RBD amplicon sequencing of wastewater reveals patterns of variant emergence and evolution

Xingwen Chen[1,2], John Balliew[3], Cici X. Bauer[1,2], Jennifer Deegan[1,2], Anna Gitter[1,2], Blake M. Hanson[1,2], Anthony W. Maresso[4], Michael J. Tisza[5], Catherine L. Troisi[1,2], Janelle Rios[1,2], Kristina D. Mena[1,2], Eric Boerwinkle[1,2], Fuqing Wu[1,2]*

Affiliations:

1: School of Public Health, University of Texas Health Science Center at Houston, TX, USA.

2: Texas Epidemic Public Health Institute (TEPHI), UTHealth Houston, Houston, TX, USA.

3: El Paso Water Utility, El Paso, TX, United States

4: TAILOR Labs, Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX, USA

5: The Alkek Center for Metagenomics and Microbiome Research, Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX, USA

Correspondence to: fuqing.wu@uth.tmc.edu.

Abstract

Rapid evolution of SARS-CoV-2 has resulted in the emergence of numerous variants, posing significant challenges to public health surveillance. Clinical genome sequencing, while valuable, has limitations in capturing the full epidemiological dynamics of circulating variants in the general population. This study utilized receptor-binding domain (RBD) amplicon sequencing of wastewater samples to monitor the SARS-CoV-2 community dynamics and evolution in El Paso, TX. Over 17 months, we identified 91 variants and observed waves of dominant variants transitioning from BA.2 to BA.2.12.1, BA.4&5, BQ.1, and XBB.1.5. Our findings demonstrated early detection of variants and identification of unreported outbreaks, while showing strong consistency with clinical genome sequencing data at the local, state, and national levels. Alpha diversity analyses revealed significant periodical variations, with the highest diversity observed in winter and the outbreak lag phases, likely due to lower competition among variants before the outbreak growth phase. The data underscores the importance of low transmission periods for rapid mutation and variant evolution. This study highlights the effectiveness of integrating RBD amplicon sequencing with wastewater surveillance in tracking viral evolution, understanding variant emergence, and enhancing public health preparedness.

Keywords: Receptor-binding domain, amplicon sequencing, Wastewater surveillance, SARS-CoV-2, variant emergence, viral evolution

Introduction

SARS-CoV-2, the virus responsible for COVID-19, has undergone significant evolution since its emergence, leading to waves of outbreaks by multiple variants such as Alpha, Beta, Gamma, Delta, and Omicrons. Each of these variants exhibits distinct characteristics in terms of transmissibility, pathogenicity, and ability to evade immune responses. A key driver of this evolution is the mutation of the spike-protein-encoding gene, particularly in the receptor-binding domain (RBD), which is crucial for the virus's ability to bind to host cell receptors and initiate infection[1–3]. Mutations in the RBD can enhance the virus's ability to spread and escape immune detection[4], making it a critical focus for variant tracking and public health monitoring. For example, the Delta variant harbors the L452R and T478K mutations which increases binding affinity to ACE2 receptor on host cells and contributes to reduced vaccine efficacy[5–9]. Understanding and tracking these variants are crucial for managing ongoing and future outbreaks.

Traditional surveillance methods, such as genome sequencing of clinically ascertained samples, have been instrumental in identifying and tracking SARS-CoV-2 variants. However, these methods are often time-consuming, expensive, and resource-intensive[10–12]. Moreover, clinical testing primarily captures data from symptomatic individuals, potentially missing large portions of the infected population who are asymptomatic or pre-symptomatic[13–16]. In contrast, wastewater surveillance has emerged as a promising approach to monitor community-level prevalence and diversity of SARS-CoV-2. By analyzing wastewater samples, one can detect viral RNA shed by infected individuals, regardless of their symptom status[17–19]. Wastewater surveillance offers several advantages, including non-invasiveness, cost-effectiveness, and the ability to capture data from a broad segment of the population. This method has now been widely implemented globally, providing early warnings of outbreaks and offering a comprehensive view of community-level viral transmission.

Wastewater sequencing is a crucial tool for early detection of viral variants. By collecting and analyzing the nucleic acids in wastewater using next-generation sequencing, researchers can identify mutations and new variants weeks before they appear in clinical samples[20–22]. This method allows continuous tracking of variant dynamics and provides a more complete picture of the viral landscape[23–26]. However, the complex nature of wastewater, containing a mixture of genetic material from various sources, complicates data analysis and interpretation, posing challenges in genome assembly, variant assignment/classification, and sequencing accuracy[27–32]. Typically, wastewater sequencing relies on tiled amplicon amplification to recover the whole

genome due to low concentrations. A simpler approach targets the spike genes, particularly the RBD, with specific target areas varying among studies [21,33–38]. While existing research has primarily focused on the detection and surveillance of variants, there has been limited exploration into the epidemiological dynamics and patterns of variant emergence and evolution within community wastewater.

In this study, we targeted the receptor-binding domain region and developed a streamlined bioinformatic pipeline to track the variant dynamics of SARS-CoV-2 over 17 months using wastewater samples from El Paso, a border city in west Texas, US, and Ciudad Juárez, Mexico. We compared our wastewater data with clinical genome sequencing results at local, state, and national levels to assess consistency and reliability. Additionally, we conducted epidemiological dynamic analyses to uncover patterns of variant evolution and emergence. By integrating RBD amplicon sequencing with wastewater surveillance, this research offers an innovative approach to understanding variant emergence and evolution, enriching the surveillance toolkit to anticipate and mitigate emerging threats in the post-pandemic era.

2. Materials and Methods

2.1 Wastewater sampling

Composite wastewater samples (24-hr) were collected weekly from March 14th, 2022, to August 8th, 2023, from three wastewater treatment plants (WWTPs) in the City of El Paso, Texas. The WWTPs A, B, and C serve 98,754, 125,462, and 128,003 residents, respectively. The daily total wastewater flow volume was measured and provided by El Paso Water. Samples were shipped overnight in ice boxes to the laboratory and processed on the day of receipt. The rest of the raw samples were stored at -80 °C.

2.2 Wastewater sample processing, RNA extraction, and RT-qPCR

Samples were processed based on previous methods[39–41]. Briefly, 50 mL wastewater samples were first centrifuged at 3000 g for 10 min at 4°C. The supernatant was collected and further filtered with MilliporeSigma™ Steriflip™ Sterile Disposable Vacuum Filter Units. Then, the 15 mL filtrates were concentrated to ~200uL with MilliporeSigma™ Amicon™ Ultra-15 Centrifugal Filter Units. The enriched samples were subject to RNA extraction with QIAamp Viral RNA Mini Kit according to the manufacturer's protocol. RNA was eluted with 100 uL nuclease-free water and stored at -20°C.

SARS-CoV-2 was tested and quantified using the US CDC N1 real-time PCR assay (Biorad, CFX96 C1000 Touch) using the following program: 50 °C 10 min for reverse transcription, 95 °C 20 s for RT inactivation and initial denaturation, and 48 cycles of denature (95 °C, 1 s) and anneal/extend (60 °C, 30 s). At least two negative controls were included in every PCR run. Pepper mild mottle virus (PMMoV) was also measured as an internal reference for sample processing and variations in wastewater flow and/or fecal materials, as previously reported [40,42]. Two technical replicates were performed for all RT-qPCR reactions. SARS-CoV-2 RNA concentrations were adjusted by the corresponding PMMoV concentrations in the sample using the method described in previous work[40,42]. The total viral load per capita was calculated for data comparison across sewersheds with the following formula:

$$Viral\ load = RNA\ concentraion \times Dilution\ factor \times Flow\ rate$$

$$Viral\ load\ per\ capita = \frac{Viral\ concentration}{Population\ covered\ by\ the\ WWTP} * 100{,}000$$

The Dilution Factor is the scale of enrichment during the experiment, which is 6.7 (100 uL/15 mL). This indicates that 15 mL of raw composite wastewater sample was concentrated into 100 uL of RNA.

2.3 Preparation of synthetic SARS-CoV-2 variant communities

The SARS-CoV-2 Wuhan-Hu strain and Omicron BA.4&5 genomes were purchased from TWIST Bioscience. Original stocks were diluted one thousand-fold before mixing with four different ratios: 10%:90%, 25%:75%, 50%:50%, and 90%:10%. 10 ul of each mixture sample were then subjected to RNA extraction with QIAamp Viral RNA Mini Kit.

2.4 RBD amplicon amplification of wastewatwer samples

Primers for amplifying the receptor-binding domain (RBD) of the SARS-CoV-2 spike glycoprotein were provided in Table S1 and adapted from a previous study[20]. This amplicon amplifies 168 amino acids from 337 to 504 aa, covering almost the entire receptor-binding motif of SARS-CoV-2 (from 437 to 508 aa) and most receptor-binding domain (from 319 to 541 aa). Briefly, wastewater samples and positive control RNA were transcribed into cDNA using SuperScript™ IV Reverse Transcriptase (ThermoFisher Scientific, USA). The RBD amplicon was then amplified from the cDNA using primers S008 and S012 in a first PCR with Q5® High-Fidelity 2X Master Mix (New England Biolabs, USA), under the following conditions: polymerase activation at 95°C for 2 minutes, followed by 30 cycles of denaturation at 95°C for 1 second,

annealing at 72°C for 30 seconds, and extension at 60°C for 30 seconds. A second PCR was performed to add Illumina overhangs to the RBD amplicon using primers S009 and S013, with the same conditions but for 40 cycles. The PCR product was purified using AMPure XP beads and eluted with 30 µL of nuclease-free water. For testing the protocol, synthetic SARS-CoV-2 Wuhan-Hu strain and Omicron BA.4&5 genomes from TWIST Bioscience were diluted 1,000-fold, and four mixtures were prepared in ratios of 10%:90%, 25%:75%, 50%:50%, and 90%:10%. 10 µL of each mixture was used for RNA extraction, amplicon amplification, library preparation, and sequencing.

2.5 Library preparation and sequencing data analysis

The purified PCR amplicons were dual-indexed using NEBNext Multiplex Oligos (New England Biolabs, USA) for Illumina and further purified using AMPure XP beads. Libraries were pooled based on their concentrations and sequenced on a 300 paired-end Miseq run at the ATGC Facility at MD Anderson Cancer Center. The raw sequencing data in fastq format were de-multiplexed. DADA2 package (version 4.3.0)[43] was used to filter and trim sequences, merge paired ends, and remove chimeras in Rstudio version 4.3.2. The unique amplicon sequence variants (ASVs) were aligned and identified using the BLAST online tool[44] using the Betacoronavirus database. For each ASV, the top 10 BLAST results containing the Pango Lineage name with the highest alignment scores (ranging from 99% to 100%) were downloaded, and the most frequent variant was identified as the variant for that sequence. The relative abundance of each variant was calculated by the ratio of the count of each variant to the total sequence count. Variants with relative abundance higher than 10% on at least one sample were designated as major variants. The remaining variants were grouped as 'others' for plotting and visualization purposes.

