Surveillance of the SARS-CoV-2 Spike Gene.
8 doi:10.1101/2021.07.22.21259587.
- 9 95. Nelson, C. W., Moncla, L. H. & Hughes, A. L. SNPGenie: Estimating evolutionary parameters to
0 detect natural selection using pooled next-generation sequencing data. *Bioinformatics* **31**, 3709–
1 3711 (2015).
- 2 96. Heiler, G. *et al.* Country-wide Mobility Changes Observed Using Mobile Phone Data during
3 COVID-19 Pandemic. in *Proceedings - 2020 IEEE International Conference on Big Data, Big Data*
4 *2020* 3123–3132 (Institute of Electrical and Electronics Engineers Inc., 2020).
5 doi:10.1109/BigData50022.2020.9378374.
6
7

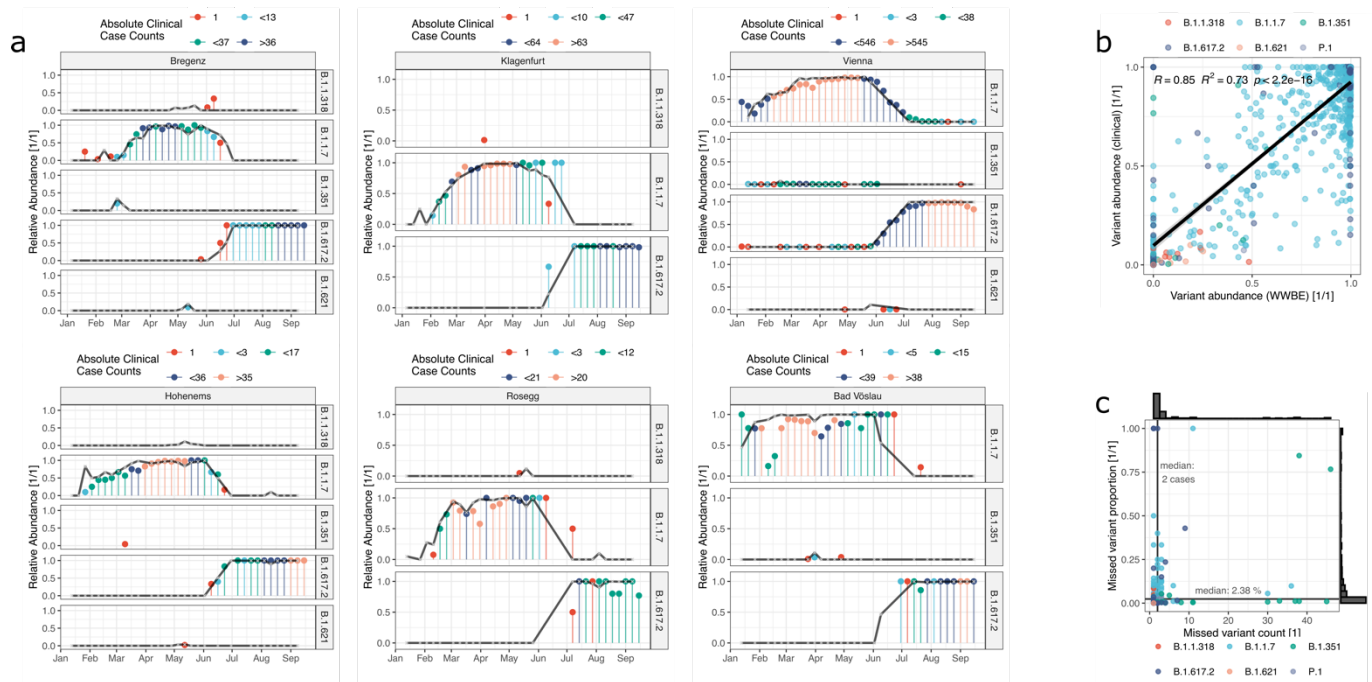
8 **Figures**

9
0



1

2 **Fig. 1 | Variant Quantification.** **a**, Location of Austrian WWTP considered. WWTP labelled with names
 3 are used to showcase analysis details throughout the manuscript. **b**, Results of the *VaQuERo* variant
 4 quantification analysis for the ten most abundant variants for all WWTP with more than 10 timepoints. **c**,
 5 Spatiotemporal spread of selected variants. Time is projected on the vertical axis, longitude is projected on
 6 the horizontal axis, colour indicates relative variant abundance, B.1.160, Beta (B.1.351), Alpha (B.1.1.7),
 7 Delta (B.1.617.2), from left to right, as deduced from the measured points and interpolated with B-splines.
 8 White crosses indicate measured data points.



9

0 **Fig. 2 | Validation.** Comparison between relative variant abundance as deduced from WW variant
 1 surveillance and from the aggregated statistics of the epidemiological case-based variant monitoring
 2 program by the Austrian Agency for Health and Food Safety. Data are aggregated over a week and
 3 catchment areas. **a**, Comparison in detail for a selection of six WWTP from west, central and eastern
 4 Austria, for each a small and large population size are selected. Black line represents the signal from the
 5 WW. The capped bars (lollipop plot) represent the results of the epidemiological monitoring program,
 6 whereas the respective colour indicates the range of absolute case numbers for the respective catchment
 7 area and week. The chosen breaks correspond to the 0, 0.2, 0.5, 0.75 quantile of the observed zero-truncated
 8 absolute case count distribution. **b**, depicting the overall correlation across all detected variants and all
 9 surveyed WWTP. **c**, depicting the absolute case counts and the relative case count for all data points for
 0 which no corresponding variant specific signal was detected in the WW. Black bars and depict distribution
 1 for both variables, black line the corresponding median value.

