Impacts of receptor binding and immunity on SARS-CoV-2 fitness landscape: a modeling study

Zhaojun Ding [a], Hsiang-Yu Yuan [a,b,*]

[a] Department of Biomedical Sciences, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong

[b] Centre for Applied One Health Research and Policy Advice, City University of Hong Kong, Hong Kong

*Correspondence: sean.yuan@cityu.edu.hk

Abstract

Despite numerous studies investigating SARS-CoV-2 ACE2 binding affinity and its transmissibility, their relationship concerning varying immunity remains unclear. Daily ACE2 binding and immune escape in SARS-CoV-2 were calculated by summing the effects of all amino acid mutations in the receptor binding domain for each viral sequence per day, based on deep mutational scanning data. We developed an infectious disease transmission model that decomposed the effective reproduction number into viral infectiousness, host protection, and contact patterns. Viral infectiousness was used to represent a component of the fitness determined by both ACE2 binding and the immunity of infected hosts. By fitting the model to daily reported cases, immune escape, vaccine rollout, and population mobility, both viral infectiousness and effective immunity (i.e. the protection against the circulating variants) among infected individuals were quantified. A rugged fitness landscape, spanned by ACE2 binding and hosts' effective immunity, was observed with peaks corresponding to individual VOCs (alpha, delta, and omicron (BA.1* and BA.2*)). We found that higher effective immunity was associated with a lower level of virus fitness peak initially; and lower effective immunity with weaker receptor binding was associated with an optimal virus fitness. The finding helps understand SARS-CoV-2 evolution and predict future dominant variants.

Introduction

The COVID-19 pandemic was characterized by repeated infection waves, driven by SARS-CoV-2 variants with new and advantageous genomic mutations. These mutations enhanced fitness in variants of concern (VOC) via a complex interplay between viral intrinsic traits and population immunity, shaped by vaccinations and prior infections [1,2]. Therefore, understanding the SARS-CoV-2 fitness landscape can provide insights into evolutionary virus dynamics, knowing whether the virus undergoes gradual or punctuated evolution, which is important for predicting and preparing for future outbreaks.

The SARS-CoV-2 fitness is associated with virus's ability to bind to ACE2 receptors and evade immune defenses[1]. Upon entering the human body, the virus uses its spike protein to bind to the ACE2 receptor on the surface of host cells. This interaction occurs primarily through the receptor-binding domain (RBD). Simultaneously, the host immune system produces neutralizing antibodies, which prevent the virus from binding to ACE2. However, SARS-CoV-2 can evade immune recognition by mutating amino acid residues within the RBD, a phenomenon known as immune escape [1,3]. How ACE2 binding and immune escape interact to influence virus fitness under varying immunity levels among infected hosts remains unclear.

Certain accumulated mutations in the RBD of SARS-CoV-2 spike protein confer selective advantages toward a higher fitness, resulting in antigenic drift [2]. For example, the E484K mutation in the delta (B.1.617.2) sub-lineage appeared to affect antibody recognition, and hence declined vaccine effectiveness [4]. The N501Y mutation in the alpha (B.1.1.7) sub-lineage may have altered spike protein binding affinity to ACE2, resulting in higher transmission rates [5,6]. After the end of the pandemic, whether COVID-19 tends to be endemic (more stable incidence) or epidemic (with repeated outbreaks) is unknown.

The fitness landscape has been developed to represent relationships between genetic variations and their corresponding fitness values, providing a better understanding of the pathways that populations might take during adaptation [2,7]. A rugged landscape occurs due to interactions between mutations (epistasis) [7,8]. Immunity changes can also exert high selection pressure on viruses, resulting in antigenic drift [2,9]. When a virus evolves through fitness valleys toward a new

peak, it generates boom-and-bust patterns, likely resulting in repeated outbreaks [10]. However, when SARS-CoV-2 evolutionary and transmission dynamics are intertwined, reconstructing the fitness landscape is challenged. For example, not only virus traits, but also host immunity (e.g. vaccination) and host behaviors (e.g. social distancing measures and public awareness) can influence transmissibility (or effective reproduction number) [11]. This results in a time-varying fitness even when viral traits are fixed. In order assess the impact of ACE2 binding on virus fitness under different host immunity levels, separating the effects from the factors not associated with ACE2 receptor will be required.

In this study, we aimed to assess the impact of RBD mutations on virus fitness that is associated with ACE2 receptor binding using data from Italy, a country with high viral incidence rates and high vaccine coverage. To disentangle intervention effects (i.e. vaccinations and non-pharmaceutical interventions (NPIs)) on virus fitness, we incorporated mobility and vaccination data in an infectious disease transmission model to calibrate fitness estimates. From simulated results, and for the first time, we constructed a fitness landscape after mapping the virus's evolutionary trajectory traits (i.e. receptor binding and immune escape) over different periods across different immunity levels.

Material and methods

Sequence data

Italian SARS-CoV-2 spike protein sequence data (31st January 2020–17th October 2022) were downloaded from the GISAID database [12]. To ensure data quality control, two exclusion criteria were applied: (1) sequences with > 5% ambiguous amino acids [13] and (2) sequences < 1,263 bp or > 1,273 bp were filtered out. Ultimately, 112,011 sequences were included for virus ACE2 binding and immune escape measurements. The proportion of beta and gamma sequences are very low, only 2.07% and 0.12%, respectively (see Fig.S1). Therefore, only alpha, delta, and omicron variants were reported as the main VOCs in subsequent analyses.

Measuring virus ACE2 binding and immune escape

Deep mutational scanning data were used to calculate virus ACE2 binding by mapping binding affinities between ACE2 and SARS-CoV-2 RBD amino acid mutation sites[14]. Similarly, escape

fraction data were used to determine viral immune escape by mapping reduced binding levels between the virus and antibodies[3]. We first used the Clustal Omega tool to align all sequences [15]. Then, to calculate virus ACE2 binding for each sequence, binding changes across individual amino acid mutations of RBD region (residues from N331 to T531) were summed [16]. For immune escape, we first calculated mean reductions in antibody binding for each amino acid mutation site across 10 human antibodies. Subsequent steps were identical to ACE2 binding calculations. Finally, we computed daily average values for these viral traits to indicate the daily levels of ACE2 binding and immune escape. Further details are shown in Supplementary Methods 2.1. In addition, logistic regression model was used to explore the relationship between ACE2 binding and immune escape.

Mobility data and calculating the mobility index

The mobility index, which indicated the degree of population contact, was estimated from Google Mobility and Apple Mobility data [17,18]. Specifically, three Apple mobility data subtypes (driving, walking, and transit movement) and four Google subtypes (grocery and pharmacy stores, transit, retail and recreation, and workplaces) were used to calculate mobility indices based on previous research [19].

