bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

Marie Cariou[1], Léa Picard[π, 1,2], Laurent Guéguen[π, 2], Stéphanie Jacquet[∂, 1,2], Andrea Cimarelli[∂, 1], Oliver I Fregoso[∂, 3], Antoine Molaro[∂, 4], Vincent Navratil[∂, 5,6,7], Lucie Etienne[1,*]

Affiliations:

[1], CIRI, Centre International de Recherche en Infectiologie, Univ Lyon, Inserm, U1111, Université Claude Bernard Lyon 1, CNRS, UMR5308, ENS de Lyon, F-69007, Lyon, France.

[2], Laboratoire de Biométrie et Biologie Evolutive, UMR5558, Univ Lyon, Université Lyon 1, F-69622 Villeurbanne, France

[3], UCLA, University of California Los Angeles, Department of Microbiology, Immunology, and Molecular Genetics, 615 Charles E Young Dr. S., Los Angeles, CA, 90095, USA

[4], Genetics, Reproduction & Development Institute, UMR INSERM 1103 - CNRS 6293 - UCA, Clermont-Ferrand, France.

[5], PRABI, Rhône-Alpes Bioinformatics Center, Université Lyon 1, 69622 Villeurbanne, France

[6], European Virus Bioinformatics Center, 07743 Jena, Germany

[7], Institut Français de Bioinformatique, IFB-core, UMS 3601, 91057 Évry, France

[π] and [∂], Contributed equally

* Contact: lucie.etienne@ens-lyon.fr

Lucie ETIENNE, CIRI, ENS-Lyon, 46 allée d'Italie, 69364 LYON Cedex 7, France

Author Contributions:

Designed research: LE

Performed research: All authors

Contributed new analytic tools: LP, LG, VN

Analyzed data: All authors

Wrote the paper: LE, with edits and approval by all authors.

No competing interests.

Classification: Biological Sciences. Evolution, Immunology and Inflammation, Microbiology, Medical Sciences, Genetics.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Keywords (3-5):

SARS-CoV-2; COVID-19; bats; primates; viral reservoir; comparative genetics; positive selection; virus-host molecular arms-races; virus-host coevolution

Key findings:

- Evolutionary history of 334 SARS-CoV-2 interacting proteins (VIPs) in bats and primates identifying how the past has shaped modern viral reservoirs and humans – results publicly-available in an online resource.
- Identification of 81 primate and 38 bat VIPs with signatures of adaptive evolution. The common ones among species delineate a core adaptive interactome, while the ones displaying distinct evolutionary trajectories enlighten host lineage-specific determinants.
- Evidence of primate specific adaptation of the entry factor TMPRSS2 pointing to its host-specific *in vivo* importance and predicting molecular interfaces.
- FYCO1 sites associated with severe COVID-19 in human (GWAS) display hallmarks of ancient adaptive evolution in primates, highlighting its importance in SARS-CoV-2 replication or pathogenesis and differences with the bat reservoir.
- Identification of adaptive evolution in the bat's multifunctional RIPK1 at residues that may differentially regulate inflammation.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

The COVID-19 pandemic is caused by SARS-CoV-2, a novel coronavirus that spilled 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. Here, 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* important 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 functional adaptations involved in SARS-CoV-2 infection in bats and primates, critically enlightening modern genetic determinants of virus susceptibility and severity.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

The current COVID-19 pandemic already led to over six million human deaths (WHO April 2022). The causative agent is a novel severe acute respiratory syndrome coronavirus strain, SARS-CoV-2, that originated from viral cross-species transmission from the bat reservoir, directly or through an intermediate host, to human (Temmam et al., 2022). Bats naturally hosts some of the most high-profile zoonotic viruses, including SARS-CoVs, without apparent symptoms (Wang and Anderson, 2019). Despite scores of clinical trials 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 underly infections in the bat reservoir and the primate host (Christie et al., 2021).

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/KO screens, cDNA library screens or mass-spectrometry analyses (Gordon et al., 2020; Parkinson et al., 2020). 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-adaptations 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 residues and domains in VIPs (Daugherty and Malik, 2012; Duggal and Emerman, 2012; Enard et al., 2016). Thus, when combined with functional data, identifying the VIPs with signatures of positive selection is a powerful way to

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

discover virus-host interfaces (e.g. (Boys et al., 2020; Fregoso et al., 2013; Sawyer et al., 2005)).