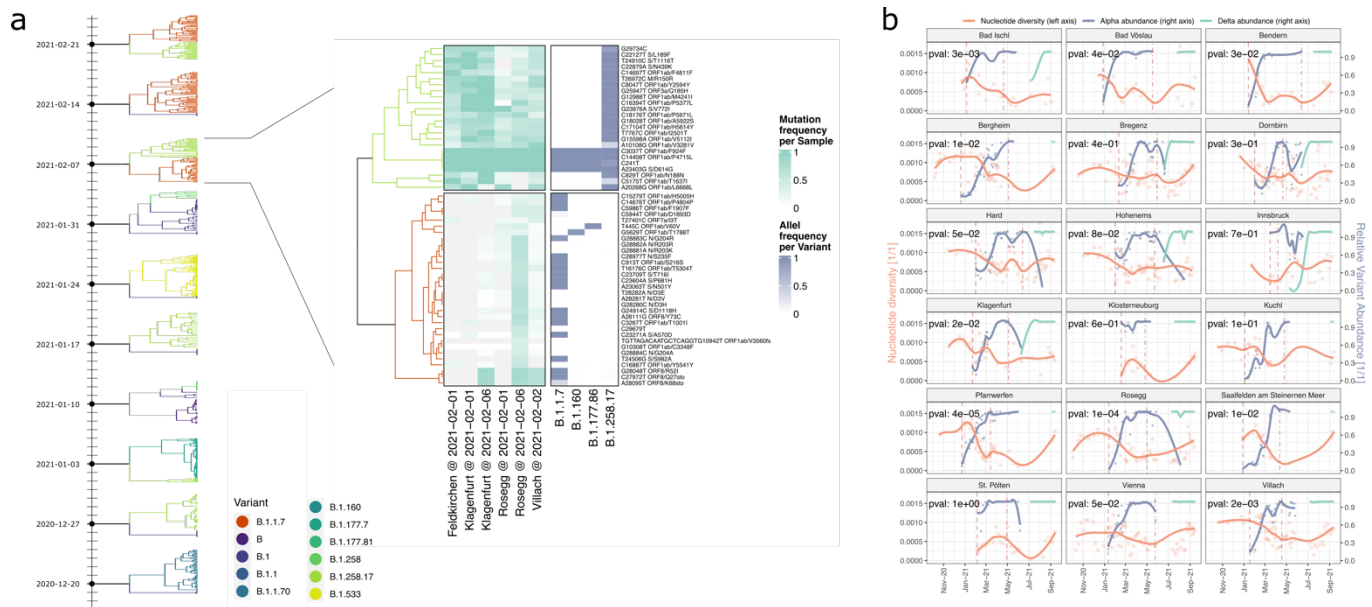
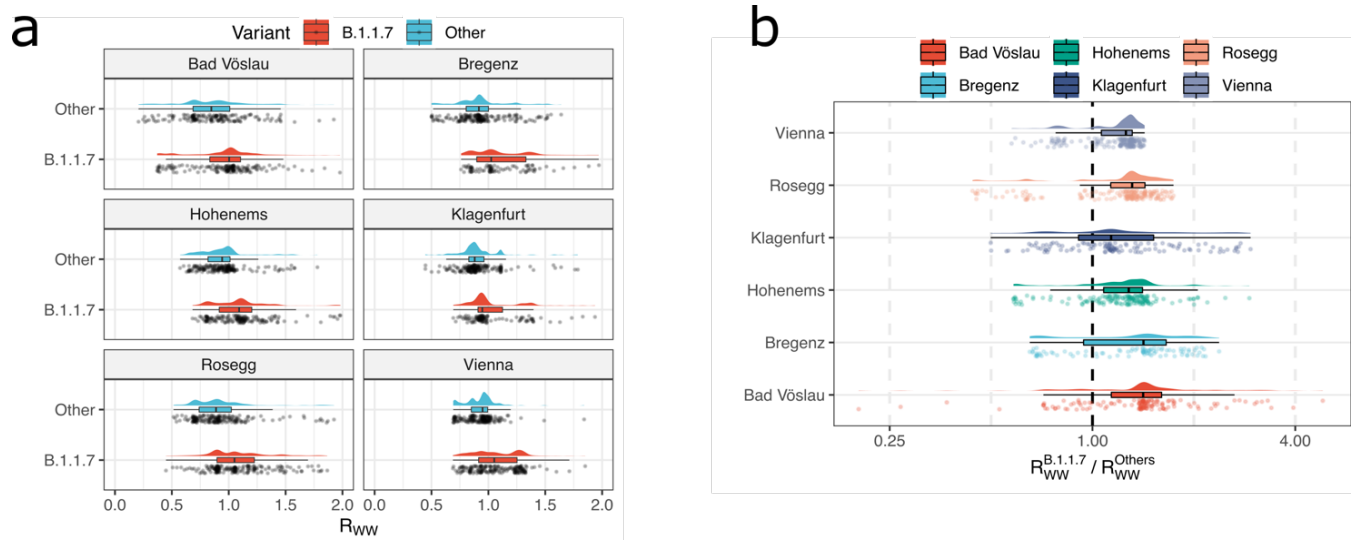


