



Distinct evolutionary trajectories of SARS-CoV-2-interacting proteins in bats and primates identify important host determinants of COVID-19

Marie Cariou^a, Léa Picard^{a,b,1}, Laurent Guéguen^{b,1}, Stéphanie Jacquet^{a,b,2}, Andrea Cimarelli^{a,2}, Oliver I. Fregoso^{c,2}, Antoine Molaro^{d,2}, Vincent Navratil^{e,f,g,2}, and Lucie Etienne^{a,3}

Edited by Jean-Laurent Casanova, The Rockefeller University, New York, NY; received April 15, 2022; accepted June 17, 2022

The coronavirus disease 19 (COVID-19) pandemic is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a coronavirus that spilled over from the bat reservoir. Despite numerous clinical trials and vaccines, the burden remains immense, and the host determinants of SARS-CoV-2 susceptibility and COVID-19 severity remain largely unknown. Signatures of positive selection detected by comparative functional genetic analyses in primate and bat genomes can uncover important and specific adaptations that occurred at virus–host interfaces. We performed high-throughput evolutionary analyses of 334 SARS-CoV-2-interacting proteins to identify SARS-CoV adaptive loci and uncover functional differences between modern humans, primates, and bats. Using DGINN (Detection of Genetic INNovation), we identified 38 bat and 81 primate proteins with marks of positive selection. Seventeen genes, including the ACE2 receptor, present adaptive marks in both mammalian orders, suggesting common virus–host interfaces and past epidemics of coronaviruses shaping their genomes. Yet, 84 genes presented distinct adaptations in bats and primates. Notably, residues involved in ubiquitination and phosphorylation of the inflammatory RIPK1 have rapidly evolved in bats but not primates, suggesting different inflammation regulation versus humans. Furthermore, we discovered residues with typical virus–host arms race marks in primates, such as in the entry factor TMPRSS2 or the autophagy adaptor FYCO1, pointing to host-specific *in vivo* interfaces that may be drug targets. Finally, we found that FYCO1 sites under adaptation in primates are those associated with severe COVID-19, supporting their importance in pathogenesis and replication. Overall, we identified adaptations involved in SARS-CoV-2 infection in bats and primates, enlightening modern genetic determinants of virus susceptibility and severity.

SARS-CoV-2 and COVID-19 | primates and bats | comparative genetics | virus–host coevolution | positive selection

The current coronavirus disease 19 (COVID-19) pandemic already led to over six million human deaths (WHO April 2022). The causative agent is severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that originated from viral cross-species transmission from the bat reservoir, directly or through an intermediate host, to humans (1). Bats naturally host some of the most high-profile zoonotic viruses, including SARS-CoVs, without apparent symptoms (2). Despite prevention measures and effective vaccines, the burden from COVID-19 remains immense in humans, and the determinants of SARS-CoV-2 susceptibility and COVID-19 severity remain largely unknown. A powerful way to identify these factors is to use comparative functional genomics to map host–virus interfaces that underlie infections in the bat reservoir and the primate host (3).

During infection, viruses interact with many host proteins, or viral-interacting proteins (VIPs). While some VIPs are usurped for viral replication away from their normal host functions, some are specifically targeting the virus as part of the host antiviral immune defense. Since the emergence of SARS-CoV-2, VIPs have been identified in hundreds of screens using *in vitro* approaches, such as CRISPR/knockout screens, complementary DNA library screens, or mass spectrometry analyses (4, 5). However, the *in vivo* importance of the identified SARS-CoV-2 VIPs remains largely unknown.

From an evolutionary standpoint, the fitness cost imposed by pathogenetic viruses triggers strong selective pressures on VIPs, such that those VIPs able to prevent, or better counteract, viral infection will quickly become fixed in host populations. In turn, host adaptations push viral proteins into recurring counter adaptation cycles creating stereotypical virus–host molecular arms races. These arms races are witnessed by signatures of accelerated rates of evolution, or positive selection, over functionally important

Significance

The coronavirus disease 19 (COVID-19) pandemic is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a coronavirus that spilled over from the bat reservoir. However, the host genetic determinants that drive SARS-CoV-2 susceptibility and COVID-19 severity are largely unknown. Understanding how cellular proteins interacting with SARS-CoV-2 have evolved in primates and bats is of primary importance to decipher differences in the infection outcome between humans and the viral reservoir in bats. Here, we performed comparative functional genetic analyses of hundreds of SARS-CoV-2-interacting proteins to study virus–host interface adaptation over millions of years, pointing to genes similarly—or differentially—engaged in evolutionary arms races and that may be at the basis of *in vivo* pathogenic differences.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹L.P. and L.G. contributed equally to this work.

²S.J., A.C., O.I.F., A.M., and V.N. contributed equally to this work.

³To whom correspondence may be addressed. Email: lucie.etienne@ens-lyon.fr.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2206610119/-/DCSupplemental>.

Published August 10, 2022.

residues and domains in VIPs (6–8). Thus, when combined with functional data, identifying the VIPs with signatures of positive selection is a powerful way to discover virus–host interfaces (e.g., refs. 9–11).

When studies of adaptive signatures in host genes are combined with human clinical studies or genome-wide association studies (GWAS), they are powerful to uncover the importance of gene evolution and variants in disease severity (e.g., refs. 12, 13). Interestingly, several genetic loci associated with COVID-19 severity and susceptibility in humans, such as OAS1 (2'-5'-Oligoadenylate Synthetase 1) or those from the interferon signaling pathway (12, 14–19), bear hallmarks of such adaptive arms races. Furthermore, dozens of VIPs that bear marks of adaptive evolution in the human lineage from ancient SARS-CoV epidemics may be important host determinants of SARS-CoV-2 (20).

Here, we aimed to identify key SARS-CoV adaptive loci and functional genomic differences between bats, which include the natural reservoir of SARS-CoVs, and primates, including humans. We performed high-throughput evolutionary and positive selection screens of 334 SARS-CoV-2-interacting proteins (4) using the Detection of Genetic INNovation (DGINN) pipeline (21), followed by comprehensive (phylo)genetic analyses of seven VIPs of interest. We provide the results in the searchable VirHostNet 2.0 web portal. Using this approach, we identified 38 bat and 81 primate genes with strong evidence of positive selection. Of these, we found 17 proteins, including the ACE2 receptor, subjected to adaptative evolution in both clades, 1) confirming that past SARS-CoV epidemics occurred during both bat and primate evolution, and 2) identifying core VIPs that shaped universal SARS-CoV–host molecular arms races. We also identified 84 VIPs with lineage-specific adaptations that likely contributed to SARS-CoV pathogenicity in different mammalian hosts. Among these, we uncover the important role of several genes, including TMPRSS2, FYCO1 (FYVE and coiled-coil domain containing 1), or RIPK1 that play important roles in entry, trafficking, or inflammatory responses, respectively. We hypothesize that these past adaptation events in bats and primates underlie differences in susceptibility to SARS-CoV-2 infections and key determinants in COVID-19 severity in modern humans.

Results

Characterization of the Evolutionary History of SARS-CoV-2 VIPs in Bats and Primates. Because pathogenic viruses and hosts are engaged in evolutionary arms races, adaptive signatures accumulate in VIP genes as a result of past epidemics (6, 8). Adaptive evolution can be identified by positive selection analyses over a set of protein-coding orthologs when their rate of nonsynonymous codon substitutions (dN) exceeds that of synonymous ones (dS) (22). To identify the proteins with such signatures of adaptive evolution, we studied the evolutionary history of the SARS-CoV-2 interactome identified in *in vitro* experiments. Furthermore, to discover key SARS-CoV-2–host determinants of replication and pathogenesis, we aimed to identify the common and different evolutions and genetics of the VIPs in the human versus the reservoir host. We therefore performed comparative phylogenetics of the VIPs in primates and bats. Specifically, we studied the 332 host proteins identified by Gordon et al. (4) in mass spectrometry assays of SARS-CoV-2 proteins in human cells, in addition to the angiotensin converting enzyme 2 (ACE2) receptor and the transmembrane protease serine 2 (TMPRSS2), which are both necessary for virus entry into the cells.

To perform the phylogenetic and positive selection analyses, we used the DGINN bioinformatic pipeline (21) that entirely automates the analyses and combines several methods to test for selection across large datasets (Fig. 1*A*). Briefly, from each of the 334 human reference gene sequences, DGINN automatically retrieved bat and primate homologs (from National Center for Biotechnology Information [NCBI] nonredundant database), curated the coding sequences, and performed a codon-alignment followed by a gene phylogenetic reconstruction (Fig. 1*A* and *SI Appendix, Table S1*). The pipeline then screened for duplication events and identified orthologs and potential paralogs, as well as recombination events. This mainly allows correct phylogenetic and positive selection analyses of VIPs from gene families, and with recombination events. Finally, each aligned set of orthologs was used to measure rates of codon substitutions and to estimate whether the whole gene, as well as any codon, are evolving under positive selection. For this, DGINN uses a combination of methods from the following selection tools: Hypothesis Testing using Phylogenies HYPHY (Branch-Site Unrestricted Statistical Test for Episodic Diversification [BUSTED]), and Mixed Effects Model of Evolution [MEME]), Phylogenetic Analysis by Maximum Likelihood PAML (codeml M0, M1, M2, M7, M8, and associated Bayesian Empirical Bayes [BEB] for codon-specific analyses), and Bio++ (M0^{NS}, M1^{NS}, M2^{NS}, M7^{NS}, M8^{NS}, and associated Posterior Probabilities [PP] for codon-specific analyses) (Fig. 1*A*, *Methods* for details) (21, 23–25).

We found that the DGINN pipeline, previously validated on 19 primate genes (21), was efficient at screening hundreds of genes and at analyzing other mammalian orders (here, chiroptera) (*SI Appendix, Figs. S1 and S2*). Overall, our bioinformatic screen allowed us to obtain the bat and primate evolutionary history of 324 common SARS-CoV-2 VIP genes (i.e., 330 in bats and 329 in primates). We compiled the resulting sequence alignments, phylogenetic trees, and gene- and site-specific positive selection results to an open-access and searchable web application (<https://virhostnet.prabi.fr/virhostevol/>), which constitutes a public resource to visualize and download the evolution of SARS-CoV-2 VIPs in primates and bats.

Identification and Comparative Analysis of SARS-CoV-2 VIPs with Signatures of Positive Selection during Bat and Primate Evolution. To characterize the overall trend in the evolution of each VIP in primates and in bats, we compared their omega parameter, which is positively correlated with the strength of the natural selection acting on the protein (Fig. 1*B*). We found a similar trend in the natural selection of bat and primate genes; those with an elevated omega in primates had an overall rapid protein evolution in bats too. Beyond this trend, we cannot compare the omega values quantitatively between the two mammalian orders and the reasons include, for example, differences in the number of analyzed species (i.e., 12 and 24 median number of species in bats and primates, respectively), the population sizes, and the genetic distances.

