



# Antigenic variation of SARS-CoV-2 in response to immune pressure

Diego Forni<sup>1</sup> | Rachele Cagliani<sup>1</sup> | Chiara Pontremoli<sup>1</sup> | Alessandra Mozzi<sup>1</sup> |  
Uberto Pozzoli<sup>1</sup> | Mario Clerici<sup>2,3</sup> | Manuela Sironi<sup>1</sup>

<sup>1</sup>Scientific Institute IRCCS E. MEDEA, Bioinformatics, Bosisio Parini, Italy

<sup>2</sup>Department of Physiopathology and Transplantation, University of Milan, Milan, Italy

<sup>3</sup>Don C. Gnocchi Foundation ONLUS, IRCCS, Milan, Italy

## Correspondence

Diego Forni, Scientific Institute IRCCS E. MEDEA, Bioinformatics, Bosisio Parini, Italy.  
Email: diego.forni@lanostrafamiglia.it

## Funding information

This work was supported by the Italian Ministry of Health ("Ricerca Corrente 2019–2020" to M.S., "Ricerca Corrente 2018–2020" to D.F.).

## Abstract

Analysis of the bat viruses most closely related to SARS-CoV-2 indicated that the virus probably required limited adaptation to spread in humans. Nonetheless, since its introduction in human populations, SARS-CoV-2 must have been subject to the selective pressure imposed by the human immune system. We exploited the availability of a large number of high-quality SARS-CoV-2 genomes, as well as of validated epitope predictions, to show that B cell epitopes in the spike glycoprotein (S) and in the nucleocapsid protein (N) have higher diversity than nonepitope positions. Similar results were obtained for other human coronaviruses and for sarbecoviruses sampled in bats. Conversely, in the SARS-CoV-2 population, epitopes for CD4<sup>+</sup> and CD8<sup>+</sup> T cells were not more variable than nonepitope positions. A significant reduction in epitope variability was instead observed for some of the most immunogenic proteins (S, N, ORF8 and ORF3a). Analysis over longer evolutionary time frames indicated that this effect is not due to differential constraints. These data indicate that SARS-CoV-2 evolves to elude the host humoral immune response, whereas recognition by T cells is not actively avoided by the virus. However, we also found a trend of lower diversity of T cell epitopes for common cold coronaviruses, indicating that epitope conservation per se is not directly linked to disease severity. We suggest that conservation serves to maintain epitopes that elicit tolerizing T cell responses or induce T cells with regulatory activity.

## KEY WORDS

B cell epitope, COVID-19, human coronavirus, sarbecovirus, SARS-CoV-2, T cell epitope

## 1 | INTRODUCTION

The COVID-19 pandemic has been caused by a novel coronavirus named SARS-CoV-2 (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020). SARS-CoV-2 probably originated and evolved in bats, eventually spilling

over to humans, either directly or through an intermediate host (Killerby et al., 2020; Lam et al., 2020; Liu et al., 2020; Sironi et al., 2020; Wong et al., 2020; Xiao et al., 2020; Zhou et al., 2020). Sustained human-to-human transmission had led to global spread of the virus, which has now resulted in an unprecedented global sanitary crisis. Although the majority of COVID-19 cases are relatively mild, a significant proportion of patients develop a serious,

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. Molecular Ecology published by John Wiley & Sons Ltd

often fatal illness, characterized by acute respiratory distress syndrome (Wu & McGoogan, 2020). Both viral-induced lung pathology and overactive immune responses are thought to contribute to this disease severity (St John & Rathore, 2020; Vabret et al., 2020).

Ample evidence suggests that coronaviruses can easily cross species barriers and have high zoonotic potential. Indeed, seven coronaviruses are known to infect humans and all of them originated in animals (Cui et al., 2019; Forni et al., 2017; Ye et al., 2020). Among these, HCoV-OC43, HCoV-HKU1, HCoV-NL63, and HCoV-229E have been circulating for decades in human populations and usually cause limited disease (Bucknall et al., 1972; Forni et al., 2017; Woo et al., 2005). They are thus referred to as "common cold" coronaviruses. Conversely, MERS-CoV and SARS-CoV, whose emergence in the 2000s preceded that of SARS-CoV-2, can cause serious illness and respiratory distress syndrome in a non-negligible proportion of infected individuals (Petrosillo et al., 2020). Like all coronaviruses, these human-infecting viruses have positive-sense, single stranded RNA genomes. Two-thirds of the coronavirus genome are occupied by two large overlapping open reading frames (ORF1a and ORF1b) that are translated into polyproteins. These latter are processed to generate 16 nonstructural proteins (nsP1 to nsP16). The remaining portion of the genome includes ORFs for the structural proteins (spike, envelope, membrane and nucleocapsid) and a variable number of accessory proteins (Cui et al., 2019; Forni et al., 2017).

Analyses of the bat viruses most closely related to SARS-CoV-2 have indicated that, in analogy to SARS-CoV, the virus probably required limited adaptation to gain the ability to infect and spread in humans (Boni et al., 2020; Cagliani et al., 2020). Nonetheless, since its introduction in human populations, SARS-CoV-2 must have been subject to the selective pressure imposed by the human immune system. In fact, as with most other viruses, data from COVID-19, SARS and MERS patients indicate that both B and T lymphocytes play a role in controlling infection (Channappanavar et al., 2014; St John & Rathore, 2020; Vabret et al., 2020).

Recent efforts predicted B cell and T cell epitopes in SARS-CoV-2 proteins (Grifoni, Sidney, et al., 2020) and validated such predictions using sera/lymphocytes from convalescent COVID-19 patients (Grifoni, Weiskopf, et al., 2020). These studies, as well as others (Farrera-Soler et al., 2020; Peng et al., 2020; Poh et al., 2020), revealed that the cell-mediated responses against SARS-CoV-2 are not restricted to the nucleocapsid (N) and spike (S) proteins, but rather target both structural and nonstructural viral products. In parallel, analyses of B cell responses in SARS-CoV-2-infected patients showed that the S and N proteins are the major targets of the antibody response and identified specific B cell epitopes in the S protein (Farrera-Soler et al., 2020; Jiang et al., 2020; Poh et al., 2020). We exploited this growing wealth of information to investigate whether, after a few months of sustained transmission, the selective pressure exerted by the human adaptive immune response is already detectable in the SARS-CoV-2 population.

## 2 | MATERIALS AND METHODS

### 2.1 | Epitope prediction and experimental epitopes

Epitope prediction was performed using different tools from The Immune Epitope Database (IEDB; <https://www.iedb.org/>), as previously suggested (Grifoni, Sidney, et al., 2020). Protein sequences from reference strains of human coronaviruses were used as input for all prediction analyses (SARS-CoV-2, NC\_045512; SARS-CoV, NC\_004718; Human coronavirus 229E, NC\_002645; Human coronavirus NL63, NC\_005831; Human coronavirus OC43, NC\_006213; Human coronavirus HKU1, NC\_006577). In particular, for linear B cell epitope prediction, we used the Bepipred Linear Epitope Prediction 2.0 tool (Jespersen et al., 2017) with a cutoff of 0.550 and epitope length >7. Conformational B epitopes for the S and N proteins of SARS-CoV-2 were calculated using DISCOTOPE 2.0 (Kringelum et al., 2012) with a threshold of -2.5 and published 3D protein structures (PDB IDs: 6VSB, spike; 6M3M [N-term] and 7C22 [C-term], nucleocapsid protein).

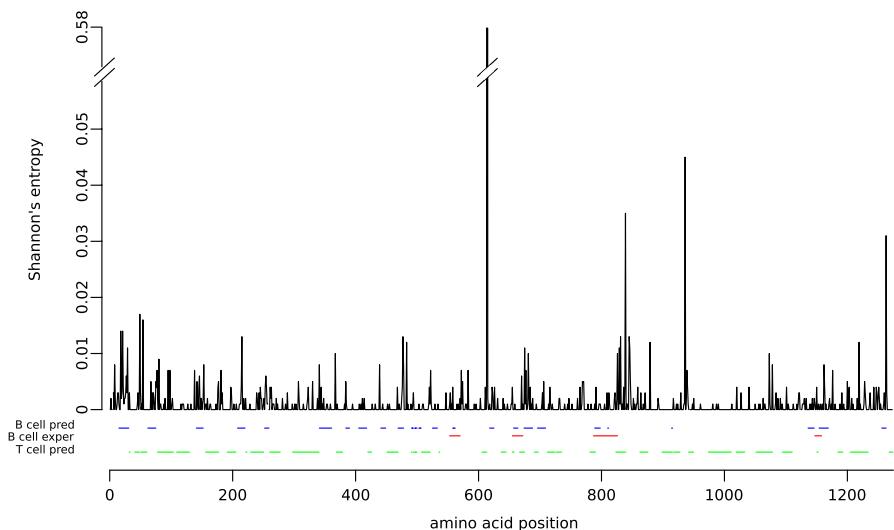
SARS-CoV-2 predicted T cell epitopes were retrieved from Grifoni, Sidney, et al. (2020). For all other coronaviruses, we applied the same methodology used by Grifoni, Sidney, et al. (2020). CD4<sup>+</sup> cell epitopes were predicted using TEPI TOOL (Paul et al., 2016) with default parameters. CD8<sup>+</sup> epitopes were predicted by using the MHC-I BINDING PREDICTIONS version 2.23 tool (<http://tools.iedb.org/mhci/>). The NETMHCPAN EL 4.0 method (Jurtz et al., 2017) was applied and the 12 most frequent HLA class I alleles in human populations (HLA-A01:01, HLA-A02:01, HLA-A03:01, HLA-A11:01, HLA-A23:01, HLA-A24:02, HLA-B07:02, HLA-B08:01, HLA-B35:01, HLA-B40:01, HLA-B44:02, HLA-B44:03) were analysed with a 8–14 kmer range. Only epitopes with a score rank ≤0.1 in one of the 12 HLA classes were selected.