Fig. 3 | Analysis of mutation patterns. **a**, Mutation constellations in tree representation (left side) as agglomerate according to their mutation frequencies. Samples from Carinthia were jointly analysed per sampling week in the time period from December 2020 to February 2021. The first constellation enriched with Alpha (B.1.1.7) mutations is showcased with the frequencies of the mutations as observed in the samples used for deconvolution (heatmap, left side) and the frequency of the respective mutation in all samples of a specific variant as deposited in GISAID (heatmap, right side). Mutations of both constellations were traced over time using *VaQuERO* (Extended Data Fig. 6). **b**, Comparison of the relative abundance of Alpha (B.1.1.7) and Delta (B.1.617.2) and the observed nucleotide diversity π in the WW samples. Time course data smoothed by locally weighted regression (loess). Upper left corner indicates p-value that nucleotide diversity is higher around the first observation of Alpha (B.1.1.7) (dashed vertical line) compared to timepoint of maximal Alpha (B.1.1.7) abundance (dot-dashed vertical line) based on a Mann-Whitney U test.

5

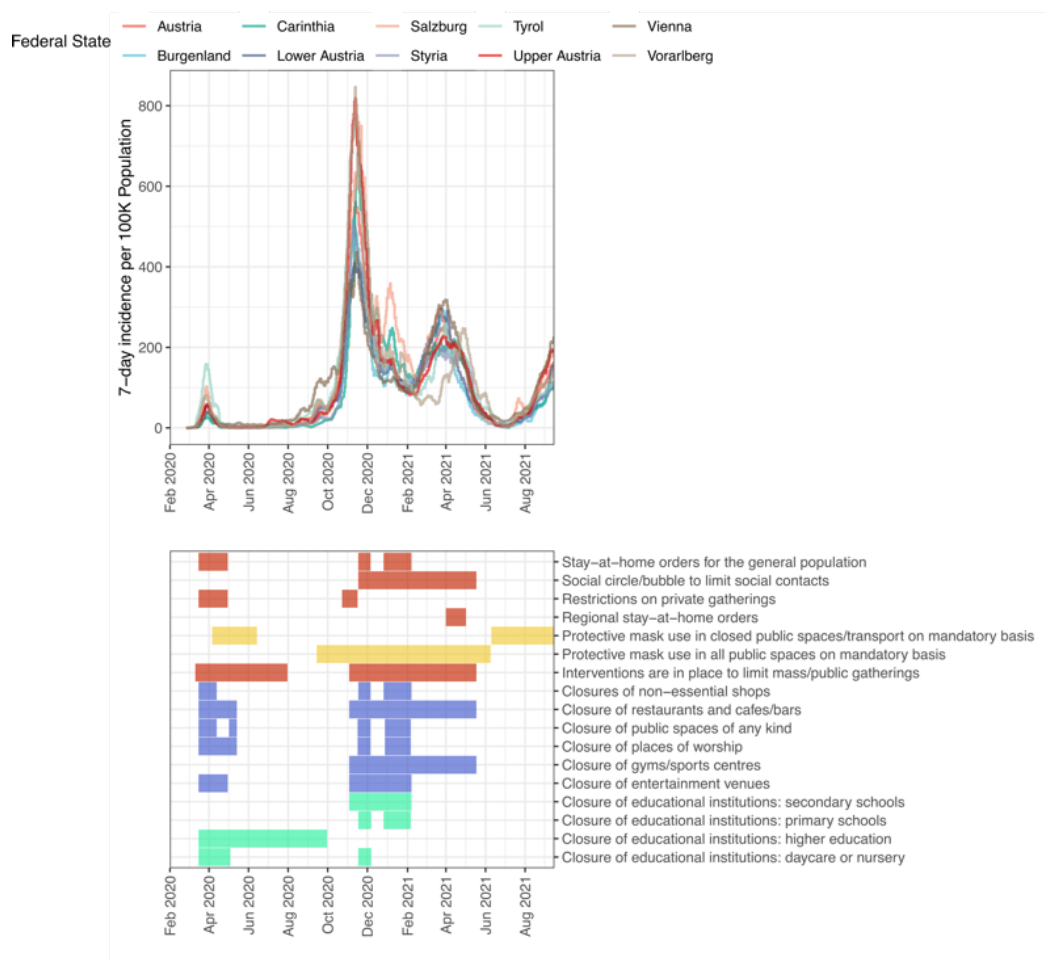


6

7 **Fig. 4 | Quantitative trend analysis of virus load interlaced with variant quantification.** a, Distribution
 8 of R_{ww} for the variant Alpha (B.1.1.7) and all other variants as calculated by an integration of absolute case
 9 estimate by RT-qPCR and variant estimate by sequencing, showcased for six WWTP across the time period
 0 from January to June, 2021. b, The ratios of the variant specific R_{ww} from the same timepoint and the same
 1 WWTP exhibit a systematic shift to the values >1 , with a median value of 1.33, indicating an 33% increased
 2 transmissibility of Alpha (B.1.1.7) over other competing variants subjected to equal environment.