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. (Wickenhagen et al., 2021; Xie et al., 2018)). 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 (Bastard et al., 2020; Crow and Stetson, 2021; Schoggins, 2021; The Severe Covid-19 GWAS Group, 2020; Wickenhagen et al., 2021; Zeberg and Pääbo, 2020; Zhang et al., 2020), 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 (Souilmi et al., 2021).

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 (Gordon et al., 2020) using the Detection of Genetic INNovation (DGINN) pipeline (Picard et al., 2020), followed by comprehensive functional-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, (i) confirming that past SARS-CoV epidemics occurred during both bat and primate evolution and (ii) identifying the 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 or RIPK1 that play important roles in entry, trafficking or inflammatory responses, respectively. We

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (Daugherty and Malik, 2012; Enard et al., 2016). Adaptive evolution can be identified by positive selection analyses over a set of protein coding orthologs when their rate of non-synonymous codon substitutions exceeds that of synonymous ones (Sironi et al., 2015). 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. in mass-spectrometry assays of SARS-CoV-2 proteins in human cells (Gordon et al., 2020), in addition to the angiotensin converting enzyme 2 (ACE2) receptor and the transmembrane protease serine 2 (TMPRSS2), both necessary for virus entry into the cells.

To perform the phylogenetic and positive selection analyses, we used the Detection of genetic innovation DGINN bioinformatic pipeline (Picard et al., 2020) that entirely automates the analyses and combines several methods to test for selection across large datasets (Figure 1A). Briefly, from each of the 334 human reference gene sequences, DGINN automatically retrieved bat and primate homologs (from NCBI nr database), curated the coding sequences, and performed a codon-alignment followed by a gene phylogenetic reconstruction (Figure 1A, Table

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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: HYPHY (BUSTED and MEME), PAML (Codeml M0, M1, M2, M7, M8, and associated Bayesian Empirical Bayes (BEB) for codon-specific analyses), and bpp ($M0^{NS}$, $M1^{NS}$, $M2^{NS}$, $M7^{NS}$, $M8^{NS}$, and associated Posterior Probabilities (PP) for codon-specific analyses) (Figure 1A, Methods for details, Picard et al 2020) (Guéguen et al., 2013; Pond et al., 2005; Yang, 2007).

We found that the DGINN pipeline, previously validated on nineteen primate genes (Picard et al., 2020), was efficient at screening hundreds of genes and at analyzing other mammalian orders (here, chiroptera) (Figures S1-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 new 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 natural selection acting on the protein (Figure 1B). 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

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

too. Beyond this trend, we cannot compare the omega values *quantitatively* between the two mammalian orders – reasons include differences in the number of analyzed species (i.e. 12 and 24 median number of species in bats and primates, respectively), the population sizes, the genetic distances, etc.

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 (Figure 1C). These include the ACE2 receptor, also reported by others as under strong positive selection in bats (Demogines et al., 2012; Frank et al., 2020). In primates, we identified 81 genes under positive selection, after discarding seven due to low-quality alignments and inclusion of erroneous sequences in the automatic steps (Figure 1D legend).

In the case of primate analyses, we identified more VIPs under positive selection than Gordon *et al*., in which they identified 40/332 genes under positive selection in primates using Codeml M8 vs M8a model (Gordon et al., 2020). One example is the Zinc finger protein ZNF318 that has some marks of positive selection during primate evolution in our analyses (Supplementary Information). However, the overall dN/dS estimate for each gene was highly similar between the two studies (Figure S3A) and we detected most of the genes they identified under positive selection: 38/40 VIPs under positive selection in Gordon *et al* also detected by $\geqslant$1 DGINN method, including 28 by $\geqslant$3 DGINN methods (Figure S3B-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 (Figure 1C-D). Beyond Gordon *et al*., 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 (Figure S5, (Parkinson et al.,

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2020)). Analyses of pathway enrichment showed that positively selected VIPs are strongly associated with cell cycle control and centrosome behavior biological pathways (Figure S6), suggesting that the control of cell division, and perhaps centrosome-regulated cell polarization, are important for SARS coronavirus *in vivo*.

