Perspective

Evolutionary Mechanisms of the Emergence of the Variants of Concern of SARS-CoV-2

Igor M. Rouzine

Sechenov Institute of Evolutionary Physiology and Biochemistry, Russian Academy of Sciences, St. Petersburg 194223, Russia; igor.rouzine@iephb.ru

Abstract: The evolutionary origin of the variants of concern (VOCs) of SARS-CoV-2, characterized by a large number of new substitutions and strong changes in virulence and transmission rate, is intensely debated. The leading explanation in the literature is a chronic infection in immunocompromised individuals, where the virus evolves before returning into the main population. The present article reviews less-investigated hypotheses of VOC emergence with transmission between acutely infected hosts, with a focus on the mathematical models of stochastic evolution that have proved to be useful for other viruses, such as HIV and influenza virus. The central message is that understanding the acting factors of VOC evolution requires the framework of stochastic multi-locus evolution models, and that alternative hypotheses can be effectively verified by fitting results of computer simulation to empirical data.

Keywords: immunosuppressed patient; chronic infection; primary mutation; compensatory mutation; recombination

1. Introduction

Similarly to HIV, HCV, and influenza virus, SARS-CoV-2 is perpetually acquiring new substitutions in its genome, with a very high mean substitution rate of $(0.6–1.6) \cdot 10^{-3}$/year/site, depending on the subtype [1,2] (Figure 1). The substitution rate of its envelope protein, Spike, is even faster, at $(5–6) \cdot 10^{-3}$/year/site, and second only to the envelope protein of HIV [3]. Millions of genomes have been sequenced to date, producing a vast amount of data on genetic variation, antigenicity, and transmission [4–8]. The variants of concern (VOCs) of SARS-CoV-2, such as Alpha (B.1.1.7), Delta (B.1.617.2), and Omicron (B.1.1.529 and its subvariants B.A1, B.A2, etc.), particularly stand out, due to their changes in virulence and transmission rate. For example, the variant Alpha had at least 43% faster transmission than the previous virus variant [9]. A VOC carries a new set of a dozen or more mostly non-synonymous mutations [10–13]. Phylogenetic analysis demonstrates that a VOC does not descend directly from any of the current dominant virus variants, but represents a sister strain, with a common ancestor that existed many months prior [10–13]. The evolutionary cause of this phenomenon remains unknown. Understanding VOC genesis demands a concerted effort of clinicians, epidemiologists, virologists, immunologists, and virus evolution experts.
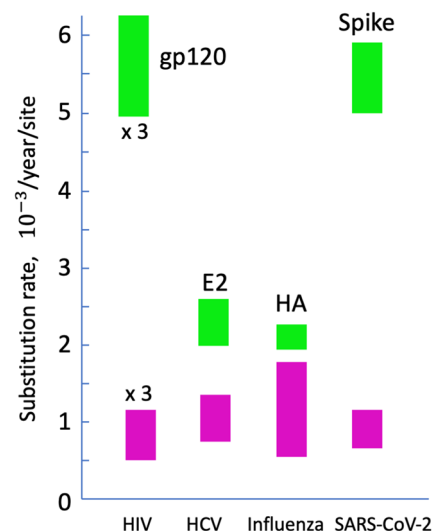
Figure 1. Viral substitution rates. The purple rectangles show the intervals of the median values for the most rapidly and most slowly evolving subtypes of HIV, HCV, influenza virus, and SARS-CoV-2 for the full genome [1,14–16]. The green rectangles correspond to the envelope proteins targeted by neutralizing antibodies [3,17–19]. To fit into the plot, HIV values are shown at a third of their actual value. Based on [20].

2. Epidemiological Reservoirs (Silent Spread)

One possible explanation postulates the existence of a hidden epidemiological reservoir, where the virus evolves un-noticed and then returns to the population. Such a reservoir could be an animal population (reverse zoonosis) [21,22] or a human population not covered by genetic surveillance [6,13,23,24]. The two hypotheses are similar. The facts in favor of them include numerous documented cases of a pathogen passing between an animal reservoir and a human population (e.g., bubonic plague, HIV, influenza), as well as certain SARS-CoV-2 variants being found only in specific locations, at least early on following their first detection [24]. A difference between the animal reservoir hypothesis and the hypothesis of silent spread in a human population is the presumed evolutionary rate. Another difference is that cross-species jumps often require additional adaptation mutations. The silent spread hypothesis could be confirmed by tracing a VOC to a specific human or animal reservoir.

3. Evolution in Immunocompromised Individuals

Much discussed in the literature is the hypothesis that VOCs may evolve within immunocompromised individuals who exhibit a long-term chronic infection [10,25–30]. Due to its adaptation to a human host in the presence of a weakened immune response, the virus could rapidly accumulate many mutations and then return to the general population. The prevalence of chronic infections in the population is small, 0.1–0.5% [31], but they could seed new highly fit variants. In the support of this idea, evolution in immunocompromised patients and VOCs share common features, including some hallmark mutations in the Spike protein and convergent evolution [10]. Several studies have found an accelerated substitution rate within some chronic individuals compared to the substitution rate of the virus in the main population [25–30], although some have challenged this [2]. However, acceleration is not specific evidence for seeding VOCs from these patients. The speed of viral evolution is often faster in chronically infected individuals due to a faster turnover; this is a common phenomenon among viruses. These data are also consistent with viral adaptation occurring in parallel and independently at two biological levels. One level is the evolution within a chronic patient, and another level is the evolution due to transmission

in the general population (in hosts with acute infections). Both processes confer viral adaptation to human species, and likely select for many (but not all) similar substitutions.

A search for drivers of VOC evolution in 27 chronic patients demonstrated that, although substitutions in chronic patients include some hallmark VOC mutations, the long subsets of mutations associated with the global transmission of VOCs are absent from these patients [32]. The difference between VOCs and the virus in chronic patients could be explained by an evolutionary conflict between two biological levels [32,33]: the level of a chronically infected host and the level of a population. The evolutionary conflict, however, does not require chronic infection, and can also result from the parallel evolution at the two biological levels. Nevertheless, the chronic patient hypothesis remains the leading hypothesis in the literature, and deserves further investigation. Identification of the source patients for at least some VOCs would allow direct confirmation. Recently, shedding of novel SARS-CoV-2 Spike sequences to wastewater was traced to a single commercial building, and has been detected and interpreted as coming from a chronically infected individual [34].