We next identified the genes with evidence of positive selection by at least three methods in the DGINN screen. In bats, we found 38 genes, roughly 12% of SARS-CoV-2-interacting proteins, with signatures of positive selection (Fig. 1*C*). These include the ACE2 receptor, also reported by others as under strong positive selection in bats (26, 27). In primates, we identified 81 genes under positive selection, after discarding 7 due to low-quality alignments and inclusion of erroneous sequences in the automatic steps (Fig. 1*D* legend).

In the case of primate analyses, we identified more VIPs under positive selection than Gordon et al. (4), in which they

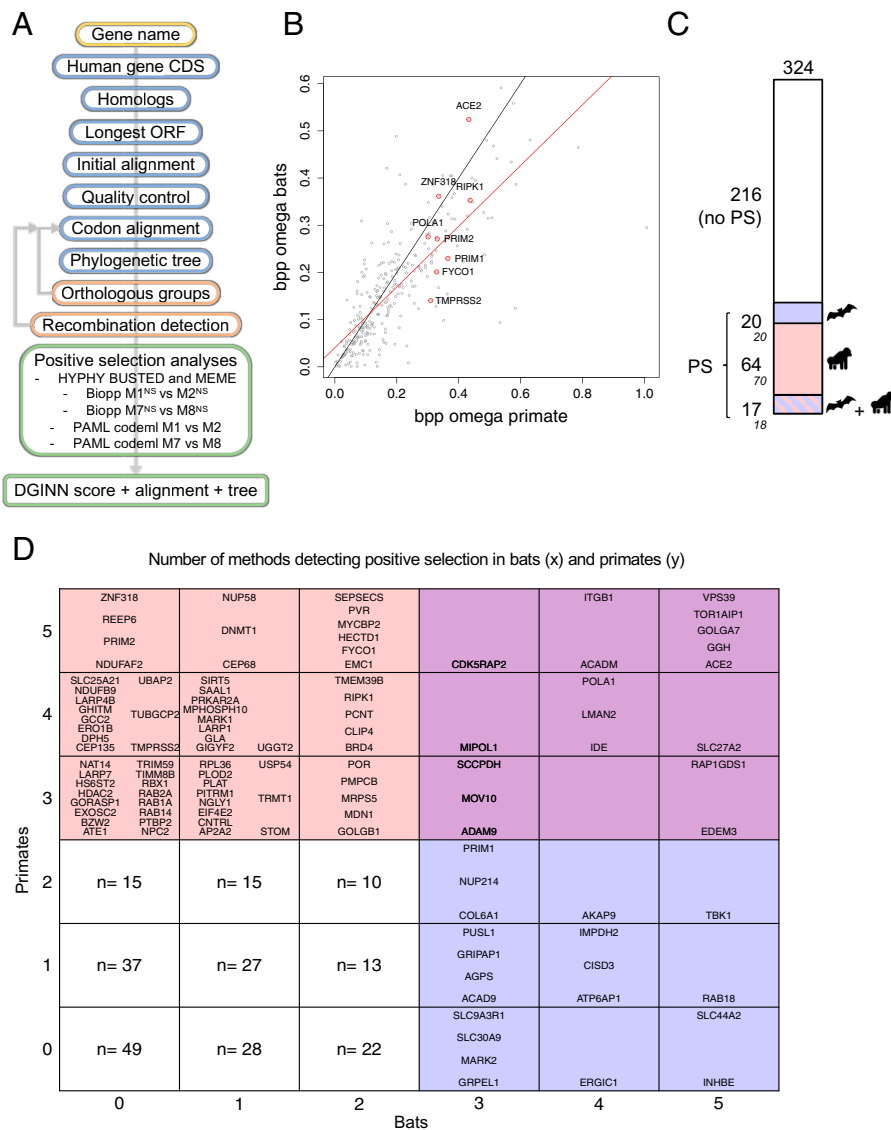


Fig. 1. Identification of the SARS-CoV-2 interactome with signatures of positive selection (PS) in bats and primates. (A) Overview of the DGINN pipeline to detect adaptive evolution in SARS-CoV-2 VIPs. CDS, coding DNA sequence; ORF, open reading frame. (B) Natural selection acting on bat and primate VIP genes. Comparison of omega (dN/dS) values of the VIPs during bat (y axis) and primate (x axis) evolution, estimated by Bio++ Model M0. In black, the bisector. In red, the linear regression. The names correspond to genes that we comprehensively analyzed (Table 1). (C) Overview of the number of VIPs under significant PS (i.e., by at least three methods in the DGINN screen) in bats and/or primates. A total of 324 genes could be fully analyzed in the two mammalian orders. Numbers represent the number of genes in the categories: No PS or PS, within each host, is represented by a pictogram. The numbers correspond to the conservative values after visual inspection of the positively selected VIP alignments, while the italic numbers are from the automated screen. (D) Table showing the genes identified by x,y DGINN methods in bats and primates, respectively. For the genes with low DGINN scores (<3), only the number of genes in each category is shown (*SI Appendix, Fig. S4* for details). Of note, seven primate genes are false positive, as follows: EMC1 (ER membrane protein complex subunit 1), MOV10 (Mov10 RISC complex RNA helicase), POR (cytochrome p450 oxidoreductase), PITRM1 (pitrilysin metalloproteinase 1), RAB14, RAB2A, and TIMM8B (translocase of inner mitochondrial membrane 8 homolog B).

identified 40/332 genes under positive selection in primates using a codeml M8 vs. M8a model. One example is the Zinc finger protein ZNF318 that has some marks of positive selection during primate evolution in our analyses (*SI Appendix*). However, the overall dN/dS estimate for each gene was highly similar between the two studies (*SI Appendix, Fig. S3A*), and we detected most of the genes they identified under positive selection, namely, 38/40 VIPs under positive selection in Gordon et al. (4) were also detected by ≥ 1 DGINN method, including 28 by ≥ 3 DGINN methods (*SI Appendix, Fig. S3 B and C*). Thus, the main advantages of DGINN were the end-to-end automatic pipeline and the combination of multiple methods, thereby increasing sensitivity and specificity of positive selection analyses in screening approaches.

Altogether, we found 81 primate VIPs and 38 bat VIPs with evidence of positive selection (Fig. 1 C and D). Beyond Gordon et al. (4), other SARS-CoV-2 in vitro and clinical studies also identified many of these positively selected genes as SARS-CoV-2 VIPs, thus confirming their suspected role as SARS-CoV-2 regulators or interacting proteins (*SI Appendix, Fig. S5*) (5). Gene Ontology (GO)-term enrichment analyses with GO enrichment analysis and visualization tool (GORilla) (28) did not show major differences between genes with and without evidence of positive selection (*Dataset S1*). Analyses of pathway

enrichment over the entire genome using Reactome showed that positively selected VIPs are associated with cell cycle control and centrosome behavior biological pathways (*SI Appendix, Fig. S6*), suggesting that the control of cell division and perhaps centrosome-regulated cell polarization are important for SARS-CoV in vivo. Analyzing the expression patterns of the VIPs from 29 human tissues, we found that the mean expression of the genes with and without evidence of positive selection was overall similar (*SI Appendix, Fig. S7A*). Importantly, the vast majority of the genes with evidence of positive selection are expressed in lungs at control conditions (*SI Appendix, Fig. S7 B and C*).

We found 17 rapidly evolving genes shared between bats and primates, corresponding to 16% of all SARS-CoV-2 VIPs with evidence of positive selection (i.e., 17 genes in common over 108 in total) (Fig. 1 C and D). This list notably includes the ACE2 receptor of SARS-CoVs that has undergone positive selection in both primates and bats (Fig. 1 B and D). It also includes known drug targets, such as the metalloprotease ADAM9 (29), the ITGB1 integrin (30), and POLA1 from the Prim-Pol primase complex (31) (*SI Appendix*) (Fig. 1D). Therefore, these genes may represent the core SARS-CoV VIPs that have been subjected to positive selection pressure during both primate and bat evolution.

However, we also identified 84 genes that have evolved through distinct selective pressures during primate and bat

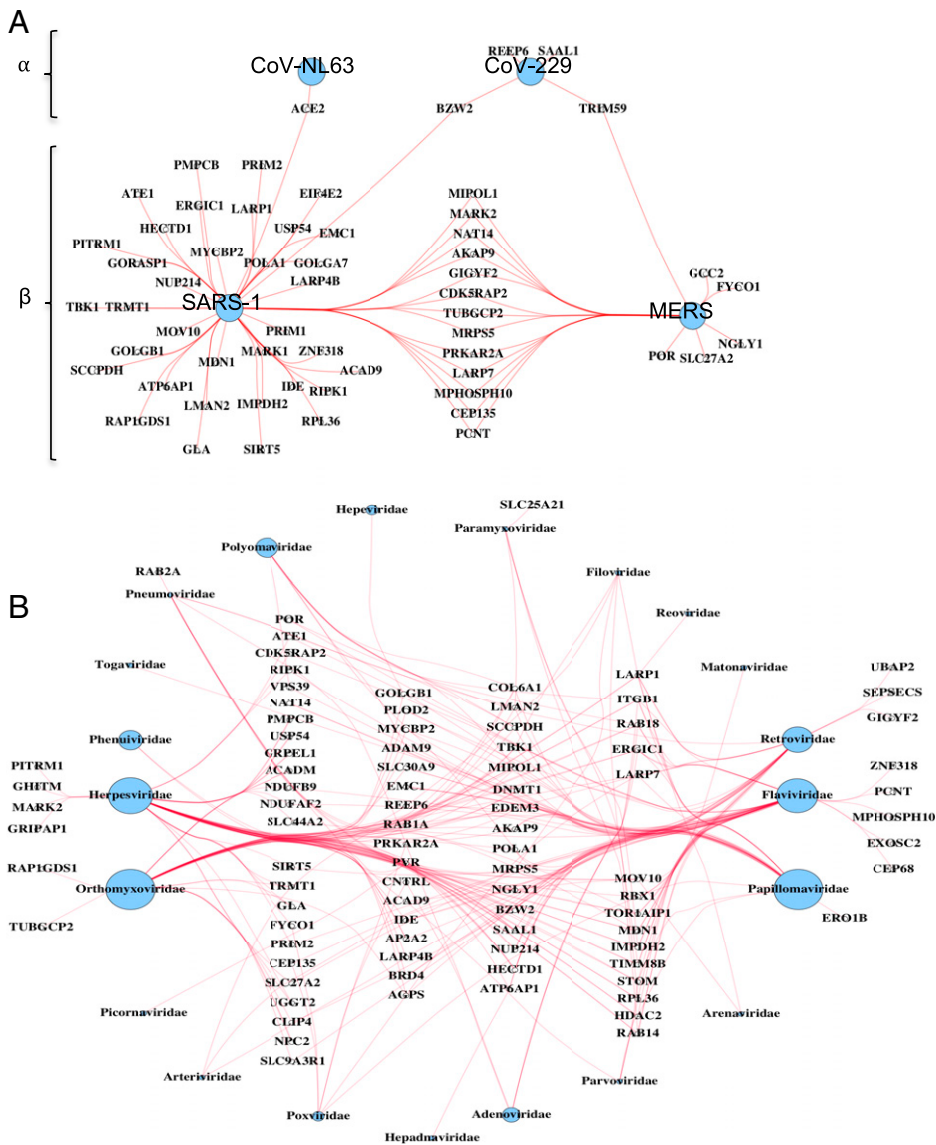


Fig. 2. SARS-CoV-2 VIPs under PS are interacting proteins of other coronaviruses, as well as other viral families. Virus-host protein-protein interaction network of VIP genes under PS and interconnected with (A) other coronaviruses (from alpha- or beta-coronavirus genus), and (B) viral families other than coronaviruses. VIPs interacting with more than one additional viral family are in the *Center* and arranged in columns (from *Left to Right*, interconnected with 2 to 6 different viral families). Node sizes at the virus families are proportional to the number of edges. The VIPs not interconnected are shown in [SI Appendix, Table S1](#).

evolution—being under positive selection only in primates (64 VIPs) or bats (20 VIPs) (Fig. 1 *C* and *D*)—including TMPRSS2, FYCO1, RIPK1, ZNF318 and the Prim-Pol primase complex ([SI Appendix](#)) that we will focus on. These genes represent VIPs with different evolutionary trajectories in bats and primates.