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) (Figure 1C-D). This list notably includes the ACE2 receptor of SARS coronaviruses that has undergone positive selection in both primates and bats (Figure 1B,D). It also includes known drug targets, such as the metalloprotease ADAM9 (Carapito et al.), the ITGB1 integrin (Sigrist et al., 2020), and POLA1 from the Prim-Pol primase complex (Chaudhuri, 2021) (Supplementary Information) (Figure 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 evolution – being under positive selection only in primates (64 VIPs) or bats (20 VIPs) (Figure 1C-D) – including TMPRSS2, FYCO1, RIPK1, ZNF318 and the Prim-Pol primase complex (Supplementary Information) 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 (Guirimand et al., 2015) for interconnection with SARS-1 and MERS (beta coronaviruses), and 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

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

least another coronavirus (Figure 2A). 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 (Mitchell et al., 2013), we similarly investigated whether rapidly evolving SARS-CoV-2 VIPs were also functionally linked to other viral families (Figure 2B). 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 co-factors of viral replication (Figure 2B). On the other hand, several genes, such as FYCO1, ZNF318 or TMPRSS2, are interconnected with only 1-2 other viral families and may therefore represent more specialized VIPs (Figure 2B). Of note, although the TMPRSS2 co-entry factor has no other interactor in this analysis (Figure 2B, Table S2), it is a host factor for influenza virus entry (Böttcher et al., 2006; Limburg et al.). Lastly, the ACE2 receptor and other genes (Table S2) were not known to interact with other viruses, and therefore likely represent coronavirus-specific VIPs (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 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 (Figure 3A) (Bestle et al., 2020; Hoffmann et al., 2020). In addition to SARS-CoV-2, other coronaviruses, including HCoV-229E, MERS-CoV, SARS-CoV-1, enter human cells in a TMPRSS2-dependent manner (Bertram et al., 2013; Iwata-Yoshikawa et al.,

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2019; Matsuyama et al., 2010). Whilst the genetic and functional adaptation of ACE2 has been studied (Demogines et al., 2012; Frank et al., 2020), 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). 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: 0.0095 and < $4.27\ 10^{-06}$, 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: 1 and 1, respectively) (Table 1).

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 412 – numbering from the human TMPRSS2 sequence) that were significantly detected under positive selection by at least two independent methods (Table 1, Figure 3B). Of note, position 197, which is polymorphic in human TMPRSS2 (rs12329760, V197M) and may be associated with COVID-19 severity ((Jeon et al., 2021) p-value around 10^{-5} above the 10^{-8} significance threshold commonly used in GWAS multiple testing), encoded for a conserved valine in all non-human primate sequences. Because the SARS-CoV-2 – TMPRSS2 interface is currently unknown, only *in silico* molecular docking studies have predicted the substrate binding region (Brooke and Prischi, 2020; Hoffmann et al., 2020; Rangel et al., 2020; SENAPATI et al., 2021). Remarkably, the sites under positive selection cluster nearby or within the predicted

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SARS-CoV-2-host interface (Figure 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 α1-antitrypsin (α1AT), Camostat mesylate, Nafamostat and Bromhexine hydrochloride inhibitors (Hoffmann et al., 2020; Li et al., 2021; Wettstein et al., 2021) and newly reported N-0385 (Shapira et al., 2022), 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 412 was particularly high in Hominoids, but low in Old World monkeys (Figure 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 which exhibited high variability at positions 263 and 360 (Figure 3E). In bats, although none of the models identified significant positive selection in TMPRSS2, the sites 260, 360 and 412 were also variable (Figure 3D). Comparing the variability between resistant and susceptible (naturally or experimentally) species to SARS-CoVs and MERS-CoVs did not reveal any clear pattern (Figure 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 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 interspecies variability in TMPRSS2.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Evidence that FYCO1 is involved in SARS coronavirus pathogenesis or replication at different time scales during primate evolution

