The Rise and Fall of SARS-CoV-2 Variants and Ongoing Diversification of Omicron

Tanner Wiegand 1, Artem Nemudryi 1, Anna Nemudraia 1, Aidan McVey 1, Agusta Little 1, David N. Taylor 2, Seth T. Walk 1 and Blake Wiedenheft 1,*

1 Department of Microbiology and Cell Biology, Montana State University, Bozeman, MT 59715, USA
2 Clinical Research Department, Bozeman Health, Bozeman, MT 59715, USA
* Correspondence: bwiedenheft@gmail.com

Abstract: In late December of 2019, high-throughput sequencing technologies enabled rapid identification of SARS-CoV-2 as the etiological agent of COVID-19, and global sequencing efforts are now a critical tool for monitoring the ongoing spread and evolution of this virus. Here, we provide a short retrospective analysis of SARS-CoV-2 variants by analyzing a subset (n = 97,437) of all publicly available SARS-CoV-2 genomes (n = ~11.9 million) that were randomly selected but equally distributed over the course of the pandemic. We plot the appearance of new variants of concern (VOCs) over time and show that the mutation rates in Omicron (BA.1) and Omicron sub-lineages (BA.2–BA.5) are significantly elevated compared to previously identified SARS-CoV-2 variants. Mutations in Omicron are primarily restricted to the spike and nucleocapsid proteins, while 24 other viral proteins—including those involved in SARS-CoV-2 replication—are generally conserved. Collectively, this suggests that the genetic distinction of Omicron primarily arose from selective pressures on the spike, and that the fidelity of replication of this variant has not been altered.

Keywords: SARS-CoV-2; Omicron; BA.4; BA.5; COVID-19; viral surveillance

1. Introduction

All viruses, including SARS-CoV-2, accumulate mutations as they replicate and spread. Most of these changes have little or no impact on the transmissibility, disease severity, or effectiveness of current vaccines and diagnostics. However, selective pressures that act on phenotypes produced by random mutations efficiently enrich rare variants with enhanced viral fitness (e.g., replication and transmissibility). Initially, SARS-CoV-2 variants were named according to their geographic origin (e.g., Wuhan), but naming based on geography can be culturally insensitive, stigmatizing, and is often inaccurate. Thus, this naming scheme was quickly replaced by a unique combination of letters and numbers known as the Pangolin nomenclature (e.g., B.1.1.7), but alphanumeric codes are cumbersome and sometimes confusing for scientists and the public alike. Part of this confusion arises from the use of aliases, which are used to limit the growing string of letters and numbers that accumulate with successive differentiation [1]. With increasing frequency, we now complement the Pangolin nomenclature with the use of a single Greek letter. Omicron (B.1.1.529 or its alias BA.1) is the 15th letter of the Greek alphabet but only the fifth variant designated as a variant of concern (VOC) by the World Health Organization (WHO).

2. Omicron, the Newest Variant of Concern

Omicron was first identified from a specimen collected on 9 November 2021 in South Africa and was designated as a VOC on 26 November [2]. This designation was based on the number of mutations (26–32) in the spike protein relative to previously sequenced isolates, as well as concerning epidemiological reports from South Africa [3,4]. Omicron has since diversified into five phylogenetically distinct sub-lineages (BA.1 to BA.5) (Figure 1) [1,5].
which are frequently associated with vaccine breakthrough or reinfection of previously infected individuals [6,7]. Here, we take a step back and provide a short retrospective overview of how SARS-CoV-2 genomes have changed over the course of the pandemic. We analyze synonymous mutations that have no impact on the protein sequence and non-synonymous mutations that change the SARS-CoV-2 proteome (Figure 2A,B). Non-synonymous mutations, particularly those in the spike, tend to attract the most attention due to the prominent role of this protein in viral attachment and entry into the host. However, synonymous mutations also accumulate over time. Many of these mutations are neutral, but some alter expression patterns [8,9] or codon usage in ways that may improve fitness [10,11]. Collectively, we analyze a high-quality subset (n = 97,437) of all SARS-CoV-2 genomes (n = ~11.3 million) that were randomly selected but equally distributed over the course of the pandemic—starting with the Wuhan reference genome (2019) and ending with high-quality Omicron genomes from the GISAID database (June 17, 2022) [12]. Overall, this analysis reveals a trend of increasing mutations over time, with a statistically significant jump in both synonymous and non-synonymous mutations in Omicron sub-lineages, relative to all the previous strains of the virus (p < 2 × 10^{-16} and p < 2 × 10^{-16}, respectively; one-way ANOVA) (Figure 2A,B).