Vaccination data

COVID-19 vaccine administration data provided by the Italian Ministry of Health were used to summarize daily vaccination numbers for first, second, and third doses [20]. During the study period (i.e. model fitting period, 31st January 2020–5th March 2022), 134,611,131 vaccine doses were recorded and sourced from five major vaccine manufacturers: Janssen, Moderna, Novavax, Pfizer, and AstraZeneca (Supplementary Table S1). Messenger RNA (mRNA) vaccines (i.e. Pfizer and Moderna) and viral vector vaccines (i.e. Janssen and AstraZeneca) accounted for almost 99% of the Italian vaccination campaign (Supplementary Fig. S3). Consequently, we only considered mRNA (Pfizer and Moderna) and viral vector vaccines (Janssen and AstraZeneca).

Calculating virus fitness

We extended a stochastic Susceptible - Exposed - Infectious - Recovered (SEIR) model, coupled with vaccination, mobility, and immune escape data, to reproduce COVID-19 transmission dynamics and estimate virus fitness across different VOC periods (see Fig. 1A, B). The model was

used to estimate the part of the transmission rate that were determined by viral ACE2 receptor binding only over time (denoted as $\beta_v(t)$) under different immunity levels, representing the change of infectiousness. This infectiousness was referred to as ACE2 binding-dependent viral fitness. The relative virus fitness was defined as $\beta_v(t)/\beta_0$, where β_0 is the average $\beta_v(t)$ within the first month since the pandemic began, representing the wild-type.

The effective reproduction number (R_e) was represented as two components, intrinsic factor: infectiousness, and extrinsic factor (Fig. 2D): susceptibility and contact rate as below.

$$\begin{aligned} R_e &= Infectiousness \times (Susceptibility \times Contact) \\ &= \beta_v(t) \cdot T_g \times ((1 - P_{imm_i}(t) \cdot e^{-c \cdot immune_escape_t}) \cdot \frac{S_i(t)}{N} \times M(t)) \quad (1) \end{aligned}$$

where T_g is the generation time, which is 4 days [21–23]; $P_{imm_i}(t)$ refers to vaccine protection from different doses against the wild-type (see Supplementary Methods 2.2.3 for calculating $P_{imm_i}(t)$) and subscript i indicates vaccine doses (i.e. 0, 1, 2 or 3); $P_{imm_i}(t) \cdot e^{-c \cdot immune_escape(t)}$ represents protection against circulating strains, which considers immune escape as defined by effective immunity. $Immune_escape_t$ is the daily average immune escape of viruses; c is the scaling factor for daily immune escape; $\frac{S_i(t)}{N}$ represents the fraction of individuals at each susceptibility level (i.e. S_0, S_1, S_2 or S_3); $M(t)$ is the daily mobility index, representing the effect of the change in contact rate on reducing overall fitness [23]. $\beta_v(t) \cdot (1 - P_{imm_i}(t) \cdot e^{-c \cdot immune_\mathrm{escape_t}}) \times M(t)$ represents the transmission rate among susceptible groups with different protection levels (i.e. $\beta_{imm_i}(t)$ in Fig. 1B). The effective reproduction number R_e was obtained after fitting the transmission model (Fig. 1B) to daily reported cases, when daily vaccine doses were considered.

Exposed (E_i) individuals who came into contact with infected individuals will become infectious (I_i) after $\frac{1}{\sigma}$ days. As not all cases were reported correctly and immediately, reported cases (IR) were considered based on the detection rate (d) and the confirmation delay (ρ_t) in the transition between E_i and I_i, which was used to fit daily reported cases. Individuals with an infectious status would subsequently enter into a removed status (i.e. R_0, R_1, R_2 and R_3) due to self-isolation or recovery after $1/\gamma(\mathrm{t})$ days. Considering waning immunity among these individuals, we assumed that

recovered and vaccinated individuals would lose their viral immunity after ω days ($\omega = log(0.85)/182.5$, corresponding to exponential waning with a 15% loss in protection after 6 months, more detail about the parameters set can be found in the "2.2 Model details" of Supplementary Material) and then returning to original susceptibility levels (i.e. S0) [24].

Model fitting for reproduce of infection dynamics and fitness landscape constructed

The number of individuals in compartment IR, as the modeling output, was used to fit daily numbers of reported cases. The Particle Markov Chain Monte Carlo method [25,26], which combines particle filtering and Markov Chain Monte Carlo approaches, was used to calibrate model output using daily reported cases and then estimate model parameters. All model parameter estimations were obtained using the 'mcstate' package in R [25]. A fitness landscape was fitted with virus fitness and assumed a multivariate normal distribution, which helped visualize SARS-CoV-2 adaptive evolutionary trajectories.

Results

The relationship between daily mobility, vaccine coverage, and daily incidence

A large variety in daily incidence and mobility was observed (Fig. 2A, C and Fig. S2). Despite ongoing vaccination efforts (Fig. 2B), viral outbreaks reoccurred and cumulative cases increased rapidly during the omicron period (Fig. 2A). The effective reproduction number was likely influenced by both virus infectiousness and susceptibility (Fig. 2D). In our study, we considered susceptibility (determined by population immunity and immune escape) and contact rate (determined by mobility) as extrinsic factors while virus infectiousness as an intrinsic factor (i.e. ACE2 binding-dependent viral fitness; see Methods for the explanation).

Temporal dynamics of virus ACE2 binding and immune escape

We focused on the changes in ACE2 binding-dependent viral fitness and traits across main VOCs, such as alpha, delta, and omicron (Fig. 3A). There was a slight decrease in ACE2 binding during the alpha to delta variant transition, followed by a larger reduction during the delta to omicron transition. During the omicron wave, virus ACE2 binding continued decreasing. These reductions occurred along with increases in immune escape (Fig. 3B, C and Fig. S4). Rare mutations which appeared during late delta or early omicron appeared to have binding values within the gap

196 between the transition. Some of these rate mutations eventually became dominant. For example,
197 G446V and K417N variants (Fig. 3B, purple text) coexisted at the end of the delta period. The
198 K417N variant, with lower ACE2 binding and higher immune escape, was also appeared in
199 omicron. G446V was replaced by G446S, which had lower immune escape. N440K, G446S and
200 G496S coexisted at the early of the omicron period. The N440K, with higher binding and higher
201 immune escape, was found frequently in omicron. G446S and G496S were replaced by other
202 mutation set eventually (see Fig. S6).
203
204 The overall trend with an inverse relationship (i.e. Pearson correlation coefficient (R-value) = -
205 0.98; P-value < 0.001; see Material and Methods: Measuring virus ACE2 binding and immune
206 escape) between ACE2 binding and immune escape throughout the whole period suggested that
207 the cost for enhancing the immune escape can lead to a decrease in the ACE2 binding (Fig. 3D).
208 This finding aligns with previous studies, showing that Omicron prefers a more closed
209 conformation of its spike protein, which aids in immune evasion and reduces its binding affinity
210 to ACE2 [27,28]. Mutations (N501Y and E484K) generated notable increases in ACE2 binding, with
211 N501Y showing a peak increase of approximately 0.25 in relative binding affinity, aligned with its
212 significant prevalence across certain variants (Fig. 3E, F). This highlighted the substantial role that
213 this mutation played in enhancing viral infectivity and transmissibility, consistent with previous
214 studies [29–31]. T478K and L452R prevalence was significantly higher in the delta variant (Fig. 3F),
215 highlighting their contribution to heightened transmissibility and potential resistance to antibody
216 neutralization when compared with earlier strains. Compared with previous VOCs, more mutations
217 occurred in omicron, resulting in lower ACE2 binding (e.g. S371F, S375F, T376A, K417N, E484A,
218 Q493R, and Y505H) and higher immune escape (e.g. K417N, E484A, and Q493R) (Fig. 3E).
219
220 **SARS-CoV-2 transmission dynamics**
221 To understand how ACE2 receptor binding influenced virus fitness, we incorporated vaccine
222 coverage, mobility and virus immune escape data into the model to calibrate model fitting for daily
223 reported cases (Fig. 1B). Our model successfully reproduced the observed infection dynamics,
224 including Non-VOC and other VOC waves (Fig. 4A). The relative ACE2 binding-dependent virus
225 fitness did not show very similar pattern during late delta and omicron periods (Fig. 4B). The
226 immunity against the wild-type strain (such as vaccine-induced (Fig. 4C, green line) and natural