2.6 COVID-19 cases and genome data

The reported COVID-19 case data by lab-testing diagnostics was downloaded from healthdata.gov, and the 7-day rolling average was used for further analysis and data visualization. The Texas-specific data was extracted according to the state name label. SARS-CoV-2 genome data for the study period was obtained from the GISAID EpiCoV™ database using the GISAIDR (version 0.9.9) package in R[45]. The variants were assigned by Pango v.4.3 by GISAID. Only records with complete genomes were downloaded. The number of each variant per week was counted, and the relative abundance was calculated accordingly.

Results

3.1 RBD amplicon sequencing of wastewater reveals temporal dynamics of SARS-CoV-2 variants circulating in the city.

To validate the RBD amplicon sequencing approach for variant identification, we utilized synthetic mock variant communities. Four mock communities were prepared by mixing the original SARS-CoV-2 Wuhan-Hu strain (Wild Type) and the BA.4&5 strain at ratios of 10%:90%, 25%:75, 50%:50%, and 90%:10%. Each mock community underwent RNA extraction, reverse transcription, amplicon amplification, library preparation, and sequencing, followed by bioinformatic analysis. As a negative control, nuclease-free water was processed identically, except for RNA extraction. The sequencing results showed that nearly all reads (98.5%~100%) were accurately identified as either Wild Type or Omicron BA.4&5 strains across all four mock communities. The estimated relative abundances for each variant in the mock communities were 6.7%:92.3%, 21.9%:76.6%, 51.3%:47.5%, and 93.1%:6.9% (**Figure 1B**), closely matching the original mixing ratios with a deviation range of -3.1% to 3.1%. These results demonstrate that the RBD amplicon sequencing method is effective for identifying variants and estimating the composition and structure of SARS-CoV-2 variants in a sample.

Next, we applied the RBD amplicon sequencing approach to weekly wastewater samples collected from the three wastewater treatment plants. All samples tested positive for SARS-CoV-2. Consistent with clinically reported case data, viral load results revealed three major waves of infection in the city: from April 2022 to September 2022, from November 2022 to December 2022, and from February 2023 to the end of the experiment period (August 2023) (**Figure 1C and Figure S1**).

Across the 216 samples analyzed, 91 SARS-CoV-2 variants were identified, each with over 99% identity to the reported variant sequences. For visualization, we plotted the related abundance of the variants of interest (VOIs) listed by WHO and those with an abundance of 10% or higher in at least one wastewater sample. All other variants were grouped under 'others'. **Figures 1D-F** show the relative abundance of SARS-CoV-2 variants over the 17 months across the three sewersheds, highlighting the evolving composition of variant communities. Overall, each wastewater sample contained multiple variants, and we observed waves of dominant variants transitioning from BA.2 to BA.2.12.1, BA.4&5, BQ.1, and to XBB.1.5. While there were some differences, the overall variant profiles for each WWTP were similar, indicating that the spread and evolution of the virus followed a comparable trajectory across the three sewersheds

in the city. These results demonstrate that RBD amplicon sequencing of wastewater samples can be used to monitor the temporal dynamics and community structure of SARS-CoV-2 variants circulating in the population.

3.2 Consistency and novel insights from RBD amplicon sequencing data compared to clinical genome sequencing data.

We further validated the wastewater findings using reported clinical genome sequencing data from the USA, Texas, and El Paso. As shown in **Figure 1G**, the overall trends between the wastewater RBD amplicon sequencing data and clinical genome sequence data in Texas are consistent, with multiple waves of major variants emerging sequentially. Moreover, the composition of variants in wastewater samples is more diverse, suggesting that some variants circulating in the city may go undetected in clinical testing. We further compared the temporal dynamics of major variants during the study period: BA.2, BA.2.12.1, BA.4&5, BQ.1, and XBB.1.5 (**Figure 2**) in both wastewater and clinical sequencing data at the El Paso, Texas, and national levels. All five variant waves were consistently found in both datasets, showing similar transmission waves and relative abundance at the peaks. We also noticed some differences between Texas and national trends. For example, the BA.2 peak occurred later, and the BA.4&5 peak arrived earlier and lasted longer in Texas compared to the national data. However, wastewater data in El Paso closely matched the clinical sequencing data in Texas.

In addition, we found that some variants circulated in the population earlier than their prevalence observed in clinical tests. For example, the XBB.1.5 wave began in December 2022, as shown in both clinical and wastewater data (Figure 2). However, this variant was detectable in wastewater samples before the major wave, with small peaks observed in March and September 2022 (**Figure 2**). Similarly, the waves of EG.5.1 and XBB.1.16 started in May and March 2023, respectively (**Figure 3A and 3B**), yet these variants were detectable from the beginning of our experiment. These findings strengthen the existing idea that wastewater surveillance could provide early warnings to the community [28,46,47].

We also identified some uncommon variants that were not detected in the clinical data. Those variants, including EF.1.1, GD.1, EG.6.1, and FE.1.2, were barely noticeable in clinical records due to low abundance but were significant enough to exceed 10% in at least one wastewater sample (**Figure 3C and 3D, and Figure S2**). The EF.1.1 variant even had a small outbreak in all of the three sewersheds, with a relative abundance of up to 25% from October 2022 to

February 2023. The EG.6.1, FE.1.2, and GD.1 variants were detected throughout the experiment period, but very few records were found in clinical data. This discrepancy likely occurs because these variants have a moderate transmission rate but do not cause severe symptoms, leading to fewer clinical tests.

In summary, the results from RBD amplicon sequencing of wastewater samples are consistent with clinical genome sequencing data in capturing the variants' dynamics. This consistency highlights the effectiveness and accuracy of this approach in monitoring SARS-CoV-2 variants circulating in the sewershed. Additionally, the method allows for the early detection of emerging variants and provides a comprehensive view of variant diversity, emphasizing its importance for community-level surveillance efforts.

3.3 RBD amplicon sequencing data highlights off-season and lag-phase as crucial periods for variant emergency and evolution.

To better understand the epidemiological dynamics of SARS-CoV-2, we performed alpha diversity analyses of the variants identified from the RBD amplicon sequencing of wastewater samples. Specifically, we examined how variant diversity changes with the seasons, during high transmission periods (in outbreak waves) versus low transmission periods (out of wave), and across different outbreak phases including lag, growth, stationary, and decline. Metrics including Observed, Chao1, Shannon, and Inverse Simpson indices were used to measure the richness and evenness of SARS-CoV-2 variants in the wastewater samples. The Observed index counts the actual number of unique variants, Chao1 estimates total species richness, Shannon quantifies the diversity considering both abundance and evenness, and Inverse Simpson emphasizes the dominance of the most abundant variants.

Results revealed significant seasonal variations in alpha diversity in all four metrics (**Figure 4A**). An increasing trend of alpha diversity was observed from spring to winter, with the highest alpha diversity observed in winter, suggesting that many variants emerged during the winter months. And summer months exhibited the lowest diversity. Moreover, higher diversity was also observed during the low transmission (out of wave) period across the four metrics (**Figure 4B and 4C**). This suggests that during these low transmission periods, a more diverse pool of variants circulates at low levels in the population, potentially due to reduced selective pressures compared to the outbreak period. When grouping data into different outbreak phases, we found that the highest diversity emerged in the lag phase, followed by the decline and growth phases,

with the lowest diversity observed in the stationary phase (**Figure 4D and 4E**). This result suggests an interesting phenomenon: new variants mostly emerged not during the stationary phase, where the highest viral shedding and new cases occur, but in the lag phase, before a new outbreak starts. Beta diversity analysis showed no significant difference across the three sewersheds (**Figure 4F**). Together, these results showed that RBD amplicon sequencing of wastewater is a powerful tool for uncovering the patterns of variant emergence and evolution in the sewershed.

Discussion

One significant insight from this study is the ability of RBD amplicon sequencing data to reveal patterns of viral evolution and variant emergence. Previous studies using whole genome sequencing of SARS-CoV-2 in wastewater have highlighted the early-warning potential of this approach and its ability to uncover variant dynamics[48–51]. However, our study emphasizes that wastewater data can also be viewed from an ecological and evolutionary perspective, reflecting the community's viral variant landscape.

Our seasonal analysis revealed significant variations in alpha diversity, with higher diversity observed in winter and lower diversity in summer. This suggests that colder months facilitate the emergence and spread of more variants, likely due to increased indoor activities, enhanced viral stability, and seasonal variations in immune responses. Additionally, higher alpha diversity during low transmission periods and the lag phase of outbreaks indicates that a diverse pool of variants circulates at low levels during these times. These periods allow for the accumulation and persistence of a wider range of variants, which may not immediately lead to outbreaks but could contribute to future waves if conditions become favorable. Ecologically, the lag phase represents a period of low competition, allowing new variants to compete until a strain with the highest fitness prevails. These insights underscore the importance of continuous monitoring of viral diversity to predict future outbreaks and implement timely public health interventions.

Our findings also demonstrate that RBD amplicon sequencing of wastewater samples effectively captures the dynamics of SARS-CoV-2 variants, showing consistent results with clinical genome sequencing data at local, state, and national levels. Notably, our study identified variants circulating in the population earlier than observed in clinical sequencing data and detected variants and outbreaks that were not reported or noticed locally. This consistency with clinical genome sequencing data and the earlier detection of variants underscore the accuracy and

reliability of RBD amplicon sequencing of wastewater in variant surveillance. It also strengthens the role of wastewater surveillance as a complementary tool to clinical testing[24–26,52–55], especially valuable in areas with limited access to clinical sequencing resources.

Despite its advantages, RBD amplicon sequencing has limitations. While the RBD region is rich in information and useful for targeted analysis, it represents only a small portion of the viral genome. This limitation means that mutations outside the RBD region will be missed, and closely related variants with mutations in other regions may not be differentiated. Additionally, evolutionary changes occurring outside the targeted region, including recombination events, are not captured. The PCR amplification step can also introduce biases[56–58], potentially skewing the relative abundance of certain variants. While our analyses using synthetic variant communities (**Figure 1B**) suggest that this bias may not significantly impact our results, further validation is necessary. On the other hand, RBD amplicon sequencing is faster, easier, and cheaper than whole genome sequencing[20,59,60], making it accessible for widespread use. Its high throughput capability with relatively lower costs allows for the rapid analysis of numerous samples, which is critical during pandemic conditions. By balancing these trade-offs, RBD amplicon sequencing provides a practical and efficient tool for ongoing variant surveillance, particularly when rapid data collection is necessary and resources are constrained.

Viruses have various mechanisms to enhance transmission between hosts. One significant strategy is mutating the receptor-binding domain in the spike gene, as observed with SARS-CoV-2, which has produced multiple waves of variants from Alpha to Omicron. These mutations in RBD enable the virus to evade adaptive immune responses from prior infections and/or vaccinations. Consequently, monitoring the RBD region is crucial for tracking viral evolution, predicting the emergence of new variants, and informing updated vaccine composition. This study highlights the utility of RBD amplicon sequencing and the integration of wastewater surveillance to track SARS-CoV-2 evolution and variant emergence. Our findings demonstrate that this tool captures the dynamics of variant community, provides early detection of variants, aligns well with clinical genome sequencing data, and reveals epidemiological patterns in viral diversity. The concept of monitoring the receptor-binding domain can also be generalized to other viruses, providing a broader framework for variant surveillance. This tool can improve our ability to predict future outbreaks and enhance public health preparedness.