4. Evolutionary Forces in the General Population

The above class of potential explanations was based on virus evolution in and seeding from a special reservoir of population, other than the general population. Another direction of the search is the detailed analysis of biological mechanisms shaping virus evolution in the general population. In principle, the same forces can also act in reservoirs, but the presence of reservoirs is not important for this line of inquiry. Viruses evolve due to a complex interplay between various biological factors, including random mutation, natural selection, epistatic interaction, random genetic drift, linkage between loci due to common phylogeny, recombination, migration, and inhomogeneous population structure (Box 1). The interplay between these evolutionary factors, which can be readily simulated in silico, generates a set of specific evolutionary patterns, which can be compared to genomic data in the public domain. All hypotheses of this class are testable, and their confirmation or falsification would help with understanding the origins of VOCs. Two examples follow.

Box 1. Terminology.

Allele—a variant of a locus.
Compensatory mutation—a mutation at another locus, compensating the deleterious effect of a mutation at the given locus.
Convergent evolution—evolution towards the same genetic sequence in independent populations.
Fitness (reproduction number)—the average number of cells infected by a virus from one infected cell, or the average number of individuals infected by one individual.
Fitness valley—a sequence of mutations with a transient fitness decrease in the middle.
Genetic drift—random fluctuation in the progeny number of a virus genome, measured by the number of infected cells for within-host evolution and the number of infected hosts for population-level evolution, respectively. It causes random fluctuations in allelic frequency over time, and may cause their eventual loss.
Genetic linkage—co-inheritance of alleles at different loci because they are close together on a chromosome (i.e., low chance of recombination splitting them up)
Immune escape—a mutation causing a loss of recognition.
Locus—a small variable part of a genome (usually, a nucleotide or amino acid position)
Natural selection—an increase in the number of genetic variants with a larger fitness over generations.
Recombination—a mosaic mix of two individual genomes, due to them switching templates during replication.

5. Fitness Valley Hypothesis

VOCs may arise early on, while still undetectable, due to a cascade of compensatory mutations emerging in a population after a primary mutation, conferring a major change

in viral properties [35]. A population-level simulation study proposed this effect in conjunction with the evolution in immunocompromised patients [35]. However, under certain conditions, a fitness valley could also be crossed in the general population, in direct patient-to-patient transmission. The authors of [36] used computer simulation to interpret the origin of VOCs using simple fitness landscapes (a single mutation, several additive mutations, and a fitness plateau followed by a single mutation). The fitness valley, in contrast, includes a transient decrease in fitness forced by some changes in the environment (tissue, host, species, immune response). This "fitness valley" effect has been previously observed or inferred for various viruses, including HIV [37,38] and influenza [39,40], and studied using mathematical models [40,41]. For example, a primary mutation can help a virus to escape the immune response, adapt to a species or a tissue, or expand the pH range in which it can replicate. Mutations conferring these benefits to a virus, but reducing its overall fitness, are followed by compensatory mutations that elevate its fitness above the initial level.

Compensatory mutations (positive epistasis) have also been observed for SARS-CoV-2. Mutations in neutralizing antibody-binding regions that help SARS-CoV-2 to escape immune recognition during transmission [3,4,6,20,42] negatively affect the function of receptor binding, and may also have to be compensated for elsewhere. K417N and E484K in Spike seem to help it to avoid antibody recognition, but they also remove two salt-bridges with ACE2 (the target receptor on our cells); however, N501Y appears to increase affinity for ACE2, and might compensate for the loss of the salt-bridges [43]. The tradeoff between antibody escape and transmissibility has been inferred in immunocompromised patients [32]. The same is likely true for mutations conferring escape from innate responses to SARS-CoV-2 [44,45].

A classic example of the fitness valley effect is the rapid formation of drug resistance in HIV in patients treated with early antiretroviral inhibitors, where the initial primary mutation decreasing the viral sensitivity to a drug is followed by many compensatory mutations [46,47]. For example, HIV-1 develops resistance to protease inhibitors due to a protease mutation that reduces binding affinity [38,48,49]. The fitness cost of drug resistance is high, and requires one or more additional mutations in protease to compensate for reduced fitness in order for resistance to evolve. Compensatory cleavage-site mutations in the capsid protein are required as well [50].

Another example is mutations in the HIV genome that help it to escape the cytotoxic immune response in a host [51,52] (Figure 2A). The viral replication capacity lowered by an escape mutation is later gradually restored by compensatory mutations [53–55]. Compensatory alleles with the highest degree of compensation grow the fastest, creating a cascade of compensatory mutations over time. This process is thought to be responsible for the high diversity and rapid HIV evolution in untreated patients [56]. The loss of fitness due to escape mutations, and the partial loss of recognition by the immune system, together determine the order, the number, and the evolution rates of escape mutations [57,58]. Compensatory mutations prevent the reversion of escape mutations after the decay of CD8 T cell clones against escaped epitopes. Bacteria developing antibiotic resistance also pass through such fitness valleys [59].

To test the "fitness valley" hypothesis, one has to measure the network of epistatic interactions responsible for the transition between the initial and final viral strains. The fitness landscape comprising the epistatic network, and the costs of separate mutations, can be inferred from genetic samples using various methods [40,60–64]. The methods that take into account the effects of genetic linkage are more reliable. The signature of a fitness valley is a simple epistatic network, with a shape close to that of a star, and approximately one path connecting any two nodes (connectivity close to 1). Indeed, most