**Figure 1.** Phylogenetic relationship of named SARS-CoV-2 variants. Variants of concern (VOC) are represented by a colored node. The phylogenetic tree was adapted from data provided by NextStrain, CoVariants (i.e., covariants.org, http://covariants.org (accessed on 18 July 2022)), and Pangolin (i.e., cov-lineages.org, http://cov-lineages.org (accessed on 18 July 2022)) [1,8,13].
Figure 2. (A) Non-synonymous mutations acquired over time in 26 SARS-CoV-2 protein sequences extracted from 97,437 genomes from 19 December 2021, to 17 June 2022 (GISAID accessions available at: https://github.com/WiedenheftLab/Omicron (accessed on 18 July 2022) and DOI: https://doi.org/10.55876/gis8.220721mv (accessed on 18 July 2022)). Genomes included in this analysis are a random sampling of 11,336,176 million SARS-CoV-2 from GISAID, which were quality filtered using NextClade (“good” overall QC status) then sampled with the Filter utility in NextStrain (selecting up to 120 genome per country per month of the pandemic) [1,14]. Variants of concern (VOC) are shown in bold and are colored as in Figure 1. Vertical lines mark the date the first sequence for each lineage was identified. The time elapsed between first detection and VOC designation by the WHO is shown as dotted lines above the graph. Dots are colored similar to variant names, and grey circles represent non-VOC lineage genomes. A linear regression line is shown in black. Omicron variants deviate from the trend. (B) Synonymous mutations in the SARS-CoV-2 genomes shown in panel (A). (C) Non-synonymous mutations in the Omicron RNA replication proteins (n = 13,094 genomes) are shown as red lines on schematic representations of each protein, and frequencies of each mutation shown as vertical lines (red). Most non-synonymous mutations are found in less than 1% of Omicron sequences. (D) Schematic depiction of SARS-CoV-2 protein coding sequences, with each gene colored according to its respective amino acid mutation rate. These rates are normalized to account for the length of each protein (i.e., substitutions/amino acids in protein/month of the pandemic).

Based on the anomalous mutational profiles of Omicron viruses, we hypothesized that Omicron genomes have mutations that affect the fidelity of viral RNA replication machinery. To test this hypothesis, we analyzed the sequences of replication-associated proteins (i.e., Nsp7, Nsp8, Nsp12, Nsp13 and Nsp14) to determine whether they had acquired mutations that would be likely to affect the fidelity of replication (Figure 2C). This analysis revealed...
only one widespread mutation in the RNA-dependent RNA polymerase (i.e., P314L in Nsp12), which arose prior to the emergence of the Omicron and co-occurred with a spike mutation (D614G) that swept to global fixation in mid-2020 and is present in 97.8% of all the SARS-CoV-2 sequences we analyzed [15,16]. The only other widespread mutations we detected in the RNA replication components of Omicron were mutations in the RNA helicase (R392C in Nsp13) and a conservative amino acid substitution (i.e., I42V) in the 3′–5′ exoribonuclease protein (Nsp14) that is responsible for proof-reading (Figure 2C) [17,18]. Since these mutated residues are distal from the Nsp13 and Nsp14 active sites [19,20] and the P314L mutation in Nsp12 is not unique to Omicron variants, it seems unlikely that they are responsible for the increase in mutations acquired by Omicron lineage viruses.