Declaration of Competing Interest

The authors declare no competing interest.

Data and Code Availability

Data in this study will be shared with the paper publication. Scripts will be shared on GitHub.

Acknowledgments

We are grateful to Teresa T. Alcala at El Paso Water for sample collection and shipping. This work is supported by the PRIME award from the UTHealth Houston School of Public Health, National Institute of Health RADxUP (1U01TR004355-01), and the Texas Epidemic Public Health Institute (TEPHI).

References

(1) Shin, H. J.; Ku, K. B.; Kim, H. S.; Moon, H. W.; Jeong, G. U.; Hwang, I.; Yoon, G. Y.; Lee, S.; Lee, S.; Ahn, D.-G.; Kim, K.-D.; Kwon, Y.-C.; Kim, B.-T.; Kim, S.-J.; Kim, C. Receptor-Binding Domain of SARS-CoV-2 Spike Protein Efficiently Inhibits SARS-CoV-2 Infection and Attachment to Mouse Lung. *Int. J. Biol. Sci.* **2021**, *17* (14), 3786–3794. https://doi.org/10.7150/ijbs.61320.

(2) Tai, W.; He, L.; Zhang, X.; Pu, J.; Voronin, D.; Jiang, S.; Zhou, Y.; Du, L. Characterization of the Receptor-Binding Domain (RBD) of 2019 Novel Coronavirus: Implication for Development of RBD Protein as a Viral Attachment Inhibitor and Vaccine. *Cell. Mol. Immunol.* **2020**, *17* (6), 613–620. https://doi.org/10.1038/s41423-020-0400-4.

(3) Shang, J.; Wan, Y.; Luo, C.; Ye, G.; Geng, Q.; Auerbach, A.; Li, F. Cell Entry Mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci.* **2020**, *117* (21), 11727–11734. https://doi.org/10.1073/pnas.2003138117.

(4) Chakraborty, C.; Bhattacharya, M.; Sharma, A. R. Present Variants of Concern and Variants of Interest of Severe Acute Respiratory Syndrome Coronavirus 2: Their Significant Mutations in S-Glycoprotein, Infectivity, Re-Infectivity, Immune Escape and Vaccines Activity. *Rev. Med. Virol.* **2022**, *32* (2), e2270. https://doi.org/10.1002/rmv.2270.

(5) Dhawan, M.; Sharma, A.; Priyanka; Thakur, N.; Rajkhowa, T. K.; Choudhary, O. P. Delta Variant (B.1.617.2) of SARS-CoV-2: Mutations, Impact, Challenges and Possible Solutions. *Hum. Vaccines Immunother. 18* (5), 2068883. https://doi.org/10.1080/21645515.2022.2068883.

(6) Adhikari, P.; Jawad, B.; Rao, P.; Podgornik, R.; Ching, W.-Y. Delta Variant with P681R Critical Mutation Revealed by Ultra-Large Atomic-Scale Ab Initio Simulation: Implications for the Fundamentals of Biomolecular Interactions. *Viruses* **2022**, *14* (3), 465. https://doi.org/10.3390/v14030465.

(7) Bhattacharya, M.; Chatterjee, S.; Sharma, A. R.; Lee, S.-S.; Chakraborty, C. Delta Variant (B.1.617.2) of SARS-CoV-2: Current Understanding of Infection, Transmission, Immune Escape, and Mutational Landscape. *Folia Microbiol. (Praha)* **2023**, *68* (1), 17–28. https://doi.org/10.1007/s12223-022-01001-3.

(8) Goher, S. S.; Ali, F.; Amin, M. The Delta Variant Mutations in the Receptor Binding Domain of SARS-CoV-2 Show Enhanced Electrostatic Interactions with the ACE2. *Med. Drug Discov.* **2022**, *13*, 100114. https://doi.org/10.1016/j.medidd.2021.100114.

(9) Pondé, R. A. A. Physicochemical Effect of the N501Y, E484K/Q, K417N/T, L452R and T478K Mutations on the SARS-CoV-2 Spike Protein RBD and Its Influence on Agent Fitness and on Attributes Developed by Emerging Variants of Concern. *Virology* **2022**, *572*, 44–54. https://doi.org/10.1016/j.virol.2022.05.003.

(10) Schwarze, K.; Buchanan, J.; Taylor, J. C.; Wordsworth, S. Are Whole-Exome and Whole-Genome Sequencing Approaches Cost-Effective? A Systematic Review of the Literature. *Genet. Med.* **2018**, *20* (10), 1122–1130. https://doi.org/10.1038/gim.2017.247.

(11) Christensen, K. D.; Dukhovny, D.; Siebert, U.; Green, R. C. Assessing the Costs and Cost-Effectiveness of Genomic Sequencing. *J. Pers. Med.* **2015**, *5* (4), 470–486. https://doi.org/10.3390/jpm5040470.

(12) Nurchis, M. C.; Riccardi, M. T.; Damiani, G. Health Technology Assessment of Whole Genome Sequencing in the Diagnosis of Genetic Disorders: A Scoping Review of the Literature. *Int. J. Technol. Assess. Health Care* **2022**, *38* (1), e71. https://doi.org/10.1017/S0266462322000496.

(13) Long, Q.-X.; Tang, X.-J.; Shi, Q.-L.; Li, Q.; Deng, H.-J.; Yuan, J.; Hu, J.-L.; Xu, W.; Zhang, Y.; Lv, F.-J.; Su, K.; Zhang, F.; Gong, J.; Wu, B.; Liu, X.-M.; Li, J.-J.; Qiu, J.-F.; Chen, J.; Huang, A.-L. Clinical and Immunological Assessment of Asymptomatic SARS-CoV-2 Infections. *Nat. Med.* **2020**, *26* (8), 1200–1204. https://doi.org/10.1038/s41591-020-0965-6.

(14) Sah, P.; Fitzpatrick, M. C.; Zimmer, C. F.; Abdollahi, E.; Juden-Kelly, L.; Moghadas, S. M.; Singer, B. H.; Galvani, A. P. Asymptomatic SARS-CoV-2 Infection: A Systematic Review and Meta-Analysis. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118* (34), e2109229118. https://doi.org/10.1073/pnas.2109229118.

(15) You, Y.; Yang, X.; Hung, D.; Yang, Q.; Wu, T.; Deng, M. Asymptomatic COVID-19 Infection: Diagnosis, Transmission, Population Characteristics. *BMJ Support. Palliat. Care* **2024**, *14* (e1), e220–e227. https://doi.org/10.1136/bmjspcare-2020-002813.

(16) Kimball, A. Asymptomatic and Presymptomatic SARS-CoV-2 Infections in Residents of a Long-Term Care Skilled Nursing Facility — King County, Washington, March 2020. *MMWR Morb. Mortal. Wkly. Rep.* **2020**, *69*. https://doi.org/10.15585/mmwr.mm6913e1.

(17) Thompson, J. R.; Nancharaiah, Y. V.; Gu, X.; Lee, W. L.; Rajal, V. B.; Haines, M. B.; Girones, R.; Ng, L. C.; Alm, E. J.; Wuertz, S. Making Waves: Wastewater Surveillance of SARS-CoV-2 for Population-Based Health Management. *Water Res.* **2020**, *184*, 116181. https://doi.org/10.1016/j.watres.2020.116181.

(18) Xagoraraki, I. Can We Predict Viral Outbreaks Using Wastewater Surveillance? *J. Environ. Eng.* **2020**, *146* (11), 01820003. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001831.

(19) Kilaru, P.; Hill, D.; Anderson, K.; Collins, M. B.; Green, H.; Kmush, B. L.; Larsen, D. A. Wastewater Surveillance for Infectious Disease: A Systematic Review. *Am. J. Epidemiol.* **2023**, *192* (2), 305–322. https://doi.org/10.1093/aje/kwac175.

(20) Kuroiwa, M.; Gahara, Y.; Kato, H.; Morikawa, Y.; Matsui, Y.; Adachi, T.; Kurosawa, S.; Kuroita, T.; Ando, Y.; Rokushima, M. Targeted Amplicon Sequencing of Wastewater Samples for Detecting SARS-CoV-2 Variants with High Sensitivity and Resolution. *Sci. Total Environ.* **2023**, *893*, 164766. https://doi.org/10.1016/j.scitotenv.2023.164766.

(21) Iwamoto, R.; Yamaguchi, K.; Katayama, K.; Ando, H.; Setsukinai, K.; Kobayashi, H.; Okabe, S.; Imoto, S.; Kitajima, M. Identification of SARS-CoV-2 Variants in Wastewater Using Targeted Amplicon Sequencing during a Low COVID-19 Prevalence Period in Japan. *Sci. Total Environ.* **2023**, *887*, 163706. https://doi.org/10.1016/j.scitotenv.2023.163706.

(22) Lou, E. G.; Sapoval, N.; McCall, C.; Bauhs, L.; Carlson-Stadler, R.; Kalvapalle, P.; Lai, Y.; Palmer, K.; Penn, R.; Rich, W.; Wolken, M.; Brown, P.; Ensor, K. B.; Hopkins, L.; Treangen, T. J.; Stadler, L. B. Direct Comparison of RT-ddPCR and Targeted Amplicon Sequencing for SARS-CoV-2 Mutation Monitoring in Wastewater. *Sci. Total Environ.* **2022**, *833*, 155059. https://doi.org/10.1016/j.scitotenv.2022.155059.

(23) *Environmental surveillance for SARS-COV-2 to complement public health surveillance – Interim Guidance*. https://www.who.int/publications/i/item/WHO-HEP-ECH-WSH-2022.1 (accessed 2024-06-26).

(24) Bisseux, M.; Debroas, D.; Mirand, A.; Archimbaud, C.; Peigue-Lafeuille, H.; Bailly, J.-L.; Henquell, C. Monitoring of Enterovirus Diversity in Wastewater by Ultra-Deep Sequencing: An Effective Complementary Tool for Clinical Enterovirus Surveillance. *Water Res.* **2020**, *169*, 115246. https://doi.org/10.1016/j.watres.2019.115246.

(25) Fantilli, A.; Cola, G. D.; Castro, G.; Sicilia, P.; Cachi, A. M.; de los Ángeles Marinzalda, M.; Ibarra, G.; López, L.; Valduvino, C.; Barbás, G.; Nates, S.; Masachessi, G.; Pisano, M. B.; Ré, V. Hepatitis A Virus Monitoring in Wastewater: A Complementary Tool to Clinical Surveillance. *Water Res.* **2023**, *241*, 120102. https://doi.org/10.1016/j.watres.2023.120102.