FYCO1 (FYVE and coiled-coil domain containing 1) is involved in microtubule transport and autophagy (Figure 4A). Autophagy is an important degradation process of cytoplasmic proteins and organelles, which may be dysregulated during aging, diseases, and by pathogens. FYCO1 acts as an adaptor protein allowing the microtubule transport of autophagosomes in a STK4-LC3B-FYCO1 axis (Cheng et al., 2016; Nieto-Torres et al., 2021). Mutations of the human *FYCO1* gene cause autosomal-recessive congenital cataract, a major cause of vision dysfunction and blindness (Chen et al., 2011; Satoh et al., 2021). 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 genome-wide association studies (GWAS) to be significantly associated with severe COVID-19 (Pairo-Castineira et al., 2021; The Severe Covid-19 GWAS Group, 2020)(The COVID-19 Host Genetics Initiative, 2020). 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 MAIC score (Figure S5B, (Parkinson et al., 2020)), indicating that several studies suspect its involvement in SARS-CoV-2 pathogenesis or replication, including Gordon and colleagues that identified human FYCO1 interaction with SARS-CoV-2 NSP13 (Gordon et al., 2020).

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 new comprehensive positive selection analyses confirmed that FYCO1 has undergone positive selection in primates, but not in bats (Table 1).

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Site-specific selection analyses identified four residues with strong evidence of significant positive selection in primates in at least two independent methods: 447, 471, 552, and 928 (Figure 4B-C, Table 1). Though 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 (Figure 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 from Pairo-Castineira and colleagues (https://genomicc.org/data/), we identified five codons in FYCO1 with polymorphisms associated with severe COVID-19 in humans (Figure 4D). By comparing these positions to the four positively selected sites in primates, we found one position in common, site 447 (genome position 45967996) (Figure 4C-E). This shows that residue 447, whose alleles are 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 430-555 and 910-1005 both have several residues associated with severe COVID-19 in humans and residues under adaptive evolution in primates (Figure 4D-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 coronavirus 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

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (Figure 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 pro-survival, apoptotic or necroptotic signals (Figure 5A) (Delanghe et al., 2020; Liu et al., 2018). 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; Figure S8A). RIPK1 interacts with SARS-CoV-2 NSP12 (RdRp) (Gordon et al., 2020), and is further involved in several bacterial and viral infections, being usurped by pathogens or involved in anti-microbial immunity (Figure S8B).

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 (n=18 species) and primate (n=29 species) RIPK1 and performed new codon alignments and analyses. Here, we found strong evidence of positive selection in bat RIPK1, but not in primates (Table 1). This is different from our screen results, and this discrepancy was mostly due to (i) 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 (ii) the high-quality codon alignments, which are crucial for positive selection studies.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Next, using site-specific analyses, we identified five residues in bat RIPK1 that have evolved under significant positive selection (Figure 5B, Table 1). 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 (Delanghe et al., 2020; Liu et al., 2018). To determine where the positively selected sites fall in the three-dimensional protein, we used a structure prediction of bat RIPK1 from *Rhinolophus ferrumequinum,* which genus is a SARS-CoV-2 reservoir. We found that the rapidly evolving sites are exposed at the protein surface (Figure 5B-C; and Figure S8B for a comparison with the predicted 3D structure of human RIPK1, (Mompeán et al., 2018)). Therefore, physicochemical variations at sites 662 and 665 (Figure 5D) in the death domain could modulate interactions with DD-bearing proteins, and thus influence the ability of bat RIPK1 to drive cell death (Grimm et al., 1996). 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 (Delanghe et al., 2020; Simpson et al., 2021) (Figure 5B in red). The posttranslational modifications at these sites are very important for the balance between the pro-survival and the pro-cell death functions of human RIPK1. It is thus possible that variation at these residues (Figure 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 (i) as determinants of virus-host interfaces, and (ii) as regulators of the protein balance between pro-survival and pro-cell death activities. The latter may allow certain bat species to tolerate viral infections and regulate the associated inflammation.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (i) past epidemics of pathogenic coronaviruses in bats and primates shaping mammalian genomes, and (ii) 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 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 these rapidly evolving residues in SARS-CoV-2 pathogenesis and replication. Overall, our study identified several host factors that (i) have been driven by ancient epidemics of pathogenic SARS coronaviruses, (ii) are different between the bat reservoir and the primate host, and (iii) 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