While the origins and conditions leading to the emergence of Omicron remain uncertain, some scientists suspect that Omicron arose in chronically infected, immunocompromised patients in which the immune response was too weak to clear the virus but strong enough to select for variants with increased fitness [21]. To identify other proteins that might be evolving under similar selective pressures, we evaluated mutational trends for each SARS-CoV-2 protein over the course of the pandemic (Figure 2D). Consistent with what has been observed previously, this analysis highlights the unique evolutionary signatures of the structural (i.e., S, E, M, N) and accessory proteins (i.e., 7a, 7b, 9, and 9b), which are evolving more rapidly than the non-structural proteins (NSP1-16) (Figure 2D) [5,22–24]. Such a concentration of mutations in some but not all loci is consistent with at least two different evolutionary scenarios. First, Omicron may have emerged from a dichotomy of strong purifying selection on NSP1-16 genes and strong diversifying selection on S, E, M, N, and accessory genes. Second, Omicron may represent a recombinant between a more ancestral (e.g., B1.1) lineage and a yet-to-be-discovered hyper-mutated virus (i.e., a virus with many mutations distributed more evenly across all loci) [25–27]. While there are examples of both scenarios among other viral emerging infectious diseases [28], the second scenario (recombination) is simpler and quicker compared to the necessary coincident rounds of opposing mutation-selection pressures of the first scenario. Additional sequencing of rare variants should help clarify Omicron’s evolutionary history.

3. The Future of SARS-CoV-2

Since most diagnostics target the N gene (PCR) or N protein (antigen), continued evolution of N is expected to complicate ongoing efforts to provide rapid and reliable diagnostics [29–31]. Similarly, changes in the spike will continue to have major impacts on the efficacy of natural or vaccine-induced immunity [6,7,32,33]. Increased incidence of breakthrough infections may indicate that a globally distributed and coordinated rollout of new vaccine cocktails that simultaneously targets the spike of several distinct variants [34–37] or multiple different viral genes that go beyond the spike [38,39] are necessary to establish durable immunity.

Accurately predicting the future is improbable, but it is safe to assume that the virus will continue to evolve and that we will continue to work our way through the Greek alphabet. However, at a time when Omicron variants are driving infection rates up (i.e., 3.3 million weekly cases on 5 June versus 6.7 million on 17 July) [40,41], viral surveillance is in rapid decline. Over a million SARS-CoV-2 genome sequences were deposited in GISAID in the months of November (2021) through February (2022), but submissions have dropped steadily over the last four months, and in June, the number of genomes contributed dipped below 500,000 (Figure 3) [12]. Wastewater surveillance helps fill in some of the gaps by providing community level assessments of the viral load [42], which is especially important as diagnostic testing moves out the clinic, where numbers are regularly tracked, to at-home testing, which is less sensitive and rarely reported to public health departments. There is no easy solution to this situation, but as infections increase, and tracking decreases, we are wandering blind into an uncertain future and we are at greater risk of being caught off guard by an unlucky role of the evolutionary dice, which may result in a new variant with distinct clinical outcomes.
Figure 3. SARS-CoV-2 genomic sequences submitted to the GISAID data repository between January 2020 and June 2022 [12].

Author Contributions: Conceptualization, T.W., A.M., A.N. (Artem Nemudryi), A.N. (Anna Nemudraia), and B.W.; investigation, T.W., A.M., A.L., A.N. (Artem Nemudryi), and A.N. (Anna Nemudraia); methodology, T.W., A.M., A.N. (Artem Nemudryi), A.N. (Anna Nemudraia), and B.W.; visualization, T.W., A.M., A.L., A.N. (Artem Nemudryi), A.N. (Anna Nemudraia); project administration, B.W.; writing—original draft T.W., B.W.; writing—review and editing, T.W., A.M., A.L., A.N. (Artem Nemudryi), A.N. (Anna Nemudraia), D.N.T., S.T.W. and B.W. All authors have read and agreed to the published version of the manuscript.