epistatic connections exist due to compensation for the same primary mutation. Tertiary mutations compensating for the compensatory mutations can add some loops to the star, raising the connectivity slightly above 1. In one example, to explain the origin of the 2009 pandemic of Influenza A H1N1, the epistatic network was inferred for the neuraminidase protein [40] (Figure 2B). The epistatic network has a mostly a star-like appearance due to direct compensatory mutations, with a few loops due to secondary compensations. The primary mutation 248 is well known to decrease the sensitivity of the virus to pH [65]. Observation of a similar network for SARS-CoV-2 would represent evidence in favor of the "fitness valley" hypothesis. At the present moment, the observation of strong compensation in SARS-CoV-2 [43] represents indirect evidence, because it proves that compensatory mutations required for the fitness valley effect, in principle, exist for this virus.
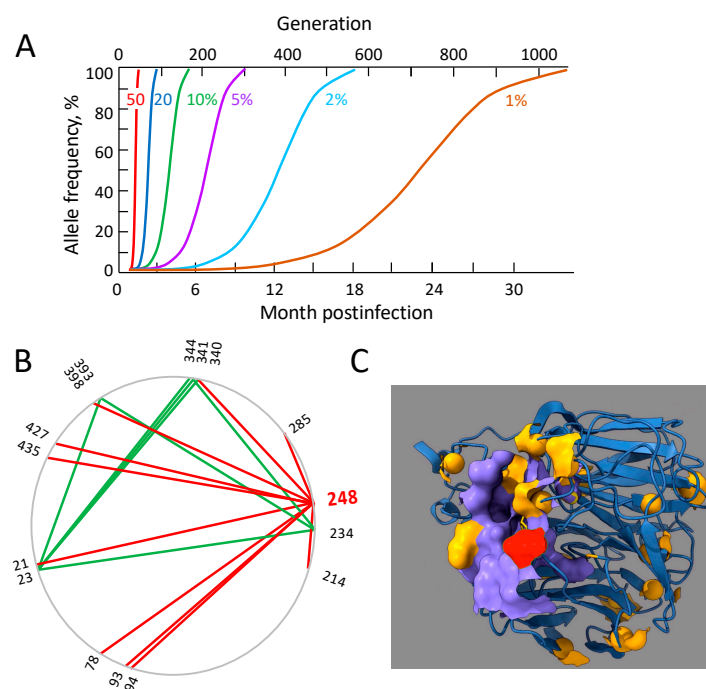

Figure 2. The fitness valley hypothesis. (**A**) Consecutive fixation of alleles, compensating for the fitness effect of the primary immune escape mutation in an HIV-infected patient. The percent selection coefficients are shown on the curves. (**B**) An example of an epistatic network predicted from genomic sequence data for the surface protein sequences of Influenza A H1N1, to explain the pandemics of 2009 [40]. The circular diagram shows the network of interaction between variable amino acid sites in the neuraminidase protein, as inferred by the 3-site haplotype method. The red lines show residue 248 with the primary mutation and the direct compensatory residues. The green lines show secondary compensatory mutations. (**C**) The structural location of the predicted epistasis network in a neuraminidase monomer of influenza virus. The three-dimensional structure of Influenza A H1N1 neuraminidase (PDB ID code 4QVZ). The colored spheres represent the predicted epistatic residues from (**B**). Red sphere: Predicted primary mutation, residue 248 in (**B**). Yellow spheres: Compensatory residues from (**B**). Note that all compensatory residues are located in alpha helixes on the surface of protein, representing a case of allosteric interaction. Based on [40].

Another test of the "fitness valley" hypothesis is that locations of the inferred compensatory residues must follow a structural pattern that is consistent with a specific mechanism of compensation. In this example, all compensatory sites are located on the protein surface in alpha helixes. Apparently, they serve to re-adjust the angles between beta-sheets that are changed by the primary mutation 248, representing a case of allosteric interaction (Figure 2C).

6. Multi-Locus Evolution with Recombination

Variants with large sets of newly acquired mutations could also arise due to multiple recombination events, with fitter recombinants then rising in frequency through natural selection. Indeed, SARS-CoV-2 has observable crossover recombination due to co-infection events [66–69]. In this hypothesis, highly fit sequences are formed, when still at undetectable levels, by rare recombination events (Figure 3A). This possibility is supported by computational and mathematical models, that have not yet been applied to SARS-CoV-2, which take advantage of the fact that, due to random recombination crossovers, some rare progeny genomes can have an unusually large number of beneficial alleles compared to their parents [70–73]. These models demonstrate that the evolution of a viral population can be understood as a distribution of genomes in fitness traveling along the fitness coordinate. The front of the distribution moves forward towards a higher fitness, because these rare recombinants, with an excess of beneficial alleles, are the most fit in a population [70]. Starting from a single copy, although not detectable initially, such recombinant lineages eventually outgrow all others, and become a new dominant strain in a population (Figure 3A). The time delay with which this process occurs, and the rapid growth of the lineage, could be an explanation for the "sudden" emergence of VOCs. Interestingly, a computer simulation based on such a model has generated a polyphyletic tree and a distribution of beneficial mutations among branches (Figure 3B) resembling that of a VOC [74] (Figure 3C). Initially, the genetic distance from the original virus strain increased linearly with time. When the variant Alpha appeared in the autumn of 2021, it deviated from this gradual increase by exhibiting a jump by 10 distinct mutations, including three short deletions, one knockout, and six amino acid replacements (Figure 3C). Phylogeny hinted at a common ancestor of Alpha (B.1.1.7) and the previous strain (20E or B.1.177) occurring sometime around March. The simulation (Figure 3B) reproduced these features qualitatively, including independent subtrees at different and even the same moments of time, and jumps in the mutation number. For the simulation details, see the caption of Figure 3 and [74].
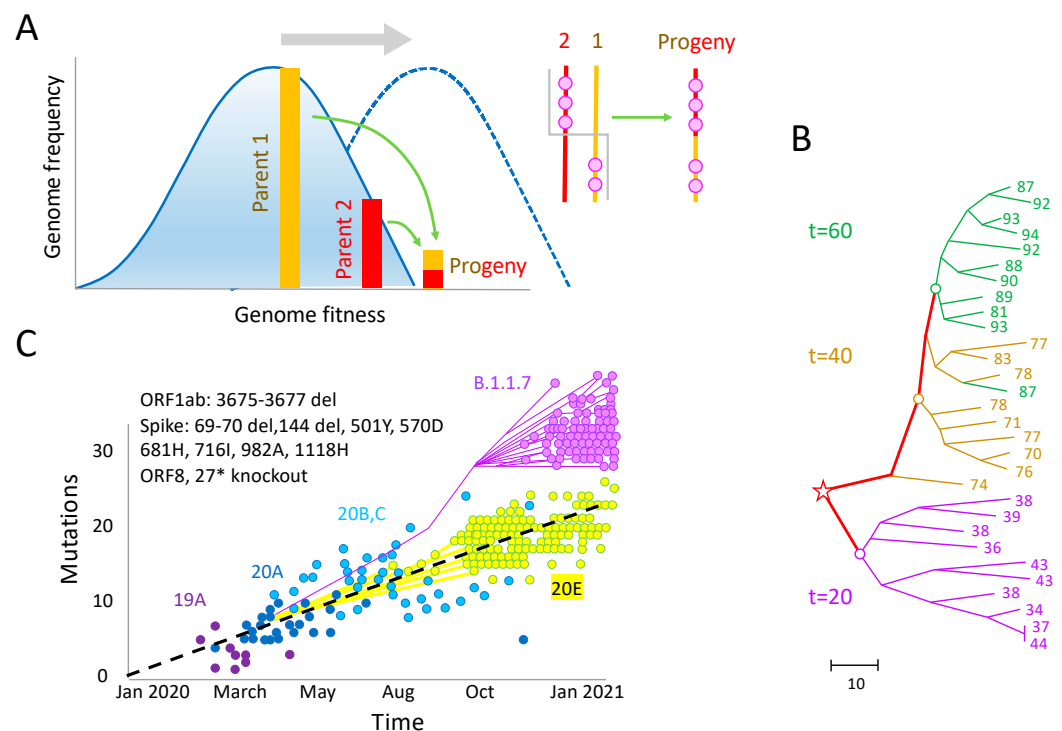