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

hosts (i.e. specialization). Therefore, SARS-CoVs may have adapted to usurp and/or antagonize different cellular proteins in the primates versus the bats. This is exemplified by the evolution of the entry factor TMPRSS2 (amongst others). We identified strong evidence of virus-host arms-races in primates, but not in bats, suggesting that SARS-CoVs may not strongly rely on TMPRSS2 for entry in bat cells, as opposed to primates. 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 intra-host species plasticity at these interfaces (Meng et al., 2022; Peacock et al., 2022; Pia and Rowland-Jones, 2022; Willett et al., 2022). 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 (Soveg et al., 2021; Wickenhagen et al., 2021), but *Rhinolophidae* bats do not encode for an OAS1 capable to interact with SARS-CoV-2 (Wickenhagen et al., 2021). 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 (Christie et al., 2021). For example, the inflammatory protein RIPK1 showed signatures of adaptive evolution in bat residues that correspond with 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

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

S358 phosphorylation site in bat STING that participates in a dampened inflammation response in bats (Xie et al., 2018), 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 (Ahn et al., 2019; De La Cruz-Rivera et al., 2018; Irving; Pavlovich et al., 2018; Prescott et al.; Xie et al., 2018; Zhang et al., 2013).

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 (Emerman and Malik, 2010; Irving). However, evidence of strong positive selection in the bat ACE2 receptor driven by ancient pathogenic SARS-CoVs (this study, and (Demogines et al., 2012; Frank et al., 2020)) 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 inter-species 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 (Souilmi et al., 2021). 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 found that marks of positive selection have been particularly strong at non-coding regions of SARS-CoV interacting proteins (Souilmi et al., 2021). However, these analyses are challenging at the interspecies level, and more methods and high-quality genome alignments would be

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (Duggal and Emerman, 2012; Sawyer et al., 2005; Sironi et al., 2015). These 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, amongst others, as potential targets of interest. Primidone, an FDA-approved RIPK1 inhibitor, has proven ineffective as a direct inhibitor of viral replication in established cell lines (Gordon et al., 2020; Riebeling et al., 2021). 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-pathogens 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., 2020. Briefly, CCDS identifiers were downloaded from HGNC Biomart (http://biomart.genenames.org/) for all 334 genes of interest. If there were 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 (Löytynoja and Goldman, 2008); PhyML v3.2 (Guindon et al., 2010)). Duplication events were detected through the combined use of Long Branch Detection and Treerecs (Comte et al., 2020) as implemented in DGINN. Recombination

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

events were detected through the use of GARD (Kosakovsky Pond et al., 2006) 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 (Murrell et al., 2012, 2015; Pond et al., 2005) and codon substitution models from PAML codeml (M0, M1, M2, M7, M8) (Yang, 2007), and from Bio++ (M0, $M1^{NS}$, $M2^{NS}$, $M7^{NS}$, $M8^{NS}$) (Guéguen et al., 2013) as implemented in DGINN (Picard et al., 2020). 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 re-analyzed with DGINN starting at the “alignment” step and included in the final results (Figure S1A, Table S1).

MAIC Scores. MAIC scores were obtained from the database for COVID-19 (https://baillielab.net/maic/covid19/, 2020-11-25 release) (Parkinson et al., 2020). 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 (Löytynoja and Goldman, 2008) or Muscle Translate (Edgar, 2004), and phylogenetic trees were built using PhyML with HKY+I+G model and aLRT for branch support (Guindon et al., 2010). Each curated gene alignment and tree were then submitted to positive selection analyses using the DGINN pipeline: 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 screen”). To test for

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

statistical significance of positive selection in Codeml and Bio++, we ran a chi-squared test of the LRT from models disallowing positive selection versus models allowing for positive selection (M1 versus M2, and M7 versus M8) to derive p-values. To identify the sites under positive selection, we used HYPHY MEME (p value < 0.05), the Bayes Empirical Bayes statistics (BEB) from the codeml M2 and M8 models (BEB > 0.95), and the Bayesian Posterior Probabilities (PP) from the $M2^{NS}$ and $M8^{NS}$ models in Bio++ (PP > 0.95). Three other web-based methods were used for this set of genes: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection (FUBAR) method to detect site-specific positive selection (PP > 0.90) (Murrell et al., 2013) and An adaptive branch-site REL test for episodic diversification (aBS-REL) to detect branch/lineage-specific positive selection (p-value < 0.1) (Smith et al., 2015).