Funding: Research in the Wiedenheft lab is supported by the National Institutes of Health (1R35GM134867), the Montana State University (MSU) Agricultural Experimental Station, the MJ Murdock Charitable Trust, the Gianforte Foundation, the State of Montana, the City of Bozeman, a charitable donation from the Rosolowsky family, and the MSU Office of the Vice President for Research. A. Nemudryi is supported by the National Institute of Allergy and Infectious Diseases K99/R00 award (K99AI171893-01).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Accessions of GISAID sequences used in this work are available at https://github.com/WiedenheftLab/Omicron (accessed on 18 July 2022) and DOI: https://doi.org/10.55876/gis8.220721mv (accessed on 18 July 2022).

Acknowledgments: We thank the GISAID’s EpiCoV Database and contributing laboratories. The analyses in this paper would not have been possible without their willingness to share data. Accessions of sequences utilized in this work and information about the researchers who contributed these sequences are accessible at https://doi.org/10.55876/gis8.220721mv (accessed on 18 July 2022), https://doi.org/10.55876/gis8.220721mv (accessed on 18 July 2022) or github.com/WiedenheftLab/Omicron http://github.com/WiedenheftLab/Omicron (accessed on 18 July 2022).

Conflicts of Interest: B.W. is the founder of SurGene LLC, and VIRIS Detection Systems Inc. B.W., A. Nemudryi and A. Nemudraia, and A.S.-F. are inventors on patents related to CRISPR-Cas systems and applications thereof.

References
1. Rambaut, A.; Holmes, E.C.; O’Toole, A.; Hill, V.; McCrone, J.T.; Ruis, C.; du Plessis, L.; Pybus, O.G. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat. Microbiol. 2020, 5, 1403–1407. [CrossRef] [PubMed]
2. World Health Organization. Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern. Available online: https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern (accessed on 12 December 2021).
3. Callaway, E.; Ledford, H. How bad is Omicron? What scientists know so far. Nature 2021, 600, 197–199. [CrossRef] [PubMed]
4. Chen, J.; Wang, R.; Gilby, N.B.; Wei, G.W. Omicron Variant (B.1.1.529): Infectivity, Vaccine Breakthrough, and Antibody Resistance. J. Chem. Inf. Model. 2022, 62, 412–422. [CrossRef] [PubMed]
5. Hadfield, J.; Megill, C.; Bell, S.M.; Huddleston, J.; Potter, B.; Callender, C.; Sagulenko, P.; Bedford, T.; Neher, R.A. Nextstrain: Real-time tracking of pathogen evolution. Bioinformatics 2018, 34, 4121–4123. [CrossRef]
6. Cao, Y.; Yisimayi, A.; Jian, F.; Song, W.; Xiao, T.; Wang, L.; Du, S.; Wang, J.; Li, Q.; Chen, X.; et al. BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection. *Nature* **2022**, *608*, 593–602. [CrossRef]

7. Tuekprakhon, A.; Nutalai, R.; Djokjate-Guralic, A.; Zhou, D.; Ginn, H.M.; Selvaraj, M.; Liu, C.; Mentzer, A.J.; Supasa, P.; Duyvvesteyn, H.M.E.; et al. Antibody escape of SARS-CoV-2 Omicron BA.4 and BA.5 from vaccine and BA.1 serum. *Cell* **2022**, *185*, 2422–2433.e13. [CrossRef]

8. Thorne, L.G.; Bouhaddou, M.; Reuschl, A.K.; Zuliani-Alvarez, L.; Polacco, B.; Pelin, A.; Batra, J.; Whelan, M.V.X.; Hosmillo, M.; Fossati, A.; et al. Evolution of enhanced innate immune evasion by SARS-CoV-2. *Nature* **2022**, *602*, 487–495. [CrossRef]

9. Zhu, L.; Wang, Q.; Zhang, W.; Hu, H.; Xu, K. Evidence for selection on SARS-CoV-2 RNA translation revealed by the evolutionary dynamics of mutations in UTRs and CDSs. *RNA Biol.* **2022**, *19*, 866–876. [CrossRef]