Several SARS-CoV-2 VIPs under Positive Selection Are VIPs of Other Coronaviruses and May Also Be Interconnected with Other Viral Families. To investigate whether the SARS-CoV-2 VIPs under positive selection are also known to interact with other coronaviruses, we interrogated the VirHostNet database (32) for interconnection with SARS-1, Middle East respiratory syndrome (MERS) (beta coronaviruses), CoV-NL63 and CoV-229 (alpha coronaviruses). We found 58 genes (i.e., 54% of 108 genes under positive selection in bats or primates) that are adaptive SARS-CoV-2 VIPs and also known interacting proteins of at least another coronavirus (Fig. 2*A*). The positive selection marks in these VIPs therefore likely represent adaptations on host proteins that have regulated or interacted with coronaviruses over million years of coevolution with mammals. These coronavirus VIPs therefore represent an evolutionarily common set of coronavirus-interacting proteins.

Because positive selection may be driven by several viruses (33), we similarly investigated whether rapidly evolving SARS-CoV-2 VIPs were also functionally linked to other viral families (Fig. 2*B*). We found that 82% of them (89 of 108 genes under positive selection in bats or primates) interconnected with one or more additional viral families beside coronaviruses. A number of proteins, including LARP1 and LARP7, ITGB1, Rab18, and ERGIC1, interconnected with six distinct viral families, highlighting their likely involvement as broad cofactors of viral replication (Fig. 2*B*). On the other hand, several genes, such as FYCO1, ZNF318, or TMPRSS2, are interconnected with only 1 to 2 other viral families and may therefore represent more specialized VIPs (Fig. 2*B*). Of note, although the TMPRSS2 coentry factor has no other interactor in this analysis (Fig. 2*B* and [SI Appendix, Table S2](#)), it is a host factor for influenza virus entry (34, 35). Lastly, the ACE2 receptor and other genes ([SI Appendix, Table S2](#)) were not known to interact with other viruses and therefore likely represent coronavirus-specific VIPs ([SI Appendix, Table S2](#)).

The SARS-CoV-2 Predicted Interface in TMPRSS2 Has Evolved under Adaptive Evolution in Primates But Not in Bats. Although the intrinsic role of TMPRSS2 in the cell is poorly

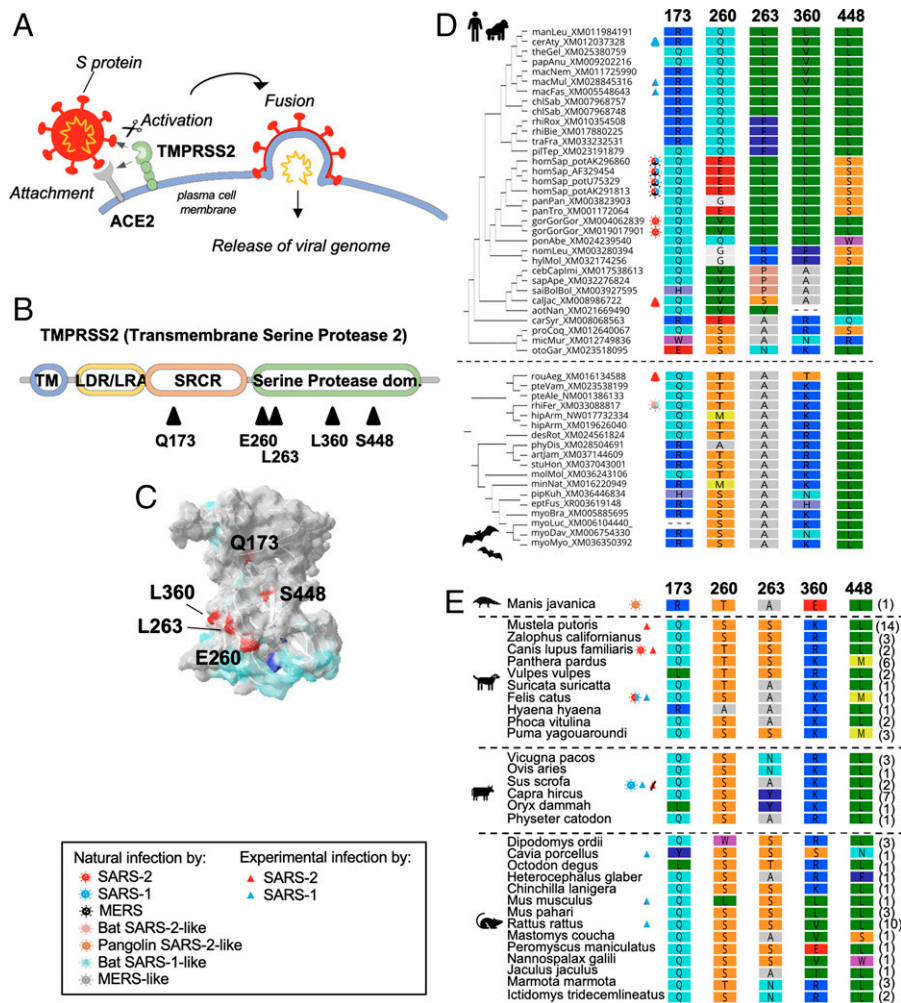


Fig. 3. TMPRSS2 has evolved under strong PS in primates but not in bats. (A) Role of TMPRSS2 in SARS-CoV-2 entry. (B) Diagram of TMPRSS2 predicted domains, with sites under PS in primates represented by triangles (Table 1). Codon numbering and amino acid residue based on *Homo sapiens* TMPRSS2. (C) 3D structure modeling of human TMPRSS2 (amino acids 1 to 492) with the positively selected sites (red), the SARS-CoV-2 predicted interface (light blue), and the catalytic site (dark blue). (D) The positively selected sites identified in primate TMPRSS2 are highly variable in primates (Top) but more conserved in bats (Bottom) where they are not identified as under adaptive evolution. Left, cladograms of primate and bat TMPRSS2 with species abbreviation and accession number of sequences. Amino acid color-coding, RasMol properties (Geneious, Biomatters). Icon legend is embedded in the figure, with multicolored pictograms/triangles showing cases fulfilling multiple conditions. (E) Positively selected sites in primates exhibit different patterns of variability in other mammals, as follows: pangolin, carnivores, artiodactyls, and rodents. Right, numbers in brackets correspond to the number of species within the order with the same TMPRSS2 haplotype at these positions (e.g., the QSSKL motif in *Mustela putorius* was found in 14 rodent species). The corresponding motif in species/cells susceptible or permissive to coronaviruses is shown in *SI Appendix*, Fig. S8.

known, this serine protease is a key factor for the cellular entry of SARS-CoV-2. TMPRSS2 is responsible for the priming of the viral spike S protein, an essential step for the ACE2 receptor recognition and the plasma cell membrane fusion process (Fig. 3A) (36, 37). In addition to SARS-CoV-2, other coronaviruses, including HCoV-229E, MERS-CoV, and SARS-CoV-1, enter human cells in a TMPRSS2-dependent manner (38–40). While the genetic and functional adaptations of ACE2 have been studied (26, 27), the genetic diversification of mammalian TMPRSS2 is currently unknown. Our screen identified positive selection in TMPRSS2 in primates, but not bats, indicating that its functional diversification is specific to coronavirus adaptation in primates.

To validate the screen results and further characterize TMPRSS2 evolution in both orders, we obtained sequences from additional primate and bat species that were not included in the automated DGINN screen. We therefore obtained two new high-quality codon alignments of TMPRSS2 from 18 bat species and from 33 primate species (<https://virhostnet.prabi.fr/virhostevol/>, Genes of focus; Table 1; *SI Appendix*, Table S3). From these comprehensive alignments, we first confirmed that TMPRSS2 has experienced significant and strong positive selection during primate evolution (Bio++ and PAML codeml M1 vs. M2 *P* values of 0.0095 and $< 4.27 \cdot 10^{-06}$, respectively). This was in contrast to its evolution in bats, in which we did not find evidence of selective pressure (Bio++ and codeml M1 vs. M2 *P* values of 1) (Table 1 and *SI Appendix*, Table S3).

To identify the precise residues that have diversified during primate evolution, we performed site-specific positive selection analyses. We identified five residues (173, 260, 263, 360, and 448, numbering from the human TMPRSS2 sequence) that were significantly detected under positive selection by at least two independent methods (Table 1, *SI Appendix*, Table S3, and Fig. 3B). Of note, position 197, which is polymorphic in human TMPRSS2 (rs12329760, V197M) and may be associated with COVID-19 severity (41) (*P* value around 10^{-5} , above the 10^{-8} significance threshold commonly used in GWAS multiple testing), encoded for a conserved valine in all nonhuman primate sequences. The SARS-CoV-2–TMPRSS2 interface is currently unknown. Only *in silico* molecular docking studies have predicted the SARS-CoV-2 binding region on TMPRSS2 (37, 42–44). Remarkably, the sites under positive selection cluster nearby or within the predicted SARS-CoV-2–host interface (Fig. 3C), suggesting that SARS-CoVs played a significant role in TMPRSS2 diversification. These regions of TMPRSS2 are also the target of several drugs, such as α 1-antitrypsin, Camostat mesylate, Nafamostat and Bromhexine hydrochloride inhibitors (37, 45, 46), and newly reported *N*-0385 (47) and could therefore be prioritized in functional studies.

