1 Introduction

In contrast to the seasonal Influenza virus, where evolution has been rather gradual and linear [4], SARS-CoV2 has seen unexpected patterns of viral evolution that puzzled scientists [10].

This is particularly true for the viral strains – known as variants – associated with the "Omicron" lineage, which has proven to be only a distant relative to the previously dominating "Delta" variant [2]. Accordingly, variants from the Omicron lineage tend to evade immunity acquired from other variants as well as (non-adapted) vaccinations [8].

Various plausible explanations have been put forward to explain this unexpected phenomenon of non-linear evolution including a prolonged evolution in reservoirs that are not under surveillance, such as immunocompromised hosts, or animal reservoirs [10]. Some studies suggest that a predecessor of Omicron has developed in mice [11, 9], while other studies show that the virus is able to survive and evolve in some HIV patients, leading to idiosyncratic patterns of viral evolution [3]. Phylogenetic analysis suggests that Omicron shares common ancestry with the "Lambda" variant [2].

We aim to shed light on the co-dynamics of the spread and evolution of viruses using a theoretical framework. To this end, we couple a version of the standard epidemiological SIR model [6] with evolutionary dynamics as suggested by [12]. We transform an evolutionary-epidemiological agent-based model [7] back into a SIR-based approach, taking into account important findings of [4].

Our results do not preclude explanations to the origins of Omicron based on animal reservoirs or immunocompromised human carriers, but illustrate that there may be no need for them to explain a highly non-linear viral evolution as observed of the SARS-CoV2 Omicron variant. Instead, we show that such a non-linearity can emerge endogenously through interventions common during the Covid-19 pandemic.

2 Methodology

Despite introducing the assumption of perfect mixing the results of [4] regarding linear evolutionary patterns can still be reproduced. In turn, we gain the ability for a more detailed sensitivity analysis. In addition, we simplify the model for intervention policies during a pandemic of [7]. To ease the evaluation of our model results we introduce a metric for tree linearity outlined below.

2.1 Mutations

The simulated RNA strand in each strain is divided into 12 codons. Each consists of 3 nucleotide bases. If one base mutates a new strain is added to the phylogeny.

The mutability is one of the core parameters in [4]'s analysis. They find the most linear mutation pattern in a low mutability of $1e-6$ mutations per day

and infected individual [4]. Hence, the lowest mutability is of high interest to test our hypothesis.

Like in [7] the strain specific parameters of the epidemiological model are also subject to mutation.

2.2 Epidemiological model

The ODE system derived from [7]'s agent-based model is a classic SEPIARD model with cross immunity and an additional compartment for the short lived immunity of [4]. Two compartments are global: That of the susceptible S and that of the short lived immune F. The other compartments are local to each variant i in the phylogenetic tree.

Cross immunity between strains is also based on [4]'s model. The codons of the simulated RNA mentioned above can be translated into a sequence of proteins. An antigenic distance $d_{i,j}$ between two strains i and j can be derived by comparing their protein sequences. It serves as a proxy for the probability to evade a previously infected host immune system. For further details on the cross immunity model refer to [4] where we also find the basis for our parameters.

2.3 Interventions

Our intervention model is rooted in the assumption that the threshold of symptomatic infections for interventions is inversely correlated to the observed fatality of the virus. We employ a base threshold τ_b giving the threshold at a base death rate λ_b. This actual threshold τ_i is then found as the product of τ_b and an acute death rate dependent factor:

$$\tau_i = \tau_b * \lambda_b * N / \sum_i \dot{(}D)_i \tag{1}$$

In case an intervention is in place, the infectiousness β_s of all strains s is multiplied by a reduction factor $\beta_i \in [0, 1]$

2.4 Metrics

To quantify the effects of the intervention we introduce a metric targeted at the key feature of phylogenetic linearity. This way we can also extend the qualitative analysis previously done by [4]. We base this metric on the impact a strain at s in tree t has within the phylogeny. Our quantification can be weighted with a weight function $f_w \geq 0$, where $f_w(s) > 0$ for some s in t. This impact I is defined as the weight of a strain and the sum of the weights of all its descendants:

$$I(t, s) = f_w(t_s) + \sum_c f_w(t_c) \tag{2}$$

Where c covers all children of s. The mean linearity Λ of a tree can then be based on the ratio of maximum child impact and total strain impact.

$$\Lambda(t) = \frac{\sum_i f_w(t_i) + \sum_i \max_{c(i)} I(t,c)}{\sum_i I(t,i)} \tag{3}$$

In effect, Λ is defined in the interval $]0,1]$. A linearity of 1 represents a tree where each node barring a single final leaf bears exactly 1 child. In this case Λ is independent of the weight function f_w. In contrast, a low linearity can be achieved by distributing the weight uniformly among a large number of children.