Experimentally identified CD4<sup>+</sup> and/or CD8<sup>+</sup> T cell epitopes in S, N, M, ORF3a and ORF7a were retrieved from Peng et al. (2020). Epitopes were defined as being recognized by CD4<sup>+</sup> or CD8<sup>+</sup> T cells following indications in the original publication (Peng et al., 2020). When this information was not available, epitopes were only included in the overall analysis of T cell epitopes. Experimental B cell epitopes were obtained from two studies that systematically mapped antibody responses against the S protein (Farrera-Soler et al., 2020; Poh et al., 2020).

### 2.2 | Sequences and alignments

SARS-CoV-2 protein sequences were downloaded from the GISAID Initiative (<https://www.gisaid.org>) database (as of June 5, 2020). All protein sequences were retrieved and several filters were applied. Only complete genomes flagged as "high coverage only" and "human" were selected. Positions recommended to be masked by DeMaio and coworkers (<https://virological.org/t/masking-strategies-for-sars-cov-2-alignments/480>, last accessed June 5, 2020) were also removed.

**FIGURE 1** Amino acid variability of the SARS-CoV-2 spike protein. Shannon's entropy ( $H$ ) values for each amino acid position calculated using 10,000 SARS-CoV-2 spike proteins are shown. B cell predicted epitopes and T cell predicted epitopes are also reported in blue and green, respectively. B cell epitopes identified in the sera of COVID-19 patients (Farrera-Soler et al., 2020; Poh et al., 2020) are also reported in red



Finally, for each SARS-CoV-2 protein, we selected only strains that had the same length as the protein in the SARS-CoV-2 reference strain (NC\_045512), generating a set of at least 23,625 sequences for each ORF. Proteins with <60 amino acids were excluded from the analyses.

The list of GISAID IDs along with the list of laboratories which generated the data is provided in Table S1.

For all the other human coronaviruses, as well as for a set of nonhuman infecting sarbecoviruses, sequences of either complete genomes or single ORFs (i.e., nucleocapsid and spike protein) were retrieved from the National Center for Biotechnology Information database (NCBI, <http://www.ncbi.nlm.nih.gov/>). For all human coronaviruses, the only filter we applied was host identification as "human". SARS-CoV strains sampled during the second outbreak were excluded from the analyses. NCBI ID identifiers are listed in Tables S2 and S3.

Alignments were generated using MAFFT (Katoh & Standley, 2013).

### 2.3 | Protein variability and statistical analysis

Variability at each amino acid position was estimated using the Shannon's entropy ( $H$ ) index using the Shannon Entropy-One tool from the HIV database (<https://www.hiv.lanl.gov/content/index>), with ambiguous character (e.g., gaps) excluded from the analysis. For SARS-CoV-2 strains,  $H$  was calculated on alignments of 10,000 randomly selected sequences for each protein. For each protein we evaluated the difference  $D$  between average  $H$  values at epitope and nonepitope positions.

Most positions of analysed viruses are invariable along the alignments, so the distribution of  $H$  is zero-inflated. We thus calculated statistical significance by permutations. For each protein, the predicted epitope intervals were collapsed to a single position while nonepitope intervals were left unchanged. After randomly shuffling this collapsed sequence it was expanded back to full length and the difference between shuffled epitope and nonepitope  $H$  values was

calculated. This procedure was repeated 1,000 times and the proportion of repetitions showing a difference more extreme than  $D$  was reported as the  $p$ -value. An in-house R script was written and is available as Appendix S1.

## 3 | RESULTS

### 3.1 | Antigenic variability of SARS-CoV-2 proteins

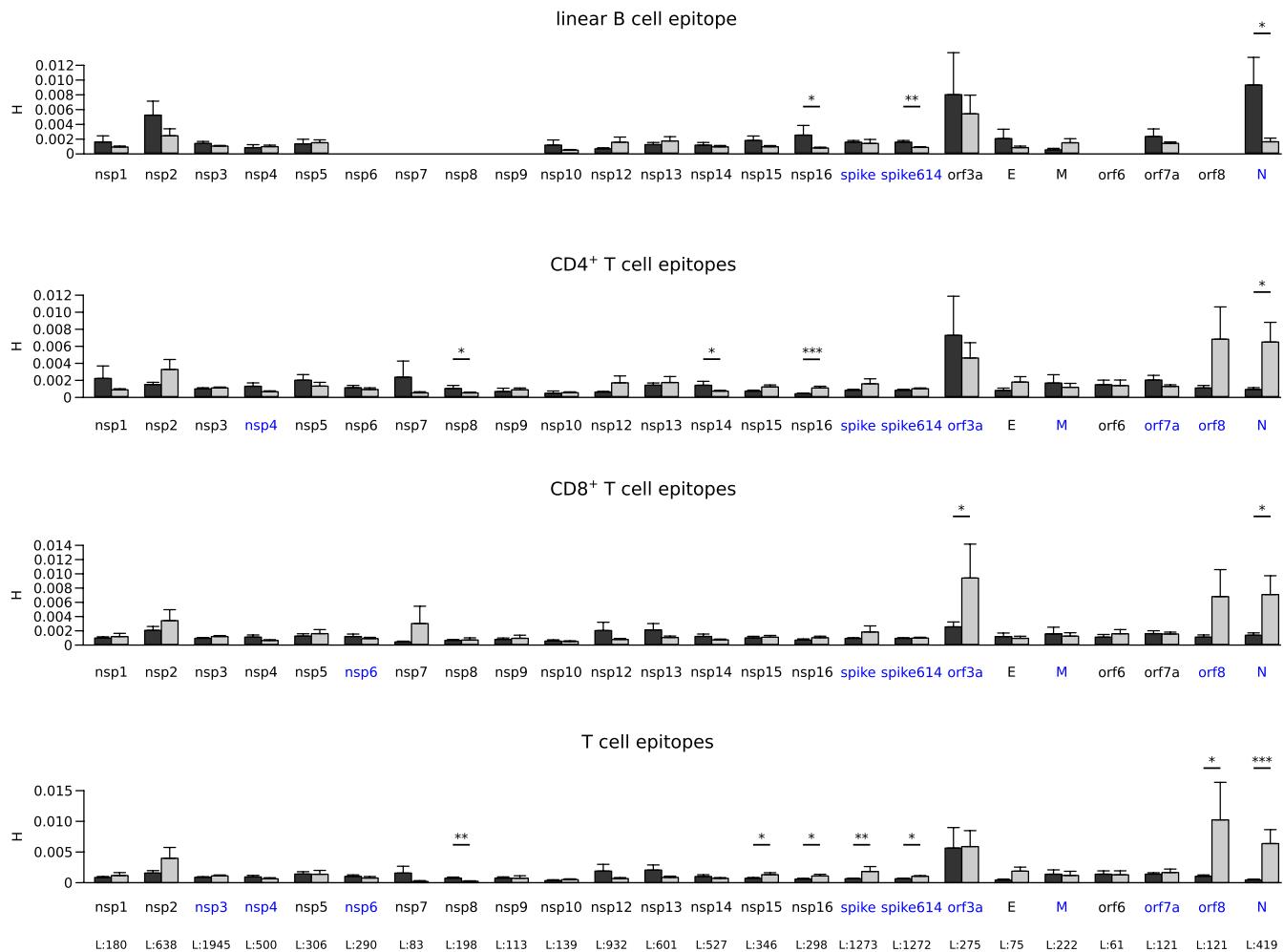
To analyse B cell epitope diversity in SARS-CoV-2, we randomly selected 10,000 high-quality viral genomes from those available in the GISAID database (as of June 5, 2020) (Elbe & Buckland-Merrett, 2017). Potential epitopes were predicted using IEDB tools, as previously described (Grifoni, Sidney, et al., 2020). Specifically, because they are the major targets of the humoral immune response (Channappanavar et al., 2014; St John & Rathore, 2020; Vabret et al., 2020), we predicted both linear and conformational B epitopes for the S and N proteins, whereas only linear epitopes were predicted for the other viral proteins (Table S4). A good correspondence was observed between B cell epitope predictions for the S protein and epitopes identified in two studies that systematically mapped antibody responses in the sera of convalescent COVID-19 patients (Farrera-Soler et al., 2020; Poh et al., 2020; Figure 1).

Variability at each amino acid site of the proteins encoded by SARS-CoV-2 was quantified using Shannon's entropy ( $H$ ). Specifically, only predicted proteins longer than 60 amino acids were analysed. Because most positions in SARS-CoV-2 genomes are invariable across the sampled genomes, the distribution of  $H$  is zero-inflated, making the use of conventional statistical tests inappropriate (McElduff et al., 2010). We thus calculated statistical significance by permutations, that is by reshuffling epitope positions as amino acid stretches of the same size as the predicted epitopes. This approach also has the advantage of accounting for the possibility that, as a result of locally varying selective constraints,  $H$  is not independent among continuous protein positions.

Using this methodology, we found that, for the N and nsp16 proteins, positions mapping to predicted B cell linear epitopes are significantly more variable than those not mapping to these epitopes. A higher diversity of B cell epitopes was also observed for S, although it did not reach statistical significance (Figure 2). However, the  $H$  distribution for the spike protein includes a clear outlier represented by position 614 (Figure 1). Recent studies have indicated that the D614G variant, which is now prevalent worldwide, enhances viral infectivity (Hou et al., 2020; Korber et al., 2020; Plante et al., 2020; Yurkovetskiy et al., 2020; Zhang, Jackson, et al., 2020). Although contrasting results were obtained, it seems that the variant either does not change or modestly affects virus neutralization by antibodies (Beaudoin-Bussières et al., 2020; Hassan et al., 2020; Hou et al., 2020; Korber et al., 2020; Plante et al., 2020; Weissman et al., 2020; Yurkovetskiy et al., 2020; Zhang, Jackson, et al., 2020). Hence, the frequency increase of D614 is unlikely to be primarily related to immune evasion. Moreover, the modulation of resistance to antibodies is thought to be mediated by a change in the exposure

of neutralizing epitopes in the receptor-binding domain, rather than by the creation/destruction of an epitope by D614G itself (Plante et al., 2020; Weissman et al., 2020). We thus repeated the analyses after excluding position 614 and we observed that predicted B cell linear epitopes in the spike protein are significantly more variable than nonepitope positions (Figure 2). The same analysis for B cell conformational epitopes in the N and S proteins indicated a similar trend, although statistical significance was not reached (data not shown). This is probably due to the small number of positions in these epitopes. Overall, these data fit very well with the observation that most humoral immune responses against SARS-CoV-2 and other human coronaviruses are directed against the S and N proteins (Farrera-Soler et al., 2020; Jiang et al., 2020; Poh et al., 2020). These results also support the idea that the selective pressure exerted by the human antibody response is already detectable in the SARS-CoV-2 population.