Figure 3. Multi-locus evolution with recombination. (**A**) The recombination of two viruses (parents, yellow and red bars) sometimes generates a highly fit recombinant virus (progeny), which extends

the edge of genome distribution (blue solid curve) in fitness. It becomes the dominant strain later, when the distribution of genomes over fitness, the travelling fitness wave, moves over (gray arrow, blue dashed curve). Inset: the pink dots are beneficial alleles in two parental genomes that recombine into the progeny. (**B**) A phylogenetic tree for 10 genomes sampled at three time points from a population, shown in generations in 3 colors, the evolution of which is simulated by the Monte Carlo method (Fisher–Wright process). Th open circles show the roots of the three subtrees, which correspond to the previously best-fit genomes. The tree was obtained by neighbor-joining analysis in the MEGA11 package (version 11.0.10). The red star shows the root of the tree, chosen from the sampled sequences by MEGA11. The numbers at the leaves show the number of beneficial alleles in the genome. The genetic distance scale is below. Simulation parameters: population size $N = 1000$, outcrossing rate $r = 1$, locus number $L = 500$, selection coefficient $s_0 = 0.1$, initial allele frequency $f_0 = 0.02$, $M = 3$ crossovers. (**C**) The dependence of the mutation number on time for various clades of SARS-CoV-2 during the formation of the VOC Alpha (B.1.1.7). Phylogeny is shown for two clades competing in UK in the Fall 2020, clade 20E (B.1.177) and variant B.1.1.7. (**B,C**) is based on [74,75], respectively.

The simplest way to rule out the recombination hypothesis is to demonstrate, with the use of a computer simulation, that the recombination incidence of SARS-CoV-2 is too low to cause such a frequent emergence of VOCs. The unknown input parameter is the effective outcrossing rate, defined as the probability that a genome undergoes recombination with another genome. This quantity depends on the effective rate of co-infection with two viral variants [71,73]. The last value is unknown, and can be either higher or lower than the average rate of individual co-infection, because transmission rate varies strongly among individuals for a variety of reasons, including superspreading events [76–78].

Fortunately, there are methods of inference of the outcrossing rate from genomic data. For example, the outcrossing rate in an average untreated HIV-infected individual was demonstrated, by two independent methods, to be around 1.5% [79,80]. Such techniques could be re-applied to SARS-CoV-2 data. If the estimated outcrossing rate is too low, and a Monte Carlo simulation based on it does not predict the emergence of VOCs with a frequency similar to that observed in real data, the hypothesis is falsified. This would not mean, of course, that recombination events do not occur, but only that they are insufficient to explain the emergence of the VOC under investigation. If, however, it is quantitatively consistent with the observed emergence, the hypothesis will be elevated to the rank of a working model. The discovery of the two parental strains of a VOC would provide its ultimate confirmation, and bioinformatics tools could help with wide screening.

Note that multi-locus models of evolution predict that the fittest strains are formed by sequential multiple recombination events [70,73,74], which may be difficult to detect. Genetic evidence demonstrates the importance of recombination in the formation of recent genetic variants of SARS-CoV-2 [81–85]. However, the fact that a new strain is a recombinant does not prove that recombination is the primary driver of their emergence, but only that it participates in its formation. The most important evolutionary factor can be hiding behind the scene. Only by performing computer simulation for alternative dynamic models, and comparing their predictions to genomic data, can it be determined which evolutionary mechanism is the primary driver [56,79,86–88].

The above hypotheses are not mutually excluding. For example, fitness valley crossing or recombination may occur within chronically infected patients, and then transfer to the population. A simulation study at the level of a population [35] has found that, when the pathogen does not have to cross a fitness valley for immune escape to occur (no epistasis), this has no principal effect on antigenic evolution in immunocompromised individuals. However, if a fitness valley exists at the between-host level (epistasis), persistent infections of immunocompromised individuals allow mutations to accumulate in a population. Usually, a fitness valley is crossed under a change of conditions, such as in the presence of an

immune response or an antiviral, or during transmission between species, individual hosts, or organs; these factors need to be investigated.

An interesting question is whether the evolutionary pressure from infection prevention and control (lockdowns, vaccines, masks) affects the hypotheses presented in this manuscript. Because these hypotheses are poorly studied in the context of SARS-CoV-2, and we do not know which one is true, only future simulation and data analysis will be able to answer this question.

7. Conclusions

To summarize, silent evolution in population reservoirs or chronically infected patients is not the only possibility that one should focus on in the search of explanations for VOCs. The explanations based on the interplay between different factors driving virus evolution in the general population, such as a fitness valley and multi-locus evolution with recombination, which have received much less attention in the literature on SARS-CoV-2 but are well studied for other viruses, deserve more careful investigation. Although these explanations do not require evolution in chronically infected patients, they are compatible with it. Once one of these hypotheses about evolutionary mechanisms is confirmed, as described above, and the relevant parameters are estimated, the probabilistic prediction of future VOCs in the next few years, by Monte Carlo simulation, will become possible.

Conflicts of Interest: The author declares no conflict of interests.