227 infection combined with vaccine-induced immunity (Fig. 4C, red line)) increased due to the
228 initiation of the vaccination campaign (beginning in the late December 2020) in Italy around the
229 same time [32]. A similar trend was observed for effective immunity against VOCs during alpha and
230 delta periods, but a significant decrease was recorded in the omicron period (Fig. 4D). This decline
231 highlighted the variant's robust immune escape capabilities of the omicron, despite the ongoing
232 vaccination campaign.
233
234 **SARS-CoV-2's fitness landscape**
235 The phenotype-fitness landscape spanned by ACE2 binding and host' effective immunity showed
236 rugged patterns with four major peaks, corresponding to B.1.1.7 (alpha), B.1.617.2 (delta), BA.1*
237 (first omicron peak) and BA.2* (second omicron peak). One minor peak, corresponding to the
238 transition period between delta and omicron, and four valleys between these peaks. Before the
239 emergence of omicron, ACE2 binding remained strong. A higher effective immunity was
240 associated with a lower fitness peak. Within the omicron period, ACE2 binding and effective
241 immunity decreased simultaneously, associated with an optimal fitness peak (Fig. 5A and
242 Supplementary Fig. S5). To help visualize the rugged patterns, a smoothed fitness landscape was
243 constructed after fitting the actual data (Fig. 5B; see Methods).
244
245 Potential key RBD mutations that may characterize its evolution were explored (Fig. 5). Besides
246 the early period during alpha variant, the sequential change in mutations, characterized by partial
247 replacement, appeared through the fitness valleys toward the peaks:
248 **Alpha -> delta:** Initially, the E484K mutation (ACE2 binding changes (denoted as Δ1): 0.06,
249 immune escape changes (denoted as Δ2): 0.27, see Fig. 3E) emerged as an attempt at adaptation
250 during the transition between the fitness peaks at alpha and delta. However, its prevalence
251 decreased after passing through the first valley. In contrast, the prevalence of L452R (Δ1: 0.02, Δ2:
252 0.18) and T478K (Δ1: 0.02, Δ2: 0.004) began to increase during this fitness valley and replaced a
253 high binding N501Y (Δ1: 0.24, Δ2: 0.01) (see Fig. S6), the dominated mutation previously.
254 **Delta-> Omicron (BA.1*):** After the delta variant, the virus's evolutionary pathway moved
255 through two fitness valleys (i.e. Valley 2 and Valley 3) with a local peak (i.e. a minor peak during
256 the transition period) and then reach to BA.1*, the next peak. Meanwhile, only T478K was present
257 in the subsequent omicron while L452R was gradually decreased. Next, the Set 1 (G339D, S373P,

S375F, K417N, S477N, E484A, Q498R and Y505H; $\Delta 1_{total}$ = -1.8, $\Delta 2_{total}$ = 0.51), Set 2 mutations (S371L, G446S and G496S; $\Delta 1_{total}$ = -0.97, $\Delta 2_{total}$ = 0.037) and few other mutations (N440K, Q493R and N501Y, see Fig. S6) gradually increased. These mutations appeared during the transition from delta to omicron through the fitness valleys and cause immune escape with decreases in binding. The effective immunity at the first omicron peak period is still larger than delta, together with a low binding value, showing a suboptimal fitness.

Omicron (BA.1*) -> Omicron (BA.2*): The virus continued to evolve after the emergence of omicron (BA.1*), moving through Valley 4 towards the next fitness peak associated with omicron (BA.2*). During this transition, the Set 1 and other additional mutations (N440K, Q493R and N501Y) was retained while Set 2 (transient mutations) were gradually decreased. In the end, the Set 3 (S371F, T376A, D405N and R408S; $\Delta 1_{total}$ = -0.61, $\Delta 2_{total}$ = 0.051) was gradually increased and dominated. As the virus approached BA.2* and overcame the fitness valley, a shift in the fitness landscape was observed, coinciding with a lower level of effective immunity and a slight bounce back of binding with an optimized fitness (Fig. 5B).

Discussion

We introduced a mathematical framework to elucidate the transmission dynamics of different SARS-CoV-2 variants in Italy. For the first time, the fitness landscape was constructed for the virus from alpha to omicron. This was achieved by incorporating both intrinsic viral factors (e.g., infectiousness) and extrinsic factors (e.g., host susceptibility and contact rates) into an SEIR-based model, which helps reproduce virus transmission dynamics and allows us to quantify virus fitness. While identifying major substitutions resulting in antigenic change is important[33], our research suggests that some mutations may act as transient intermediates playing an important role in the evolutionary pathway via fitness valleys, thereby facilitating the transition to new fitness peaks. Epistasis plays a crucial role as that the interaction between mutations can lead to non-linear effects on fitness [7,34,35]. The results provide valuable insights into the real-world evolutionary constraints of SARS-CoV-2.

Inverse relationship between ACE2 binding and immune escape was found among the virus traits. Especially, when omicron arrived, a highest fitness in both low binding and low population immunity against the circulating strains was present. Even though the N501Y mutation, which can

increase virus ACE2 binding, reemerged, the RBD binding affinity in omicron was extremely lower than other variants. The combined effects of these factors illustrate potential evolutionary paths alternating between low and high fitness peaks. For example, if a virus population is on a "low fitness peak" (characterized by strong binding to ACE2 receptors together with a high population immunity), certain mutations may help the virus transit to a "high fitness peak" (characterized by weak binding to ACE2 receptors, but the virus exhibits strong immune escape, leading to a lower population immunity). Once this peak is reached, further moderate mutations, such as those increasing binding, may occur to help the virus adapt as population immunity gradually increases, driving toward a low fitness peak again. Hence, the constantly changing population immunity may play an important role in evolutionary shifts between different fitness peaks.