We next assessed whether epitopes for cell-mediated immune responses are also more variable than nonepitope positions. We



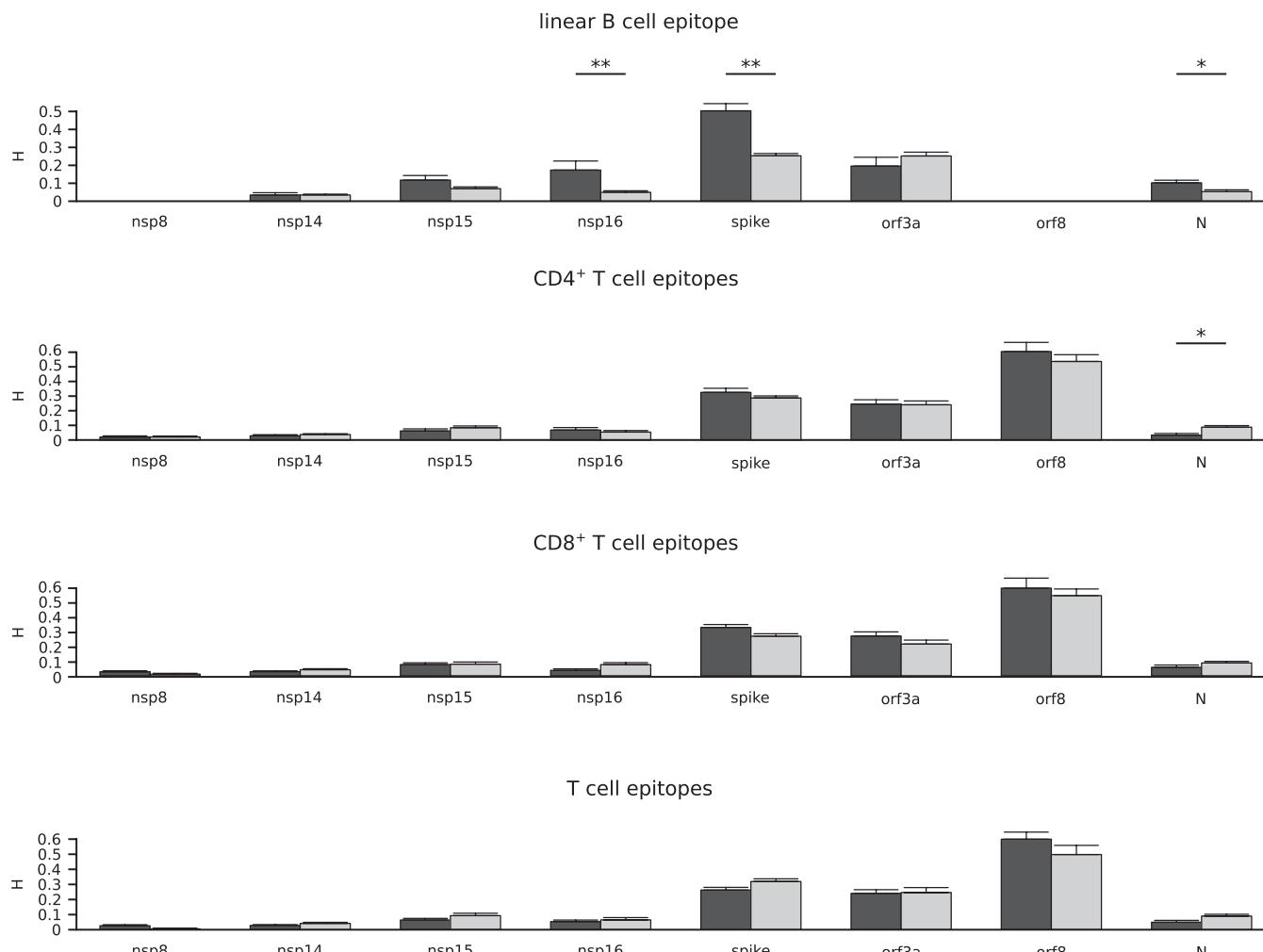
**FIGURE 2** Variability of epitope and nonepitope positions among SARS-CoV-2 proteins. Shannon's entropy ( $H$ ) mean values along with standard errors are shown for all SARS-CoV-2 proteins longer than 60 residues. Epitope positions are shown in dark grey and nonepitopes in light grey. Significant comparisons, calculated by a permutation approach, are indicated with asterisks (\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ ). Immunogenic proteins are shown in blue and the length of each protein is reported in the bottom panel

thus retrieved predicted CD4<sup>+</sup> and CD8<sup>+</sup> T cell epitopes from Grifoni, Sidney, et al. (2020). These epitope predictions were shown to be reliable, as they capture a significant proportion of T cell responses in convalescent COVID-19 patients (Grifoni, Weiskopf, et al., 2020). Analysis of entropy values indicated that CD4<sup>+</sup> T cell epitopes are significantly less variable than nonepitope positions for the N and nsp16 proteins (Figure 2). A similar trend was observed for ORF8, E and S, although significance was not reached. Reduced variability was also observed for CD8<sup>+</sup> T cell epitopes for the N protein, as well as for ORF3a. Higher variability in epitope positions was observed for nsp8 and nsp14 for CD4<sup>+</sup> T cells alone (Figure 2). Because several epitopes for T cells compair with B cell epitopes, which tend to show higher diversity, we compared positions within CD4<sup>+</sup> or CD8<sup>+</sup> T cell epitopes only (not overlapping with B cell epitopes) with positions not mapping to any of these epitopes. A significant reduction of variability was observed for S, N, ORF8, nsp15 and nsp16, whereas higher diversity was still evident for nsp8 (Figure 2).

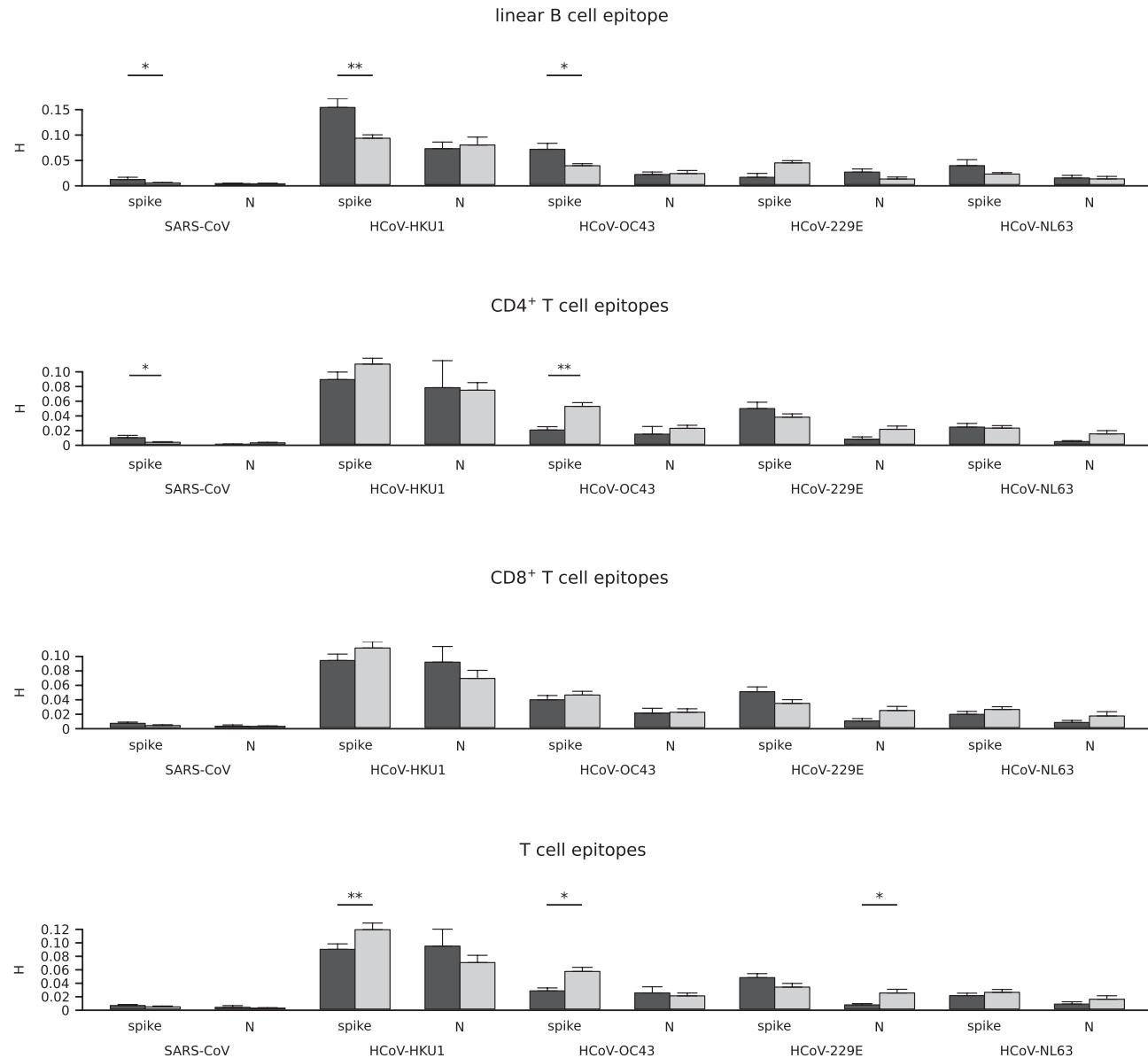
Overall, these data indicate that T cell epitopes in the most immunogenic SARS-CoV-2 proteins (S, N, ORF3a and ORF8; Grifoni,

Weiskopf, et al., 2020; Peng et al., 2020) tend to be more conserved than nonepitopes. However, this was not the case for other proteins targeted by T cell responses, namely M, ORF7a, nsp3, nsp4, and nsp6. Qualitatively similar results were obtained when a set of recently described experimental CD4<sup>+</sup> and/or CD8<sup>+</sup> T cell epitopes in S, N, M, ORF3a and ORF7a were analysed (Peng et al., 2020; Figure S1). The lack of statistical significance in some of these comparisons is probably due to the low number of epitope positions.