(26) Smith, T.; Holm, R. H.; Yeager, R.; Moore, J. B.; Rouchka, E. C.; Sokoloski, K. J.; Elliott, E. M.; Talley, D.; Arora, V.; Moyer, S.; Bhatnagar, A. Combining Community Wastewater Genomic Surveillance with State Clinical Surveillance: A Framework for SARS-CoV-2 Public Health Practice. *Food Environ. Virol.* **2022**, *14* (4), 410–416. https://doi.org/10.1007/s12560-022-09531-2.

(27) Li, J.; Hosegood, I.; Powell, D.; Tscharke, B.; Lawler, J.; Thomas, K. V.; Mueller, J. F. A Global Aircraft-Based Wastewater Genomic Surveillance Network for Early Warning of Future Pandemics. *Lancet Glob. Health* **2023**, *11* (5), e791–e795. https://doi.org/10.1016/S2214-109X(23)00129-8.

(28) Karthikeyan, S.; Levy, J. I.; De Hoff, P.; Humphrey, G.; Birmingham, A.; Jepsen, K.; Farmer, S.; Tubb, H. M.; Valles, T.; Tribelhorn, C. E.; Tsai, R.; Aigner, S.; Sathe, S.; Moshiri, N.; Henson, B.; Mark, A. M.; Hakim, A.; Baer, N. A.; Barber, T.; Belda-Ferre, P.; Chacón, M.; Cheung, W.; Cresini, E. S.; Eisner, E. R.; Lastrella, A. L.; Lawrence, E. S.; Marotz, C. A.; Ngo, T. T.; Ostrander, T.; Plascencia, A.; Salido, R. A.; Seaver, P.; Smoot, E. W.; McDonald, D.; Neuhard, R. M.; Scioscia, A. L.; Satterlund, A. M.; Simmons, E. H.; Abelman, D. B.; Brenner, D.; Bruner, J. C.; Buckley, A.; Ellison, M.; Gattas, J.; Gonias, S. L.; Hale, M.; Hawkins, F.; Ikeda, L.; Jhaveri, H.; Johnson, T.; Kellen, V.; Kremer, B.; Matthews, G.; McLawhon, R. W.; Ouillet, P.; Park, D.; Pradenas, A.; Reed, S.; Riggs, L.; Sanders, A.; Sollenberger, B.; Song, A.; White, B.; Winbush, T.; Aceves, C. M.; Anderson, C.; Gangavarapu, K.; Hufbauer, E.; Kurzban, E.; Lee, J.; Matteson, N. L.; Parker, E.; Perkins, S. A.; Ramesh, K. S.; Robles-Sikisaka, R.; Schwab, M. A.; Spencer, E.; Wohl, S.; Nicholson, L.; McHardy, I. H.; Dimmock, D. P.; Hobbs, C. A.; Bakhtar, O.; Harding, A.; Mendoza, A.; Bolze, A.; Becker, D.; Cirulli, E. T.; Isaksson, M.; Schiabor Barrett, K. M.; Washington, N. L.; Malone, J. D.; Schafer, A. M.; Gurfield, N.; Stous, S.; Fielding-Miller, R.; Garfein, R. S.; Gaines, T.; Anderson, C.; Martin, N. K.; Schooley, R.; Austin, B.; MacCannell, D. R.; Kingsmore, S. F.; Lee, W.; Shah, S.; McDonald, E.; Yu, A. T.; Zeller, M.; Fisch, K. M.; Longhurst, C.; Maysent, P.; Pride, D.; Khosla, P. K.; Laurent, L. C.; Yeo, G. W.; Andersen, K. G.; Knight, R. Wastewater Sequencing Reveals Early Cryptic SARS-CoV-2 Variant Transmission. *Nature* **2022**, *609* (7925), 101–108. https://doi.org/10.1038/s41586-022-05049-6.

(29) Jiang, M.; Wang, A. L. W.; Be, N. A.; Mulakken, N.; Nelson, K. L.; Kantor, R. S. Evaluation of the Impact of Concentration and Extraction Methods on the Targeted Sequencing of Human Viruses from Wastewater. *Environ. Sci. Technol.* **2024**, *58* (19), 8239–8250. https://doi.org/10.1021/acs.est.4c00580.

(30) Levy, J. I.; Andersen, K. G.; Knight, R.; Karthikeyan, S. Wastewater Surveillance for Public Health. *Science* **2023**, *379* (6627), 26–27. https://doi.org/10.1126/science.ade2503.

(31) Parkins, M. D.; Lee, B. E.; Acosta, N.; Bautista, M.; Hubert, C. R. J.; Hrudey, S. E.; Frankowski, K.; Pang, X.-L. Wastewater-Based Surveillance as a Tool for Public Health Action: SARS-CoV-2 and Beyond. *Clin. Microbiol. Rev.* **2023**, *37* (1), e00103-22. https://doi.org/10.1128/cmr.00103-22.

(32) Li, Y.; Miyani, B.; Childs, K. L.; Shiu, S.-H.; Xagoraraki, I. Effect of Wastewater Collection and Concentration Methods on Assessment of Viral Diversity. *Sci. Total Environ.* **2024**, *908*, 168128. https://doi.org/10.1016/j.scitotenv.2023.168128.

(33) Gregory, D. A.; Trujillo, M.; Rushford, C.; Flury, A.; Kannoly, S.; San, K. M.; Lyfoung, D. T.; Wiseman, R. W.; Bromert, K.; Zhou, M.-Y.; Kesler, E.; Bivens, N. J.; Hoskins, J.; Lin, C.-H.; O'Connor, D. H.; Wieberg, C.; Wenzel, J.; Kantor, R. S.; Dennehy, J. J.; Johnson, M. C. Genetic Diversity and Evolutionary Convergence of Cryptic SARS- CoV-2 Lineages Detected via Wastewater Sequencing. *PLOS Pathog.* **2022**, *18* (10), e1010636. https://doi.org/10.1371/journal.ppat.1010636.

(34) Baaijens, J. A.; Zulli, A.; Ott, I. M.; Nika, I.; van der Lugt, M. J.; Petrone, M. E.; Alpert, T.; Fauver, J. R.; Kalinich, C. C.; Vogels, C. B. F.; Breban, M. I.; Duvallet, C.; McElroy, K. A.; Ghaeli, N.; Imakaev, M.; Mckenzie-Bennett, M. F.; Robison, K.; Plocik, A.; Schilling, R.; Pierson, M.; Littlefield, R.; Spencer, M. L.; Simen, B. B.; Altajar, A.; Brito, A. F.; Watkins, A. E.; Muyombwe, A.; Neal, C.; Liu, C.; Castaldi, C.; Pearson, C.; Peaper, D. R.; Laszlo, E.; Tikhonova, I. R.; Razeq, J.; Rothman, J. E.; Wang, J.; Bilguvar, K.; Niccolai, L.; Wilson, M. S.; Anderson, M. L.; Landry, M. L.; Adams, M. D.; Hui, P.; Downing, R.; Earnest, R.; Mane, S.; Murphy, S.; Hanage, W. P.; Grubaugh, N. D.; Peccia, J.; Baym, M.; Yale SARS-CoV-2 Genomic Surveillance Initiative. Lineage Abundance Estimation for SARS-CoV-2 in Wastewater Using Transcriptome Quantification Techniques. *Genome Biol.* **2022**, *23* (1), 236. https://doi.org/10.1186/s13059-022-02805-9.

(35) Cancela, F.; Ramos, N.; Smyth, D. S.; Etchebehere, C.; Berois, M.; Rodríguez, J.; Rufo, C.; Alemán, A.; Borzacconi, L.; López, J.; González, E.; Botto, G.; Thornhill, S. G.; Mirazo, S.; Trujillo, M. Wastewater Surveillance of SARS-CoV-2 Genomic Populations on a Country-Wide Scale through Targeted Sequencing. *PLOS ONE* **2023**, *18* (4), e0284483. https://doi.org/10.1371/journal.pone.0284483.

(36) Gregory, D. A.; Wieberg, C. G.; Wenzel, J.; Lin, C.-H.; Johnson, M. C. Monitoring SARS-CoV-2 Populations in Wastewater by Amplicon Sequencing and Using the Novel Program SAM Refiner. *Viruses* **2021**, *13* (8), 1647. https://doi.org/10.3390/v13081647.

(37) Adamopoulos, P. G.; Diamantopoulos, M. A.; Boti, M. A.; Zafeiriadou, A.; Galani, A.; Kostakis, M.; Markou, A.; Sideris, D. C.; Avgeris, M.; Thomaidis, N. S.; Scorilas, A. Spike-Seq: An Amplicon-Based High-Throughput Sequencing Approach for the Sensitive Detection and Characterization of SARS-CoV-2 Genetic Variations in Environmental Samples. *Sci. Total Environ.* **2024**, *914*, 169747. https://doi.org/10.1016/j.scitotenv.2023.169747.

(38) Rodrigues, G. M.; Volpato, F. C. Z.; Wink, P. L.; Paiva, R. M.; Barth, A. L.; de-Paris, F. SARS-CoV-2 Variants of Concern: Presumptive Identification via Sanger Sequencing Analysis of the Receptor Binding Domain (RBD) Region of the S Gene. *Diagnostics* **2023**, *13* (7), 1256. https://doi.org/10.3390/diagnostics13071256.

(39) Oghuan, J.; Chavarria, C.; Vanderwal, S. R.; Gitter, A.; Ojaruega, A. A.; Monserrat, C.; Bauer, C. X.; Brown, E. L.; Cregeen, S. J.; Deegan, J.; Hanson, B. M.; Tisza, M.; Ocaranza, H. I.; Balliew, J.; Maresso, A. W.; Rios, J.; Boerwinkle, E.; Mena, K. D.; Wu, F. Wastewater Analysis of Mpox Virus in a City with Low Prevalence of Mpox Disease: An Environmental Surveillance Study. *Lancet Reg. Health – Am.* **2023**, *28*. https://doi.org/10.1016/j.lana.2023.100639.

(40) Wu, F.; Xiao, A.; Zhang, J.; Moniz, K.; Endo, N.; Armas, F.; Bonneau, R.; Brown, M. A.; Bushman, M.; Chai, P. R.; Duvallet, C.; Erickson, T. B.; Foppe, K.; Ghaeli, N.; Gu, X.; Hanage, W. P.; Huang, K. H.; Lee, W. L.; Matus, M.; McElroy, K. A.; Nagler, J.; Rhode, S. F.; Santillana, M.; Tucker, J. A.; Wuertz, S.; Zhao, S.; Thompson, J.; Alm, E. J. SARS-CoV-2 RNA Concentrations in Wastewater Foreshadow

Dynamics and Clinical Presentation of New COVID-19 Cases. *Sci. Total Environ.* **2022**, *805*, 150121. https://doi.org/10.1016/j.scitotenv.2021.150121.

(41) Wu, F.; Xiao, A.; Zhang, J.; Moniz, K.; Endo, N.; Armas, F.; Bushman, M.; Chai, P. R.; Duvallet, C.; Erickson, T. B.; Foppe, K.; Ghaeli, N.; Gu, X.; Hanage, W. P.; Huang, K. H.; Lee, W. L.; McElroy, K. A.; Rhode, S. F.; Matus, M.; Wuertz, S.; Thompson, J.; Alm, E. J. Wastewater Surveillance of SARS-CoV-2 across 40 U.S. States from February to June 2020. *Water Res.* **2021**, *202*, 117400. https://doi.org/10.1016/j.watres.2021.117400.