Constraints on mutational effects are pronounced in the presence of rugged landscape. Punctuated change in immune escape or receptor binding but may potentially allow the virus move beyond the fitness peak. For example, we observed the first omicron peak is at a suboptimal level (see Fig. 5). The binding level might be too low for the effective population immunity. The shift from Set 2 to Set 3 mutations, coupled with the retention of previous mutations, such as Set 1 and N501Y, slightly increased the binding and achieved the highest fitness (Fig. S5). If low binding results in high virus fitness, as shown in a previous influenza virus serial passaging study[36], one important question is how easy strong immune escape from polyclonal antibodies can occur. For omicron, multiple mutations that exhibit high immune escape.

The fitness valley maybe also related to deleterious effects among immune escape mutations [37,38]. Those mutations occurred in the fitness valley may be detrimental to SARS-CoV-2's evolution, requiring additional time to reach a fitness peak when compensatory mutations occur.

Implications

Firstly, continuous surveillance efforts of SARS-CoV-2 is important. Even within the same VOC, such as omicron, virus still experienced rugged fitness landscape, leading to repeated epidemics of varying sizes. This is exemplified by the transition from omicron BA.1* to BA.2*, where the

virus's evolutionary path traversed a fitness valley before reaching a new peak. Knowing the change in traits will be important to understand why a new outbreak occur.

Secondly, forecasting when a large outbreak will occur is still very challenging. Our research showed that the virus ACE2 binding and immune escape along with population immunity, jointly shaped its evolutionary path. An optimal fitness was likely to be reached until a set of mutations having strong immune escape with required reduction in binding.

Moreover, the relationship between virus fitness and immunity, influenced by ACE2 binding, is not simply linear. Using more complex model, such as the deep neural network, maybe more powerful to capture the relationship between the complex interplay of multiple viral traits and fitness. This complex and dynamic evolutionary landscape also underscores the necessity for regular updates to SARS-CoV-2 vaccines to ensure they remain effective against evolving strains.

Study limitations

There are certain limitations in this study. Firstly, the study only focused on RBD region in the spike protein and we assumed the impact of each amino acid on viral traits is additive. This approach, while pragmatic, oversimplified potentially complex interactions between sites. Nonetheless, several reports have suggested that synergistic binding effects may occur at multiple sites [16]. Secondly, the population mobility index was an NPI proxy as it assumed social contacts were proportional to population mobility, which not considered the awareness-driven behavior changes on the dynamic of mobility. Further studies are required to explore the temporal variation of the number of contacts, public awareness and behavior changes under the implementation of NPIs during the pandemic and its relationship with the population mobility index. Thirdly, our model categorized the population based on administered numbers of vaccine doses, without considering age and gender heterogeneity in the population. This simplification overlooked potential differences in vaccine responses and exposure risks across different demographic groups. Lastly, but critically, the seasonal or climatic factors were not considered. Future research should introduce the impact of these factors on virus fitness to provide a more comprehensive understanding of virus transmission dynamics.

Conclusions

SARS-CoV-2 mutations along with immunity influences generated a rugged fitness landscape, which shaped viral evolutionary trajectories in Italy. Knowing the landscape can help provide a better understanding of the complex dynamics of viral evolution and adaptation, particularly in drug resistance and vaccine development. More importantly, host immune systems and viruses are constantly engaged in an arms race, with each adapting to changes in the other. Updating vaccines is crucial to winning future battles against COVID-19.

Ethics approval and consent to participate

Ethics approval was not needed as we used publicly available reported cases, virus amino acid sequences, and vaccination data.

Declaration of interests

All authors declare no competing interests.

Data and code availability

All data files and code scripts used in this study were uploaded to the GitHub website (https://github.com/DING219/bind_escape_code)

Acknowledgments

We are grateful to the support from grants funded by City University of Hong Kong (#9610416 and #7005748) and General Research Fund (#9043514). We gratefully acknowledge all researchers for contributing to, generating, and openly sharing genome data via the Global Initiative on Sharing All Influenza Data, as most of our analyses were based on this invaluable resource. We thank Dr. Jingbo Liang for helpful discussions, Ms. Fatema Khairunnasa for useful editing suggestions. We greatly appreciate Prof. Joshua S. Weitz at University of Maryland for his meaningful suggestions on model fitting.

References

1. Carabelli, A. M. *et al.* SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nat Rev Microbiol* **21**, 162–177 (2023).

2. Yewdell, J. W. Antigenic drift: Understanding COVID-19. *Immunity* **54**, 2681–2687 (2021).

3. Greaney, A. J. *et al.* Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. *Cell Host Microbe* **29**, 44-57.e9 (2021).

4. Baj, A. *et al.* Breakthrough infections of E484K-Harboring SARS-CoV-2 delta variant, Lombardy, Italy. *Emerg Infect Dis* **27**, (2021).

5. Fiorentini, S. *et al.* First detection of SARS-CoV-2 spike protein N501 mutation in Italy in August. *Lancet Infect Dis* **21**, 16 (2021).

6. Tian, F. *et al.* N501Y mutation of spike protein in SARS-CoV-2 strengthens its binding to receptor ACE2. **10**, 69091 (2021).

7. Wright, S. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Sixth International Congress on Genetics* **1**, (1932).

8. Kauffman, S. & Levin, S. Towards a general theory of adaptive walks on rugged landscapes. *J Theor Biol* **128**, (1987).

9. Grenfell, B. T. *et al.* Unifying the Epidemiological and Evolutionary Dynamics of Pathogens. *Science* **303**, 327-332 (2004).

10. Koelle, K., Cobey, S., Grenfell, B. & Pascual, M. Epochal evolution shapes the phylodynamics of interpandemic influenza a (H3N2) in humans. *Science (1979)* **314**, (2006).

11. Weitz, J. S., Park, S. W., Eksin, C. & Dushoff, J. Awareness-driven behavior changes can shift the shape of epidemics away from peaks and toward plateaus, shoulders, and oscillations. *Proc Natl Acad Sci U S A* **117**, (2020).

12. Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges* **1**, 33–46 (2017).

13. Zhao, S., Ran, J. & Han, L. Exploring the interaction between E484K and N501Y substitutions of SARS-CoV-2 in shaping the transmission advantage of COVID-19 in Brazil: A modeling study. *American Journal of Tropical Medicine and Hygiene* **105**, 1247–1254 (2021).

14. Starr, T. N. *et al.* Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **182**, 1295-1310.e20 (2020).

15. Bodenhofer, U., Bonatesta, E., Horejš-Kainrath, C. & Hochreiter, S. Msa: An R package for multiple sequence alignment. *Bioinformatics* **31**, 3997–3999 (2015).

16. Bozdaganyan, M. E., Shaitan, K. V., Kirpichnikov, M. P., Sokolova, O. S. & Orekhov, P. S. Computational Analysis of Mutations in the Receptor-Binding Domain of SARS-CoV-2 Spike and Their Effects on Antibody Binding. *Viruses* **14**, (2022).