In addition, we provide a metric h based on entropy. This metric is less sensitive to the specificity of linearity. However, it can be rooted in literature and yields a better representation of order within the tree [5, 1]. The path entropy is defined on the probability p that a path in a tree takes a node n [5]. Again, we allow for a weighting of the metric with a weight function f_w inspired by [1]. Using a weighted entropy we find a weighted mean of the marginal information components in the tree:

$$h(t) = -\frac{\sum_i \sum_c f_w(t_c) * 1/n_i * \log 1/n_i}{\sum_i f_w(t_i)} \tag{4}$$

3 Results

To test our hypothesis, we perform a sensitivity analysis of multiple parameters of the intervention model. We consider base thresholds τ_b of $5,000$, $10,000$ and $15,000$, with base death rates λ_b of $5e-6$, $1e-5$ and $5e-5$ and reduction factors β_i of 0.1, 0.3 and 0.5. The mutability is kept low with $1e-6$ mutations per day and infected agent, to provide for the most linear baseline.

These scenarios are compared against a baseline scenario with no interventions. In contrast to [4]'s simulations, our scenarios deal with a novel pathogen. Thus, our simulations do not start at equilibrium. Reproductions of [4]'s results were successfully performed with prior equilibration though.

Figure 1 pane a shows the linearity metric weighted by total infections of a strain for different levels of intervention. It depicts the mean and standard deviation of a total of 35 simulations for each parameter combination. As visible, the phylogenetic linearity is reduced in all intervention scenarios. This effect is strongest in the beginning of the simulated epidemic. The interventions successfully reduce the infections. However, this also reduces the amount of agents with short lived immunity. Although there are fewer mutations, new strains face a more vulnerable public in a high number of susceptible agents. As a result, the phylogenetic linearity is decreased.

The similarly weighted tree path entropy is depicted in Figure 1 pane b. The entropy data stems from a total of 30 simulations per parameter combination. It shows a similar short term effect of the interventions. In contrast to the linearity,

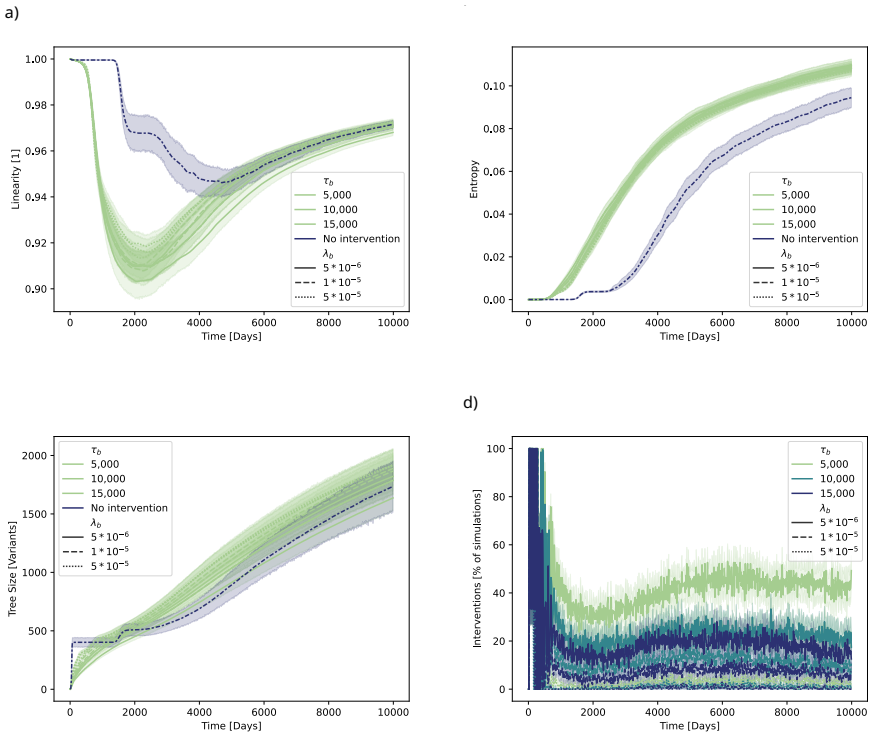

Figure 1: Simulation results as a mean of 35 simulations per parameter combination. Pane a) shows the linearity in scenarios with and without interventions. The entropy metric is plotted in b) and pane c) shows the evolution of the phylogenetic tree size. In d) we depict the prevalence of interventions among the simulation runs for different intervention scenarios.

the entropy remains higher throughout the rest of the simulation. This could be due to the larger size of the trees in intervention scenarios shown in pane *c*.

Figure 1 pane *d* shows the prevalence of interventions for different scenarios. Despite differences in the abundance of interventions the effects on tree linearity do not differ by much. It appears that initial interventions are key in the tree shape heterogeneity. This may have important implications for future pandemic management.