Finally, by analyzing the physicochemical nature of the positively selected sites, we found that they encode for residues with very different properties, which would significantly impact the TMPRSS2 protein structure over primate evolution and lead to species specificity at the virus–host interface. In particular, variation at key residues 260 and 448 was particularly high

Table 1. Results from the comprehensive PS analyses of the genes of interest

Gene	Seq alignment info			Identified under PS by x/7 methods	MEME ($P < 0.05$)	FUBAR ($PP > 0.9$)	Bio++				codeml		aBSREL ($P < 0.1$)	PSS aln	PSS in human ref	
	Order	Size	n sp.				M2 PS ω	M2 PSS	M8 PS ω	M8 PSS	M2 PSS	M8 PSS				
FYCO1	bats	1481	18	1	375, 504, 566, 688, 790, 1059, 1092	—	—	—	—	—	—	607	—	—	—	
FYCO1	primates	1500	29	6	355, 416, 472 , 484, 601, 629, 728, 869, 919, 1102, 1218, 1219, 1242, 1259, 1267, 1407	472	4.42	448, 553	1.73	$n = 206^*$	—	448, 930	rhiBie	448, 472, 553, 930	R447, R471, M552, C928	
POLA1	bats	1484	17	7	147, 246, 249 , 317, 1314, 1449	109, 196, 246, 249, 250, 276, 284, 315, 1080 , 1270	59.91	—	2.05	$n = 186^*$	249, 296, 1080	246, 249, 250, 284, 296, 315, 1080	phyDis, pteAle	246, 249, 250, 284, 296, 315, 1080	V235, E239, E240, W273, Q285, V304, V1069	
POLA1	primates	1451	25	5	590, 707, 718, 1082	817	—	—	14.36	221	—	221, 226, 227, 817, 1058	N	221, 817	D232, N828	
PRIM1	bats	421	17	3	—	257, 258, 275, 278, 289, 291, 292	—	—	—	—	—	289, 291, 292	N	289, 291, 292	Y288, P290, W291	
PRIM1	primates	420	26	0	277, 361	—	—	—	—	—	—	—	—	—	—	
PRIM2	bats	517	16	6	5, 11, 20, 71, 168, 176	11, 69, 81, 168, 176, 476	23.14	186	12.34	186, 187	186	186	na	11, 168, 176, 186	L11, V167, L173, K177	
PRIM2	bats	517	14	6	11, 20, 71, 168, 176	11, 69, 168, 176, 185, 476	10.31	186	2.51	$n = 69^*$	186	186	myoBra, myotis	11, 168, 176, 186	—	
PRIM2	primates	513	20	3	5, 8, 73, 278	—	—	—	—	—	8, 10	8, 10, 76, 173, 176, 178	—	—	—	
RIPK1	bats	669	18	6	127, 230, 282 , 289, 370, 400, 480, 497, 668	294, 370	54.40	—	1.61	$n = 87^*$	—	294, 370, 662, 665	myoDav	294, 370, 662, 665	K284, S296, P372, S664, Y667	
RIPK1	bats	669	15	7	127, 230, 282, 370, 400, 480, 668	16, 277, 282, 294, 370	70.02	—	10.77	—	370, 665	370, 665	myoDav	282, 370, 665	—	
RIPK1	primates	671	29	1	491, 493, 592	—	—	—	—	—	—	664	—	—	—	
TMPRSS2	bats	496	17	2	19, 136, 249, 364, 412, 416	—	—	—	1.29	—	49, 216, 364, 413, 416, 435	—	—	—	—	
TMPRSS2	primates	563	28	5	168, 232, 299, 312, 316, 424, 467, 500	117, 224, 312, 412, 500	5.70	224, 412	5.19	—	224, 312, 315, 412, 500	224, 412	224, 312, 315, 412, 500	N	224, 312, 315, 412, 500	Q173, E260, L263, L360, S448
ZNF318	bats	2142	16	1	127, 586, 851, 1647, 1665, 2015, 2016	—	—	—	na	na	—	—	—	—	—	
ZNF318	bats	2395	11	1	78, 370, 830, 1107, 1588, 2153, 2267	—	—	—	—	—	—	—	—	—	—	
ZNF318	primates	2108	29	3	93, 843, 1302 , 1436, 1449, 1884, 2095	1302, 1878	—	—	—	—	1302	1302, 1589, 1733	N	1302	V1481	
ZNF318	primates	2496	24	6	38, 75, 104, 174, 1372, 1636, 1814, 1929, 2166	1636, 1817, 2098, 2243	501.94	—	1.45	$n = 68^*$	1636, 2098	1636, 1817, 1929, 1954, 2098, 2243	rhiBie, micMur, homSap	1636, 1817, 1929, 2098, 2243	—	

For each gene, are presented the results of the comprehensive phylogenetic and PS analyses, as follows: BUSTED, MEME, FUBAR, aBSREL from HYPHY/Datamonkey.com, M1vsM2, M7vsM8, M8avsM8 from Bio++, and M1vsM2, M7vsM8, M8avsM8 from PAML codeml. The genes identified under PS are highlighted in gray. The sites considered under PS after the analyses are in "PSS aln" and "PSS in human ref", corresponding to the site number in the codon alignment and the corresponding amino acid site in the human reference sequence. Alignments, trees, and interactive table are available at <https://virhostnet.prabi.fr/virhostevol/>. Table with extended results including statistical P values for each test is in *SI Appendix, Table S3*. Legend details: size, length of the codon alignment; n. sp., number of species included in the alignment; PSS, PS sites; the cutoff for each method is given in the table; PS omega, corresponds to the omega value in the PS class ($dN/dS > 1$) of the given model M2 or M8. ZNF318 and the proteins from the Primase complex are in *SI Appendix* and in *SI Appendix, Figs. S11 and S12*, respectively. For aBSREL, the branch identified under PS is given by the DGINN nomenclature (three letters from the genus and three from the species). na, not available.

*For Bio++ M8 PSS analyses there were dozens of sites under PS due to the low omega value in the class $w > 1$.

in Hominoids but low in Old World monkeys (Fig. 3D), suggesting lineage-specific adaptations within primates. To determine whether this domain of TMPRSS2 has been rapidly evolving in other mammals, we extended our analyses by retrieving other mammalian sequences. We found that most of these sites were overall conserved, except in rodents that exhibited high variability at positions 263, 360 and 448 (Fig. 3E). In bats, although none of the models identified significant positive

selection in TMPRSS2, the sites 260 and 360 were also variable (Fig. 3D). Comparing the variability between resistant and susceptible (naturally or experimentally) species to SARS-CoVs and MERS-CoVs did not reveal any clear pattern (Fig. 3E). However, the location and extreme variability of the positively selected sites appear lineage specific across mammals (with high amino acid toggling in some clades and conservation in others) and suggest that these residues, combined

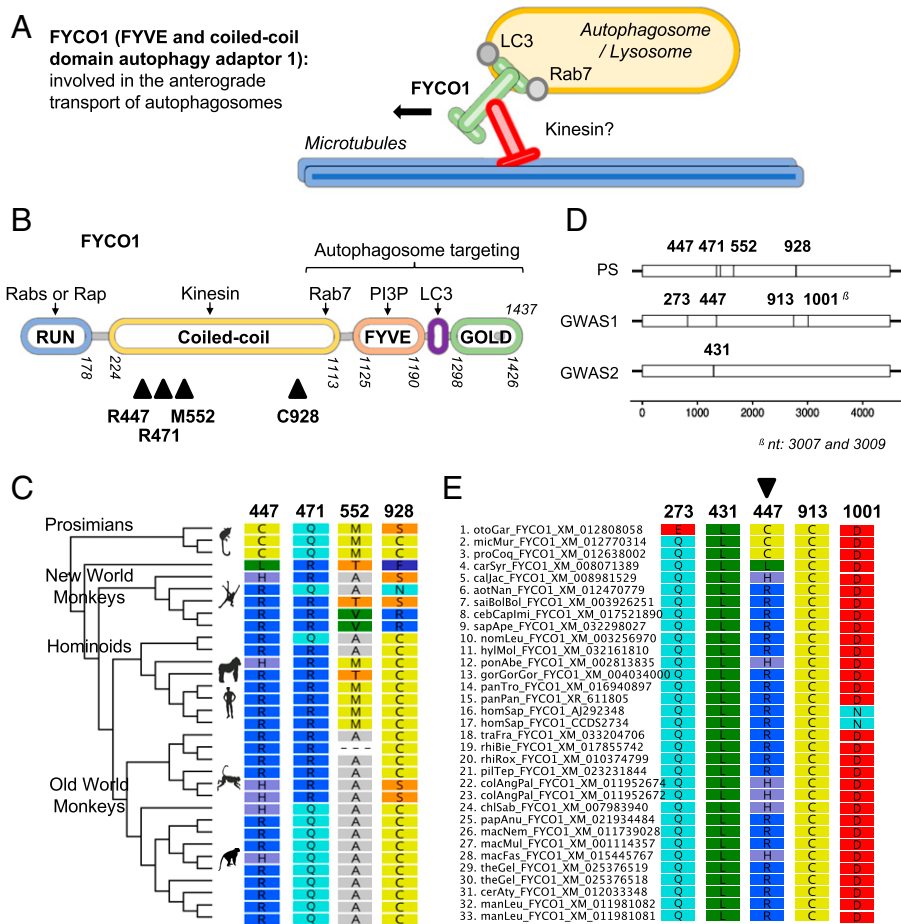


Fig. 4. Domains of FYCO1 that are associated with severe COVID-19 in human have also evolved under significant PS in primates but not in bats. (A) Known cellular role of FYCO1. (B) Diagram of FYCO1 predicted domains, with sites under PS in primates represented by triangles (Table 1). Codon numbering and amino acid residue based on *Homo sapiens* FYCO1. (C) Amino acid variation at the positively selected sites in primates. *Left*, cladogram of primate FYCO1 with major clades highlighted. The exact species and accession number of sequences are shown in *E*. Amino acid color-coding, RasMol properties (Geneious, Biomatters). (D) Sites identified in the coding sequence of FYCO1 as under PS in primates (*Top*) and as associated with severe COVID-19 in human from two GWAS studies (*Middle*: GWAS1, COVID-19 Host Genetics Initiative, 2021; *Bottom*: GWAS2, Pairo-Castineira et al., 2020). x axis, nucleotide numbering. (E) Amino acid variations in primate species at the sites associated with severe COVID-19 in GWAS.

with ACE2 receptor variability, may contribute to SARS-CoV susceptibility and species specificity.

Altogether, our findings support that the positive selection signatures in TMPRSS2 are reminiscent of ancient SARS-CoV-driven selective pressures during primate evolution. Mutagenesis studies of TMPRSS2, guided by the evolutionary analyses, are now required to identify the exact and relevant SARS-CoV determinants, as well as the functional implication of the inter-species variability in TMPRSS2.

Evidence That FYCO1 Is Involved in SARS-CoV Pathogenesis or Replication at Different Time Scales during Primate Evolution.

FYCO1 is involved in microtubule transport and autophagy (Fig. 4A). Autophagy is an important degradation process of cytoplasmic proteins and organelles, which may be dysregulated during aging, by diseases, and by pathogens. FYCO1 acts as an adaptor protein allowing the microtubule transport of autophagosomes in a STK4-LC3B-FYCO1 axis (48, 49). Mutations of the human *FYCO1* gene cause autosomal-recessive congenital cataract, a major cause of vision dysfunction and blindness (50, 51). Until the COVID-19 pandemic, there was no report of FYCO1 involvement in viral infection. However, FYCO1 is among the very few genes identified in human GWAS to be significantly associated with severe COVID-19 (17, 52, 53). GWAS correlates natural genetic variants in human populations to phenotypic traits, here, COVID-19 severity. Therefore, genes identified in GWAS may directly be involved in SARS-CoV-2 replication or pathogenesis. Furthermore, FYCO1 had a high matching-adjusted indirect comparison (MAIC) score (*SI Appendix*, Fig. S5B) (5), indicating that several studies

suspect its involvement in SARS-CoV-2 pathogenesis or replication, including Gordon and colleagues (4) that identified a human FYCO1 interaction with SARS-CoV-2 NSP13.