10. Ramazzotti, D.; Angaroni, F.; Maspéro, D.; Mauri, M.; D’Aliberti, D.; Fontana, D.; Antoniotti, M.; Elli, E.M.; Graudenzi, A.; Piazza, R. Large-scale analysis of SARS-CoV-2 synonymous mutations reveals the adaptation to the human codon usage during the virus evolution. *Virus Evol.* **2022**, *8*, vea026. [CrossRef]

11. Wang, H.; Pipes, L.; Nielsen, R. Synonymous mutations and the molecular evolution of SARS-CoV-2 origins. *Virus Evol.* **2021**, *7*, vea098. [CrossRef]

12. Elbe, S.; Buckland-Merrett, G. Data, disease and diplomacy: GISAID’s innovative contribution to global health. *Glob. Chall.* **2017**, *1*, 33–46. [CrossRef] [PubMed]

13. Hodcroft, E. CoVariants: SARS-CoV-2 Mutations and Variants of Interest. Available online: https://covariants.org/ (accessed on 18 July 2022).

14. Aksamentov, I.; Roemer, C.; Hodcroft, E.B.; Neher, R.A. Nextclade: Clade assignment, mutation calling and quality control for viral genomes. *J. Open Source Softw.* **2021**, *6*, 3773. [CrossRef]

15. Kannan, S.R.; Spratt, A.N.; Quinn, T.P.; Heng, X.; Lorson, C.L.; Sonnerborg, A.; Byrareddy, S.N.; Singh, K. Infectivity of SARS-CoV-2: There Is Something More than D614G. *J. Neuroimmune Pharmacol.* **2020**, *15*, 574–577. [CrossRef] [PubMed]

16. Korber, B.; Fischer, W.M.; Gnankardan, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E.E.; Bhattacharya, T.; Foley, B.; et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **2020**, *182*, 812–827.e19. [CrossRef]

17. Hsu, J.C.; Laurent-Rolle, M.; Pawlak, J.B.; Wilen, C.B.; Cresswell, P. Translational shutdown and evasion of the innate immune response by SARS-CoV-2 NSP14 protein. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2101161118. [CrossRef]

18. Moeller, N.H.; Shi, K.; Demir, O.; Belica, C.; Banerjee, A.; Mossman, K. Evolutionary trajectory of SARS-CoV-2 and emerging variants. *Virus Evol.* **2020**, *6*, 3645. [CrossRef] [PubMed]

19. Colson, P.; Fournier, P.E.; Delerce, J.; Million, M.; Bedotto, M.; Houhamdi, L.; Yah, N.; Bayette, J.; Levasseur, A.; Fantini, J.; et al. Culture and identification of a “Deltamicron” SARS-CoV-2 in a three cases cluster in southern France. *J. Med. Virol.* **2022**, *94*, 3739–3749. [CrossRef] [PubMed]

20. Dolan, P.T.; Whitfield, Z.J.; Andino, R. Mechanisms and Concepts in RNA Virus Population Dynamics and Evolution. *Annu. Rev. Virol.* **2018**, *5*, 69–92. [CrossRef]

21. Dinnes, J.; Deeks, J.J.; Berhane, S.; Taylor, M.; Adrian, A.; Davenport, C.; Dittrich, S.; Emperador, D.; Takwoingi, Y.; Cunningham, J.; et al. Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. *Cochrane Database Syst. Rev.* **2021**, *3*, CD013705. [CrossRef]

22. Osterman, A.; Badell, I.; Basara, E.; Stern, M.; Kriesel, F.; Eletrroby, M.; Oztan, G.N.; Huber, M.; Autenrieth, H.; Knaibe, R.; et al. Impaired detection of omicron by SARS-CoV-2 rapid antigen tests. *Med. Microbiol. Immunol.* **2022**, *211*, 105–117. [CrossRef]