References

1. Duchene, S.; Featherstone, L.; Haritopoulou-Sinanidou, M.; Rambaut, A.; Lemey, P.; Baele, G. Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evol.* **2020**, *6*, veaa061. [CrossRef] [PubMed]
2. Översti, S.; Gaul, E.; Jensen, B.E.O.; Kühnert, D. Phylogenetic meta-analysis of chronic SARS-CoV-2 infections in immunocompromised patients shows no evidence of elevated evolutionary rates. *bioRxiv* **2023**. [CrossRef]
3. Yewdell, J.W. Antigenic drift: Understanding COVID-19. *Immunity* **2021**, *54*, 2681–2687. [CrossRef] [PubMed]
4. Carabelli, A.M.; Peacock, T.P.; Thorne, L.G.; Harvey, W.T.; Hughes, J.; Consortium, C.-G.U.; Peacock, S.J.; Barclay, W.S.; de Silva, T.I.; Towers, G.J.; et al. SARS-CoV-2 variant biology: Immune escape, transmission and fitness. *Nat. Rev. Microbiol.* **2023**, *21*, 162–177. [CrossRef] [PubMed]
5. Flores-Vega, V.R.; Monroy-Molina, J.V.; Jimenez-Hernandez, L.E.; Torres, A.G.; Santos-Preciado, J.I.; Rosales-Reyes, R. SARS-CoV-2: Evolution and Emergence of New Viral Variants. *Viruses* **2022**, *14*, 653. [CrossRef] [PubMed]
6. Shrestha, L.B.; Foster, C.; Rawlinson, W.; Tedla, N.; Bull, R.A. Evolution of the SARS-CoV-2 omicron variants BA.1 to BA.5: Implications for immune escape and transmission. *Rev. Med. Virol.* **2022**, *32*, e2381. [CrossRef] [PubMed]
7. Telenti, A.; Hodcroft, E.B.; Robertson, D.L. The Evolution and Biology of SARS-CoV-2 Variants. *Cold Spring Harb. Perspect. Med.* **2022**, *12*. [CrossRef] [PubMed]
8. Singh, D.; Yi, S.V. On the origin and evolution of SARS-CoV-2. *Exp. Mol. Med.* **2021**, *53*, 537–547. [CrossRef] [PubMed]
9. Davies, N.G.; Jarvis, C.I.; Group, C.C.-W.; Edmunds, W.J.; Jewell, N.P.; Diaz-Ordaz, K.; Keogh, R.H. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature* **2021**, *593*, 270–274. [CrossRef] [PubMed]
10. Corey, L.; Beyrer, C.; Cohen, M.S.; Michael, N.L.; Bedford, T.; Rolland, M. SARS-CoV-2 Variants in Patients with Immunosuppression. *N. Engl. J. Med.* **2021**, *385*, 562–566. [CrossRef]
11. Tay, J.H.; Porter, A.F.; Wirth, W.; Duchene, S. The emergence of SARS-CoV-2 variants of concern is driven by acceleration of the substitution rate. *Mol. Biol. Evol.* **2022**, *39*, msac013. [CrossRef] [PubMed]
12. Martin, D.P.; Lytras, S.; Lucaci, A.G.; Maier, W.; Gruning, B.; Shank, S.D.; Weaver, S.; MacLean, O.A.; Orton, R.J.; Lemey, P.; et al. Selection Analysis Identifies Clusters of Unusual Mutational Changes in Omicron Lineage BA.1 That Likely Impact Spike Function. *Mol. Biol. Evol.* **2022**, *39*, msac061. [CrossRef] [PubMed]

13. Markov, P.V.; Ghafari, M.; Beer, M.; Lythgoe, K.; Simmonds, P.; Stilianakis, N.I.; Katzourakis, A. The evolution of SARS-CoV-2. *Nat. Rev. Microbiol.* **2023**, *21*, 361–379. [CrossRef] [PubMed]

14. Patino-Galindo, J.A.; Gonzalez-Candelas, F. The substitution rate of HIV-1 subtypes: A genomic approach. *Virus Evol.* **2017**, *3*, vex029. [CrossRef]

15. Yuan, M.; Lu, T.; Li, C.; Lu, L. The evolutionary rates of HCV estimated with subtype 1a and 1b sequences over the ORF length and in different genomic regions. *PLoS ONE* **2013**, *8*, e64698. [CrossRef] [PubMed]

16. Nobusawa, E.; Sato, K. Comparison of the mutation rates of human influenza A and B viruses. *J. Virol.* **2006**, *80*, 3675–3678. [CrossRef]

17. Dapp, M.J.; Kober, K.M.; Chen, L.; Westfall, D.H.; Wong, K.; Zhao, H.; Hall, B.M.; Deng, W.; Sibley, T.; Ghorai, S.; et al. Patterns and rates of viral evolution in HIV-1 subtype B infected females and males. *PLoS ONE* **2017**, *12*, e0182443. [CrossRef]

18. Gray, R.R.; Parker, J.; Lemey, P.; Salemi, M.; Katzourakis, A.; Pybus, O.G. The mode and tempo of hepatitis C virus evolution within and among hosts. *BMC Evol. Biol.* **2011**, *11*, 131. [CrossRef] [PubMed]

19. Strelkowa, N.; Lassig, M. Clonal interference in the evolution of influenza. *Genetics* **2012**, *192*, 671–682. [CrossRef]

20. Rouzine, I.M.; Rozhnova, G. Evolutionary implications of SARS-CoV-2 vaccination for the future design of vaccination strategies. *Commun. Med.* **2023**, *3*, 86. [CrossRef] [PubMed]

21. Kane, Y.; Wong, G.; Gao, G.F. Animal Models, Zoonotic Reservoirs, and Cross-Species Transmission of Emerging Human-Infecting Coronaviruses. *Annu. Rev. Anim. Biosci.* **2023**, *11*, 1–31. [CrossRef] [PubMed]

22. McBride, D.; Garushyants, S.; Franks, J.; Magee, A.; Overend, S.; Huey, D.; Williams, A.; Faith, S.; Kandeil, A.; Trifkovic, S.; et al. Accelerated evolution of SARS-CoV-2 in free-ranging white-tailed deer. *Res. Sq.* **2023**, *14*, 5105. [CrossRef] [PubMed]

23. Kemp, S.A.; Collier, D.A.; Datir, R.P.; Ferreira, I.; Gayed, S.; Jahun, A.; Hosmillo, M.; Rees-Spear, C.; Mlcochova, P.; Lumb, I.U.; et al. SARS-CoV-2 evolution during treatment of chronic infection. *Nature* **2021**, *592*, 277–282. [CrossRef] [PubMed]