As for TMPRSS2, the DGINN screen identified signatures of positive selection in primate FYCO1 but not in bat FYCO1. We then retrieved all FYCO1 sequences available for primates (29 species) and bats (18 species) and performed comprehensive phylogenetic and positive selection analyses. This comprehensive positive selection analysis confirmed that FYCO1 has undergone positive selection in primates but not in bats (Table 1 and *SI Appendix*, Table S3).

Site-specific selection analyses identified the following four residues with strong evidence of significant positive selection in primates in at least two independent methods: 447, 471, 552, and 928 (Fig. 4B and C and Table 1 and *SI Appendix*, Table S3). Although no crystal structure is available for full-length FYCO1, these rapidly evolving sites fall into the coiled-coil domain of FYCO1, which is important for interaction with Kinesin. In addition, the different primate species encode for amino acids with very different physicochemical properties at these sites (Fig. 4C), indicating potential structural and functional plasticity in this region. These positive selection marks may therefore represent virus–host interplays and be the result of selective pressure by ancient epidemics during primate evolution.

To correlate primate natural genetic variants with ongoing human polymorphisms and association with COVID-19 severity, we compared FYCO1 variations in primates with the human polymorphisms associated with increased SARS-CoV-2 pathogenicity (GWAS). Using the COVID-19 Host Genetics Initiative data (<https://www.covid19hg.org/results/r6/>) as well as the data

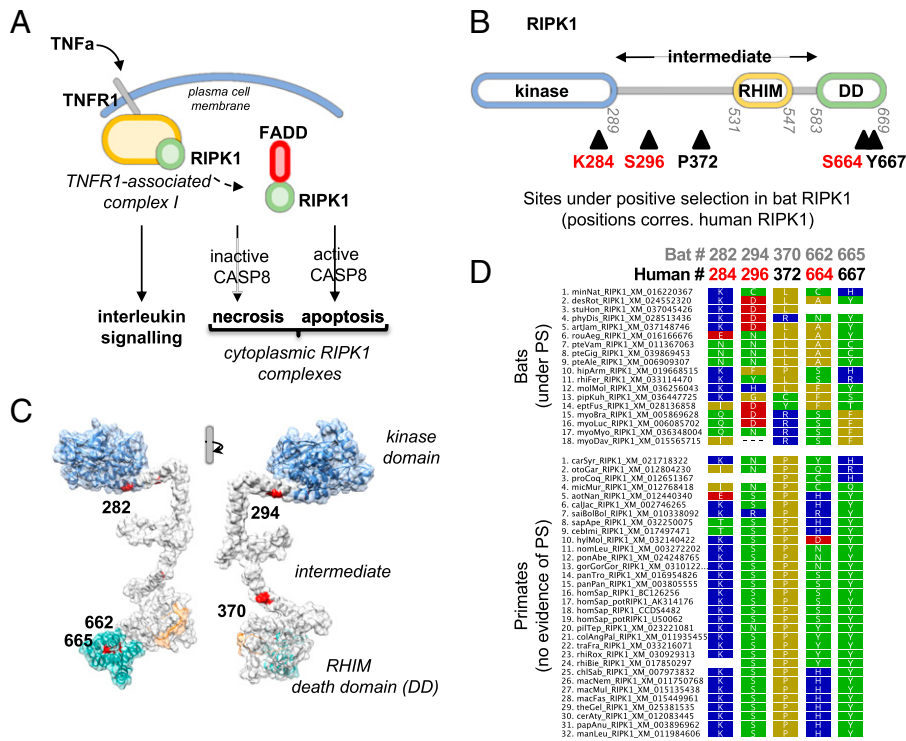


Fig. 5. The multifunctional and inflammatory RIPK1 protein exhibits strong evidence of adaptation in bats at key regulatory residues. (A) Schematic diagram of the three main functions associated to human RIPK1 in TNF signaling. As part of the TNFR1-associated complex, RIPK1 induces prosurvival signals that notably lead to NF κ B activation. When dissociating from this complex, as a result of multiple events involving both phosphorylation and ubiquitination, RIPK1 can associate to FADD and lead to apoptosis or necrosis. (B) Diagram of RIPK1 domains with the residues under PS in bats (black triangles) with the corresponding position and amino acid residue in human RIPK1 (Table 1). (C) 3D structure prediction of bat (*Rhinolophus ferrumequinum*) RIPK1, using RaptorX. The protein domains are color coded as in B. Residues under PS are in red and numbered according to their position in bat RIPK1. (D) The positively selected sites identified in bat RIPK1 are highly variable in bats (Top), but more conserved in primates (Bottom), where they are not identified as under adaptive evolution. Left, bat and primate RIPK1 with species abbreviation and accession number of sequences. Amino acid color coding, polarity properties (Geneious, Biomatters). The correspondence of residues from *Rhinolophus ferrumequinum* bat RIPK1 (gray) to human numbering (black) is shown at the Top. Detailed representation is shown in *SI Appendix, Fig. S10*.

from Païro-Castineira and colleagues (<https://genomicc.org/data/>), we identified five codons in FYCO1 with polymorphisms associated with severe COVID-19 in humans (Fig. 4D). Comparing these positions to the four positively selected sites in primates, we found one common site (site 447, genome position 45967996) (Fig. 4 C–E). This shows that residue 447, whose allele is correlated with COVID-19 severity in human, has also been subjected to adaptive evolution in primate history. In addition, at the protein domain level, the regions 430 to 555 and 910 to 1005 both have several residues associated with severe COVID-19 in humans and residues under adaptive evolution in primates (Fig. 4 D and E). Therefore, our combined positive selection and GWAS analysis identified FYCO1 regions that may be key host determinants of SARS-CoV-2 and COVID-19.

Overall, our results support the importance of FYCO1 in SARS-CoV pathogenesis or replication in primates, in both ancient (our positive selection analysis) and modern (GWAS) times. Furthermore, observed differences in positive selection between the susceptible primate hosts and bats (where no positive selection was observed and no disease is known to be associated with CoV infection) may highlight key differences in pathogenesis. We have two main hypotheses for the role of FYCO1 in SARS-CoV infection. First, given its known cellular role (Fig. 4A), FYCO1 may play a role in facilitating viral egress and replication. Second, FYCO1 may be involved in COVID-19 pathogenesis, potentially through an indirect mechanism by affecting the autophagy process or vesicle trafficking necessary to resolve viral infection.

RIPK1 Has Been under Adaptive Evolution in Bats at Residues That Are Crucial for Human RIPK1 Regulation. Human RIPK1 is an adaptor protein involved in inflammation through the tumor necrosis factor alpha receptor 1 (TNFR1) and the Toll-like receptors 3 and 4 (TLR3/4), leading to prosurvival, apoptotic, or necroptotic signals (Fig. 5A) (54, 55). A curated analysis of RIPK1 interactors showed that it is a central hub for 79 cellular

partners involved in key inflammatory and cell survival/death processes (Reactome database; *SI Appendix, Fig. S9A*). RIPK1 interacts with SARS-CoV-2 NSP12 (RdRp) (4) and is further involved in several bacterial and viral infections, being usurped by pathogens or involved in antimicrobial immunity (*SI Appendix, Fig. S9B*).

In our DGINN screens, we only identified signatures of positive selection in primate RIPK1. As previously, to obtain comprehensive phylogenetic and positive selection analyses, we retrieved all available coding sequences of bat (18 species) and primate (29 species) RIPK1 and performed codon alignments and analyses. Here, we found strong evidence of positive selection in bat RIPK1 but not in primates (Table 1 and *SI Appendix, Table S3*). This is different from our screen results, and this discrepancy was mostly due to 1) the addition of sequences as compared to our screens (i.e., from 12 to 18 bat species sequences, and from 24 to 29 primates) and 2) the high-quality codon alignments, which are crucial for positive selection studies.

Next, using site-specific analyses, we identified five residues in bat RIPK1 that have evolved under significant positive selection (Fig. 5B and Table 1 and *SI Appendix, Table S3*). These are located in the intermediate domain (282, 294, 370) and in the C-terminal death domain (DD; 662, 665) of RIPK1. The latter domain can interact with other DD-containing proteins, such as FADD, and has determinants for host–pathogen interactions (54, 55). To determine where the positively selected sites fall in the three-dimensional protein, we used a structure prediction of bat RIPK1 from *Rhinolophus ferrumequinum* (bats from the *Rhinolophus* genus naturally host viruses close to SARS-CoV-2). We found that the rapidly evolving sites are exposed at the protein surface (Fig. 5 B and C; *SI Appendix, Fig. S10A* for a comparison with the predicted three-dimensional [3D] structure of human RIPK1) (56). Therefore, physicochemical variations at sites 662 and 665 (Fig. 5D) in the DD could modulate interactions with DD-bearing proteins and thus influence the ability of bat RIPK1 to drive cell death (57). Alternatively, these variations

may affect interactions between bat RIPK1 and viral antagonists and thus may be directly involved in host–pathogen evolutionary conflicts.

Interestingly, using comparative analyses of bat and human RIPK1s, we found that the positively selected sites 282, 294, and 662 in bat RIPK1 correspond to sites K284 and S296 and S664 in human RIPK1, which are ubiquitinated and phosphorylated, respectively (54, 58) (Fig. 5B in red, and *SI Appendix, Fig. S10 B and C* for logo plots and comparative analyses). The posttranslational modifications at these sites are very important for the balance between the cell survival and the cell death functions of human RIPK1. It is thus possible that variation at these residues (Fig. 5D) affects how bat RIPK1 is regulated.

Overall, our evolutionary analyses indicate that RIPK1 is an important SARS-CoV-2 (and other virus)-interacting protein and suggest that residues undergoing positive selection in bats may be important 1) as determinants of virus–host interfaces, and 2) as regulators of the protein balance between prosurvival and procell death activities. The latter may allow certain bat species to tolerate viral infections and regulate the associated inflammation.