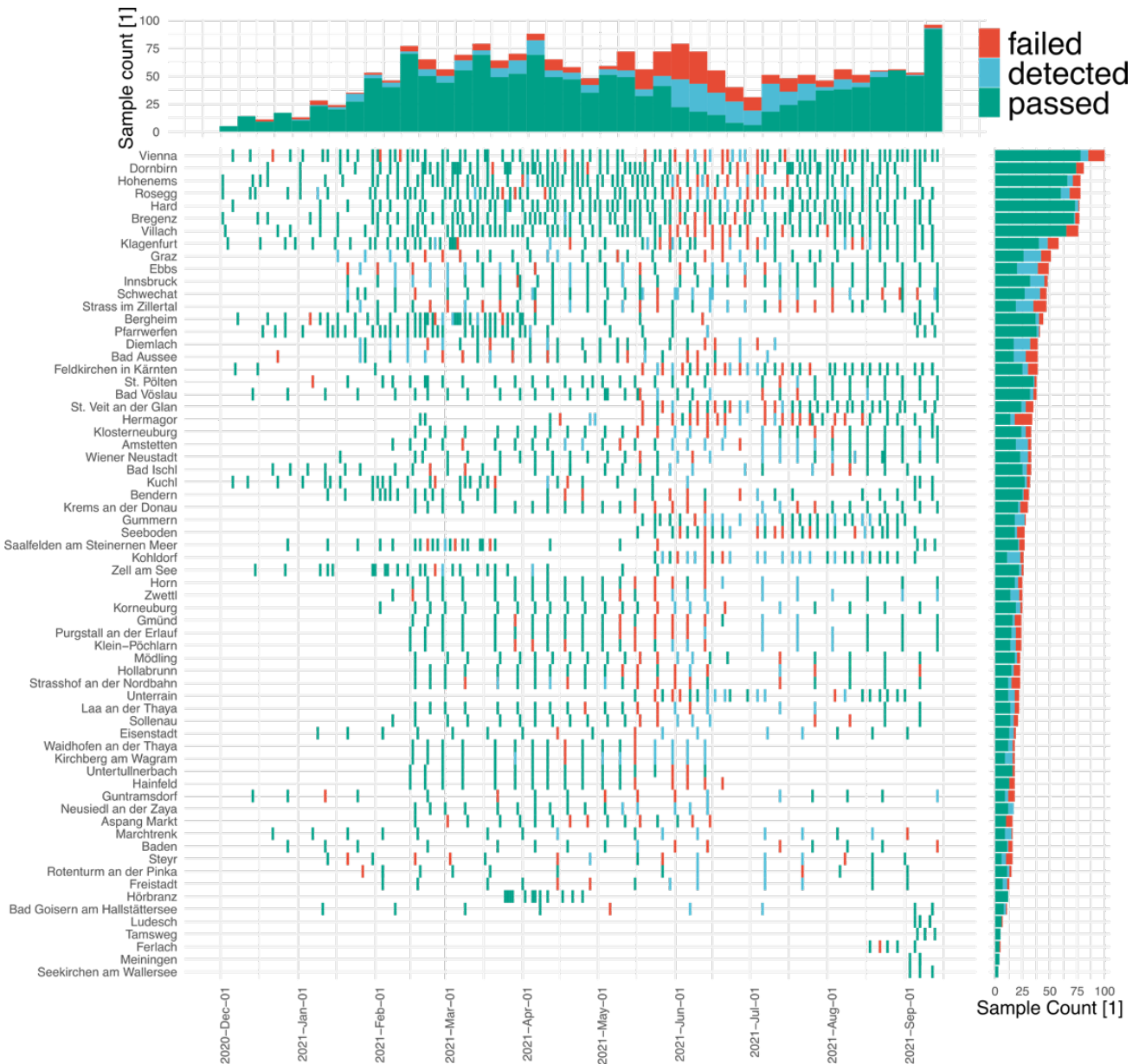
3 **Extended Data**

4



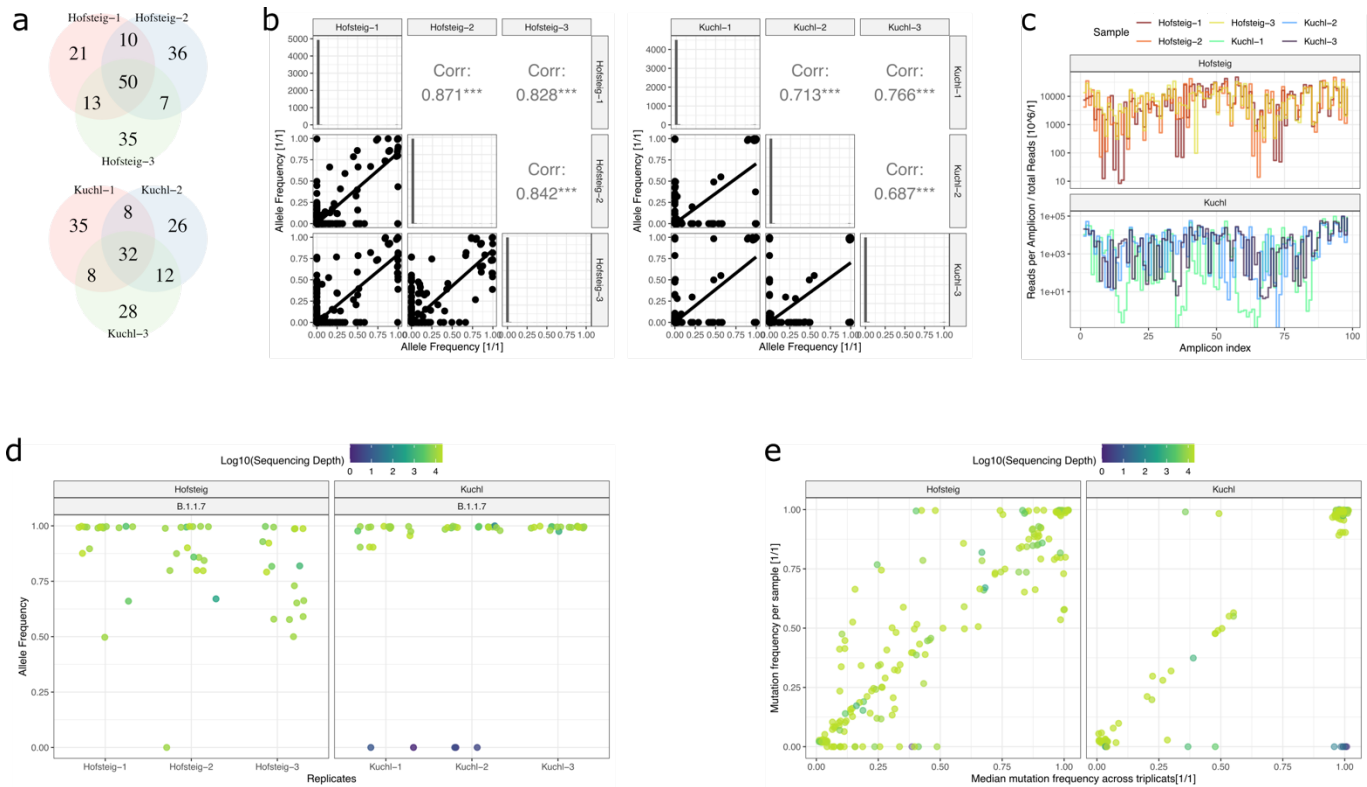
5

6 **Extended Data Fig. 1 | Course of the COVID-19 pandemic in Austria.** Upper panel depicts the average
 7 daily cases per 100K population in last 7 days for the course of the COVID-19 pandemic for each Austrian
 8 province and for all of Austria (<https://covid19-dashboard.ages.at>). The lower panel indicates the non-
 9 pharmaceutical interventions in the same time period (adapted from
 0 https://www.ecdc.europa.eu/sites/default/files/documents/response_graphs_data_2021-10-21.csv).