24. Smyth, D.S.; Trujillo, M.; Gregory, D.A.; Cheung, K.; Gao, A.; Graham, M.; Guan, Y.; Guldenpfennig, C.; Hoxie, I.; Kannoly, S.; et al. Tracking cryptic SARS-CoV-2 lineages detected in NYC wastewater. *Nat. Commun.* **2022**, *13*, 635. [CrossRef] [PubMed]

25. Borges, V.; Isidro, J.; Cunha, M.; Cochicho, D.; Martins, L.; Banha, L.; Figueiredo, M.; Rebelo, L.; Trindade, M.C.; Duarte, S.; et al. Long-Term Evolution of SARS-CoV-2 in an Immunocompromised Patient with Non-Hodgkin Lymphoma. *mSphere* **2021**, *6*, e00244-21. [CrossRef] [PubMed]

26. Chaguza, C.; Hahn, A.M.; Petrone, M.E.; Zhou, S.; Ferguson, D.; Breban, M.I.; Pham, K.; Pena-Hernandez, M.A.; Castaldi, C.; Hill, V.; et al. Accelerated SARS-CoV-2 intrahost evolution leading to distinct genotypes during chronic infection. *Cell Rep. Med.* **2023**, *4*, 100943. [CrossRef] [PubMed]

27. Choi, B.; Choudhary, M.C.; Regan, J.; Sparks, J.A.; Padera, R.F.; Qiu, X.; Solomon, I.H.; Kuo, H.H.; Boucau, J.; Bowman, K.; et al. Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. *N. Engl. J. Med.* **2020**, *383*, 2291–2293. [CrossRef] [PubMed]

28. Ciuffreda, L.; Lorenzo-Salazar, J.M.; Alcoba-Florez, J.; Rodriguez-Perez, H.; Gil-Campesino, H.; Inigo-Campos, A.; Garcia-Martinez de Artola, D.; Valenzuela-Fernandez, A.; Hayek-Peraza, M.; Rojo-Alba, S.; et al. Longitudinal study of a SARS-CoV-2 infection in an immunocompromised patient with X-linked agammaglobulinemia. *J. Infect.* **2021**, *83*, 607–635. [CrossRef] [PubMed]

29. Hettle, D.; Hutchings, S.; Muir, P.; Moran, E.; The COVID-19 Genomics UK (COG-UK) Consortium. Persistent SARS-CoV-2 infection in immunocompromised patients facilitates rapid viral evolution: Retrospective cohort study and literature review. *Clin. Infect. Pract.* **2022**, *16*, 100210. [CrossRef] [PubMed]

30. Stanevich, O.V.; Alekseeva, E.I.; Sergeeva, M.; Fadeev, A.V.; Komissarova, K.S.; Ivanova, A.A.; Simakova, T.S.; Vasilyev, K.A.; Shurygina, A.P.; Stukova, M.A.; et al. SARS-CoV-2 escape from cytotoxic T cells during long-term COVID-19. *Nat. Commun.* **2023**, *14*, 149. [CrossRef] [PubMed]

31. Ghafari, M.; Hall, M.; Golubchik, T.; Ayoubkhani, D.; House, T.; MacIntyre-Cockett, G.; Fryer, H.R.; Thomson, L.; Nurtay, A.; Kemp, S.A.; et al. Prevalence of persistent SARS-CoV-2 in a large community surveillance study. *Nature* **2024**, *626*, 1094–1101. [CrossRef] [PubMed]

32. Harari, S.; Tahor, M.; Rutsinsky, N.; Meijer, S.; Miller, D.; Henig, O.; Halutz, O.; Levytskyi, K.; Ben-Ami, R.; Adler, A.; et al. Drivers of adaptive evolution during chronic SARS-CoV-2 infections. *Nat. Med.* **2022**, *28*, 1501–1508. [CrossRef] [PubMed]

33. Rast, L.I.; Rouzine, I.M.; Rozhnova, G.; Bishop, L.; Weinberger, A.D.; Weinberger, L.S. Conflicting selection pressures will constrain viral escape from interfering particles: Principles for designing resistance-proof antivirals. *PLoS Comput. Biol.* **2016**, *12*, e1004799. [CrossRef] [PubMed]

34. Shafer, M.M.; Bobholz, M.J.; Vuyk, W.C.; Gregory, D.A.; Roguet, A.; Haddock Soto, L.A.; Rushford, C.; Janssen, K.H.; Emmen, I.E.; Ries, H.J.; et al. Tracing the origin of SARS-CoV-2 omicron-like spike sequences detected in an urban sewershed: A targeted, longitudinal surveillance study of a cryptic wastewater lineage. *Lancet Microbe* **2024**, *5*, e335–e344. [CrossRef] [PubMed]

35. Smith, C.A.; Ashby, B. Antigenic evolution of SARS-CoV-2 in immunocompromised hosts. *Evol. Med. Public. Health* **2023**, *11*, 90–100. [CrossRef]

36. Ghafari, M.; Liu, Q.; Thilon, A.; Katzurakis, A. Investigating the evolutionary origins of the first three SARS-CoV-2 variants of concern. *Front. Virol.* **2022**, *2*, 942555. [CrossRef]

37. Bonhoeffer, S.; Chappey, C.; Parkin, N.T.; Whitcomb, J.M.; Petropoulos, C.J. Evidence for positive epistasis in HIV-1. *Science* **2004**, *306*, 1547–1550. [CrossRef] [PubMed]

38. Zhang, T.H.; Dai, L.; Barton, J.P.; Du, Y.; Tan, Y.; Pang, W.; Chakraborty, A.K.; Lloyd-Smith, J.O.; Sun, R. Predominance of positive epistasis among drug resistance-associated mutations in HIV-1 protease. *PLoS Genet.* **2020**, *16*, e1009009. [CrossRef] [PubMed]

39. Kryazhimskiy, S.; Dushoff, J.; Bazykin, G.A.; Plotkin, J.B. Prevalence of epistasis in the evolution of influenza A surface proteins. *PLoS Genet.* **2011**, *7*, e1001301. [CrossRef] [PubMed]

40. Pedruzzi, G.; Rouzine, I.M. An evolution-based high-fidelity method of epistasis measurement: Theory and application to influenza. *PLoS Pathog.* **2021**, *17*, e1009669. [CrossRef] [PubMed]

41. Weissman, D.B.; Desai, M.M.; Fisher, D.S.; Feldman, M.W. The rate at which asexual populations cross fitness valleys. *Theor. Popul. Biol.* **2009**, *75*, 286–300. [CrossRef] [PubMed]