Discussion

This study of the evolution of SARS-CoV-2-interacting proteins in mammals help us to understand how the bat reservoir and the primate host have adapted to past coronavirus epidemics and may shed light on modern genetic determinants of virus susceptibility and COVID-19 severity. Here, among the 334 genes encoding for SARS-CoV-2 VIPs, we identified 38 and 81 genes with strong signatures of adaptive evolution in bats and primates, respectively. Results are available at <https://virhostnet.prabi.fr/virhostevol/>. First, we found a core set of 17 genes, including the ACE2 receptor and POLA1, with strong evidence of selective pressure in both mammalian orders, suggesting 1) past epidemics of pathogenic coronaviruses in bats and primates shaping mammalian genomes and 2) common virus–host molecular and adaptive interfaces between these two mammalian host orders. This represents a list of host genes that should be prioritized and studied for roles in broad SARS-CoV replication. We also found several genes under positive selection only in bats or primates (such as RIPK1 or TMPRSS2), which could highlight important differences in the coevolution of primate and bat with SARS-CoVs. Furthermore, we discovered specific residues within the VIPs with typical marks of virus–host arms races, which may point to precise SARS-CoV–host interfaces that have been important in vivo and may therefore represent key SARS-CoV-2 drug targets (such as TMPRSS2 or FYCO1). Finally, we found that FYCO1 sites with hallmarks of positive selection during primate evolution are those associated with severe COVID-19 in humans, supporting the importance of these rapidly evolving residues in SARS-CoV-2 pathogenesis and replication. Overall, our study identified several host proteins 1) whose evolution may have been driven by ancient epidemics of pathogenic SARS-CoVs, 2) are different between the bat reservoir and the primate host, and 3) may represent key in vivo virus–host determinants and drug targets.

The difference in adaptive VIPs in primates and bats suggests that, beyond the common virus–host interfaces, SARS-CoVs have an intrinsically different interactome in these distant hosts (i.e., specialization). Therefore, SARS-CoVs may have adapted to usurp and/or antagonize different cellular proteins in primates versus bats. This is exemplified by the evolution of the entry factor TMPRSS2 (among others). We identified strong evidence of virus–host arms races in primates but not in bats.

This suggests that SARS-CoVs may not strongly rely on TMPRSS2 for entry in bat cells, as opposed to primates, but only functional studies on SARS-CoV natural entry pathways into bat cells would firmly determine this. Interestingly, the recent SARS-CoV-2 Omicron variant has evolved to enter the human cell through a TMPRSS2-dependent and -independent route, showing also intrahost species plasticity at these interfaces (59–62). Lastly, the importance of lineage specificity of SARS-CoV-2 VIPs has previously been highlighted for OAS1. Indeed, humans rely on prenylated OAS1 to inhibit SARS-CoV-2 replication and prevent COVID-19 severity (12, 63), but *Rhinolophidae* bats do not encode for an OAS1 capable to interact with SARS-CoV-2 (12). Thus, in addition to genes such as TMPRSS2, FYCO1, or RIPK1, our findings provide dozens of genes that represent host-specific interfaces and may be critical in vivo SARS-CoV VIPs.

The differences between primate and bat evolution of the SARS-CoV-2 interactome may further result from important differences in the adaptation at the virus–host interface in a reservoir host versus a recipient host. In this model, beyond the core SARS-CoV-2 interactome of bats and primates, the genes under positive selection would correspond to host-specific adaptations to SARS-CoV. This could underlie important immunomodulatory differences between primates and bats (3). For example, the inflammatory protein RIPK1 showed signatures of adaptive evolution in bat residues that correspond with the loss of important RIPK1 regulatory phosphorylation and ubiquitination residues in humans. With the caveats that no functional studies exist on bat RIPK1, the extrapolation of the functions ascribed to the corresponding residues in human RIPK1 suggests that positive selection in bat RIPK1 may result from an advantageous decrease of RIPK1-driven inflammation in bats. This is analogous to the loss of the S358 phosphorylation site in bat STING that participates in a dampened inflammation response in bats (13) and supports a model where hosts that are more tolerant to viral infection contribute to the establishment of a host reservoir, such as hypothesized for bats (13, 64–69).

It is also possible that there are fewer signatures of adaptation in SARS-CoV-interacting proteins in bats over primates because coronaviruses may have been less pathogenic in the former host and therefore less selective (66, 70). However, evidence of strong positive selection in the bat ACE2 receptor driven by ancient pathogenic SARS-CoVs (this study and refs 26 and 27) supports a model in which past SARS-CoV epidemics have been sufficiently potent to shape bat genomes.

Our work also tries to bridge studies of ancient and recent evolution of genes, which can help us better understand past epidemics and adaptive genes and ultimately develop evolutionary medicine. This study over millions of years of evolution (at the interspecies level) shows evidence of very ancient epidemics of SARS-CoVs that have shaped both primate and bat genomes. Marks of adaptation in SARS-CoV-2 VIPs at the human population level further identified evidence of past SARS-CoV epidemics in more recent human history (20). Bridging these ancient and more recent evolutionary analyses with GWAS studies would bring more direct confirmation of the causal role of viral interacting proteins in pathogenesis. This is here exemplified by the FYCO1 gene that may be a central protein in SARS-CoV-2 pathogenesis and disease etiology.

A limitation of our study is that we did not quantify the selective pressures occurring at (regulatory) noncoding regions of the VIPs. Using human population genomics, Souilmi et al. (20) found that marks of positive selection have been particularly strong at noncoding regions of SARS-CoV-interacting proteins.

However, these analyses are challenging at the interspecies level, and more methods and high-quality genome alignments would be necessary for state-of-the-art mammalian genomic analyses. Our findings are therefore conservative and other marks of adaptation in the same, and in more, VIPs are certainly at play.

At the heart of our study analyzing the coding sequences of SARS-CoV-2 VIPs is the identification of site-specific adaptations at multiple SARS-CoV-2-interacting proteins, which may reflect the exact sites of molecular arms races of proviral and antiviral VIPs with SARS-CoVs (7, 11, 22). It is further possible that other protein functions and physicochemical constraints may also have driven some of these evolutions. For example, the signatures of positive selection in the mammalian hepadnavirus receptor NTCP (Na⁺-taurocholate cotransporting polypeptide) may have been driven by pathogenic hepadnaviruses as well as metabolic changes in bats due to its role as bile acid transporter (71). The identified rapidly evolving sites are therefore of primary importance to investigate in functional assays to firmly identify key SARS-CoV-2-cell determinants and drug targets. For example, our study highlights TMPRSS2 and RIPK1, among others, as potential targets of interest. Primidone, a Food and Drug Administration–approved RIPK1 inhibitor, has proven ineffective as a direct inhibitor of viral replication in established cell lines (4, 72). However, our findings suggest that RIPK1 inhibitors will more likely exert an effect on the virus-induced hyperinflammation rather than on viral replication itself. As such, the evaluation of the effectiveness of RIPK1–kinase inhibitors will require a more complex cellular setup. Lastly, other viruses may also have driven adaptation at these VIPs, which therefore represent essential host–pathogen interfaces. Targeting the identified VIPs with strong marks of virus–host arms races may be an effective broad antiviral strategy.

Methods

DGINN Screens. Analyses were performed as previously described in Picard et al. (21). Briefly, consensus coding sequence (CCDS) identifiers were downloaded from HUGO Gene Nomenclature Committee Biomart (biomart.genenames.org/) for all 334 genes of interest. If there was more than one CCDS, the longest was selected. Initial codon alignments and phylogenetic trees were obtained using DGINN with default parameters (prank -F -codon; version 150803, HKY+G+I model [73]; PhyML v3.2 [74]). Duplication events were detected through the combined use of Long Branch Detection and Treerecs (75) as implemented in DGINN. Recombination events were detected through the use of Genetic Algorithm for Recombination Detection (GARD) (76) from the HyPhy suite as implemented in DGINN. For each VIP gene, the analyses of primate evolution and of bat evolution were separately run. The species trees employed for the tree reconciliation with Treerecs are accessible at <https://virhostnet.prabi.fr/virhostevol/>. Positive selection analyses were then run using models from BUSTED and MEME from the HyPhy suite (24, 77, 78) and codon substitution models from PAML codeml (M0, M1, M2, M7, M8) (25) and from Bio++ (M0, M1^{NS}, M2^{NS}, M7^{NS}, M8^{NS}) (23) as implemented in DGINN (21). For the chiroptera screen, a visual inspection of the resulting gene alignments was performed, and we refined 28 of them to delete erroneous ortholog sequences, erroneous isoforms, or sequencing errors. These 28 curated alignments were then reanalyzed with DGINN starting at the alignment step and included in the final results (SI Appendix, Fig. S1A and Table S1).

MAIC Scores. MAIC scores were obtained from the database for COVID-19 (<https://baillielab.net/maic/covid19/>, 2020-11-25 release) (5). The 334 VIP genes were cross-referenced against the 10,000 best hits of the MAIC database.

Detailed Phylogenetic Analyses on Genes of Interest. Alignments from the DGINN screens were retrieved, and sequences that appeared erroneous were taken out. To obtain a maximum number of species along primate and bat phylogenies, further sequences were retrieved from NCBI databases using BLASTn.

Final codon alignments were then made using PRANK (73) or Muscle Translate (79), and phylogenetic trees were built using PhyML with HKY+I+G model and approximate Likelihood Ratio Test (aLRT) for branch support (74). Each curated gene alignment and tree were then submitted to positive selection analyses using the DGINN pipeline, as follows: HYPHY BUSTED and MEME, PAML codeml (M0, M1, M2, M7, M8, M8a), and Bio++ (M0, M1^{NS}, M2^{NS}, M7^{NS}, M8^{NS}, M8a^{NS}) (references in *DGINN Screens*). To test for statistical significance of positive selection in codeml and Bio++, we ran a χ^2 test of the LRT from models disallowing positive selection versus models allowing for positive selection (M1 versus M2, M7 versus M8, and M8a versus M8) to derive *P* values. To identify the sites under positive selection, we used HYPHY MEME (*P* value < 0.05), the BEB statistics from the codeml M2 and M8 models (BEB > 0.95) and the Bayesian PP from the M2^{NS} and M8^{NS} models in Bio++ (PP > 0.95). Two other web-based methods were used for this set of genes, as follows: a Fast, Unconstrained Bayesian Approximation for Inferring Selection (FUBAR) method to detect site-specific positive selection (PP > 0.90) (80) and an adaptive Branch-Site Random Effect Likelihood test for episodic diversification (aBS-REL) to detect branch/lineage-specific positive selection (*P* value < 0.1) (81).

GWAS Analyses. Using the COVID-19 Host Genetics Initiative data (<https://www.covid19hg.org/results/r6/>), we extracted the positions of human polymorphisms associated with “very severe respiratory confirmed covid vs. population” that are within FYCO1 coding sequence and have a *P* value below 10^{−8}. We similarly retrieved the positions found associated to severe COVID-19 by Pairo-Castineira et al. (52) from the data publicly available at <https://genomicc.org/data/>. We then matched the coordinates of polymorphic sites significantly associated with severe COVID-19 to the alignment of coding sequences of FYCO1 (using transcript FYCO1-205). To note, none of the other genes under positive selection contained polymorphism significantly associated with “very severe respiratory confirmed COVID” by the COVID-19 Host Genetics Initiative (see online browser).