(42) Wu, F.; Xiao, A.; Zhang, J.; Gu, X.; Lee, W. L.; Kauffman, K.; Hanage, W.; Matus, M.; Ghaeli, N.; Endo, N.; Duvallet, C.; Moniz, K.; Erickson, T.; Chai, P.; Thompson, J.; Alm, E. SARS-CoV-2 Titers in Wastewater Are Higher than Expected from Clinically Confirmed Cases. *mSystems* **2020**, *5* (4), 2020.04.05.20051540. https://doi.org/10.1128/mSystems.00614-20.

(43) Callahan, B. J.; McMurdie, P. J.; Rosen, M. J.; Han, A. W.; Johnson, A. J. A.; Holmes, S. P. DADA2: High-Resolution Sample Inference from Illumina Amplicon Data. *Nat. Methods* **2016**, *13* (7), 581–583. https://doi.org/10.1038/nmeth.3869.

(44) Sayers, E. W.; Bolton, E. E.; Brister, J. R.; Canese, K.; Chan, J.; Comeau, D. C.; Connor, R.; Funk, K.; Kelly, C.; Kim, S.; Madej, T.; Marchler-Bauer, A.; Lanczycki, C.; Lathrop, S.; Lu, Z.; Thibaud-Nissen, F.; Murphy, T.; Phan, L.; Skripchenko, Y.; Tse, T.; Wang, J.; Williams, R.; Trawick, B. W.; Pruitt, K. D.; Sherry, S. T. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2021**, *50* (D1), D20–D26. https://doi.org/10.1093/nar/gkab1112.

(45) Wirth, W.; Duchene, S. GISAIDR, 2022. https://doi.org/10.5281/zenodo.6474693.

(46) Amman, F.; Markt, R.; Endler, L.; Hupfauf, S.; Agerer, B.; Schedl, A.; Richter, L.; Zechmeister, M.; Bicher, M.; Heiler, G.; Triska, P.; Thornton, M.; Penz, T.; Senekowitsch, M.; Laine, J.; Keszei, Z.; Klimek, P.; Nägele, F.; Mayr, M.; Daleiden, B.; Steinlechner, M.; Niederstätter, H.; Heidinger, P.; Rauch, W.; Scheffknecht, C.; Vogl, G.; Weichlinger, G.; Wagner, A. O.; Slipko, K.; Masseron, A.; Radu, E.; Allerberger, F.; Popper, N.; Bock, C.; Schmid, D.; Oberacher, H.; Kreuzinger, N.; Insam, H.; Bergthaler, A. Viral Variant-Resolved Wastewater Surveillance of SARS-CoV-2 at National Scale. *Nat. Biotechnol.* **2022**, *40* (12), 1814–1822. https://doi.org/10.1038/s41587-022-01387-y.

(47) Gupta, P.; Liao, S.; Ezekiel, M.; Novak, N.; Rossi, A.; LaCross, N.; Oakeson, K.; Rohrwasser, A. Wastewater Genomic Surveillance Captures Early Detection of Omicron in Utah. *Microbiol. Spectr.* **2023**, *11* (3), e00391-23. https://doi.org/10.1128/spectrum.00391-23.

(48) Oude Munnink, B. B.; Nieuwenhuijse, D. F.; Stein, M.; O'Toole, Á.; Haverkate, M.; Mollers, M.; Kamga, S. K.; Schapendonk, C.; Pronk, M.; Lexmond, P.; van der Linden, A.; Bestebroer, T.; Chestakova, I.; Overmars, R. J.; van Nieuwkoop, S.; Molenkamp, R.; van der Eijk, A. A.; GeurtsvanKessel, C.; Vennema, H.; Meijer, A.; Rambaut, A.; van Dissel, J.; Sikkema, R. S.; Timen, A.; Koopmans, M. Rapid SARS-CoV-2 Whole-Genome Sequencing and Analysis for Informed Public Health Decision-Making in the Netherlands. *Nat. Med.* **2020**, *26* (9), 1405–1410. https://doi.org/10.1038/s41591-020-0997-y.

(49) Mao, K.; Zhang, K.; Du, W.; Ali, W.; Feng, X.; Zhang, H. The Potential of Wastewater-Based Epidemiology as Surveillance and Early Warning of Infectious Disease Outbreaks. *Curr. Opin. Environ. Sci. Health* **2020**, *17*, 1–7. https://doi.org/10.1016/j.coesh.2020.04.006.

(50) Ahmed, W.; Tscharke, B.; Bertsch, P. M.; Bibby, K.; Bivins, A.; Choi, P.; Clarke, L.; Dwyer, J.; Edson, J.; Nguyen, T. M. H.; O'Brien, J. W.; Simpson, S. L.; Sherman, P.; Thomas, K. V.; Verhagen, R.; Zaugg, J.; Mueller, J. F. SARS-CoV-2 RNA Monitoring in Wastewater as a Potential Early Warning System for COVID-19 Transmission in the Community: A Temporal Case Study. *Sci. Total Environ.* **2021**, *761*, 144216. https://doi.org/10.1016/j.scitotenv.2020.144216.

(51) Mackuľak, T.; Gál, M.; Špalková, V.; Fehér, M.; Briestenská, K.; Mikušová, M.; Tomčíková, K.; Tamáš, M.; Butor Škulcová, A. Wastewater-Based Epidemiology as an Early Warning System for the

Spreading of SARS-CoV-2 and Its Mutations in the Population. *Int. J. Environ. Res. Public. Health* **2021**, *18* (11), 5629. https://doi.org/10.3390/ijerph18115629.

(52) Gitter, A.; Oghuan, J.; Godbole, A. R.; Chavarria, C. A.; Monserrat, C.; Hu, T.; Wang, Y.; Maresso, A. W.; Hanson, B. M.; Mena, K. D.; Wu, F. Not a Waste: Wastewater Surveillance to Enhance Public Health. *Front. Chem. Eng.* **2023**, *4*. https://doi.org/10.3389/fceng.2022.1112876.

(53) CDC. *About CDC's National Wastewater Surveillance System (NWSS)*. Centers for Disease Control and Prevention. https://www.cdc.gov/nwss/about.html (accessed 2024-06-26).

(54) DeJonge, P. M.; Adams, C.; Pray, I.; Schussman, M. K.; Fahney, R. B.; Shafer, M.; Antkiewicz, D. S.; Roguet, A. Wastewater Surveillance Data as a Complement to Emergency Department Visit Data for Tracking Incidence of Influenza A and Respiratory Syncytial Virus - Wisconsin, August 2022-March 2023. *MMWR Morb. Mortal. Wkly. Rep.* **2023**, *72* (37), 1005–1009. https://doi.org/10.15585/mmwr.mm7237a2.

(55) Bonanno Ferraro, G.; Veneri, C.; Mancini, P.; Iaconelli, M.; Suffredini, E.; Bonadonna, L.; Lucentini, L.; Bowo-Ngandji, A.; Kengne-Nde, C.; Mbaga, D. S.; Mahamat, G.; Tazokong, H. R.; Ebogo-Belobo, J. T.; Njouom, R.; Kenmoe, S.; La Rosa, G. A State-of-the-Art Scoping Review on SARS-CoV-2 in Sewage Focusing on the Potential of Wastewater Surveillance for the Monitoring of the COVID-19 Pandemic. *Food Environ. Virol.* **2022**, *14* (4), 315–354. https://doi.org/10.1007/s12560-021-09498-6.

(56) Acinas, S. G.; Sarma-Rupavtarm, R.; Klepac-Ceraj, V.; Polz, M. F. PCR-Induced Sequence Artifacts and Bias: Insights from Comparison of Two 16S rRNA Clone Libraries Constructed from the Same Sample. *Appl. Environ. Microbiol.* **2005**, *71* (12), 8966–8969. https://doi.org/10.1128/AEM.71.12.8966-8969.2005.

(57) Kanagawa, T. Bias and Artifacts in Multitemplate Polymerase Chain Reactions (PCR). *J. Biosci. Bioeng.* **2003**, *96* (4), 317–323. https://doi.org/10.1016/S1389-1723(03)90130-7.

(58) Krehenwinkel, H.; Wolf, M.; Lim, J. Y.; Rominger, A. J.; Simison, W. B.; Gillespie, R. G. Estimating and Mitigating Amplification Bias in Qualitative and Quantitative Arthropod Metabarcoding. *Sci. Rep.* **2017**, *7* (1), 17668. https://doi.org/10.1038/s41598-017-17333-x.

(59) Techera, C.; Tomás, G.; Grecco, S.; Williman, J.; Hernández, M.; Olivera, V.; Banda, A.; Vagnozzi, A.; Panzera, Y.; Marandino, A.; Pérez, R. A Rapid and Affordable Amplicon-Based Method for next-Generation Genome Sequencing of the Infectious Bursal Disease Virus. *J. Virol. Methods* **2023**, *322*, 114807. https://doi.org/10.1016/j.jviromet.2023.114807.

(60) Kong, X.; Gao, P.; Jiang, Y.; Lu, L.; Zhao, M.; Liu, Y.; Deng, G.; Zhu, H.; Cao, Y.; Ma, L. Discrimination of SARS-CoV-2 Omicron Variant and Its Lineages by Rapid Detection of Immune-Escape Mutations in Spike Protein RBD Using Asymmetric PCR-Based Melting Curve Analysis. *Virol. J.* **2023**, *20* (1), 192. https://doi.org/10.1186/s12985-023-02137-5.