Clearly, protein sequence variability is strongly influenced by functional and structural constraints. We thus reasoned that if the observations reported above were secondary to the incidental colocalization of T cell epitopes with more constrained regions, a similar pattern should be observed for  $H$  values calculated on an alignment of proteins from other sarbecoviruses. In fact, all these viruses, except SARS-CoV, were sampled from bats. Thus, whereas structural/functional constraints are expected to be maintained across long evolutionary time frames, the pressure exerted by the human cell-mediated immune response is not, given that (in different species) antigen processing within host cells results in the



**FIGURE 3** Variability of epitope and nonepitope positions among sarbecoviruses. Shannon's entropy ( $H$ ) mean values along with standard errors are shown for a set of sarbecovirus ORFs. SARS-CoV-2 epitope positions are shown in dark grey and nonepitopes in light grey. Significant comparisons, calculated by a permutation approach, are indicated with asterisks (\* $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ )



**FIGURE 4** Variability of epitope and nonepitope positions among human coronaviruses. Shannon's entropy ( $H$ ) mean values along with standard errors are shown for human coronavirus spike and nucleocapsid proteins. Epitope positions are shown in dark grey and nonepitopes in light grey. Significant comparisons, calculated by a permutation approach, are indicated with asterisks (\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ )

preferential presentation of diverse viral epitopes to T lymphocytes depending on the MHC gene repertoire and on distinct preferences of the antigen processing pathway (Abduriyim et al., 2019; Burgevin et al., 2008; Hammer et al., 2007; Lu et al., 2019; Wynne et al., 2016). Conversely, epitopes for antibodies tend to be conserved across species (Tse et al., 2017; Wiehe et al., 2014) and consequently the selective pressure acting on these positions is expected to be constant across time and hosts.

We thus aligned the SARS-CoV-2 reference sequences of proteins showing decreased or increased variability in T cell epitopes with those of 45 sarbecoviruses. Calculation of  $H$  indicated a significant difference only for CD4<sup>+</sup> T cell epitopes in the N protein. Conversely, B cell epitopes were more variable than nonepitope

positions for the S, N and nsp16 proteins (Figure 3). Overall, these results indicate that the variability within SARS-CoV-2 T cell epitopes is not driven primarily by functional/structural constraints, but probably results from the interaction with the human adaptive immune response.

### 3.2 | Comparison with other human coronaviruses

Given the results above we set out to determine whether the other human coronaviruses show the same tendency of reduced and increased variability at T cell and B cell epitopes, respectively. For these viruses, analyses were restricted to the N and S proteins, as

they are the most antigenic proteins and because the number of complete viral genomes is relatively limited (Tables S3 and S5).

SARS-CoV, the human coronavirus most similar to SARS-CoV-2, caused the first human outbreak in 2002/2003 after a spillover from palm civets, followed by human-to-human transmission chains (Shi & Wang, 2017). A second zoonotic transmission occurred in December 2003 and caused a limited number of cases (Shi & Wang, 2017; Wang et al., 2005). Viral genomes sampled during the second outbreak were not included in the analyses because their evolution occurred in the civet reservoir (Table S2). Four other human coronaviruses, namely HCoV-OC43, HCoV-HKU1 (members of the *Embecovirus* subgenus), HCoV-229E (*Duvinavirus* subgenus), and HCoV-NL63 (*Setracovirus* subgenus), have been transmitting within human populations for at least 70 years (Forni et al., 2017). Thus, all available S and N sequences were included in the analyses (Table S3). Conversely, MERS-CoV displays limited ability for human-to-human transmission and outbreaks were caused by repeated spillover events from the camel host (Cui et al., 2019). MERS-CoV was therefore excluded from the analyses.

Quantification of sequence variability by calculation of  $H$  indicated that B cell epitopes in the S protein are significantly more variable than nonepitopes for SARS-CoV, HCoV-OC43 and HCoV-HKU1 (Figure 4). Analysis of CD4<sup>+</sup> and CD8<sup>+</sup> T cell epitopes in these viruses indicated no increased diversity for epitope compared to nonepitope positions, except for the S protein of SARS-CoV for CD4<sup>+</sup> T cells. However, when positions within B cell epitopes were excluded from the analysis, this difference disappeared and T cell epitopes were found to be significantly less variable than nonepitopes for the spike proteins of HCoV-HKU1 and HCoV-OC43, as well as for the N protein of HCoV-229E (Figure 4). Thus, the lack of antigenic diversity at T cell epitopes is a common feature of human coronaviruses, which instead tend to maintain sequence conservation of such epitopes.

## 4 | DISCUSSION

The origin of SARS-CoV-2 remains uncertain and it is presently unknown whether the virus spilled over from a bat or another intermediate host. The hypothesis of a zoonotic origin is strongly supported by multiple lines of evidence, although it cannot be excluded that SARS-CoV-2 was transmitted cryptically in humans before gaining the ability to spread efficiently among people (Andersen et al., 2020; Sironi et al., 2020). Whatever the initial events associated with the early phases of the pandemic, it is clear that circulating SARS-CoV-2 viruses shared a common ancestor at the end of 2019 (van Dorp et al., 2020; Li et al., 2020). Due to its recent origin, the genetic diversity of the SARS-CoV-2 population remains limited. This is also the result of the relatively low mutation rate of coronaviruses (as compared to other RNA viruses), which encode enzymes with some proofreading ability (Denison et al., 2011; Forni et al., 2017). Nonetheless, the huge number of transmissions worldwide has allowed thousands of mutations to appear in the viral population and, thanks to enormous international sequencing efforts, more

than 25,000 amino acid replacements have currently been reported (<http://cov-glue.cvr.gla.ac.uk>). Irrespective of the host, most variants are expected to be deleterious for viral fitness, or to have no consequences (Cagliani et al., 2020; van Dorp et al., 2020; Grubaugh et al., 2020). However, a proportion of the replacements may favour the virus and some of these may contribute to adaptation to the human host. In particular, the recent and ongoing evolution of SARS-CoV-2 is expected to be at least partially driven by the selective pressure imposed by the human immune system. Indeed, antigenic drift or immune evasion mutations have been reported for other zoonotic viruses such as Lassa virus (Andersen et al., 2015) and Influenza A virus (Su et al., 2015). The emergence of immune evasion variants was also observed during an outbreak of MERS-CoV in South Korea, when mutations in the spike proteins were positively selected as they facilitated viral escape from neutralizing antibodies, even though the same variants decreased binding to the cellular receptor (Kim et al., 2019; Kim et al., 2019; Kleine-Weber et al., 2019; Rockx et al., 2010). This exemplifies a phenomenon often observed in other viruses, most notably HIV-1 (Liu et al., 2007; Martinez-Picado et al., 2006; Schneidewind et al., 2007, 2008), whereby the virus trades off immune evasion with a fitness cost. As a consequence, immune evasion mutations may be only transiently maintained in viral populations. We therefore decided to quantify epitope variability in terms of entropy, rather than relying on measures based on substitution rates (dN/dS), which were developed for application to variants that go to fixation in different lineages over time (Kryazhimskiy & Plotkin, 2008).

The MERS-CoV mutants responsible for the outbreak in South Korea also testify to the relevance of the antibody response in coronavirus control and the selective pressure imposed by humoral immunity on the virus (Kim et al., 2019; Kleine-Weber et al., 2019; Rockx et al., 2010). This is probably also the case for SARS-CoV-2, as a recent report indicated that the sera of most COVID-19 convalescent patients have virus-neutralization activities and that antibody titres negatively correlate with viral load (Okba et al., 2020; Vabret et al., 2020; Wu et al., 2020; Zhou et al., 2020). Nonetheless, studies on relatively large COVID-19 patient cohorts reported that patients with severe disease display stronger IgG responses than milder cases, and a negative correlation between anti-S antibody titres and lymphocyte counts was reported (Jiang et al., 2020; Vabret et al., 2020; Wu et al., 2020; Zhang, Zhou, et al., 2020; Zhao et al., 2020). Consistently, asymptomatic SARS-CoV-2-infected individuals were recently reported to have lower virus-specific IgG levels than COVID-19 patients (Long et al., 2020). These observations raised concerns that humoral responses might not necessarily be protective, but rather pathogenic, either via antibody-dependent enhancement (ADE) or other mechanisms (Cao, 2020; Iwasaki & Yang, 2020; Wu et al., 2020).

Clearly, gaining insight into the dynamic interaction between SARS-CoV-2 and the human immune system is of fundamental importance not only to understand COVID-19 immunopathogenesis, but also to inform therapeutic and preventive viral control strategies. We thus exploited the availability of a large number of fully

sequenced high-quality SARS-CoV-2 genomes, as well as validated predictions of B cell and T cell epitopes, to investigate whether the selective pressure exerted by the adaptive immune response is detectable in the global SARS-CoV-2 population, and if the virus is evolving to evade it. Results indicated that B cell epitopes in the N and S proteins, which represent the major targets of the antibody response, have higher diversity than nonepitope positions. The same was observed for the spike proteins of HCoV-HKU1, HCoV-OC43 and SARS-CoV, although data on SARS-CoV should be taken with caution as they derive from a relatively small number of sequences sampled over a short time frame. Conversely, no evidence of antibody-mediated selective pressure was evident for HCoV-229E and HCoV-NL63. The reasons underlying these differences are unclear, but recent data on a relatively small population of patients with respiratory disease indicated that the titres of neutralizing antibodies against HCoV-OC43 tend to be higher compared to those against HCoV-229E and HCoV-NL63 (HCoV-HKU1 was not evaluated), suggesting the two latter viruses elicit mainly nonneutralizing responses (Gorse et al., 2020).