Reactome and GOrrilla Enrichment Analyses. Gene pathway enrichment analyses were carried out on the Reactome biological pathways tools (<https://reactome.org/>). Interactors of RIPK1 were retrieved using the Reactome Cytoscape Plugin (82). To identify enriched GO terms, we used the GOrrilla online tools (cbl-gorilla.cs.technion.ac.il/) to compare unranked lists of genes (target and background lists), derived from our analyses (i.e., genes with evidence of positive selection, or not, in bats and in primates, and 334 VIPs) and from ENSEMBL to constitute the all human genes background set (ftp.ensembl.org/pub/release-106/gtf/homo_sapiens/Homo_sapiens.GRCh38.106.gtf.gz).

Gene Expression Analyses. We used the transcriptome atlas of 29 human tissues from (83). We first compared the global (mean fragments per kilobase of transcript per million mapped read [FPKM]) expression level of genes under and not under positive selection in 29 organs. Then, we compared the expression level in the Lung (FPKM, represented in log₁₀) for the genes under PS, for bats and primates.

Protein Structure Predictions. Protein structure prediction of human and *Rhinolophus ferrumequinum* RIPK1 were modeled using RaptorX (84), and structures were visualized using the Chimera software (85).

The 3D structure of TMPRSS2 was predicted using the Iterative-Threading ASSEMBLY Refinement (I-TASSER) server (86). A 492-amino acid sequence of human TMPRSS2 obtained from NCBI GenBank (accession number AF329454) was used as the query. The best model inferred by I-TASSER was selected using the C-score—a measure assessing the quality of the models. The final estimates are as follows: model C-score, −0.41; estimated TM-score, 0.66 ± 0.13; and root-mean-square deviation, 8.2 ± 4.5Å. The corresponding TMPRSS2 structure was generated using Swiss Protein Data Bank viewer software (87).

Sequence Logo Generation. The amino acid sequence logos of TMPRSS2 were generated Geneious (Biomatters), based on an alignment of the PS sites from mammalian species reported as naturally susceptible and/or experimentally permissive to SARS-CoV-2, SARS-CoV, and MERS-CoV. The amino acid sequence logo of bat RIPK1 was generated using WebLogo3 (88), based on the amino acid alignment of 18 bat sequences.

Data Availability. Data, codes, and any information are entirely publicly available. All codes are available in https://gitbio.ens-lyon.fr/ciri/ps_sars-cov-2/2021_dginn_covid19 (89), and the DGINN pipeline is available at <http://bioweb.me/DGINN-github> (90). All results from the DGINN screens (DGINN full dataset) and from the detailed evolutionary analyses (DGINN genes of interest) are available through the VirHostEvol web service at <https://virhostnet.prabi.fr/virhostevol/> (91). The Shiny web application open-source code is available from the IN2P3 gitlab at <https://gitlab.in2p3.fr/vincent.navratil/shinyapps-virhostevol> (92).

ACKNOWLEDGMENTS. We sincerely thank Janet Young for sharing her raw data on the positive selection of primate genes for comparative analyses, as well as for her precious feedback on the project. We also thank Michael Emerman, Pierre-Olivier Vidalain, Dominique Guyot, and the members of LP2L for helpful discussions. We thank Sara Clohisey for generously providing us with the raw data for MAIC scores. We thank all the contributors of publicly available genomic and genetic sequences and of phylogenetic programs. Some molecular graphics and analyses were performed with University of California San Francisco Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311. This work was performed using the computing facilities of the PSMN (Pôle Scientifique de Modélisation Numérique) of the ENS de Lyon, the IFB (Institut Français de Bioinformatique), and the CC LBBE/PRABI, and

with the support from the BIBS (Bioinformatic and biostatistics service) of the CIRI, Lyon. We also thank S. Delmotte and B. Spataro for the cloud openstack.

L.E. and L.G. are supported by grants from the ANR LABEX ECOFECT (ANR-11-LABX-0048 of the Université de Lyon, within the program Investissements d'Avenir [ANR-11-IDEX-0007]) and the French Agence Nationale de la Recherche, under Grant ANR-20-CE15-0020-01 (Project BATantiVIR). L.E. is further supported by the CNRS and by grants from the French Research Agency on HIV and Emerging Infectious Diseases ANRS/MIE (no. ECTZ19143 and ECTZ118944) and the CNRS France-U. Arizona Institute for Global Grand Challenges. A.M. is supported by the Fondation pour la Recherche Médicale, FRM:AJE201912009932.

Author affiliations: ^aCIRI (Centre International de Recherche en Infectiologie), Univ Lyon, Inserm, U1111, Université Claude Bernard Lyon 1, CNRS, UMR5308, ENS de Lyon, F-69007 Lyon, France; ^bLaboratoire de Biométrie et Biologie Evolutive, UMR5558, Univ Lyon, Université Lyon 1, F-69622 Villeurbanne, France; ^cDepartment of Microbiology, Immunology, and Molecular Genetics, UCLA, University of California Los Angeles, Los Angeles, CA 90095; ^dGenetics, Reproduction & Development Institute, UMR INSERM 1103 - CNRS, UMR6293 - UCA, 63001 Clermont-Ferrand, France; ^ePRABI, Rhône-Alpes Bioinformatics Center, Université Lyon 1, 69622 Villeurbanne, France; ^fEuropean Virus Bioinformatics Center, 07743 Jena, Germany; and ^gInstitut Français de Bioinformatique, IFB-core, UMS 3601, 91057 Evry, France

Author contributions: L.E. designed research; M.C., L.P., L.G., S.J., A.C., O.I.F., A.M., V.N., and L.E. performed research; L.E. supervised the study; L.P., L.G., and V.N. contributed new reagents/analytic tools; M.C., L.P., L.G., S.J., A.C., O.I.F., A.M., V.N., and L.E. analyzed data; L.E. wrote the paper; and M.C., L.P., L.G., S.J., A.C., O.I.F., A.M., and V.N. provided edits and approval of the manuscript.

1. S. Temmam *et al.*, Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature* **604**, 330–336 (2022).
2. L.-F. Wang, D. E. Anderson, Viruses in bats and potential spillover to animals and humans. *Curr. Opin. Virol.* **34**, 79–89 (2019).
3. M. J. Christie *et al.*, Of bats and men: Immunomodulatory treatment options for COVID-19 guided by the immunopathology of SARS-CoV-2 infection. *Sci. Immunol.* **6**, eabd0205 (2021).
4. D. E. Gordon *et al.*, A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).
5. N. Parkinson *et al.*, Dynamic data-driven meta-analysis for prioritisation of host genes implicated in COVID-19. *Sci. Rep.* **10**, 22303 (2020).
6. M. D. Daugherty, H. S. Malik, Rules of engagement: Molecular insights from host-virus arms races. *Annu. Rev. Genet.* **46**, 677–700 (2012).
7. N. K. Duggal, M. Emerman, Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat. Rev. Immunol.* **12**, 687–695 (2012).
8. D. Enard, L. Cai, C. Gwennap, D. A. Petrov, Viruses are a dominant driver of protein adaptation in mammals. *eLife* **5**, e12469 (2016).
9. I. N. Boys *et al.*, RTP4 is a potent IFN-inducible anti-flavivirus effector engaged in a host-virus arms race in bats and other mammals. *Cell Host Microbe* **28**, 712–723.e9 (2020).
10. O. I. Fregoso *et al.*, Evolutionary toggling of Vpx/Vpr specificity results in divergent recognition of the restriction factor SAMHD1. *PLoS Pathog.* **9**, e1003496 (2013).
11. S. L. Sawyer, L. I. Wu, M. Emerman, H. S. Malik, Positive selection of primate TRIM5 α identifies a critical species-specific retroviral restriction domain. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 2832–2837 (2005).
12. A. Wickenhagen *et al.*; ISARIC4C Investigators, A prenylated dsRNA sensor protects against severe COVID-19. *Science* **374**, eabj3624 (2021).
13. J. Xie *et al.*, Dampened STING-dependent interferon activation in bats. *Cell Host Microbe* **23**, 297–301.e4 (2018).
14. P. Bastard *et al.*; HGID Lab; NIAID-USUHS Immune Response to COVID Group; COVID Clinicians; COVID-STORM Clinicians; Imagine COVID Group; French COVID Cohort Study Group; Milieu Intérieur Consortium; CoV-Contact Cohort; Amsterdam UMC Covid-19 Biobank; COVID Human Genetic Effort; Autoantibodies against type I IFNs in patients with life-threatening COVID-19. *Science* **370**, eabd4585 (2020).
15. Y. J. Crow, D. B. Stetson, The type I interferonopathies: 10 years on. *Nat. Rev. Immunol.* **22**, 1–13 (2021).
16. J. Schoggins, Defective viral RNA sensing linked to severe COVID-19. *Science* **374**, 535–536 (2021).
17. D. Ellinghaus *et al.*; Severe Covid-19 GWAS Group, Genomewide association study of severe covid-19 with respiratory failure. *N. Engl. J. Med.* **383**, 1522–1534 (2020).
18. H. Zeberg, S. Pääbo, The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature* **587**, 610–612 (2020).
19. Q. Zhang *et al.*; COVID-STORM Clinicians; COVID Clinicians; Imagine COVID Group; French COVID Cohort Study Group; CoV-Contact Cohort; Amsterdam UMC Covid-19 Biobank; COVID Human Genetic Effort; NIAID-USUHS/TAGC COVID Immunity Group, Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science* **370**, eabd4570 (2020).
20. Y. Souilmi *et al.*, An ancient viral epidemic involving host coronavirus interacting genes more than 20,000 years ago in East Asia. *Curr. Biol.* **31**, 3504–3514.e9 (2021).
21. L. Picard *et al.*, DGINN, an automated and highly-flexible pipeline for the detection of genetic innovations on protein-coding genes. *Nucleic Acids Res.* **48**, e103 (2020).
22. M. Sironi, R. Cagliani, D. Forni, M. Clerici, Evolutionary insights into host-pathogen interactions from mammalian sequence data. *Nat. Rev. Genet.* **16**, 224–236 (2015).
23. L. Guéguen *et al.*, Bio++: Efficient extensible libraries and tools for computational molecular evolution. *Mol. Biol. Evol.* **30**, 1745–1750 (2013).
24. S. L. K. Pond, S. D. W. Frost, S. V. Muse, HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2005).
25. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
26. A. Demogins, M. Farzan, S. L. Sawyer, Evidence for ACE2-utilizing coronaviruses (CoVs) related to severe acute respiratory syndrome CoV in bats. *J. Virol.* **86**, 6350–6353 (2012).
27. H. K. Frank, D. Enard, S. D. Boyd, Exceptional diversity and selection pressure on SARS-CoV and SARS-CoV-2 host receptor in bats compared to other mammals. *bioRxiv* [Preprint] (2020). <https://doi.org/10.1101/2020.04.20.051656> (Accessed 17 July 2022).
28. E. Eden, R. Navon, I. Steinfeld, D. Lipson, Z. Yakhini, GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).
29. R. Carapito *et al.*, Identification of driver genes for critical forms of COVID-19 in a deeply phenotyped young patient cohort. *Sci. Transl. Med.* **14**, eabj7521 (2022).
30. C. J. Sigrist, A. Bridge, P. Le Mercier, A potential role for integrins in host cell entry by SARS-CoV-2. *Antiviral Res.* **177**, 104759 (2020).
31. A. Chaudhuri, Comparative analysis of non structural protein 1 of SARS-CoV2 with SARS-CoV1 and MERS-CoV: An *in silico* study. *J. Mol. Struct.* **1243**, 130854 (2021).
32. T. Guirimand, S. Delmotte, V. Navratil, VirHostNet 2.0: Surfing on the web of virus/host molecular interactions data. *Nucleic Acids Res.* **43**, D583–D587 (2015).
33. P. S. Mitchell, M. Emerman, H. S. Malik, An evolutionary perspective on the broad antiviral specificity of MxA. *Curr. Opin. Microbiol.* **16**, 493–499 (2013).
34. E. Böttcher *et al.*, Proteolytic activation of influenza viruses by serine proteases TMPRSS2 and HAT from human airway epithelium. *J. Virol.* **80**, 9896–9898 (2006).
35. H. Limburg *et al.*, TMPRSS2 is the major activating protease of influenza A virus in primary human airway cells and influenza B virus in human type II pneumocytes. *J. Virol.* **93**, e00649–e19 (2019).
36. D. Bestle *et al.*, TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Sci. Alliance* **3**, e202000786 (2020).
37. M. Hoffmann *et al.*, SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280.e8 (2020).
38. S. Bertram *et al.*, TMPRSS2 activates the human coronavirus 229E for cathepsin-independent host cell entry and is expressed in viral target cells in the respiratory epithelium. *J. Virol.* **87**, 6150–6160 (2013).
39. N. Iwata-Yoshikawa *et al.*, TMPRSS2 contributes to virus spread and immunopathology in the airways of murine models after coronavirus infection. *J. Virol.* **93**, e01815–e01818 (2019).
40. S. Matsuyama *et al.*, Efficient activation of the severe acute respiratory syndrome coronavirus spike protein by the transmembrane protease TMPRSS2. *J. Virol.* **84**, 12658–12664 (2010).
41. S. Jeon *et al.*, Regional TMPRSS2 V197M allele frequencies are correlated with COVID-19 case fatality rates. *Mol. Cells* **44**, 680–687 (2021).
42. G. N. Brooke, F. Prischi, Structural and functional modelling of SARS-CoV-2 entry in animal models. *Sci. Rep.* **10**, 15917 (2020).
43. H. R. Rangel, J. T. Ortega, S. Maksoud, F. H. Pujol, M. L. Serrano, SARS-CoV-2 host tropism: An *in silico* analysis of the main cellular factors. *Virus Res.* **289**, 198154 (2020).
44. S. Senapati, P. Banerjee, S. Bhagavatlal, P. P. Kushwaha, S. Kumar, Contributions of human ACE2 and TMPRSS2 in determining host-pathogen interaction of COVID-19. *J. Genet.* **100**, 12 (2021).
45. K. Li, D. K. Meyerholz, J. A. Bartlett, P. B. McCray Jr., The TMPRSS2 inhibitor nafamostat reduces SARS-CoV-2 pulmonary infection in mouse models of COVID-19. *MBio* **12**, e0097021 (2021).
46. L. Wettstein *et al.*, Alpha-1 antitrypsin inhibits TMPRSS2 protease activity and SARS-CoV-2 infection. *Nat. Commun.* **12**, 1726 (2021).
47. T. Shapira *et al.*, A TMPRSS2 inhibitor acts as a pan-SARS-CoV-2 prophylactic and therapeutic. *Nature* **605**, 340–348 (2022).
48. X. Cheng *et al.*, Structural basis of FYCO1 and MAP1LC3A interaction reveals a novel binding mode for Atg8-family proteins. *Autophagy* **12**, 1330–1339 (2016).
49. J. L. Nieto-Torres *et al.*, LC3B phosphorylation regulates FYCO1 binding and directional transport of autophagosomes. *Curr. Biol.* **31**, 3440–3449.e7 (2021).
50. J. Chen *et al.*, Mutations in FYCO1 cause autosomal-recessive congenital cataracts. *Am. J. Hum. Genet.* **88**, 827–838 (2011).
51. K. Satoh, Y. Takemura, M. Satoh, K. Ozaki, S. Kubota, Loss of FYCO1 leads to cataract formation. *Sci. Rep.* **11**, 13771 (2021).