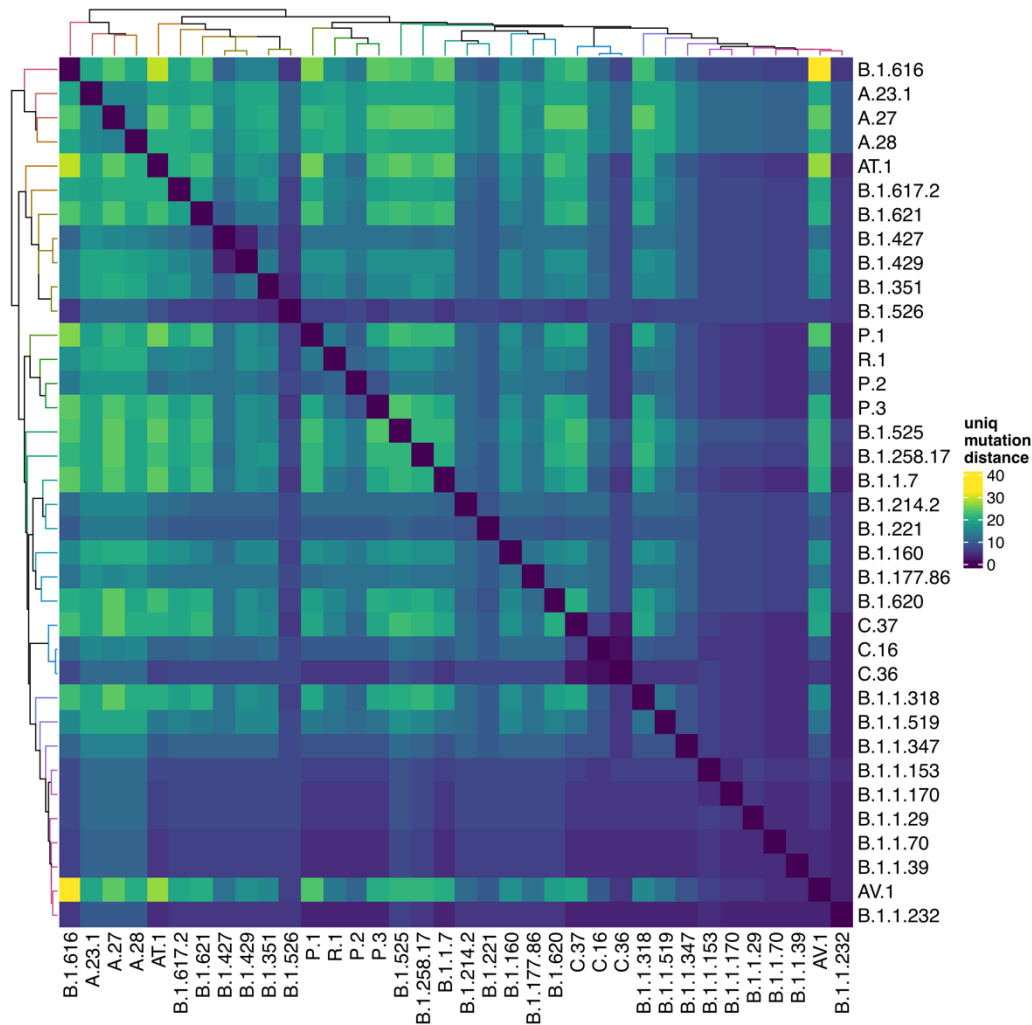
1

2 **Extended Data Fig. 2 | Data collection.** Graphic representation of all sequenced samples from WWTP
3 with more than three timepoints. Sample date and location, and sequencing success is indicated. Samples
4 are classified into *passed* and *detected* if 40% and 5% of the genome, respectively, is covered with more
5 than 10 reads. A coverage of more than 5% of the genome was considered as robust detection of the virus,
6 since at least three independent amplicons responded, but too little information is provided to reliably
7 characterize viral variants. Therefore, detected samples were not used further. If less than 5% of the genome
8 was covered the sequencing of the respective samples was categorized as *failed*. 70.7% of all samples were
9 classified as *passed*. Another 14.6% and 14.7% were classified as *detected* and *failed*, respectively.