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. 2020 from the data publicly available at https://genomicc.org/data/ (Pairo-Castineira et al., 2021). 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 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 FIV plugin in Cytoscape (Gillespie et al., 2022).

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Protein structure predictions. Protein structure prediction of human and *Rhinolophus ferrumequinum* RIPK1 were modeled using RaptorX (Wang et al., 2017) and structures were visualized using the Chimera software (Pettersen et al., 2004).

Since no crystal structure is available for TMPRSS2 protease, the 3D structure of TMPRSS2 was predicted using the Iterative-Threading ASSEmbly Refinement (I-TASSER) server (Yang and Zhang, 2015). A total of 492 amino acid sequence of human TMPRSS2 obtained from NCBI Genbank (accession number AF329454) was used as 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: model C-score, -0.41; estimated TM-score, 0.66±0.13; RMSD, 8.2±4.5Å. The corresponding TMPRSS2 structure was generated using Swiss PDB viewer software (Johansson et al., 2012).

Sequence logo generation. The amino acid sequence logos of TMPRSS2 were generated using WebLogo (V. 2.8.2, (Crooks et al., 2004)), based on an alignment of the positively sites from mammalian species reported as naturally susceptible and/or experimentally permissive to SARS-COV2, SARS-COV and MERS-COV.

Code and Data availability. All codes are available in: https://gitbio.ens-lyon.fr/ciri/ps_sars-cov-2/2021_dginn_covid19, and the DGINN pipeline is available at: http://bioweb.me/DGINN-github. 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 https://virhostnet.prabi.fr/virhostevol/ . The Shiny web application open-source code is available from the IN2P3 gitlab https://gitlab.in2p3.fr/vincent.navratil/shinyapps-virhostevol.

Acknowledgements

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 UCSF 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 Francais 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. Funding: LE and LG 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"). LE 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. AM is supported by the Fondation pour la Recherche Médicale, FRM:AJE201912009932.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Ahn, M., Anderson, D.E., Zhang, Q., Tan, C.W., Lim, B.L., Luko, K., Wen, M., Chia, W.N., Mani, S., Wang, L.C., et al. (2019). Dampened NLRP3-mediated inflammation in bats and implications for a special viral reservoir host. Nat. Microbiol. *4*, 789–799. https://doi.org/10.1038/s41564-019-0371-3.

Baranovskiy, A.G., Babayeva, N.D., Zhang, Y., Gu, J., Suwa, Y., Pavlov, Y.I., and Tahirov, T.H. (2016). Mechanism of Concerted RNA-DNA Primer Synthesis by the Human Primosome. J. Biol. Chem. *291*, 10006–10020. https://doi.org/10.1074/jbc.M116.717405.

Bastard, P., Rosen, L.B., Zhang, Q., Michailidis, E., Hoffmann, H.-H., Zhang, Y., Dorgham, K., Philippot, Q., Rosain, J., Béziat, V., et al. (2020). Autoantibodies against type I IFNs in patients with life-threatening COVID-19. Science *370*, eabd4585. https://doi.org/10.1126/science.abd4585.

Bertram, S., Dijkman, R., Habjan, M., Heurich, A., Gierer, S., Glowacka, I., Welsch, K., Winkler, M., Schneider, H., Hofmann-Winkler, H., et al. (2013). 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. https://doi.org/10.1128/JVI.03372-12.

Bestle, D., Heindl, M.R., Limburg, H., Van, T.V.L., Pilgram, O., Moulton, H., Stein, D.A., Hardes, K., Eickmann, M., Dolnik, O., et al. (2020). TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. Life Sci. Alliance *3*. https://doi.org/10.26508/lsa.202000786.

Böttcher, E., Matrosovich, T., Beyerle, M., Klenk, H.-D., Garten, W., and Matrosovich, M. (2006). Proteolytic Activation of Influenza Viruses by Serine Proteases TMPRSS2 and HAT from Human Airway Epithelium. J. Virol. *80*, 9896–9898. https://doi.org/10.1128/JVI.01118-06.