42. Cobey, S.; Larremore, D.B.; Grad, Y.H.; Lipsitch, M. Concerns about SARS-CoV-2 evolution should not hold back efforts to expand vaccination. *Nat. Rev. Immunol.* **2021**, *21*, 330–335. [CrossRef] [PubMed]

43. Moulana, A.; Dupic, T.; Phillips, A.M.; Chang, J.; Nieves, S.; Roffler, A.A.; Greaney, A.J.; Starr, T.N.; Bloom, J.D.; Desai, M.M. Compensatory epistasis maintains ACE2 affinity in SARS-CoV-2 Omicron BA.1. *Nat. Commun.* **2022**, *13*, 7011. [CrossRef] [PubMed]

44. Reuschl, A.K.; Thorne, L.G.; Whelan, M.V.X.; Ragazzini, R.; Furnon, W.; Cowton, V.M.; De Lorenzo, G.; Mesner, D.; Turner, J.L.E.; Dowgier, G.; et al. Evolution of enhanced innate immune suppression by SARS-CoV-2 Omicron subvariants. *Nat. Microbiol.* **2024**, *9*, 451–463. [CrossRef]

45. 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] [PubMed]

46. Gonzalez-Ortega, E.; Ballana, E.; Badia, R.; Clotet, B.; Este, J.A. Compensatory mutations rescue the virus replicative capacity of VIRIP-resistant HIV-1. *Antiviral Res.* **2011**, *92*, 479–483. [CrossRef] [PubMed]

47. Handel, A.; Regoes, R.R.; Antia, R. The role of compensatory mutations in the emergence of drug resistance. *PLoS Comput. Biol.* **2006**, *2*, e137. [CrossRef]

48. Piana, S.; Carloni, P.; Rothlisberger, U. Drug resistance in HIV-1 protease: Flexibility-assisted mechanism of compensatory mutations. *Protein Sci.* **2002**, *11*, 2393–2402. [CrossRef] [PubMed]

49. *HIV Drug Resistance Database*; Standford University: Stanford, CA, USA, 2024. Available online: https://hivdb.stanford.edu/dr-summary/resistance-notes/PI/ (accessed on 7 January 2025).

50. Fun, A.; Wensing, A.M.; Verheyen, J.; Nijhuis, M. Human Immunodeficiency Virus Gag and protease: Partners in resistance. *Retrovirology* **2012**, *9*, 63. [CrossRef] [PubMed]

51. Goonetilleke, N.; Liu, M.K.; Salazar-Gonzalez, J.F.; Ferrari, G.; Giorgi, E.; Ganusov, V.V.; Keele, B.F.; Learn, G.H.; Turnbull, E.L.; Salazar, M.G.; et al. The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. *J. Exp. Med.* **2009**, *206*, 1253–1272. [CrossRef] [PubMed]

52. Ganusov, V.V.; De Boer, R.J. Estimating costs and benefits of CTL escape mutations in SIV/HIV infection. *PLoS Comput. Biol.* **2006**, *2*, e24. [CrossRef] [PubMed]

53. Liu, D.; Zuo, T.; Hora, B.; Song, H.; Kong, W.; Yu, X.; Goonetilleke, N.; Bhattacharya, T.; Perelson, A.S.; Haynes, B.F.; et al. Preexisting compensatory amino acids compromise fitness costs of a HIV-1 T cell escape mutation. *Retrovirology* **2014**, *11*, 101. [CrossRef] [PubMed]

54. Crawford, H.; Prado, J.G.; Leslie, A.; Hue, S.; Honeyborne, I.; Reddy, S.; van der Stok, M.; Mncube, Z.; Brander, C.; Rousseau, C.; et al. Compensatory mutation partially restores fitness and delays reversion of escape mutation within the immunodominant HLA-B*5703-restricted Gag epitope in chronic human immunodeficiency virus type 1 infection. *J. Virol.* **2007**, *81*, 8346–8351. [CrossRef] [PubMed]

55. Song, H.; Pavlicek, J.W.; Cai, F.; Bhattacharya, T.; Li, H.; Iyer, S.S.; Bar, K.J.; Decker, J.M.; Goonetilleke, N.; Liu, M.K.; et al. Impact of immune escape mutations on HIV-1 fitness in the context of the cognate transmitted/founder genome. *Retrovirology* **2012**, *9*, 89. [CrossRef]

56. Rouzine, I.M.; Coffin, J.M. Search for the mechanism of genetic variation in the pro gene of human immunodeficiency virus. *J. Virol.* **1999**, *73*, 8167–8178. [CrossRef] [PubMed]

57. Althaus, C.L.; De Boer, R.J. Dynamics of immune escape during HIV/SIV infection. *PLoS Comput. Biol.* **2008**, *4*, e1000103. [CrossRef] [PubMed]

58. Batorsky, R.; Sergeev, R.A.; Rouzine, I.M. The route of HIV escape from immune response targeting multiple sites is determined by the cost-benefit tradeoff of escape mutations. *PLoS Comput. Biol.* **2014**, *10*, e1003878. [CrossRef] [PubMed]

59. Levin, B.R.; Perrot, V.; Walker, N. Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. *Genetics* **2000**, *154*, 985–997. [CrossRef]

60. Marsh, L. Spatial autocorrelation of amino Acid replacement rates in the vasopressin receptor family. *J. Mol. Evol.* **2009**, *68*, 28–39. [CrossRef] [PubMed]

61. Luksza, M.; Lassig, M. A predictive fitness model for influenza. *Nature* **2014**, *507*, 57–61. [CrossRef]

62. Neher, R.A.; Russell, C.A.; Shraiman, B.I. Predicting evolution from the shape of genealogical trees. *eLife* **2014**, *3*, e03568. [CrossRef]

63. Rodriguez-Rivas, J.; Croce, G.; Muscat, M.; Weigt, M. Epistatic models predict mutable sites in SARS-CoV-2 proteins and epitopes. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2113118119. [CrossRef] [PubMed]

64. Likhachev, I.V.; Rouzine, I.M. Measurement of selection coefficients from genomic samples of adapting populations by computer modeling. *STAR Protoc.* **2023**, *4*, 101821. [CrossRef] [PubMed]