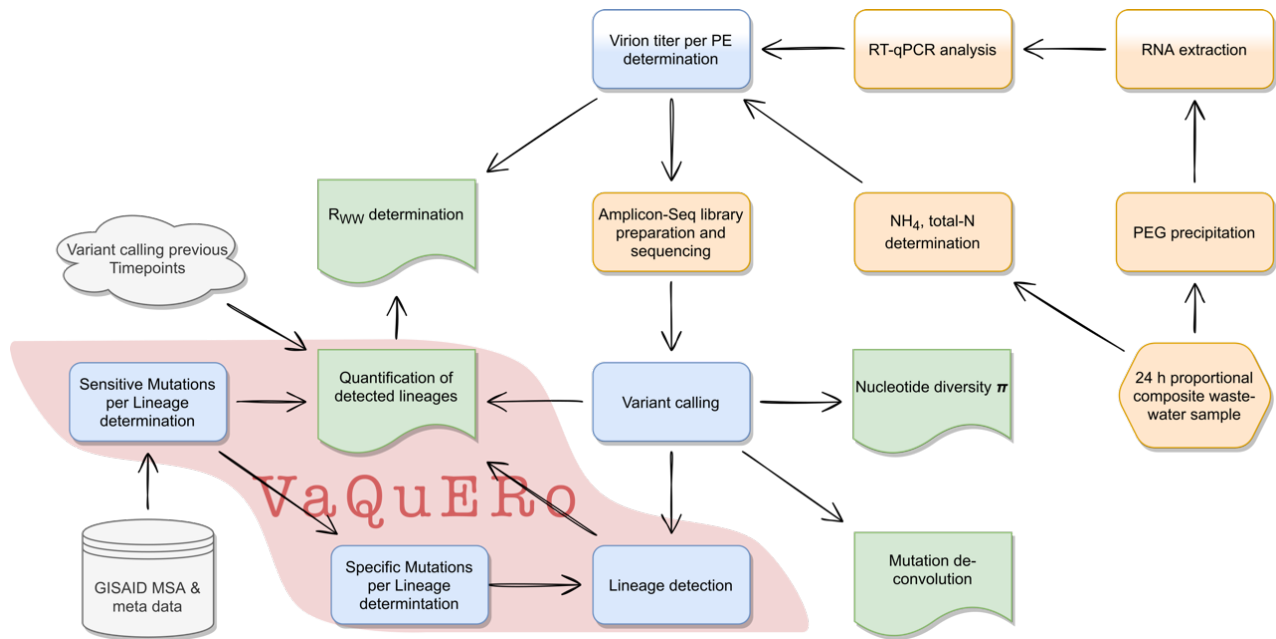
0

1 **Extended Data Fig. 3 | Reproducibility analysis.** Analysis of triplicate sample sets collected on the same
2 day at two different WWTP and processed independently. **a**, Reproducibility analysis for detection of
3 mutations. Venn diagram over the sets of mutations observed with a frequency greater zero in an individual
4 sample for all mutations called in any of the triplicates with an allele frequency greater 2%. **b**, Analysis of
5 triplicate samples with respect to the qualitative reproducibility of observed allele frequencies. Correlation
6 between allele frequencies is modest with correlation factors between 0.69 and 0.87. Generally, mutation
7 dropout in one sample contributes heavily to a reduced correlation factor. **c**, Reproducibility of normalized
8 read-coverage per amplicon between the triplicates. **d**, deduced allele frequencies of marker mutations of
9 the Alpha (B.1.1.7) variant (dominating variant according to WW and epidemiological data at the
0 corresponding time) in the triplicates. Point colour indicates read depth at the respective loci. **e**, scatter plot
1 depicting the median mutation frequency and the individual mutation frequencies for each detected
2 mutation (with median frequency greater than 2%) in the two replication sets. Samples at the baseline (y =
3 0) represent complete dropouts in one replicate. Point colour indicates read depth at the respective loci. As