414 17. Google COVID-19 Community Mobility Reports. *https://www.google.com/covid19/mobility/*
415 (accessed Sept 14, 2022).

416 18. Apple Map Mobility Trends Reports. *Apple* (2021). *https://covid19.apple.com/mobility* (accessed
417 Sept 11, 2022).

418 19. Nouvellet, P. *et al.* Reduction in mobility and COVID-19 transmission. *Nat Commun* **12**, (2021).

419 20. Open data on COVID-19 vaccination in Italy. *https://github.com/italia/covid19-opendata-vaccini*
420 (accessed Oct 11, 2022).

421 21. Hart, W. S. *et al.* Generation time of the alpha and delta SARS-CoV-2 variants: an epidemiological
422 analysis. *Lancet Infect Dis* **22**, 603–610 (2022).

423 22. Xu, X. *et al.* Assessing changes in incubation period, serial interval, and generation time of SARS-
424 CoV-2 variants of concern: a systematic review and meta-analysis. *BMC Med* **21**, (2023).

425 23. Arantes, I. *et al.* Comparative epidemic expansion of SARS-CoV-2 variants Delta and Omicron in
426 the Brazilian State of Amazonas. *Nat Commun* **14**, (2023).

427 24. Barnard, R. C. *et al.* Modelling the medium-term dynamics of SARS-CoV-2 transmission in England
428 in the Omicron era. *Nat Commun* **13**, (2022).

429 25. Knock, E. S. *et al.* Reproducible parallel inference and simulation of stochastic state space models
430 using odin, dust, and mcstate. *Wellcome Open Res* **5**, 288 (2020).

431 26. Kucharski, A. J. *et al.* Early dynamics of transmission and control of COVID-19: a mathematical
432 modelling study. *Lancet Infect Dis* **20**, 553–558 (2020).

433 27. Calvaresi, V. *et al.* Structural dynamics in the evolution of SARS-CoV-2 spike glycoprotein. *Nat*
434 *Commun* **14**, (2023).

435 28. Wu, L. *et al.* SARS-CoV-2 Omicron RBD shows weaker binding affinity than the currently dominant
436 Delta variant to human ACE2. *Sig Transduct Target Ther* **7**, (2022).

437 29. Harvey, W. T. *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol*
438 **19***, 409–424 (2021)*.

439 30. Liu, Y. *et al.* The N501Y spike substitution enhances SARS-CoV-2 infection and transmission.
440 *Nature* **602**, 294–299 (2022).

441 31. Schröder, S. *et al.* Characterization of intrinsic and effective fitness changes caused by
442 temporarily fixed mutations in the SARS-CoV-2 spike E484 epitope and identification of an
443 epistatic precondition for the evolution of E484A in variant Omicron. *Virol J* **20**, (2023).

444 32. Ferrante, P. The First 2 Years of COVID-19 in Italy: Incidence, Lethality, and Health Policies. *Front*
445 *Public Health* **10**, 986743 (2022).

446 33. Miller, N. L., Clark, T., Raman, R. & Sasisekharan, R. Mapping Potential Antigenic Drift Sites (PADS)
447 on SARS-CoV-2 Spike in Continuous Epitope-Paratope Space. *bioRxiv* (2021)
448 doi:10.1101/2021.06.07.446560.

449 34. Diaz-Colunga, J. *et al.* Global epistasis on fitness landscapes. *Phil. Trans. R. Soc. B* **378***,* 20220053
450 (2023).

451 35. Kryazhimskiy, S., Rice, D. P., Jerison, E. R. & Desai, M. M. Global epistasis makes adaptation
452 predictable despite sequence-level stochasticity. *Science (1979)* **344**, (2014).

453 36. Hensley, S. E. *et al.* Hemagglutinin receptor binding avidity drives influenza a virus antigenic drift.
454 *Science (1979)* **326**, (2009).

455 37. Koelle, K. & Rasmussen, D. A. The effects of a deleterious mutation load on patterns of influenza
456 A/H3N2's antigenic evolution in humans. *Elife* **4**, e07361 (2015).

457 38. Bloom, J. D. & Neher, R. A. Fitness effects of mutations to SARS-CoV-2 proteins. *Virus Evol* **9**,
458 vead055 (2023).