65. Takahashi, T.; Song, J.; Suzuki, T.; Kawaoka, Y. Mutations in NA that induced low pH-stability and enhanced the replication of pandemic (H1N1) 2009 influenza A virus at an early stage of the pandemic. *PLoS ONE* **2013**, *8*, e64439. [CrossRef] [PubMed]

66. Ignatieva, A.; Hein, J.; Jenkins, P.A. Ongoing recombination in SARS-CoV-2 revealed through genealogical reconstruction. *Mol. Biol. Evol.* **2022**, *39*, msac028. [CrossRef] [PubMed]

67. Jackson, B.; Boni, M.F.; Bull, M.J.; Colleran, A.; Colquhoun, R.M.; Darby, A.C.; Haldenby, S.; Hill, V.; Lucaci, A.; McCrone, J.T.; et al. Generation and transmission of interlineage recombinants in the SARS-CoV-2 pandemic. *Cell* **2021**, *184*, 5179–5188.e8. [CrossRef] [PubMed]

68. Yi, H. 2019 novel coronavirus is undergoing active recombination. *Clin. Infect. Dis.* **2020**, *71*, 884–887. [CrossRef] [PubMed]

69. Turakhia, Y.; Thornlow, B.; Hinrichs, A.; McBroome, J.; Ayala, N.; Ye, C.; Smith, K.; De Maio, N.; Haussler, D.; Lanfear, R.; et al. Pandemic-scale phylogenomics reveals the SARS-CoV-2 recombination landscape. *Nature* **2022**, *609*, 994–997. [CrossRef] [PubMed]

70. Rouzine, I.M.; Coffin, J.M. Evolution of human immunodeficiency virus under selection and weak recombination. *Genetics* **2005**, *170*, 7–18. [CrossRef]

71. Rouzine, I.M.; Coffin, J.M. Highly fit ancestors of a partly sexual haploid population. *Theor. Popul. Biol.* **2007**, *71*, 239–250. [CrossRef]

72. Neher, R.A.; Shraiman, B.I.; Fisher, D.S. Rate of adaptation in large sexual populations. *Genetics* **2010**, *184*, 467–481. [CrossRef] [PubMed]

73. Neher, R.A.; Kessinger, T.A.; Shraiman, B.I. Coalescence and genetic diversity in sexual populations under selection. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 15836–15841. [CrossRef] [PubMed]

74. Rouzine, I.M. Long-range linkage effects in adapting sexual populations. *Sci. Rep.* **2023**, *13*, 12492. [CrossRef] [PubMed]

75. Bedford, T. Evolutionary Dynamics of SARS-CoV-2. Available online: https://videocast.nih.gov/watch=43795 (accessed on 6 January 2025).

76. Otto, S.P.; Day, T.; Arino, J.; Colijn, C.; Dushoff, J.; Li, M.; Mechai, S.; Van Domselaar, G.; Wu, J.; Earn, D.J.D.; et al. The origins and potential future of SARS-CoV-2 variants of concern in the evolving COVID-19 pandemic. *Curr. Biol.* **2021**, *31*, R918–R929. [CrossRef]

77. Lau, M.S.Y.; Grenfell, B.; Thomas, M.; Bryan, M.; Nelson, K.; Lopman, B. Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA. *Proc. Natl. Acad. Sci USA* **2020**, *117*, 22430–22435. [CrossRef] [PubMed]

78. Liu, Y.; Eggo, R.M.; Kucharski, A.J. Secondary attack rate and superspreading events for SARS-CoV-2. *Lancet* **2020**, *395*, e47. [CrossRef] [PubMed]

79. Batorsky, R.; Kearney, M.F.; Palmer, S.E.; Maldarelli, F.; Rouzine, I.M.; Coffin, J.M. Estimate of effective recombination rate and average selection coefficient for HIV in chronic infection. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 5661–5666. [CrossRef] [PubMed]

80. Neher, R.A.; Leitner, T. Recombination rate and selection strength in HIV intra-patient evolution. *PLoS Comput. Biol.* **2010**, *6*, e1000660. [CrossRef] [PubMed]

81. Tamura, T.; Ito, J.; Uriu, K.; Zahradnik, J.; Kida, I.; Anraku, Y.; Nasser, H.; Shofa, M.; Oda, Y.; Lytras, S.; et al. Virological characteristics of the SARS-CoV-2 XBB variant derived from recombination of two Omicron subvariants. *Nat. Commun.* **2023**, *14*, 2800. [CrossRef]

82. Li, J.Y.; Wang, H.Y.; Cheng, Y.X.; Ji, C.; Weng, S.; Han, N.; Yang, R.; Zhou, H.Y.; Wu, A. Comprehensive detection and dissection of interlineage recombination events in the SARS-CoV-2 pandemic. *Virus Evol.* **2024**, *10*, veae074. [CrossRef] [PubMed]

83. Kodsi, I.A.; Rayes, D.E.; Koweyes, J.; Khoury, C.A.; Rahy, K.; Thoumi, S.; Chamoun, M.; Haddad, H.; Mokhbat, J.; Tokajian, S. Tracking SARS-CoV-2 variants during the 2023 flu season and beyond in Lebanon. *Virus Res.* **2024**, *339*, 199289. [CrossRef] [PubMed]

84. Branda, F.; Ciccozzi, M.; Scarpa, F. Genetic variability of the recombinant SARS-CoV-2 XEC: Is it a new evolutionary dead-end lineage? *New Microbes New Infect.* **2024**, *62*, 101520. [CrossRef] [PubMed]

85. Garcia, I.; Brate, J.; Fossum, E.; Rohringer, A.; Moen, L.V.; Hungnes, O.; Fjaere, O.; Zaragkoulias, K.; Bragstad, K. Recombinant SARS-CoV-2 Delta/Omicron BA.5 emerging in an immunocompromised long-term infected COVID-19 patient. *Sci. Rep.* **2024**, *14*, 25790. [CrossRef] [PubMed]

86. Rouzine, I.M.; Coffin, J.M. Linkage disequilibrium test implies a large effective population number for HIV in vivo. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 10758–10763. [CrossRef] [PubMed]

87. Rouzine, I.M.; Weinberger, L.S. The quantitative theory of within-host viral evolution. *J. Stat. Mech. Theory Exp.* **2013**, *2013*, P01009. [CrossRef]

88. Rouzine, I.M.; Weinberger, A.D.; Weinberger, L.S. An evolutionary role for HIV latency in enhancing viral transmission. *Cell* **2015**, *160*, 1002–1012. [CrossRef] [PubMed]