4 in panel d, some mutation with expected presence according to the replicates are absent even though the
5 position is covered by a significant number of reads.



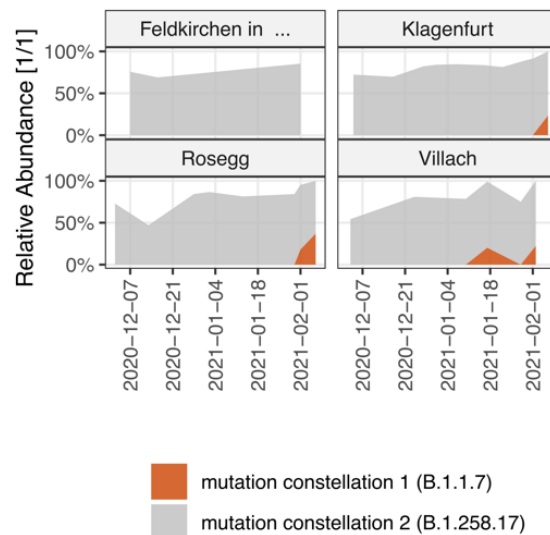
6

7 **Extended Data Fig. 4 | Variants of relevance.** Graphic representation of all variants of relevance whose
8 abundance in Austrian WW samples were examined. Variants which were defined by ECDC to be variants
9 of concern, variants of interest or variants under monitoring were included. Additionally, all variants which
0 were observed in more than ten Austrian patient samples were added. For each variant sensitive marker
1 mutations were defined based on their occurrence (in >85% of all samples of the respective lineage) and
2 specificity (unique to one lineage). The heatmap depicts the number of unique (specific) marker mutations
3 between each pair of variants of relevance.



4

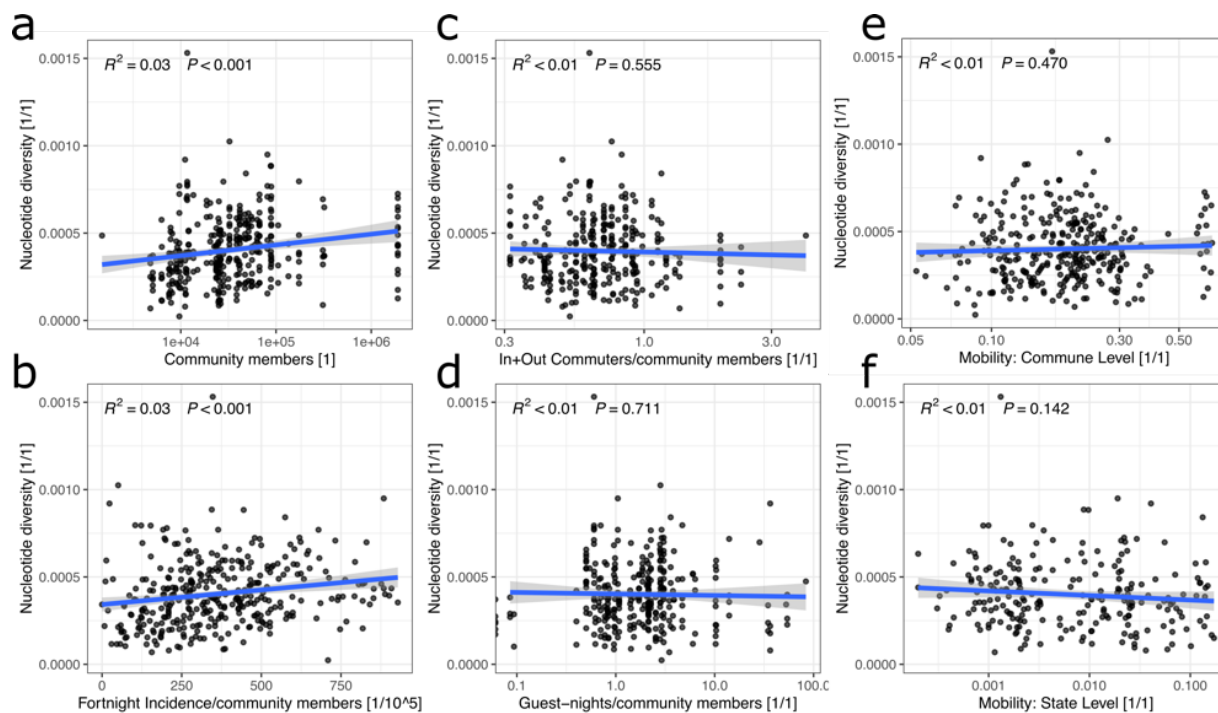
5 **Extended Data Fig. 5 | Analysis workflow.** Illustration of the main analysis workflow steps from the WW
6 sampling point, the wet lab procedures (orange panels), via the bioinformatics analysis steps (blue panels)
7 to the analysis results (green panels). Major read outs of the analysis are nucleotide diversity π , cluster of
8 correlated mutation constellations, variant quantification, and determination of reproduction number in the
9 catchment area. Analysis steps implemented in the software tool *VaQuERo* are highlighted by a red area.