B cell epitopes within nsp16 were also found to be variable, although this protein was not reported to be immunogenic (Grifoni, Weiskopf, et al., 2020). However, the antibody response to SARS-CoV-2 has presently been systematically analysed in a relatively small number of patients and most studies focused on structural proteins. It is thus possible that, during infection, antibodies against nsp16 are raised, but they have not yet been detected. An alternative possibility is that B cell epitopes in nsp16, which is highly conserved in SARS-CoV-2 strains (Cagliani et al., 2020), coincide with regions of relatively weaker constraint. This hypothesis is partially supported by the observation that these same positions also display higher diversity when entropy is calculated on an alignment of sarbecovirus nsp16 proteins. More intriguingly, this result may indicate that nsp16, together with S and N, is a target of B cell responses in the bat reservoirs. In fact, as mentioned above, antibody binding sites tend to be conserved across species (Tse et al., 2017; Wiehe et al., 2014) and thus the selective pressure exerted on B cell epitopes is likely to be constant across hosts. Although the immunogenicity of nsp16 remains to be evaluated, these data suggest that SARS-CoV-2 is evolving to elude the host humoral immune response. However, we note that this observation does not necessarily imply that antibodies against SARS-CoV-2 are protective and it does not rule out the possibility that humoral responses contribute to COVID-19 pathogenesis. We should also add that we cannot exclude that the higher diversity observed at B cell epitopes is ultimately the result of epitope regions being more exposed at protein surfaces and less constrained than other regions. However, the fact the higher values of entropy are mainly detected in the B epitope regions of proteins that are strongly targeted by the humoral system speaks against this possibility. Finally, we note that the appearance of within-host transitory mutations in B cell epitopes has previously been observed in other zoonotic, acute viral infections (Andersen et al., 2015).

In COVID-19 patients, antibody titres were found to correlate with the strength of virus-specific T cell responses (Ni et al., 2020).

Surprisingly, we found that, in the SARS-CoV-2 population, epitopes for CD4<sup>+</sup> and CD8<sup>+</sup> T cells are not more variable than nonepitope positions. Conversely, a significant reduction in epitope variability was observed for a subset of viral proteins, in particular for some of the most immunogenic ones (S, N, ORF8 and ORF3a; Grifoni, Weiskopf, et al., 2020; Peng et al., 2020). To check that the result was not due to stronger structural/functional constraints acting on epitope positions, we again used *H* values calculated on an alignment of sarbecovirus genomes, all of which, except SARS-CoV, were sampled in bats. T cell responses are initiated by the presentation of antigenic epitopes by MHC (major histocompatibility complex) class I and class II molecules. Different mammals have diverse MHC gene repertoires and thus present distinct antigens. In particular, recent data from various bat species have indicated that many MHC class I molecules have a three- or five-amino acid insertion in the peptide binding pocket, resulting in very different presented peptide repertoires compared to the MHC class I molecules of other mammals (Abduriyim et al., 2019; Lu et al., 2019; Ng et al., 2016; Papenfuss et al., 2012; Wynne et al., 2016). Thus, the selective pressure acting on T cell epitopes is probably volatile and not conserved in humans and bats. Analysis of sarbecovirus proteins indicated that, apart from CD4<sup>+</sup> T cell epitopes in the N protein, the T cell epitopes predicted in SARS-CoV-2 proteins are not less diverse than nonepitope positions, suggesting that epitope conservation is not simply secondary to structural or functional constraints, but may result from an interaction with human T cell responses. Of course, another possible explanation for this finding is that the prediction tools failed to identify real epitopes. However, we retrieved epitopes from a previous work and the authors validated their predictions using lymphocytes of 20 patients who had recovered from COVID-19 (Grifoni, Sidney, et al., 2020; Grifoni, Weiskopf, et al., 2020). Also, qualitatively similar results were observed with a small set of experimentally identified epitopes. Moreover, if a general artefact linked to epitope prediction had been introduced, we would not expect to observe significant differences and not specifically in the proteins that represent the major targets of T cell responses.

Unexpected conservation of T cell epitopes was previously observed for HIV-1 and *Mycobacterium tuberculosis* (MTB), both of which cause chronic infections in humans (Comas et al., 2010; Coscolla et al., 2015; Lindestam Arlehamn et al., 2015; Sanjuán et al., 2013). In the case of HIV-1, immune activation probably favours the virus by increasing the rate of CD4<sup>+</sup> T cell trans-infection (Sanjuán et al., 2013). Conversely, the mechanisms underlying MTB epitope conservation have not been fully elucidated. A possible explanation is that conserved epitopes generate a decoy immune response and give an advantage to the bacterium. An alternative possibility is that T cell activation results in lung tissue inflammation and damage (cavitary tuberculosis), which favours MTB transmission by aerosols (Coscolla et al., 2015; Lindestam Arlehamn et al., 2015). Although these mechanisms are unlikely to be at play in the case of SARS-CoV-2, a deregulated immune response has been associated with COVID-19 pathogenesis (Hannan et al., 2020). Specifically, recent data indicated that patients recovering from severe COVID-19 have

broader and stronger T cell responses compared to mild cases (Peng et al., 2020). This was particularly evident for responses against the S, membrane (M), ORF3a and ORF8 proteins (Peng et al., 2020). Although this observation might simply reflect higher viral loads in severe cases, the possibility that the T cell response itself is deleterious cannot be excluded. Moreover, the same authors reported that CD8<sup>+</sup> T cells targeting different virus proteins have distinct cytokine profiles, suggesting that the virus can modulate the host immune response to its benefit (Peng et al., 2020). Additionally, a post-mortem study on six patients who died from COVID-19 indicated that infection of macrophages can lead to activation-induced T cell death, which may eventually be responsible for lymphocytopenia (Chen et al., 2020). However, we also found a trend of lower diversity of T cell epitopes for common cold coronaviruses, indicating that epitope conservation per se is not directly linked to disease severity. Moreover, other SARS-CoV-2 immunogenic proteins such as M and ORF7 did not show differences in T cell epitope conservation, which was instead observed for nsp16 and nsp15. These latter proteins are not known to be T cell targets (Grifoni, Weiskopf, et al., 2020). Clearly, further analyses will be required to clarify the significance of T cell epitope conservation in SARS-CoV-2. An interesting possibility is that both for SARS-CoV-2 and for common cold coronaviruses, conservation serves to maintain epitopes that elicit tolerizing T cell responses or induce T cells with regulatory activity. Indeed, we considered T cell epitopes as a whole, but differences exist in terms of variability and, probably, antigenicity. This clearly represents a limitation of this study, but the modest amount of genetic diversity in the SARS-CoV-2 population does not presently allow for analysis of single epitope regions. Moreover, more detailed and robust analyses will certainly require the systematic, experimental definition of T and B cell epitopes in the SARS-CoV-2 proteome.

## ACKNOWLEDGMENTS

We gratefully acknowledge the authors, originating and submitting laboratories of the sequences from GISAID's EpiCoV database on which this research is based. This work was supported by the Italian Ministry of Health ("Ricerca Corrente 2019–2020" to M.S., "Ricerca Corrente 2018–2020" to D.F.).

## AUTHOR CONTRIBUTIONS

Conceptualization, D.F. and M.S.; formal analysis, M.S., U.P. and D.F.; investigation, D.F., R.C., C.P., A.M. and M.S.; visualization, D.F., R.C.; writing—original draft, M.S. and D.F.; writing—review & editing, M.S., M.C., R.C., U.P.; funding acquisition, M.S. and D.F.; supervision, M.S. and M.C.

## DATA AVAILABILITY STATEMENT

Lists of virus accession IDs are reported in Tables S1–S3. Data used for generating Figures 1–4 are reported in Tables S4 and S5. An R script for permutation analysis is reported as Appendix S1.

## ORCID

Diego Forni  <https://orcid.org/0000-0001-9291-5352>

Rachele Cagliani  <https://orcid.org/0000-0003-2670-3532>  
 Chiara Pontremoli  <https://orcid.org/0000-0002-3022-502X>  
 Manuela Sironi  <https://orcid.org/0000-0002-2267-5266>