52. E. Pairo-Castineira *et al.*; GenOMICC Investigators; ISARIC4C Investigators; COVID-19 Human Genetics Initiative; 23andMe Investigators; BRACOVIC Investigators; Gen-COVID Investigators, Genetic mechanisms of critical illness in COVID-19. *Nature* **591**, 92–98 (2021).
53. COVID-19 Host Genetics Initiative, The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur. J. Hum. Genet.* **28**, 715–718 (2020).
54. T. Delanghe, Y. Dondelinger, M. J. M. Bertrand, RIPK1 kinase-dependent death: A symphony of phosphorylation events. *Trends Cell Biol.* **30**, 189–200 (2020).
55. X. Liu *et al.*, Epstein-Barr virus encoded latent membrane protein 1 suppresses necroptosis through targeting RIPK1/3 ubiquitination. *Cell Death Dis.* **9**, 53 (2018).
56. M. Mompeán *et al.*, The structure of the necrosome RIPK1-RIPK3 core, a human hetero-amyloid signaling complex. *Cell* **173**, 1244–1253.e10 (2018).
57. S. Grimm, B. Z. Stanger, P. Leder, RIP and FADD: Two "death domain"-containing proteins can induce apoptosis by convergent, but dissociable, pathways. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10923–10927 (1996).
58. D. S. Simpson, A. Gabrielyan, R. Feltham, RIPK1 ubiquitination: Evidence, correlations and the undefined. *Semin. Cell Dev. Biol.* **109**, 76–85 (2021).
59. B. Meng *et al.*; CITIID-NIHR BioResource COVID-19 Collaboration; Genotype to Phenotype Japan (G2P-Japan) Consortium; Ecuador-COVID19 Consortium, Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity. *Nature* **603**, 706–714 (2022).
60. T. P. Peacock *et al.*, The SARS-CoV-2 variant, Omicron, shows rapid replication in human primary nasal epithelial cultures and efficiently uses the endosomal route of entry. *bioRxiv* [Preprint] (2022). <https://doi.org/10.1101/2021.12.31.474653> (Accessed 17 July 2022).
61. L. Pia, S. Rowland-Jones, Omicron entry route. *Nat. Rev. Immunol.* **22**, 144–144 (2022).
62. B. J. Willett *et al.*, The hyper-transmissible SARS-CoV-2 Omicron variant exhibits significant antigenic change, vaccine escape and a switch in cell entry mechanism. *medRxiv* [Preprint] (2022). <https://doi.org/10.1101/2022.01.03.21268111> (Accessed 17 July 2022).
63. F. W. Soveg *et al.*, Endomembrane targeting of human OAS1 p46 augments antiviral activity. *eLife* **10**, e71047 (2021).
64. M. Ahn *et al.*, Dampened NLRP3-mediated inflammation in bats and implications for a special viral reservoir host. *Nat. Microbiol.* **4**, 789–799 (2019).
65. P. C. De La Cruz-Rivera *et al.*, The IFN response in bats displays distinctive IFN-stimulated gene expression kinetics with atypical RNASEL induction. *J. Immunol. Baltim. Md* **200**, 209–217 (2018).
66. A. T. Irving *et al.*, Interferon regulatory factors IRF1 and IRF7 directly regulate gene expression in bats in response to viral infection. *Cell Rep.* **33**, 108345 (2020).
67. S. S. Pavlovich *et al.*, The Egyptian Roussette genome reveals unexpected features of bat antiviral immunity. *Cell* **173**, 1098–1110.e18 (2018).
68. J. Prescott *et al.*, Roussette bat dendritic cells overcome Marburg virus-mediated antiviral responses by upregulation of interferon-related genes while downregulating proinflammatory disease mediators. *MSphere* **4**, e00728-19 (2019).
69. G. Zhang *et al.*, Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. *Science* **339**, 456–460 (2013).
70. M. Emerman, H. S. Malik, Paleovirology—Modern consequences of ancient viruses. *PLoS Biol.* **8**, e1000301 (2010).
71. S. Jacquet *et al.*, Evolution of hepatitis B virus receptor NTCP reveals differential pathogenicities and species specificities of hepadnaviruses in primates, rodents, and bats. *J. Virol.* **93**, e01738-18 (2019).
72. T. Riebeling *et al.*, Primidone blocks RIPK1-driven cell death and inflammation. *Cell Death Differ.* **28**, 1610–1626 (2021).
73. A. Löytynoja, N. Goldman, Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* **320**, 1632–1635 (2008).
74. S. Guindon *et al.*, New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
75. N. Comte *et al.*, Treerecs: An integrated phylogenetic tool, from sequences to reconciliations. *Bioinformatics* **36**, 4822–4824 (2020).
76. S. L. Kosakovsky Pond, D. Posada, M. B. Gravenor, C. H. Woelk, S. D. W. Frost, GARD: A genetic algorithm for recombination detection. *Bioinformatics* **22**, 3096–3098 (2006).
77. B. Murrell *et al.*, Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **8**, e1002764 (2012).
78. B. Murrell *et al.*, Gene-wide identification of episodic selection. *Mol. Biol. Evol.* **32**, 1365–1371 (2015).
79. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
80. B. Murrell *et al.*, FUBAR: A fast, unconstrained bayesian approximation for inferring selection. *Mol. Biol. Evol.* **30**, 1196–1205 (2013).
81. M. D. Smith *et al.*, Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* **32**, 1342–1353 (2015).
82. M. Gillespie *et al.*, The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* **50**, D687–D692 (2022).
83. D. Wang *et al.*, A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* **15**, e8503 (2019).
84. S. Wang, S. Sun, Z. Li, R. Zhang, J. Xu, Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLoS Comput. Biol.* **13**, e1005324 (2017).
85. E. F. Pettersen *et al.*, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
86. J. Yang, Y. Zhang, I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Res.* **43**, W174–W181 (2015).
87. M. U. Johansson, V. Zoete, O. Michielin, N. Guex, Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. *BMC Bioinformatics* **13**, 173 (2012).
88. G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
89. M. Cariou, Evolutionary history of SARS-CoV-2 interactome in bats and primates identifies key virus-host interfaces and conflicts. *GitBio*. https://gitbio.ens-lyon.fr/ciri/ps_sars-cov-2/2021_dginn_covid19. Deposited 16 June 2021.
90. L. Picard, DGINN: Detection of Genetic INNovations pipeline. *GitHub*. <https://github.com/leapicard/DGINN>. Deposited 24 August 2021.
91. M. Cariou *et al.*, Data for "Distinct evolutionary trajectories of SARS-CoV-2-interacting proteins in bats and primates identify important host determinants of COVID-19." *VirHostEvol*. <https://virhostnet.prabi.fr/virhostevol/>. Accessed 1 August 2022.
92. V. Navratil, shinyapps-virhostevol. *GitLab*. <https://gitlab.in2p3.fr/vincent.navratil/shinyapps-virhostevol>. Deposited 10 May 2021.