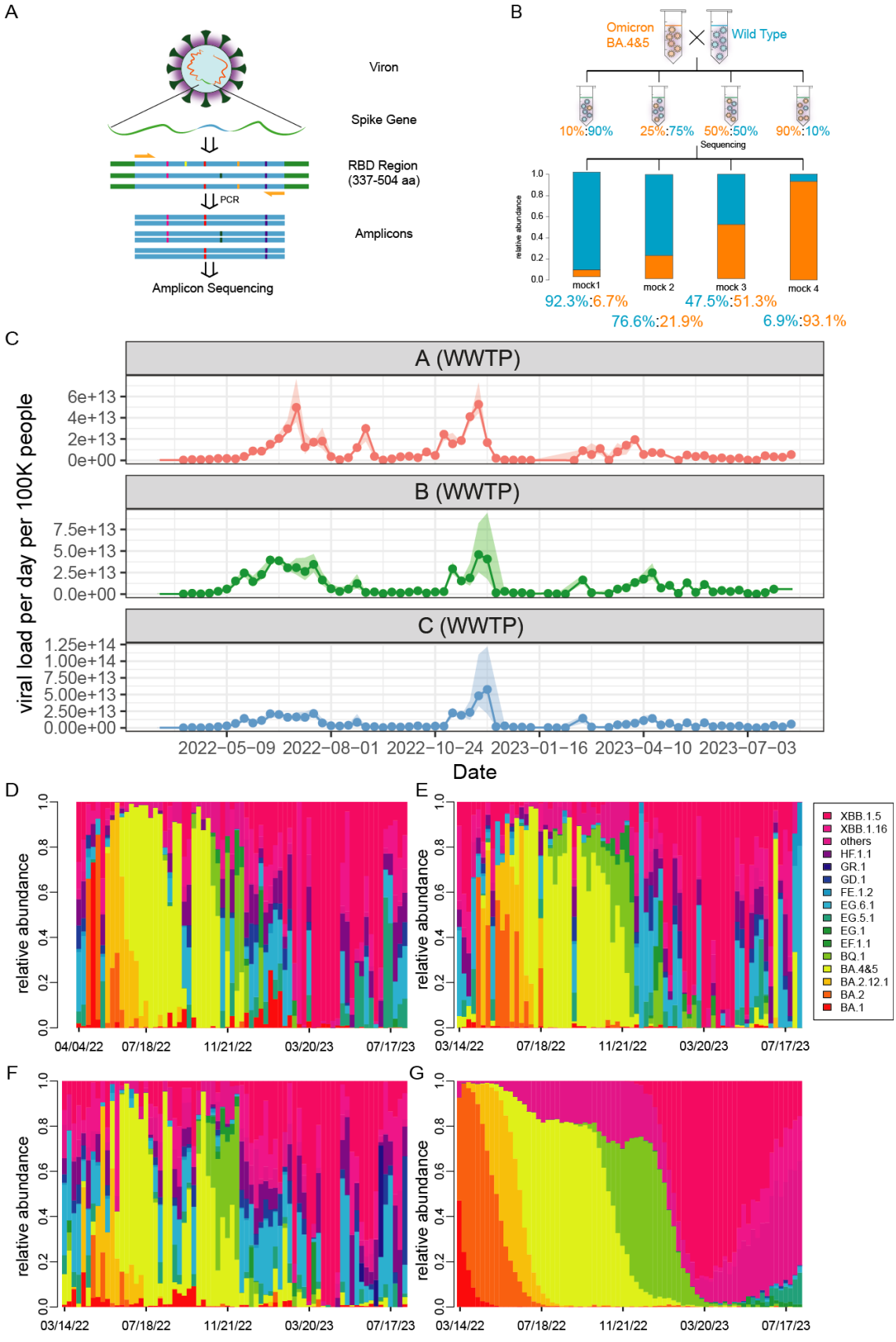


Figure 1. RBD amplicon sequencing of wastewater reveals temporal dynamics of SARS-CoV-2 variants. (A) Diagram of the RBD amplicon sequencing procedure. We specifically amplify the 337-504 aa sequence in the RBD region of SARS-CoV-2 spike gene. (B) Validation of the RBD amplicon sequencing protocol using the wild-type SARS-CoV-2 strain and BA.4&5. The bar plot shows the sequencing results of these samples, demonstrating that the resultant ratios closely match the original mixing ratios. (C) Viral load of SARS-CoV-2 per day per 100k people in the three sewersheds in the City of El Paso, TX from March 2022 to August 2023. (D-F) Relative abundance of SARS-CoV-2 variants revealed by RBD amplicon sequencing of wastewater samples from WWTPs A (D), B (E), and C (F) in the City of El Paso. (G) Relative abundance of SARS-CoV-2 variants from genomic sequencing of clinical samples in Texas.

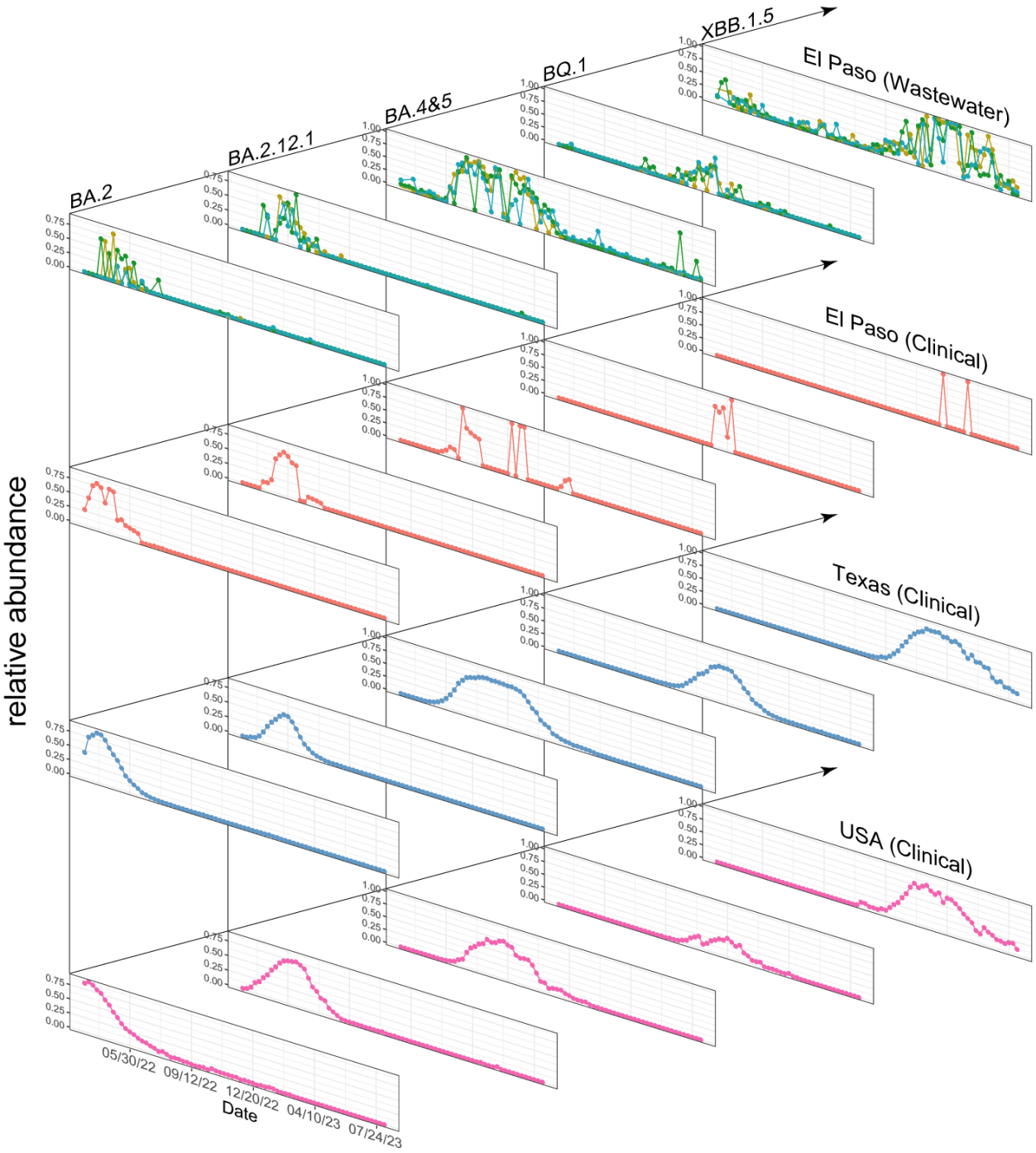


Figure 2. Time-course of major SARS-CoV-2 variants identified from RBD amplicon sequencing of wastewater. This figure presents the temporal dynamics of five major variants identified from three WWTPs alongside clinical data from El Paso, Texas, and the USA. Rows depict the progression of variant waves over time, while columns allow for a comparative analysis between the wastewater treatment plant data and clinical data, highlighting their similarities and trends.