23. Bayart, J.L.; Degosserie, J.; Favresse, J.; Gillot, C.; Didembourgh, M.; Djokoto, H.P.; Verbelen, V.; Roussel, G.; Maschietto, C.; Mullier, F.; et al. Analytical Sensitivity of Six SARS-CoV-2 Rapid Antigen Tests for Omicron versus Delta Variant. *Viruses* **2022**, *14*, 654. [CrossRef]
32. Cele, S.; Jackson, L.; Khoury, D.S.; Khan, K.; Moyo-Gwete, T.; Tegally, H.; San, J.E.; Cromer, D.; Scheepers, C.; Amoako, D.G.; et al. Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization. *Nature* 2022, 602, 654–656. [CrossRef]

33. Iketani, S.; Liu, L.; Guo, Y.; Liu, L.; Chan, J.F.; Huang, Y.; Wang, M.; Luo, Y.; Yu, J.; Chu, H.; et al. Antibody evasion properties of SARS-CoV-2 Omicron sublineages. *Nature* 2022, 604, 553–556. [CrossRef] [PubMed]

34. Hawman, D.W.; Meade-White, K.; Archer, J.; Leventhal, S.S.; Wilson, D.; Shaia, C.; Randall, S.; Khandhar, A.P.; Krieger, K.; Hsiang, T.Y.; et al. SARS-CoV2 variant-specific replicating RNA vaccines protect from disease following challenge with heterologous variants of concern. *eLife* 2022, 11, e75537. [CrossRef] [PubMed]

35. Gagne, M.; Moliva, J.I.; Foulds, K.E.; Andrew, S.F.; Flynn, B.J.; Werner, A.P.; Wagner, D.A.; Teng, I.T.; Lin, B.C.; Moore, C.; et al. mRNA-1273 or mRNA-Omicron boost in vaccinated macaques elicits similar B cell expansion, neutralizing responses, and protection from Omicron. *Cell* 2022, 185, 1556–1571.e1518. [CrossRef] [PubMed]

36. Wu, K.; Choi, A.; Koch, M.; Elbashir, S.; Ma, L.; Lee, D.; Woods, A.; Henry, C.; Palandjian, C.; Hill, A.; et al. Variant SARS-CoV-2 mRNA vaccines confer broad neutralization as primary or booster series in mice. *Vaccine* 2021, 39, 7394–7400. [CrossRef]

37. Lee, I.J.; Sun, C.P.; Wu, P.Y.; Lan, Y.H.; Wang, I.H.; Liu, W.C.; Yuan, J.P.; Chang, Y.W.; Tseng, S.C.; Tsung, S.I.; et al. A booster dose of Delta × Omicron hybrid mRNA vaccine produced broadly neutralizing antibody against Omicron and other SARS-CoV-2 variants. *J. Biomed. Sci.* 2022, 29, 49. [CrossRef] [PubMed]

38. Silva, E.; Bomfim, C.G.; Barbosa, A.P.; Noda, P.; Noronha, I.L.; Fernandes, B.H.V.; Machado, R.R.G.; Durigon, E.L.; Catanozi, S.; Rodrigues, L.G.; et al. Immunization with SARS-CoV-2 Nucleocapsid protein triggers a pulmonary immune response in rats. *PLoS ONE* 2022, 17, e0268434. [CrossRef] [PubMed]

39. Dangi, T.; Class, J.; Palacio, N.; Richner, J.M.; Penaloza MacMaster, P. Combining spike- and nucleocapsid-based vaccines improves distal control of SARS-CoV-2. *Cell Rep.* 2021, 36, 109664. [CrossRef]

40. COVID19-Map—John Hopkins Resource Center. Available online: https://coronavirus.jhu.edu/map.html (accessed on 21 July 2022).

41. Dong, E.; Du, H.; Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* 2020, 20, 533–534. [CrossRef]

42. Centers for Disease Control and Prevention: COVID Data Tracker—Wastewater Surveillance. Available online: https://covid.cdc.gov/covid-data-tracker/#/wastewater-surveillance (accessed on 20 July 2022).