Boys, I.N., Xu, E., Mar, K.B., De La Cruz-Rivera, P.C., Eitson, J.L., Moon, B., and Schoggins, J.W. (2020). 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. https://doi.org/10.1016/j.chom.2020.09.014.

Brooke, G.N., and Prischi, F. (2020). Structural and functional modelling of SARS-CoV-2 entry in animal models. Sci. Rep. *10*, 15917. https://doi.org/10.1038/s41598-020-72528-z.

Carapito, R., Li, R., Helms, J., Carapito, C., Gujja, S., Rolli, V., Guimaraes, R., Malagon-Lopez, J., Spinnhirny, P., Lederle, A., et al. Identification of driver genes for critical forms of COVID-19 in a deeply phenotyped young patient cohort. Sci. Transl. Med. *14*, eabj7521. https://doi.org/10.1126/scitranslmed.abj7521.

Chaudhuri, A. (2021). 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. https://doi.org/10.1016/j.molstruc.2021.130854.

Chen, J., Ma, Z., Jiao, X., Fariss, R., Kantorow, W.L., Kantorow, M., Pras, E., Frydman, M., Pras, E., Riazuddin, S., et al. (2011). Mutations in FYCO1 cause autosomal-recessive congenital cataracts. Am. J. Hum. Genet. *88*, 827–838. https://doi.org/10.1016/j.ajhg.2011.05.008.

Cheng, X., Wang, Y., Gong, Y., Li, F., Guo, Y., Hu, S., Liu, J., and Pan, L. (2016). Structural basis of FYCO1 and MAP1LC3A interaction reveals a novel binding mode for Atg8-family proteins. Autophagy *12*, 1330–1339. https://doi.org/10.1080/15548627.2016.1185590.

Chougui, G., Munir-Matloob, S., Matkovic, R., Martin, M.M., Morel, M., Lahouassa, H., Leduc, M., Ramirez, B.C., Etienne, L., and Margottin-Goguet, F. (2018). HIV-2/SIV viral protein X counteracts HUSH repressor complex. Nat. Microbiol. *3*, 891–897. https://doi.org/10.1038/s41564-018-0179-6.

Christie, M.J., Irving, A.T., Forster, S.C., Marsland, B.J., Hansbro, P.M., Hertzog, P.J., Nold-Petry, C.A., and Nold, M.F. (2021). Of bats and men: Immunomodulatory treatment options for COVID-19 guided by the immunopathology of SARS-CoV-2 infection. Sci. Immunol. *6*, eabd0205. https://doi.org/10.1126/sciimmunol.abd0205.

Comte, N., Morel, B., Hasić, D., Guéguen, L., Boussau, B., Daubin, V., Penel, S., Scornavacca, C., Gouy, M., Stamatakis, A., et al. (2020). Treerecs: an integrated phylogenetic tool, from sequences to reconciliations. Bioinformatics *36*, 4822–4824. https://doi.org/10.1093/bioinformatics/btaa615.

Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: A Sequence Logo Generator. Genome Res. *14*, 1188–1190. https://doi.org/10.1101/gr.849004.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Crow, Y.J., and Stetson, D.B. (2021). The type I interferonopathies: 10 years on. Nat. Rev. Immunol. 1–13. https://doi.org/10.1038/s41577-021-00633-9.

Daugherty, M.D., and Malik, H.S. (2012). Rules of Engagement: Molecular Insights from Host-Virus Arms Races. Annu. Rev. Genet. *46*, 677–700. https://doi.org/10.1146/annurev-genet-110711-155522.

De La Cruz-Rivera, P.C., Kanchwala, M., Liang, H., Kumar, A., Wang, L.-F., Xing, C., and Schoggins, J.W. (2018). The IFN Response in Bats Displays Distinctive IFN-Stimulated Gene Expression Kinetics with Atypical RNASEL Induction. J. Immunol. Baltim. Md 1950 *200*, 209–217. https://doi.org/10.4049/jimmunol.1701214.