## REFERENCES

- Abduriyim, S., Zou, D. H., & Zhao, H. (2019). Origin and evolution of the major histocompatibility complex class I region in eutherian mammals. *Ecology and Evolution*, 9, 7861–7874. <https://doi.org/10.1002/ece3.5373>
- Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nature Medicine*, 26, 450–452. <https://doi.org/10.1038/s41591-020-0820-9>
- Andersen, K. G., Shapiro, B. J., Matranga, C. B., Sealfon, R., Lin, A. E., Moses, L. M., Folarin, O. A., Goba, A., Odia, I., Ehiane, P. E., Momoh, M., England, E. M., Winnicki, S., Branco, L. M., Gire, S. K., Phelan, E., Tariyal, R., Tewhey, R., Omoniwa, O., ... Sabeti, P. C. (2015). Clinical sequencing uncovers origins and evolution of Lassa virus. *Cell*, 162, 738–750. <https://doi.org/10.1016/j.cell.2015.07.020>
- Beaudoin-Bussières, G., Laumaea, A., Anand, S. P., Prévost, J., Gasser, R., Goyette, G., Medjahed, H., Perreault, J., Tremblay, T., Lewin, A., Gokool, L., Morrisseau, C., Bégin, P., Tremblay, C., Martel-Laferrière, V., Kaufmann, D. E., Richard, J., Bazin, R., & Finzi, A. (2020). Decline of humoral responses against SARS-CoV-2 spike in convalescent individuals. *MBio*, 11, e02590-e2620. <https://doi.org/10.1128/mBio.02590-20>
- Boni, M. F., Lemey, P., Jiang, X., Lam, T.-Y., Perry, B. W., Castoe, T. A., Rambaut, A., & Robertson, D. L. (2020). Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nature Microbiology*, 5, 1408–1417. <https://doi.org/10.1038/s41564-020-0771-4>
- Bucknall, R. A., King, L. M., Kapikian, A. Z., & Chanock, R. M. (1972). Studies with human coronaviruses II. Some properties of strains 229E and OC43. *Experimental Biology and Medicine*, 139(3), 722–727. <https://doi.org/10.3181/00379727-139-36224>
- Burgevin, A., Saveanu, L., Kim, Y., Barilleau, É., Kotturi, M., Sette, A., van Endert, P., & Peters, B. (2008). A detailed analysis of the murine TAP transporter substrate specificity. *PLoS One*, 3, e2402. <https://doi.org/10.1371/journal.pone.0002402>
- Cagliani, R., Forni, D., Clerici, M., & Sironi, M. (2020). Computational inference of selection underlying the evolution of the novel coronavirus, SARS-CoV-2. *Journal of Virology*, 94, e00411-20. <https://doi.org/10.1128/JVI.00411-20>
- Cao, X. (2020). COVID-19: Immunopathology and its implications for therapy. *Nature Reviews Immunology*, 20, 269–270. <https://doi.org/10.1038/s41577-020-0308-3>
- Channappanavar, R., Zhao, J., & Perlman, S. (2014). T cell-mediated immune response to respiratory coronaviruses. *Immunologic Research*, 59, 118–128. <https://doi.org/10.1007/s12026-014-8534-z>
- Chen, Y., Feng, Z., Diao, B., Wang, R., Wang, G., Wang, C., & Wu, Y. (2020). The novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) directly decimates human spleens and lymph nodes. *medRxiv*. <https://doi.org/10.1101/2020.03.27.20045427>
- Comas, I., Chakravarti, J., Small, P. M., Galagan, J., Niemann, S., Kremer, K., Ernst, J. D., & Gagneux, S. (2010). Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nature Genetics*, 42, 498–503. <https://doi.org/10.1038/ng.590>
- Coronaviridae Study Group of the International Committee on Taxonomy of Viruses (2020). The species Severe acute respiratory syndrome-related coronavirus: Classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*, 5, 536–544. <https://doi.org/10.1038/s41564-020-0695-z>

- Coscolla, M., Copin, R., Sutherland, J., Gehre, F., de Jong, B., Owolabi, O., Mbayo, G., Giardina, F., Ernst, J. D., & Gagneux, S. (2015). *M. tuberculosis* T cell epitope analysis reveals paucity of antigenic variation and identifies rare variable TB antigens. *Cell Host & Microbe*, 18, 538–548. <https://doi.org/10.1016/j.chom.2015.10.008>
- Cui, J., Li, F., & Shi, Z. L. (2019). Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology*, 17(3), 181–192. <https://doi.org/10.1038/s41579-018-0118-9>
- Denison, M. R., Graham, R. L., Donaldson, E. F., Eckerle, L. D., & Baric, R. S. (2011). Coronaviruses: An RNA proofreading machine regulates replication fidelity and diversity. *RNA Biology*, 8, 270–279. <https://doi.org/10.4161/rna.8.2.15013>
- Elbe, S., & Buckland-Merrett, G. (2017). Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges (Hoboken, NJ)*, 1, 33–46. <https://doi.org/10.1002/gch2.1018>
- Farrera-Soler, L., Daguer, J.-P., Barluenga, S., Vadas, O., Cohen, P., Pagano, S., Yerly, S., Kaiser, L., Vuilleumier, N., & Winssinger, N. (2020). Identification of immunodominant linear epitopes from SARS-CoV-2 patient plasma. *PLoS One*, 15, e0238089. <https://doi.org/10.1371/journal.pone.0238089>
- Forni, D., Cagliani, R., Clerici, M., & Sironi, M. (2017). Molecular evolution of human coronavirus genomes. *Trends in Microbiology*, 25, 35–48. <https://doi.org/10.1016/j.tim.2016.09.001>
- Gorse, G. J., Donovan, M. M., & Patel, G. B. (2020). Antibodies to coronaviruses are higher in older compared with younger adults and binding antibodies are more sensitive than neutralizing antibodies in identifying coronavirus-associated illnesses. *Journal of Medical Virology*, 92, 512–517. <https://doi.org/10.1002/jmv.25715>
- Grifoni, A., Sidney, J., Zhang, Y., Scheuermann, R. H., Peters, B., & Sette, A. (2020). A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. *Cell Host & Microbe*, 27, 671–680.e2. <https://doi.org/10.1016/j.chom.2020.03.002>
- Grifoni, A., Weiskopf, D., Ramirez, S. I., Mateus, J., Dan, J. M., Moderbacher, C. R., Rawlings, S. A., Sutherland, A., Premkumar, L., Jadi, R. S., Marrama, D., de Silva, A. M., Frazier, A., Carlin, A. F., Greenbaum, J. A., Peters, B., Krammer, F., Smith, D. M., Crotty, S., & Sette, A. (2020). Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell*, 181, 1489–1501. <https://doi.org/10.1016/j.cell.2020.05.015>
- Grubaugh, N. D., Petrone, M. E., & Holmes, E. C. (2020). We shouldn't worry when a virus mutates during disease outbreaks. *Nature Microbiology*, 5, 529–530. <https://doi.org/10.1038/s41564-020-0690-4>
- Hammer, G. E., Kanaseki, T., & Shastri, N. (2007). The final touches make perfect the peptide-MHC class I repertoire. *Immunity*, 26, 397–406. <https://doi.org/10.1016/j.immuni.2007.04.003>
- Hannan, M. A., Rahman, M. A., Rahman, M. S., Sohag, A. A. M., Dash, R., Hossain, K. S., Farjana, M., & Uddin, M. J. (2020). Intermittent fasting, a possible priming tool for host defense against SARS-CoV-2 infection: Crosstalk among calorie restriction, autophagy and immune response. *Immunology Letters*, 6, 38–45. <https://doi.org/10.1016/j.imlet.2020.07.001>
- Hassan, A. O., Kafai, N. M., Dmitriev, I. P., Fox, J. M., Smith, B. K., Harvey, I. B., Chen, R. E., Winkler, E. S., Wessel, A. W., Case, J. B., Kashentseva, E., McCune, B. T., Bailey, A. L., Zhao, H., VanBlargan, L. A., Dai, Y.-N., Ma, M., Adams, L. J., Shrihari, S., ... Diamond, M. S. (2020). A single-dose intranasal ChAd vaccine protects upper and lower respiratory tracts against SARS-CoV-2. *Cell*, 183, 169–184.e13. <https://doi.org/10.1016/j.cell.2020.08.026>
- Hou, Y. J., Chiba, S., Halfmann, P., Ehre, C., Kuroda, M., Dinnon, K. H. 3rd, Leist, S. R., Schäfer, A., Nakajima, N., Takahashi, K., Lee, R. E., Mascenik, T. M., Edwards, C. E., Tse, L. V., Boucher, R. C., Randell, S. H., Suzuki, T., Gralinski, L. E., Kawaoka, Y., & Baric, R. S. (2020) SARS-CoV-2 D614G variant exhibits enhanced replication ex vivo and earlier transmission in vivo. *bioRxiv*. <https://doi.org/10.1101/2020.09.28.317685>
- Iwasaki, A., & Yang, Y. (2020). The potential danger of suboptimal antibody responses in COVID-19. *Nature Reviews Immunology*, 20, 339–341. <https://doi.org/10.1038/s41577-020-0321-6>
- Jespersen, M. C., Peters, B., Nielsen, M., & Marcatili, P. (2017). BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Research*, 45, W24–W29. <https://doi.org/10.1093/nar/gkx346>
- Jiang, H.-W., Li, Y., Zhang, H.-N., Wang, W., Yang, X., Qi, H., Li, H., Men, D., Zhou, J., & Tao, S.-C. (2020). SARS-CoV-2 proteome microarray for global profiling of COVID-19 specific IgG and IgM responses. *Nature Communications*, 11, 1–11. <https://doi.org/10.1038/s41467-020-17488-8>
- Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., & Nielsen, M. (2017). NetMHCpan-4.0: Improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *The Journal of Immunology*, 199(9), 3360–3368. <https://doi.org/10.4049/jimmunol.1700893>
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30, 772–780. <https://doi.org/10.1093/molbev/mst010>
- Killerby, M. E., Biggs, H. M., Midgley, C. M., Gerber, S. I., & Watson, J. T. (2020). Middle East respiratory syndrome coronavirus transmission. *Emerging Infectious Diseases*, 26, 191–198. <https://doi.org/10.3201/eid2602.190697>
- Kim, Y.-S., Aigerim, A., Park, U., Kim, Y., Rhee, J.-Y., Choi, J.-P., Park, W. B., Park, S. W., Kim, Y., Lim, D.-G., Inn, K.-S., Hwang, E.-S., Choi, M.-S., Shin, H.-S., & Cho, N.-H. (2019). Sequential emergence and wide spread of neutralization escape Middle East respiratory syndrome coronavirus mutants, South Korea, 2015. *Emerging Infectious Diseases*, 25, 1161–1168. <https://doi.org/10.3201/eid2506.181722>
- Kleine-Weber, H., Elzayat, M. T., Wang, L., Graham, B. S., Müller, M. A., Drosten, C., Pöhlmann, S., & Hoffmann, M. (2019). Mutations in the spike protein of middle East respiratory syndrome coronavirus transmitted in Korea increase resistance to antibody-mediated neutralization. *Journal of Virology*, 93, e01381–e1418. <https://doi.org/10.1128/JVI.01381-18>
- Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfaluterer, W., Hengartner, N., Giorgi, E. E., Bhattacharya, T., Foley, B., Hastie, K. M., Parker, M. D., Partridge, D. G., Evans, C. M., Freeman, T. M., de Silva, T. I., McDanal, C., Perez, L. G., Tang, H., ... Wyles, M. D. (2020). Tracking changes in SARS-CoV-2 Spike: Evidence that D614G increases infectivity of the COVID-19 virus. *Cell*, 82, 812–827. <https://doi.org/10.1016/j.cell.2020.06.043>
- Kringelum, J. V., Lundsgaard, C., Lund, O., & Nielsen, M. (2012). Reliable B cell epitope predictions: Impacts of method development and improved benchmarking. *PLoS Computational Biology*, 8, e1002829. <https://doi.org/10.1371/journal.pcbi.1002829>
- Kryazhimskiy, S., & Plotkin, J. B. (2008). The population genetics of dN/dS. *PLoS Genetics*, 4, e1000304. <https://doi.org/10.1371/journal.pgen.1000304>
- Lam, T.-Y., Jia, N. A., Zhang, Y.-W., Shum, M.-H., Jiang, J.-F., Zhu, H.-C., Tong, Y.-G., Shi, Y.-X., Ni, X.-B., Liao, Y.-S., Li, W.-J., Jiang, B.-G., Wei, W., Yuan, T.-T., Zheng, K., Cui, X.-M., Li, J., Pei, G.-Q., Qiang, X., ... Cao, W.-C. (2020). Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature*, 583, 282–285. <https://doi.org/10.1038/s41586-020-2169-0>
- Li, X., Wang, W., Zhao, X., Zai, J., Zhao, Q., Li, Y., & Chaillon, A. (2020). Transmission dynamics and evolutionary history of 2019-nCoV. *Journal of Medical Virology*, 92, 501–511. <https://doi.org/10.1002/jmv.25701>