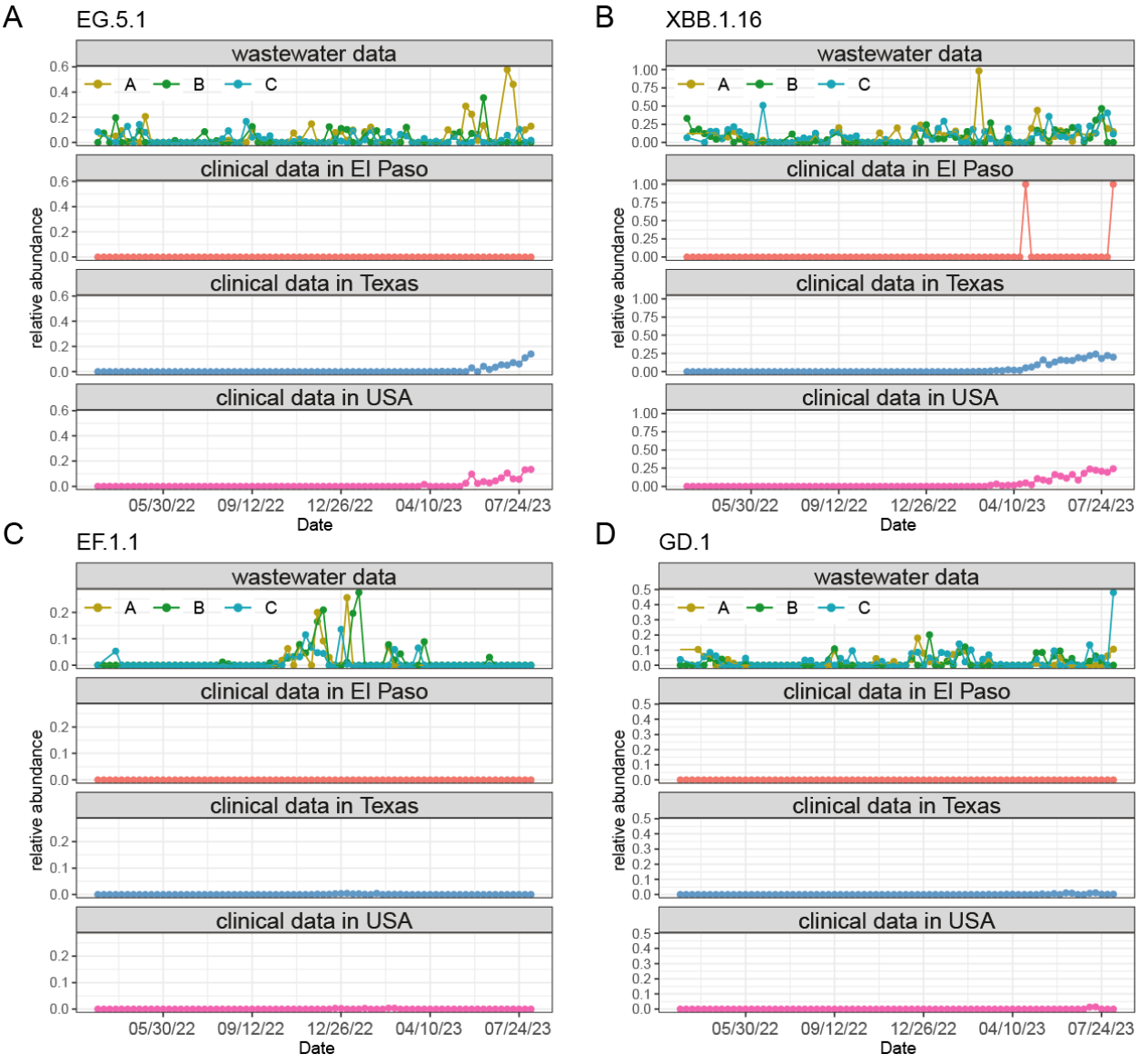
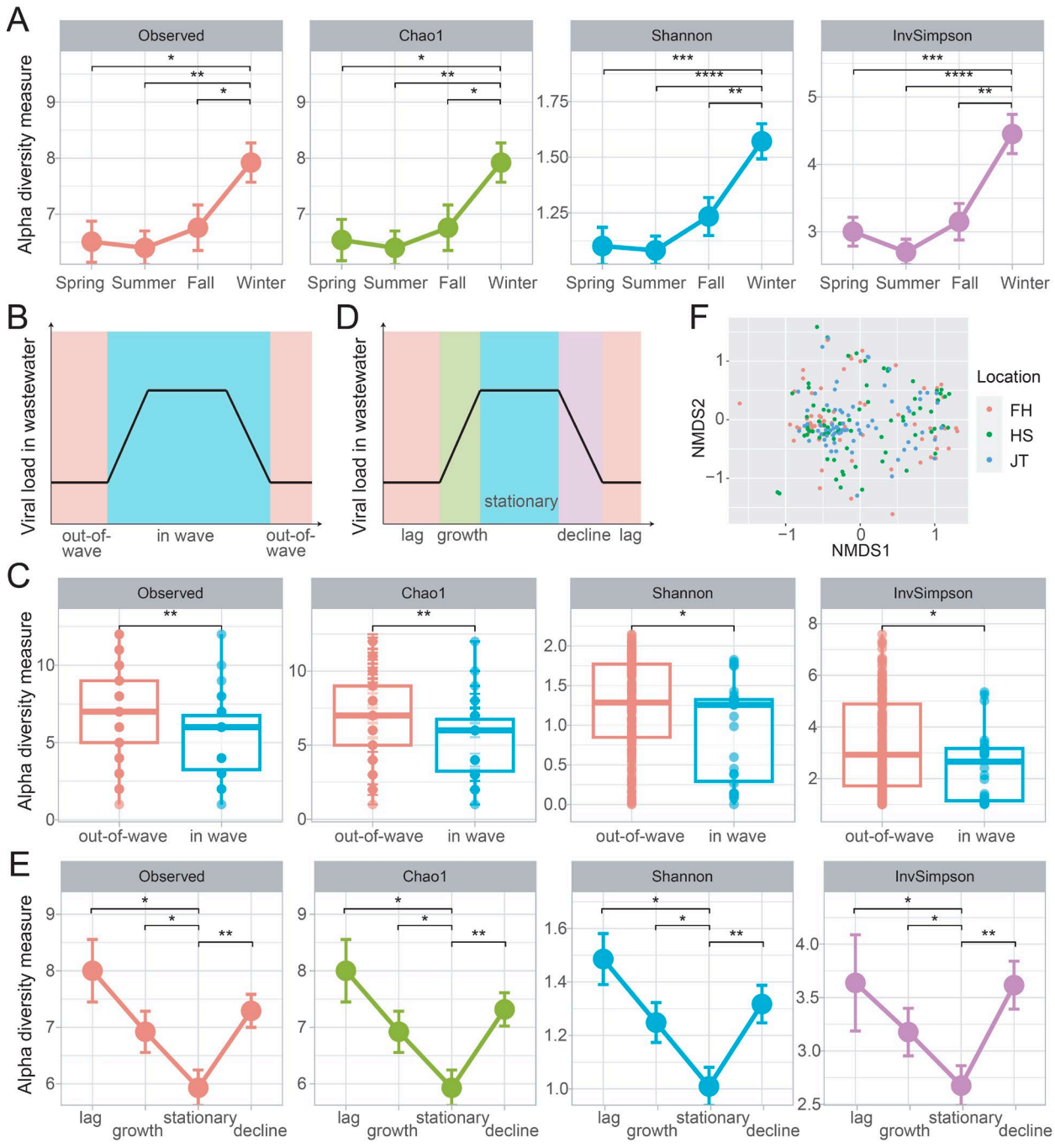


Figure 4. Early detection and identification of unreported SARS-CoV-2 variants from RBD amplicon sequencing data. (A-B) The EG.5.1 and XBB.1.16 variants were detected circulating in the city earlier than their corresponding waves in clinical genome sequencing data from El Paso, Texas, and the USA. (C-D) The EF.1.1 and GD.1 variants were found in the city but were rarely reported in clinical genome sequencing data.



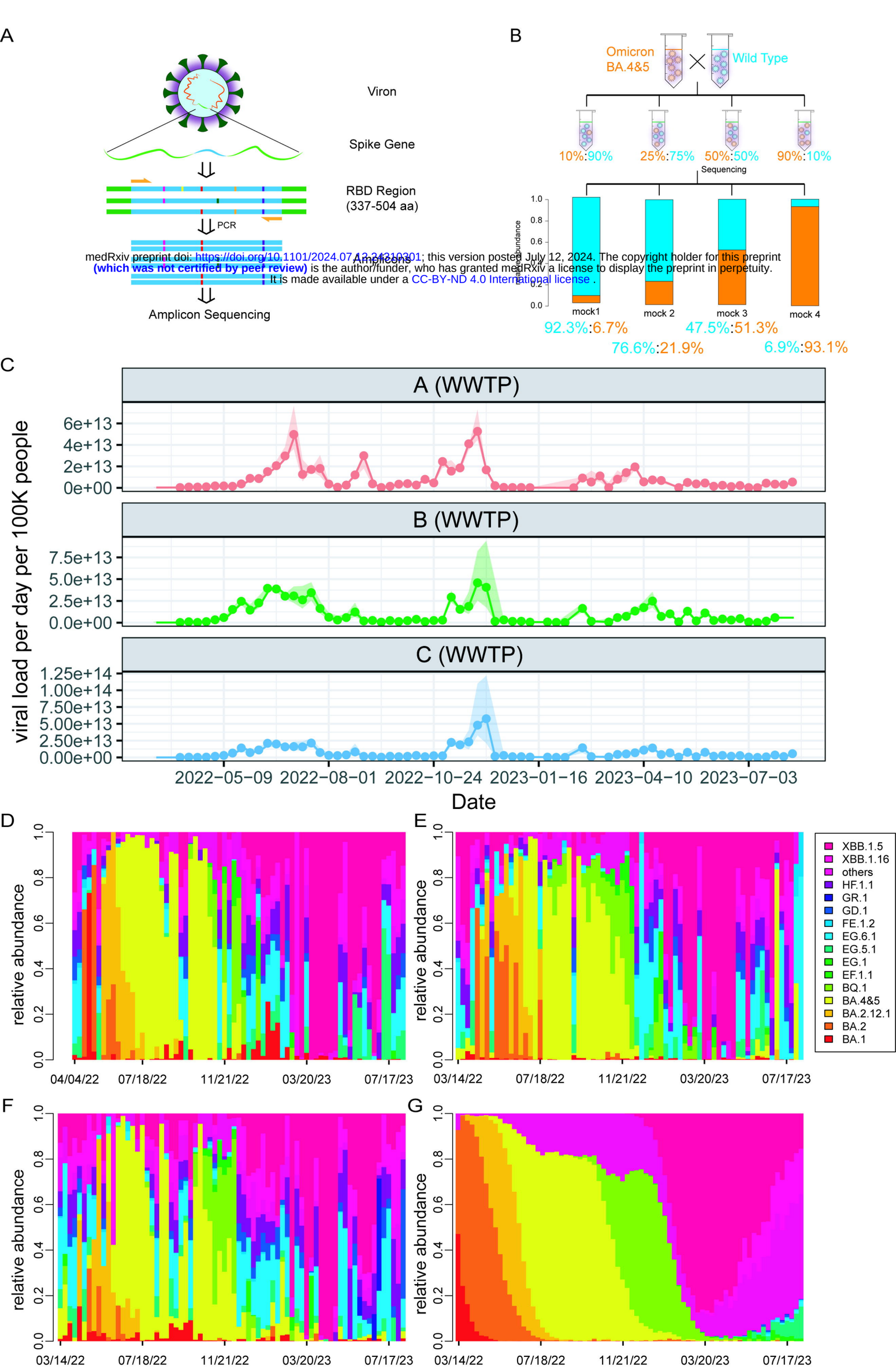
633

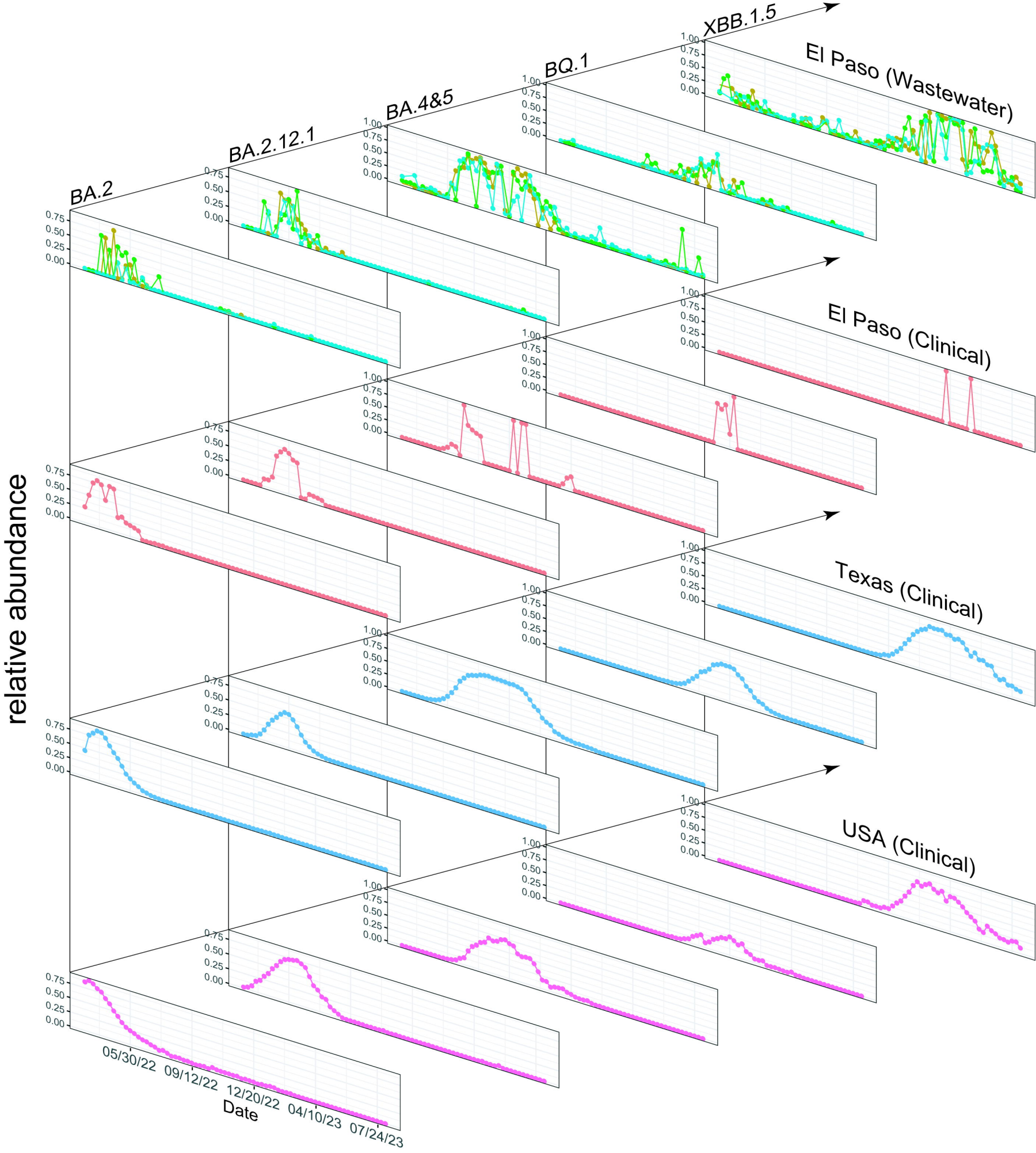
634 **Figure 4. Epidemiological patterns of variant emergence and evolution.** (A) Alpha diversity
635 comparison across seasons in the three sewersheds. (B) Schematic diagram of the high
636 transmission (in-wave) and low transmission periods (out-of-wave) in an epidemic cycle. (C)
637 Comparison of alpha diversity between in-wave and out-of-wave samples. (D) Schematic
638 diagram of an epidemic cycle: lag, growth, stationary, and decline. (E) Alpha diversity
639 comparison across the four phases: lag, growth, stationary, and decline. (F) Beta diversity