459
460

Figures

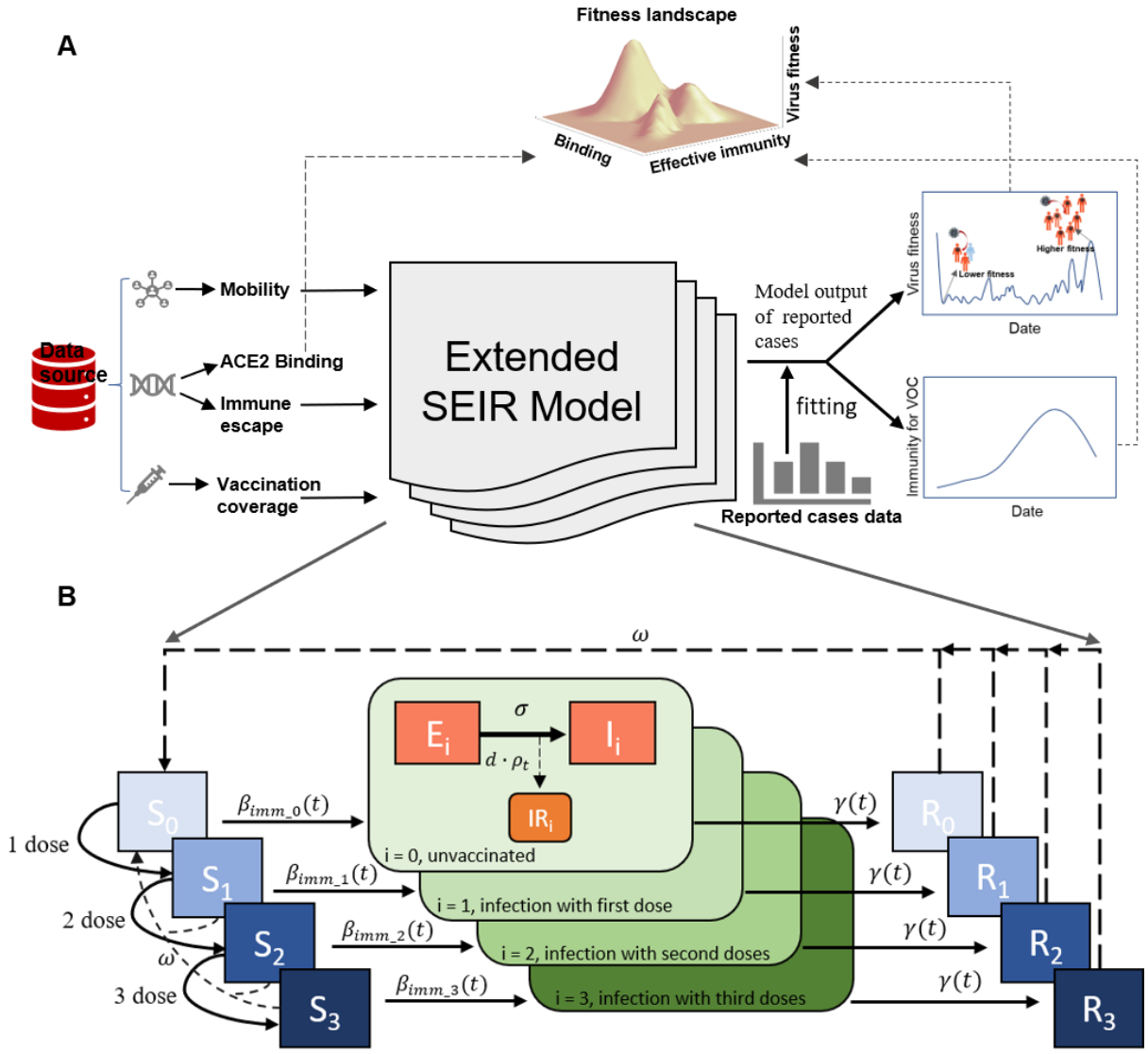

Fig. 1 Study flow and model design. (A) Schematic showing the extended SEIR model integrating multiple data sources to estimate the SARS-CoV-2 fitness landscape. Data inputs include mobility, immune escape, and vaccination coverage data. These data feed into the extended SEIR model, which was calibrated using reported case data. The model outputs include virus fitness combined with ACE2 binding data and immunity levels, and show how these factors shape virus evolution patterns. (B) Detailed structure of the SEIR compartment model. The model segments the population based on vaccination status

and infection stage, highlighting transitions between susceptible (S), exposed (E), infectious (I), reported infectious (IR) and recovered (R) states, with an emphasis on the impact of vaccination doses on these dynamics. Si, Ei, Ii, IRi and Ri subscripts denote compartments with different vaccine doses.

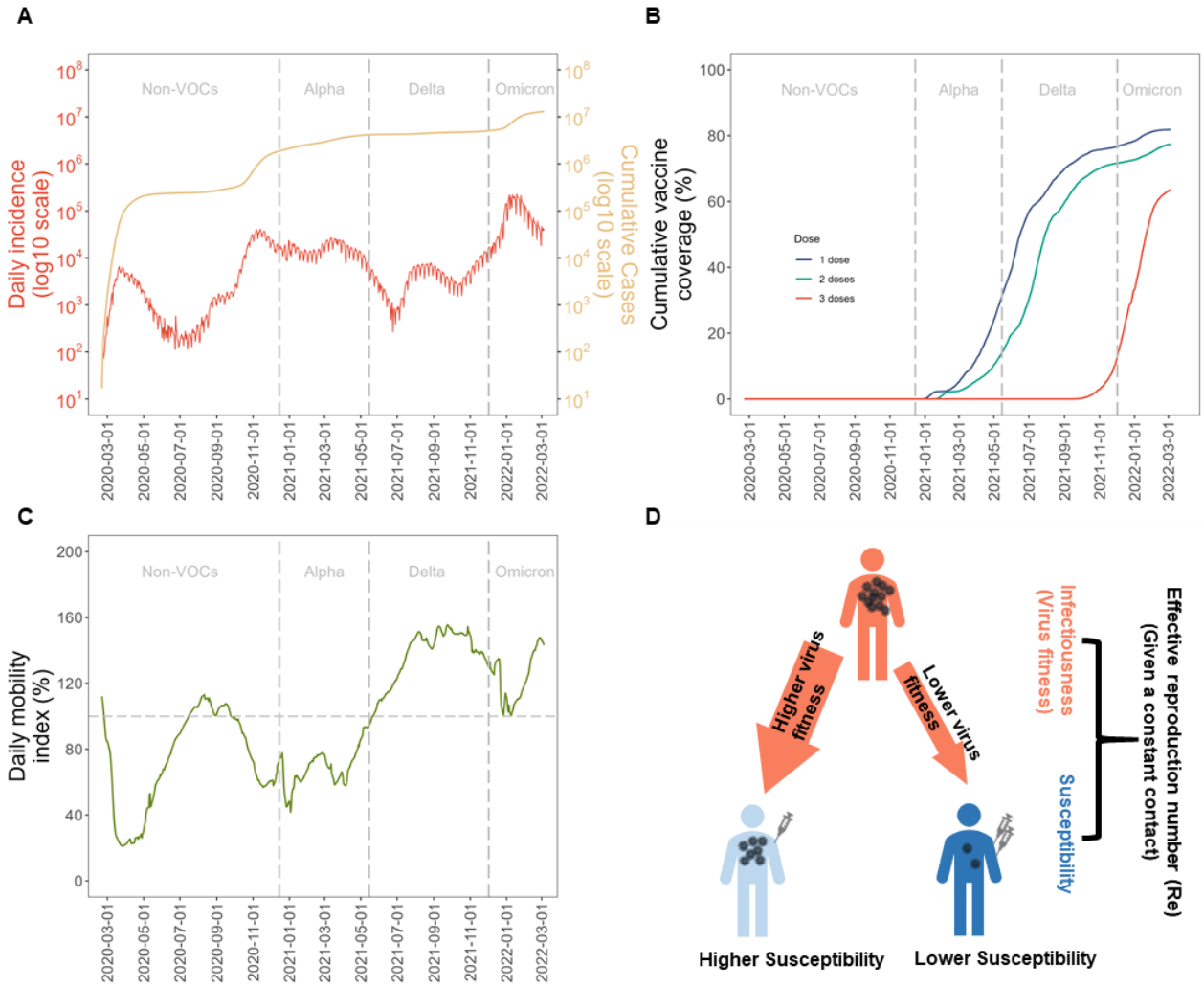

Fig. 2. Factors influencing daily COVID-19 incidences and virus fitness. (A) Daily COVID-19 incidences in Italy. New daily COVID-19 cases in Italy are measured in log10 scales. VOC emergence is highlighted by vertical dashed lines, labeled as non-VOC, alpha, delta, and omicron strains. (B) Vaccination coverage in Italy. The Italian population percentage that received 1 (blue line), 2 (green line), or 3 (red line) COVID-19 vaccine doses over time. (C) Mobility index in Italy. Population movement levels relative to pre-pandemic levels (horizontal dashed line) are shown. (D) Diagram showing virus infectiousness and Its determinants. Factors including host susceptibility and virus infectiousness contributing to effective reproduction number (Re) are shown.