- LindestamArlehamn, C. S., Paul, S., Mele, F., Huang, C., Greenbaum, J. A., Vita, R., Sidney, J., Peters, B., Sallusto, F., & Sette, A. (2015). Immunological consequences of intragenus conservation of *Mycobacterium tuberculosis* T-cell epitopes. *Proceedings of the National Academy of Sciences*, 112, E147–E155. <https://doi.org/10.1073/pnas.1416537112>
- Liu, P., Jiang, J.-Z., Wan, X.-F., Hua, Y., Li, L., Zhou, J., Wang, X., Hou, F., Chen, J., Zou, J., & Chen, J. (2020). Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? *PLoS Pathogens*, 16, e1008421. <https://doi.org/10.1371/journal.ppat.1008421>
- Liu, Y. I., McNevin, J., Zhao, H., Tebit, D. M., Troyer, R. M., McSweyn, M., Ghosh, A. K., Shriner, D., Arts, E. J., McElrath, M. J., & Mullins, J. I. (2007). Evolution of human immunodeficiency virus type 1 cytotoxic T-lymphocyte epitopes: Fitness-balanced escape. *Journal of Virology*, 81, 12179–12188. <https://doi.org/10.1128/JVI.01277-07>
- Long, Q.-X., Tang, X.-J., Shi, Q.-L., Li, Q., Deng, H.-J., Yuan, J., Hu, J.-L., Xu, W., Zhang, Y., Lv, F.-J., Su, K., Zhang, F., Gong, J., Wu, B. O., Liu, X.-M., Li, J.-J., Qiu, J.-F., Chen, J., & Huang, A.-L. (2020). Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nature Medicine*, 26, 1200–1204. <https://doi.org/10.1038/s41591-020-0965-6>
- Lu, D., Liu, K., Zhang, D. I., Yue, C., Lu, Q., Cheng, H., Wang, L., Chai, Y., Qi, J., Wang, L.-F., Gao, G. F., & Liu, W. J. (2019). Peptide presentation by bat MHC class I provides new insight into the antiviral immunity of bats. *PLoS Biology*, 17, 1–24. <https://doi.org/10.1371/journal.pbio.3000436>
- Martinez-Picado, J., Prado, J. G., Fry, E. E., Pfafferott, K., Leslie, A., Chetty, S., Thobakgale, C., Honeyborne, I., Crawford, H., Matthews, P., Pillay, T., Rousseau, C., Mullins, J. I., Brander, C., Walker, B. D., Stuart, D. I., Kiepiela, P., & Goulder, P. (2006). Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. *Journal of Virology*, 80, 3617–3623. <https://doi.org/10.1128/JVI.80.7.3617-3623.2006>
- McElduff, F., Cortina-Borja, M., Chan, S. K., & Wade, A. (2010). When t-tests or Wilcoxon-Mann-Whitney tests won't do. *Advances in Physiology Education*, 34, 128–133. <https://doi.org/10.1152/advan.00017.2010>
- Ng, J. H. J., Tachedjian, M., Deakin, J., Wynne, J. W., Cui, J., Haring, V., Broz, I., Chen, H., Belov, K., Wang, L.-F., & Baker, M. L. (2016). Evolution and comparative analysis of the bat MHC-I region. *Scientific Reports*, 6, 21256. <https://doi.org/10.1038/srep21256>
- Ni, L., Ye, F., Cheng, M.-L., Feng, Y. U., Deng, Y.-Q., Zhao, H., Wei, P., Ge, J., Gou, M., Li, X., Sun, L., Cao, T., Wang, P., Zhou, C., Zhang, R., Liang, P., Guo, H., Wang, X., Qin, C.-F., ... Dong, C. (2020). Detection of SARS-CoV-2-specific humoral and cellular immunity in COVID-19 convalescent individuals. *Immunity*, 52, 971–977.e3. <https://doi.org/10.1016/j.immuni.2020.04.023>
- Okba, N. M. A., Müller, M. A., Li, W., Wang, C., GeurtsvanKessel, C. H., Corman, V. M., Lamers, M. M., Sikkema, R. S., de Bruin, E., Chandler, F. D., Yazdanpanah, Y., Le Hingrat, Q., Descamps, D., Houhou-Fidouh, N., Reusken, C. B. E. M., Bosch, B.-J., Drosten, C., Koopmans, M. P. G., & Haagmans, B. L. (2020). Severe acute respiratory syndrome coronavirus 2 – Specific antibody responses in coronavirus disease patients. *Emerging Infectious Diseases*, 26, 1478–1488. <https://doi.org/10.3201/eid2607.200841>
- Papenfuss, A. T., Baker, M. L., Feng, Z.-P., Tachedjian, M., Crameri, G., Cowled, C., Ng, J., Janardhana, V., Field, H. E., & Wang, L.-F. (2012). The immune gene repertoire of an important viral reservoir, the Australian black flying fox. *BMC Genomics*, 13, 261. <https://doi.org/10.1186/1471-2164-13-261>
- Paul, S., Sidney, J., Sette, A., & Peters, B. (2016). TepiTool: A pipeline for computational prediction of T cell epitope candidates. *Current Protocols in Immunology*, 114, 18.19.1–18.19.24. <https://doi.org/10.1002/cipm.12>
- Peng, Y., Mentzer, A. J., Liu, G., Yao, X., Yin, Z., Dong, D., Dejnirattisai, W., Rostron, T., Supasa, P., Liu, C., López-Camacho, C., Slon-Campos, J., Zhao, Y., Stuart, D. I., Paesen, G. C., Grimes, J. M., Antson, A. A., Bayfield, O. W., Hawkins, D. E. D. P., ... Dong, T. (2020). Broad and strong memory CD4 and CD8 T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. *Nature Immunology*, 21, 1336–1345. <https://doi.org/10.1038/s41590-020-0782-6>
- Petrosillo, N., Viceconte, G., Ergonul, O., Ippolito, G., & Petersen, E. (2020). COVID-19, SARS and MERS: Are they closely related? *Clinical Microbiology and Infection*, 26, 729–734. <https://doi.org/10.1016/j.cmi.2020.03.026>
- Plante, J. A., Liu, Y., Liu, J., Xia, H., Johnson, B. A., Lokugamage, K. G., Zhang, X., Muruato, A. E., Zou, J., Fontes-Garfias, C. R., Mirchandani, D., Scharton, D., Bilello, J. P., Ku, Z., An, Z., Kalveram, B., Freiberg, A. N., Menachery, V. D., Xie, X., ... Shi, P.-Y. (2020). Spike mutation D614G alters SARS-CoV-2 fitness. *Nature*. <https://doi.org/10.1038/s41586-020-2895-3>
- Poh, C. M., Carissimo, G., Wang, B., Amrun, S. N., Lee, C.-P., Chee, R.-L., Fong, S.-W., Yeo, N.-W., Lee, W.-H., Torres-Ruesta, A., Leo, Y.-S., Chen, M.-C., Tan, S.-Y., Chai, L. Y. A., Kalimuddin, S., Kheng, S. S. G., Thien, S.-Y., Young, B. E., Lye, D. C., ... Ng, L. F. P. (2020). Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. *Nature Communications*, 11, 2806. <https://doi.org/10.1038/s41467-020-16638-2>
- Rockx, B., Donaldson, E., Frieman, M., Sheahan, T., Corti, D., Lanzavecchia, A., & Baric, R. S. (2010). Escape from human monoclonal antibody neutralization affects in vitro and in vivo fitness of severe acute respiratory syndrome coronavirus. *The Journal of Infectious Diseases*, 201, 946–955. <https://doi.org/10.1086/651022>
- Sanjuán, R., Nebot, M. R., Peris, J. B., & Alcamí, J. (2013). Immune activation promotes evolutionary conservation of T-cell epitopes in HIV-1. *PLoS Biology*, 11, 1–10. <https://doi.org/10.1371/journal.pbio.1001523>
- Schneidewind, A., Brockman, M. A., Sidney, J., Wang, Y. E., Chen, H., Suscovich, T. J., Li, B., Adam, R. I., Allgaier, R. L., Mothé, B. R., Kuntzen, T., Oniangue-Ndza, C., Trocha, A., Yu, X. G., Brander, C., Sette, A., Walker, B. D., & Allen, T. M. (2008). Structural and functional constraints limit options for cytotoxic T-lymphocyte escape in the immunodominant HLA-B27-restricted epitope in human immunodeficiency virus type 1 capsid. *Journal of Virology*, 82, 5594–5605. <https://doi.org/10.1128/JVI.02356-07>
- Schneidewind, A., Brockman, M. A., Yang, R., Adam, R. I., Li, B., Le Gall, S., Rinaldo, C. R., Crags, S. L., Allgaier, R. L., Power, K. A., Kuntzen, T., Tung, C.-S., LaBute, M. X., Mueller, S. M., Harrer, T., McMichael, A. J., Goulder, P. J. R., Aiken, C., Brander, C., ... Allen, T. M. (2007). Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication. *Journal of Virology*, 81, 12382–12393. <https://doi.org/10.1128/JVI.01543-07>
- Shi, Z., & Wang, L. F. (2017). Evolution of SARS coronavirus and the relevance of modern molecular epidemiology. *Genetics and Evolution of Infectious Diseases*, 711–728. Elsevier <https://doi.org/10.1016/B978-0-12-799942-5.00026-3>
- Sironi, M., Hasnain, S. E., Rosenthal, B., Phan, T., Luciani, F., Shaw, M.-A., Sallum, M. A., Mirhashemi, M. E., Morand, S., & González-Candelas, F. (2020). SARS-CoV-2 and COVID-19: A genetic, epidemiological, and evolutionary perspective. *Infection, Genetics and Evolution*, 84, 104384. <https://doi.org/10.1016/j.meegid.2020.104384>
- St John, A. L., & Rathore, A. P. S. (2020). Early Insights into Immune Responses during COVID-19. *The Journal of Immunology*, 205, 555–564. <https://doi.org/10.4049/jimmunol.2000526>
- Su, Y. C. F., Bahl, J., Joseph, U., Butt, K. M., Peck, H. A., Koay, E. S. C., Oon, L. L. E., Barr, I. G., Vijaykrishna, D., & Smith, G. J. D. (2015). Phylogenetics of H1N1/2009 influenza reveals the transition from host adaptation to immune-driven selection. *Nature Communications*, 6, 7952. <https://doi.org/10.1038/ncomms8952>
- Tse, L. V., Klinc, K. A., Madigan, V. J., Castellanos Rivera, R. M., Wells, L. F., Havlik, L. P., Smith, J. K., Agbandje-McKenna, M., & Asokan,