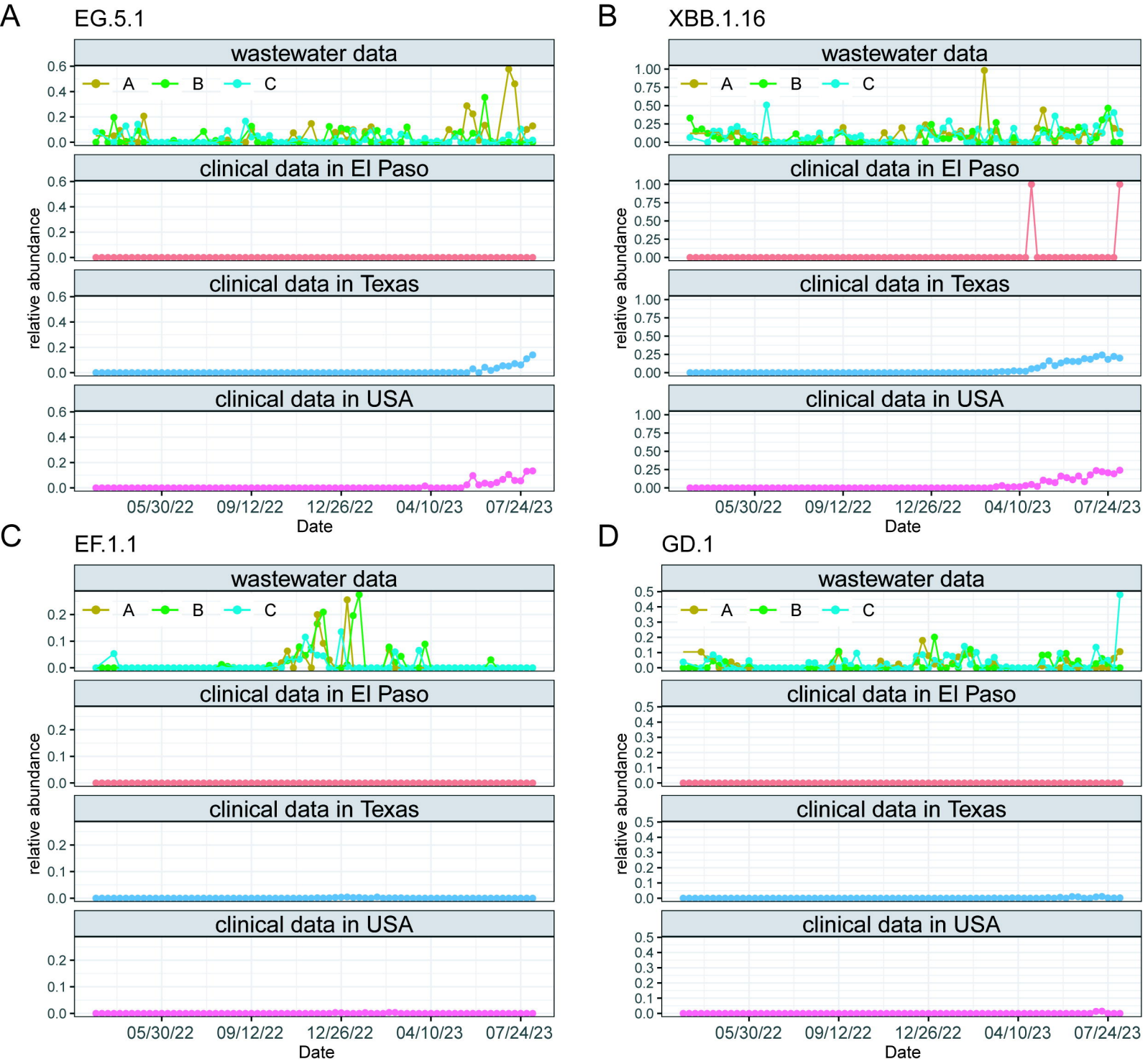
640 analysis across all three WWTPs. No significant differences in viral diversity among the three
641 sewersheds. Error bar: standard error; Wilcoxon rank-sum test was used for the group
642 comparisons with significance: *: p< 0.05; **: p<0.01; ***p<0.001. ****: p<0.0001.

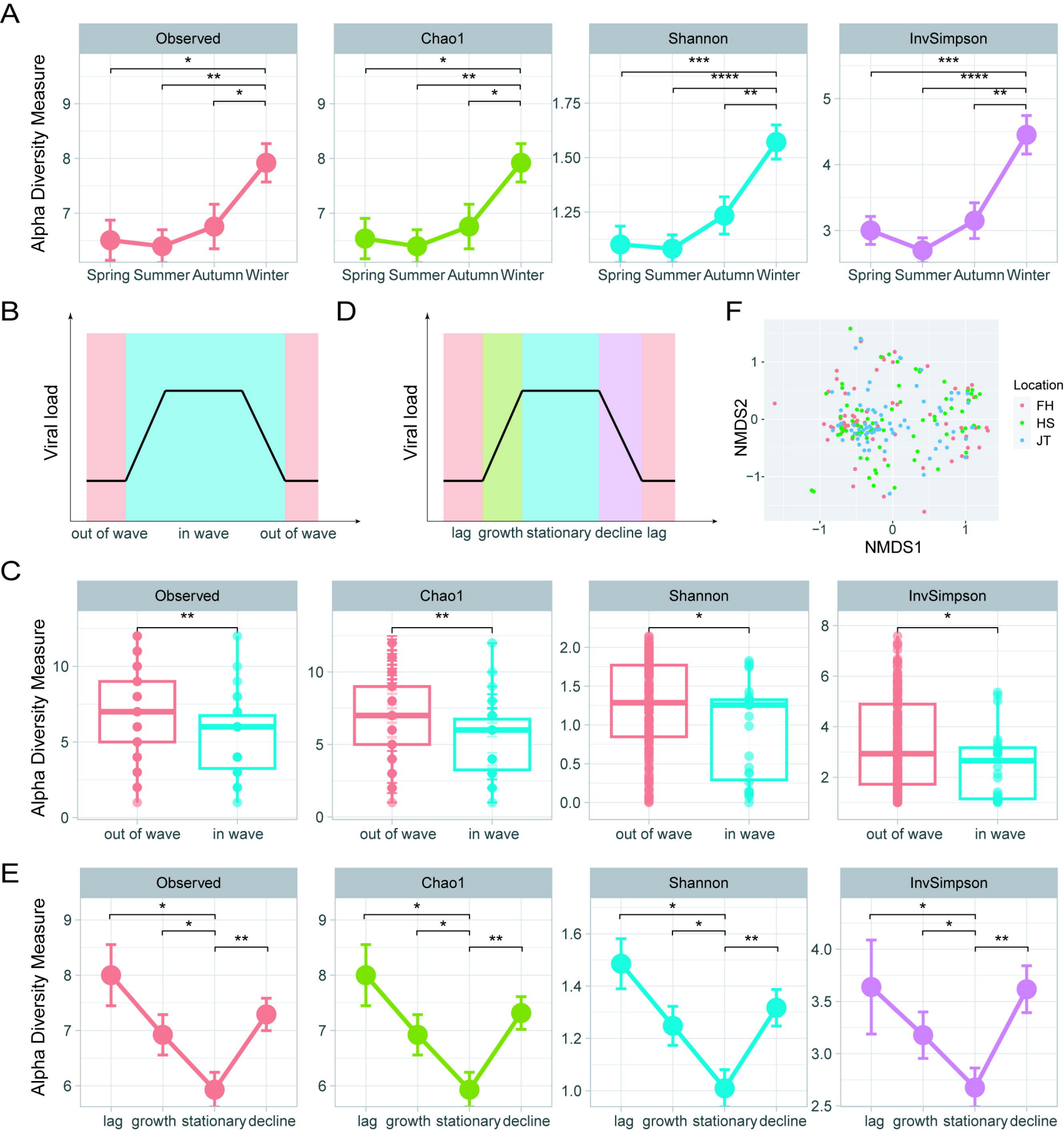
643

644




relative abundance
BA.2
BA.2.12.1
BA.4&5
BQ.1
XBB.1.5
El Paso (Wastewater)
El Paso (Clinical)
Texas (Clinical)
USA (Clinical)
0.00
0.25
0.50
0.75
1.00
05/30/22
09/12/22
12/20/22
04/10/23
07/24/23
Date


A
EG.5.1
wastewater data
A
B
C
clinical data in El Paso
clinical data in Texas
clinical data in USA
relative abundance
0.0
0.2
0.4
0.6
05/30/22
09/12/22
12/26/22
04/10/23
07/24/23
Date
B
XBB.1.16
wastewater data
clinical data in El Paso
clinical data in Texas
clinical data in USA
0.00
0.25
0.50
0.75
1.00
C
EF.1.1
D
GD.1


A
Observed
Chao1
Shannon
InvSimpson
Alpha Diversity Measure
Spring Summer Autumn Winter
B
Viral load
out of wave
in wave
out of wave
D
Viral load
lag growth stationary decline lag
F
NMDS2
NMDS1
Location
FH
HS
JT
C
Observed
Chao1
Shannon
InvSimpson
Alpha Diversity Measure
out of wave
in wave
E
Observed
Chao1
Shannon
InvSimpson
Alpha Diversity Measure
lag growth stationary decline