0

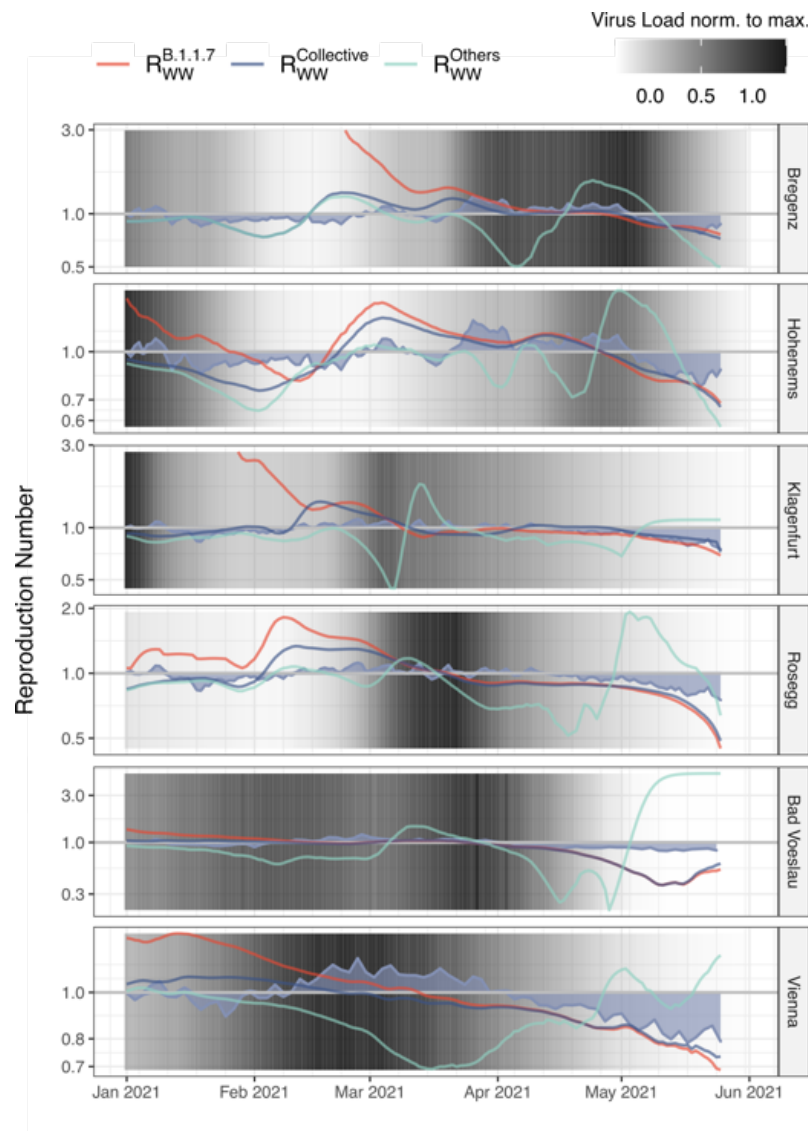
1 **Extended Data Fig. 6 | Mutation constellation deconvolution.** Mutations which are grouped into related
2 mutation constellation based on their mutation frequencies, as depicted in Fig. 3a, are used as input for
3 *VaQuERo*, instead of the normal GISAID trained variant defining mutations. The graphic illustrates the
4 geographic distribution and the dynamics of the de novo inferred mutation constellation enriched in Alpha
5 (B.1.1.7) associated mutations.

6



7

8 **Extended Data Fig. 7 | Virus population diversity and mobility.** Correlation analysis between nucleotide
9 diversity measure π of all samples with quantified Alpha variant above 95% and external data, describing
0 (from left to right) the number of people in the catchment area **a**, people connected to the WWTP in the
1 catchment area. **b**, the 14-day incidence rate per 100,000 people **c**, the proportion of in and out commuters
2 (as officially registered) to the total number of people in the catchment area **d**, the number of registered
3 over-night days in the hospitality industry in the catchment area **e**, the proportion of population leaving the
4 commune **f**, or the federal state, as observed by mobile communications network records.



5

6 **Extended Data Fig. 8 | Reproduction number R_{ww} deduced from WW.** Selected WWTP with the
7 deduced variant-specific reproduction number R_{ww} . The gauged virus load (cp/PE) as observed in the WW
8 is indicated with background shades. The blue area indicates the estimated R_{eff} per respective federal state
9 as deduced from epidemiological case monitoring.