- A. (2017). Structure-guided evolution of antigenically distinct adeno-associated virus variants for immune evasion. *Proceedings of the National Academy of Sciences*, 114, E4812–E4821. <https://doi.org/10.1073/pnas.1704766114>
- Vabret, N., Britton, G. J., Gruber, C., Hegde, S., Kim, J., Kuksin, M., Levantovsky, R., Malle, L., Moreira, A., Park, M. D., Pia, L., Risson, E., Saffern, M., Salomé, B., Esai Selvan, M., Spindler, M. P., Tan, J., van der Heide, V., Gregory, J. K., ... Sinai Immunology Review Project (2020). Immunology of COVID-19: Current state of the science. *Immunity*, 52, 910–941. <https://doi.org/10.1016/j.jimmuni.2020.05.002>
- van Dorp, L., Acman, M., Richard, D., Shaw, L. P., Ford, C. E., Ormond, L., Owen, C. J., Pang, J., Tan, C. C. S., Boshier, F. A. T., Ortiz, A. T., & Balloux, F. (2020). Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infection, Genetics and Evolution*, 83, 104351. <https://doi.org/10.1016/j.meegid.2020.104351>
- Wang, M., Yan, M., Xu, H., Liang, W., Kan, B., Zheng, B., Chen, H., Zheng, H., Xu, Y., Zhang, E., Wang, H., Ye, J., Li, G., Li, M., Cui, Z., Liu, Y.-F., Guo, R.-T., Liu, X.-N., Zhan, L.-H., ... Xu, J. (2005). SARS-CoV infection in a restaurant from palm civet. *Emerging Infectious Diseases*, 11, 1860–1865. <https://doi.org/10.3201/eid1112.041293>
- Weissman, D., Alameh, M. G., de Silva, T., Collini, P., Hornsby, H., Brown, R., LaBranche, C. C., Edwards, R. J., Sutherland, L., Santra, S., & Mansouri, K. (2020). D614G spike mutation increases SARS CoV-2 susceptibility to neutralization. *MedRxiv*. <https://doi.org/10.1101/2020.07.22.20159905>
- Wiehe, K., Easterhoff, D., Luo, K., Nicely, N. I., Bradley, T., Jaeger, F. H., Dennison, S. M., Zhang, R., Lloyd, K. E., Stolarchuk, C., Parks, R., Sutherland, L. L., Scarce, R. M., Morris, L., Kaewkungwal, J., Nitayaphan, S., Pitisuttithum, P., Rerks-Ngarm, S., Sinangil, F., ... Haynes, B. F. (2014). Antibody light-chain-restricted recognition of the site of immune pressure in the RV144 HIV-1 vaccine trial is phylogenetically conserved. *Immunity*, 41, 909–918. <https://doi.org/10.1016/j.jimmuni.2014.11.014>
- Wong, M. C., Javornik Cregeen, S. J., Ajami, N. J., & Petrosino, J. F. (2020). Evidence of recombination in coronaviruses implicating pangolin origins of nCoV-2019. *BioRxiv*. <https://doi.org/10.1101/2020.02.07.939207>
- Woo, P. C. Y., Lau, S. K. P., Tsui, H.-W., Huang, Y. I., Poon, R. W. S., Chu, C.-M., Lee, R. A., Luk, W.-K., Wong, G. K. M., Wong, B. H. L., Cheng, V. C. C., Tang, B. S. F., Wu, A. K. L., Yung, R. W. H., Chen, H., Guan, Y. I., Chan, K.-H., & Yuen, K.-Y. (2005). Clinical and molecular epidemiological features of coronavirus HKU1-associated community-acquired pneumonia. *The Journal of Infectious Diseases*, 192, 1898–1907. <https://doi.org/10.1086/497151>
- Wu, F., Wang, A., Liu, M., Wang, Q., Chen, J., Xia, S., Ling, Y., Zhang, Y., Xun, J., Lu, L., & Jiang, S. (2020). Neutralizing antibody responses to SARS-CoV-2 in a COVID-19 recovered patient cohort and their implications. *MedRxiv*. <https://doi.org/10.1101/2020.03.30.20047365>
- Wu, Z., & McGoogan, J. M. (2020). Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *JAMA*, 323, 1239–1242. <https://doi.org/10.1001/jama.2020.2648>
- Wynne, J. W., Woon, A. P., Dudek, N. L., Croft, N. P., Ng, J. H. J., Baker, M. L., Wang, L.-F., & Purcell, A. W. (2016). Characterization of the antigen processing machinery and endogenous peptide presentation of a bat MHC class I molecule. *The Journal of Immunology*, 196(11), 4468–4476. <https://doi.org/10.4049/jimmunol.1502062>
- Xiao, K., Zhai, J., Feng, Y., Zhou, N., Zhang, X. U., Zou, J.-J., Li, N. A., Guo, Y., Li, X., Shen, X., Zhang, Z., Shu, F., Huang, W., Li, Y. U., Zhang, Z., Chen, R.-A., Wu, Y.-J., Peng, S.-M., Huang, M., ... Shen, Y. (2020). Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature*, 583, 286–289. <https://doi.org/10.1038/s41586-020-2313-x>
- Ye, Z. W., Yuan, S., Yuen, K. S., Fung, S. Y., Chan, C. P., & Jin, D. Y. (2020). Zoonotic origins of human coronaviruses. *International Journal of Biological Sciences*, 16, 1686–1697. <https://doi.org/10.7150/ijbs.45472>
- Yurkovetskiy, L., Wang, X., Pascal, K. E., Tomkins-Tinch, C., Nyalile, T., Wang, Y., Baum, A., Diehl, W. E., Dauphin, A., Carbone, C., Veinotte, K., Egri, S. B., Schaffner, S. F., Lemieux, J. E., Munro, J., Rafique, A., Barve, A., Sabeti, P. C., Kyrtos, C. A., ... Luban, J. (2020). SARS-CoV-2 Spike protein variant D614G increases infectivity and retains sensitivity to antibodies that target the receptor binding domain. *BioRxiv*. <https://doi.org/10.1101/2020.07.04.187757>
- Zhang, B., Zhou, X., Zhu, C., Song, Y., Feng, F., Qiu, Y., Feng, J., Jia, Q., Song, Q., Zhu, B. O., & Wang, J. (2020). Immune phenotyping based on the neutrophil-to-lymphocyte ratio and IgG level predicts disease severity and outcome for patients with COVID-19. *Frontiers in Molecular Biosciences*, 7, 157. <https://doi.org/10.3389/fmbo.2020.00157>
- Zhang, L., Jackson, C. B., Mou, H., Ojha, A., Rangarajan, E. S., Izard, T., Farzan, M., & Choe, H. (2020) The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *bioRxiv*. <https://doi.org/10.1101/2020.06.12.148726>
- Zhao, J., Yuan, Q., Wang, H., Liu, W., Liao, X., Su, Y., Wang, X., Yuan, J., Li, T., Li, J., Qian, S., Hong, C., Wang, F., Liu, Y., Wang, Z., He, Q., Li, Z., He, B., Zhang, T., ... Zhang, Z. (2020). Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. *Clinical Infectious Diseases*, 71, 2027–2034. <https://doi.org/10.1093/cid/ciaa344>
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X. I., ... Shi, Z.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579, 270–273. <https://doi.org/10.1038/s41586-020-2012-7>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Forni D, Cagliani R, Pontremoli C, et al. Antigenic variation of SARS-CoV-2 in response to immune pressure. *Mol Ecol*. 2020;00:1–12. <https://doi.org/10.1111/mec.15730>