References

[1] Benjamin Allen, Mark Kon, and Yaneer Bar-Yam. A new phylogenetic diversity measure generalizing the shannon index and its application to phyllostomid bats. *The American Naturalist*, 174(2):236–243, 2009.

[2] Kanika Bansal and Sanjeet Kumar. Mutational cascade of SARS-CoV-2 leading to evolution and emergence of omicron variant. *Virus Research*, 315:198765, 2022.

[3] Sandile Cele, Farina Karim, Gila Lustig, James Emmanuel San, Tandile Hermanus, Houriiyah Tegally, Jumari Snyman, Thandeka Moyo-Gwete, Eduan Wilkinson, Mallory Bernstein, Khadija Khan, Shi-Hsia Hwa, Sasha W. Tilles, Lavanya Singh, Jennifer Giandhari, Ntombifuthi Mthabela, Matilda Mazibuko, Yashica Ganga, Bernadett I. Gosnell, Salim S. Abdool Karim, Willem Hanekom, Wesley C. Van Voorhis, Thumbi Ndung'u, COMMIT-KZN Team, Richard J. Lessells, Penny L. Moore, Mahomed-Yunus S. Moosa, Tulio de Oliveira, and Alex Sigal. SARS-CoV-2 prolonged infection during advanced HIV disease evolves extensive immune escape. *Cell Host & Microbe*, 30(2):154–162.e5, 2022.

[4] Neil M. Ferguson, Alison P. Galvani, and Robin M. Bush. Ecological and immunological determinants of influenza evolution. *Nature*, 422(6930):428–433, 2003. Number: 6930 Publisher: Nature Publishing Group.

[5] Christopher D. Green. A path entropy function for rooted trees. *Journal of the ACM*, 20(3):378–384, 1973.

[6] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772):700–721, 1927.

[7] Patrick Mellacher. Endogenous viral mutations, evolutionary selection, and containment policy design. *Journal of Economic Interaction and Coordination*, 17(3):801–825, 2022.

[8] Juliet R. C. Pulliam, Cari van Schalkwyk, Nevashan Govender, Anne von Gottberg, Cheryl Cohen, Michelle J. Groome, Jonathan Dushoff, Koleka Mlisana, and Harry Moultrie. Increased risk of SARS-CoV-2 reinfection associated with emergence of omicron in south africa. *Science*, 376(6593):eabn4947, 2022.

[9] Yamin Sun, Wenchao Lin, Wei Dong, and Jianguo Xu. Origin and evolutionary analysis of the SARS-CoV-2 omicron variant. *Journal of Biosafety and Biosecurity*, 4(1):33–37, 2022.

[10] Houriiyah Tegally, Monika Moir, Josie Everatt, Marta Giovanetti, Cathrine Scheepers, Eduan Wilkinson, Kathleen Subramoney, Zinhle Makatini,

Sikhulile Moyo, Daniel G. Amoako, Cheryl Baxter, Christian L. Althaus, Ugochukwu J. Anyaneji, Dikeledi Kekana, Raquel Viana, Jennifer Giandhari, Richard J. Lessells, Tongai Maponga, Dorcas Maruapula, Wonderful Choga, Mogomotsi Matshaba, Mpaphi B. Mbulawa, Nokukhanya Msomi, Yeshnee Naidoo, Sureshnee Pillay, Tomasz Janusz Sanko, James E. San, Lesley Scott, Lavanya Singh, Nonkululeko A. Magini, Pamela Smith-Lawrence, Wendy Stevens, Graeme Dor, Derek Tshiabuila, Nicole Wolter, Wolfgang Preiser, Florette K. Treurnicht, Marietjie Venter, Georginah Chiloane, Caitlyn McIntyre, Aine O'Toole, Christopher Ruis, Thomas P. Peacock, Cornelius Roemer, Sergei L. Kosakovsky Pond, Carolyn Williamson, Oliver G. Pybus, Jinal N. Bhiman, Allison Glass, Darren P. Martin, Ben Jackson, Andrew Rambaut, Oluwakemi Laguda-Akingba, Simani Gaseitsiwe, Anne von Gottberg, and Tulio de Oliveira. Emergence of SARS-CoV-2 omicron lineages BA.4 and BA.5 in south africa. *Nature Medicine*, 28(9):1785–1790, 2022. Number: 9 Publisher: Nature Publishing Group.

[11] Changshuo Wei, Ke-Jia Shan, Weiguang Wang, Shuya Zhang, Qing Huan, and Wenfeng Qian. Evidence for a mouse origin of the SARS-CoV-2 omicron variant. *Journal of Genetics and Genomics*, 48(12):1111–1121, 2021.

[12] Osman Yagan, Anirudh Sridhar, Rashad Eletreby, Simon A. Levin, Joshua B. Plotkin, and H. Vincent Poor. Modeling and analysis of the spread of COVID-19 under a multiple-strain model with mutations. *Harvard Data Science Review*, 2021.