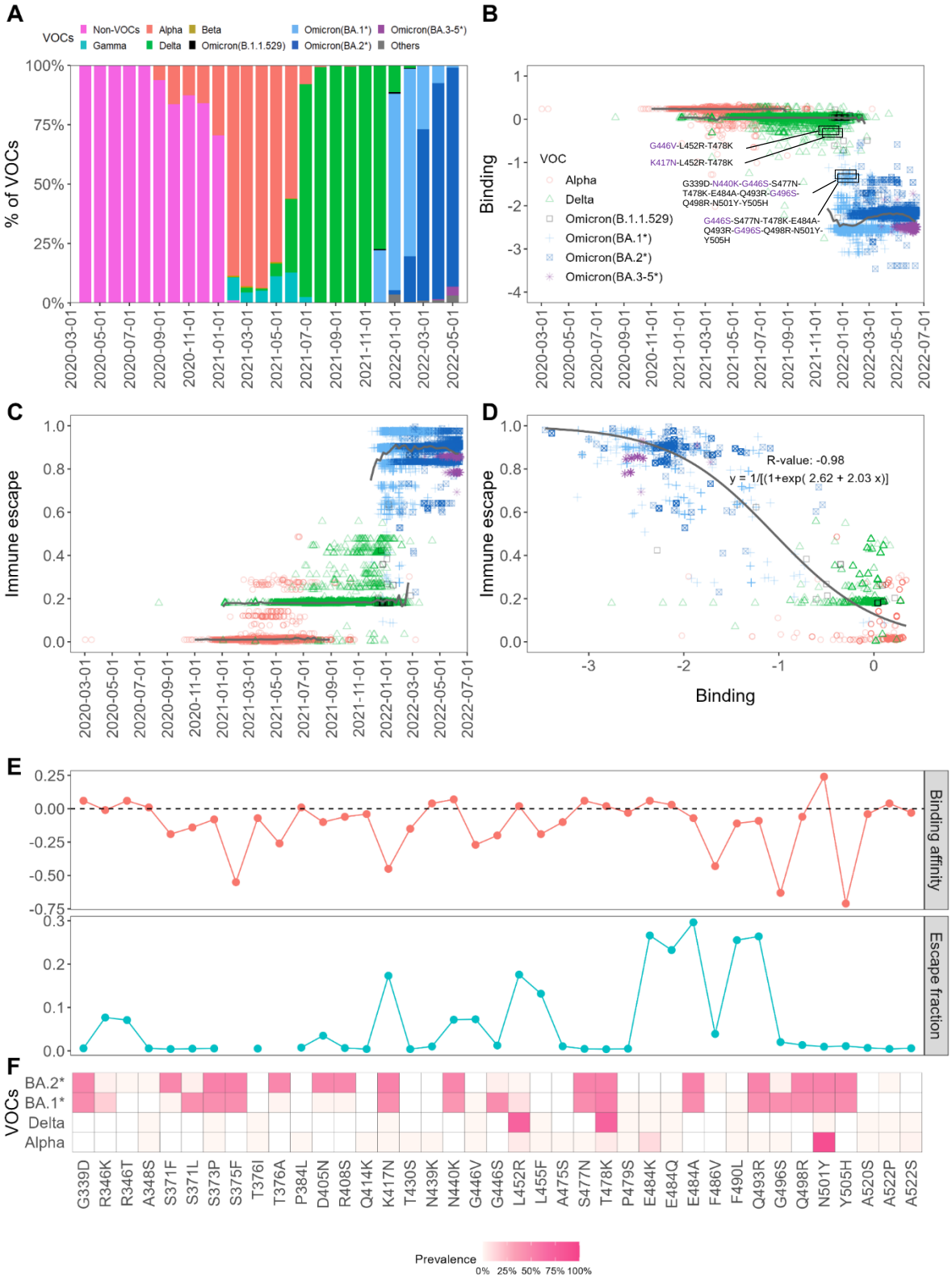

Fig. 3. Comprehensive analysis of SARS-CoV-2 VOC with ACE2 binding affinity and immune escape in Italy. (A) Variant prevalence over time. The bar chart shows the

prevalence of different SARS-CoV-2 VOC (alpha, beta, delta, gamma, and omicron strains) from early 2020 through early 2022. (B) and (C) VOC ACE2 binding and immune escape over time. Scatter plot showing ACE2 binding and alpha, delta, and omicron variant immune escape across the timeline. The gray line indicates weekly average values for each VOC with respect to ACE2 binding and immune escape. Rare mutations (purple) are indicated with prevalence < 0.1% between the delta to omicron transition period. (D) Correlation between ACE2 binding affinity and immune escape. Scatter plot showing a fitted curve (inverse logistic function) of the inverse relationship between virus ACE2 binding and immune escape across VOC. The Pearson correlation coefficient (R-value) indicates a strong inverse correlation. (E) Mutation impact on ACE2 binding affinity and immune escape. The top graph shows the effects of specific mutations on spike protein binding affinity to the ACE2 receptor (reference with wild type, the dash line), with values showing either increases or decreases in affinity. The bottom graph depicts the same mutations' impact on immune fraction, indicating the potential to escape antibody binding. (F) Mutation prevalence among each VOC. Heatmap showing specific mutation prevalence across alpha, delta, and omicron dominant periods.

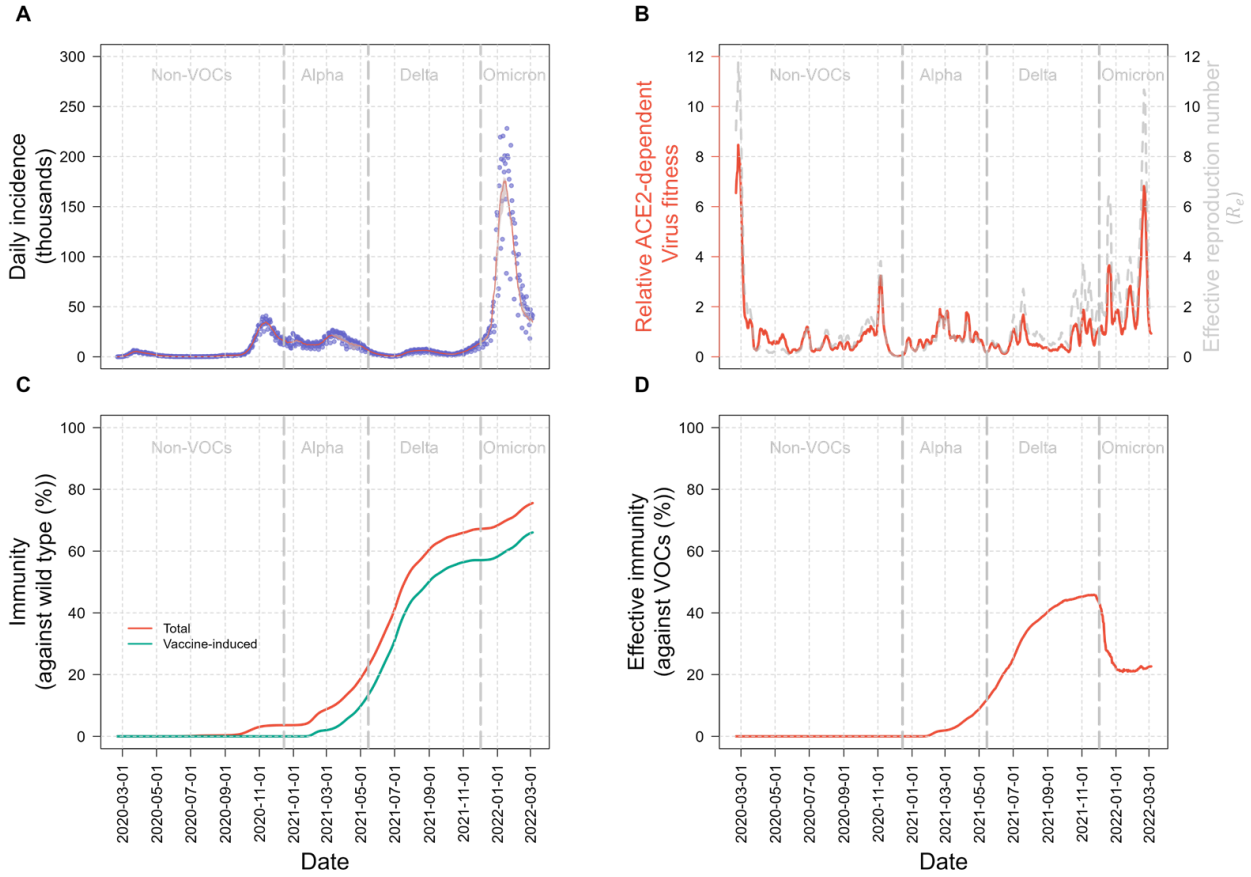

Fig. 4. Epidemic indicator and immunological response trajectories during SARS-CoV-2 variant transition in Italy. (A) Model fitting for daily confirmed cases. The red line indicates extended SEIR model estimations of daily incidences. Purple dots represent actual daily reported COVID-19 cases. (B) Virus fitness and effective reproduction number. The red line shows estimated virus fitness over time, which is relative to the wild type virus. The dashed grey line shows effective reproduction number (Overall fitness). (C) Effective immunity against wild type, defined as population immunity. The light green curve traces the population percentage that developed vaccine-induced immunity, and the red curve indicates developed immunity including vaccine-induced and recovered. (D) Effective immunity against VOCs, defined as effective immunity. The red line represents effective immunity in the infected population, combining total population immunity effects and virus immune escape.

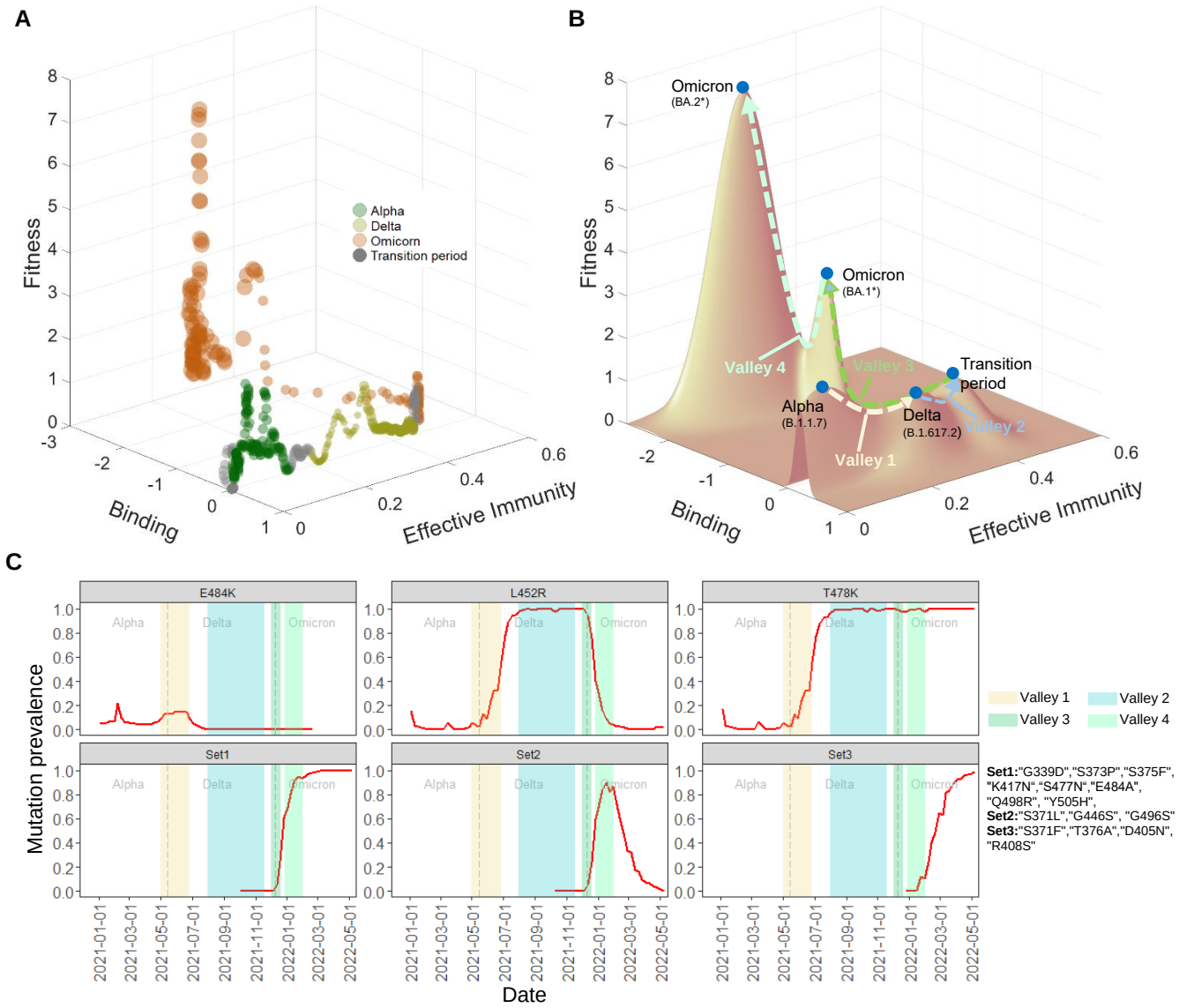

Fig. 5. Interplay between ACE2 binding, virus's fitness, and effective immunity in SARS-CoV-2 evolution. (A) Fitness vs. viral ACE2 binding and effective immunity. Color coding shows alpha (deep green), delta (light green), and omicron (brown) variants. Gray dots represent variant fitness levels during a 2-week transition period following the initial outbreak of each VOC pandemic. (B) Fitness landscape across VOC. Surface topography shows the virus's fitness landscape, which reconstructed based on multivariate normal distribution. Peaks denote maximum variant fitness values in predominant periods. (C) Mutation prevalence associated with VOC transitions. Shadowed regions corresponding to valleys in panel B are marked by different colors. Some cluster mutations with similar prevalence were termed "Set1", "Set2" and "Set3".