Delanghe, T., Dondelinger, Y., and Bertrand, M.J.M. (2020). RIPK1 Kinase-Dependent Death: A Symphony of Phosphorylation Events. Trends Cell Biol. *30*, 189–200. https://doi.org/10.1016/j.tcb.2019.12.009.

Demogines, A., Farzan, M., and Sawyer, S.L. (2012). Evidence for ACE2-Utilizing Coronaviruses (CoVs) Related to Severe Acute Respiratory Syndrome CoV in Bats. J. Virol. *86*, 6350–6353. https://doi.org/10.1128/JVI.00311-12.

Douse, C.H., Tchasovnikarova, I.A., Timms, R.T., Protasio, A.V., Seczynska, M., Prigozhin, D.M., Albecka, A., Wagstaff, J., Williamson, J.C., Freund, S.M.V., et al. (2020). TASOR is a pseudo-PARP that directs HUSH complex assembly and epigenetic transposon control. Nat. Commun. *11*, 4940. https://doi.org/10.1038/s41467-020-18761-6.

Duggal, N.K., and Emerman, M. (2012). Evolutionary conflicts between viruses and restriction factors shape immunity. Nat. Rev. Immunol. *12*, 687–695. https://doi.org/10.1038/nri3295.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797. https://doi.org/10.1093/nar/gkh340.

Emerman, M., and Malik, H.S. (2010). Paleovirology—Modern Consequences of Ancient Viruses. PLOS Biol. *8*, e1000301. https://doi.org/10.1371/journal.pbio.1000301.

Enard, D., Cai, L., Gwennap, C., and Petrov, D.A. (2016). Viruses are a dominant driver of protein adaptation in mammals. ELife *5*, e12469. https://doi.org/10.7554/eLife.12469.

Frank, H.K., Enard, D., and Boyd, S.D. (2020). Exceptional diversity and selection pressure on SARS-CoV and SARS-CoV-2 host receptor in bats compared to other mammals. 2020.04.20.051656. https://doi.org/10.1101/2020.04.20.051656.

Fregoso, O.I., Ahn, J., Wang, C., Mehrens, J., Skowronski, J., and Emerman, M. (2013). Evolutionary Toggling of Vpx/Vpr Specificity Results in Divergent Recognition of the Restriction Factor SAMHD1. PLOS Pathog. *9*, e1003496. https://doi.org/10.1371/journal.ppat.1003496.

Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., Senff-Ribeiro, A., Griss, J., Sevilla, C., Matthews, L., Gong, C., et al. (2022). The reactome pathway knowledgebase 2022. Nucleic Acids Res. *50*, D687–D692. https://doi.org/10.1093/nar/gkab1028.

Gordon, D.E., Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., White, K.M., O'Meara, M.J., Rezelj, V.V., Guo, J.Z., Swaney, D.L., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature *583*, 459–468. https://doi.org/10.1038/s41586-020-2286-9.

Grimm, S., Stanger, B.Z., and Leder, P. (1996). 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. .

Guéguen, L., Gaillard, S., Boussau, B., Gouy, M., Groussin, M., Rochette, N.C., Bigot, T., Fournier, D., Pouyet, F., Cahais, V., et al. (2013). Bio++: Efficient Extensible Libraries and Tools for Computational Molecular Evolution. Mol. Biol. Evol. *30*, 1745–1750. https://doi.org/10.1093/molbev/mst097.

Guilliam, T.A., Keen, B.A., Brissett, N.C., and Doherty, A.J. (2015). Primase-polymerases are a functionally diverse superfamily of replication and repair enzymes. Nucleic Acids Res. *43*, 6651–6664. https://doi.org/10.1093/nar/gkv625.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst. Biol. *59*, 307–321. https://doi.org/10.1093/sysbio/syq010.

Guirimand, T., Delmotte, S., and Navratil, V. (2015). VirHostNet 2.0: surfing on the web of virus/host molecular interactions data. Nucleic Acids Res. *43*, D583–D587. https://doi.org/10.1093/nar/gku1121.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.-H., Nitsche, A., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell *181*, 271-280.e8. https://doi.org/10.1016/j.cell.2020.02.052.

Irving, A.T. Interferon Regulatory Factors IRF1 and IRF7 Directly Regulate Gene Expression in Bats in Response to Viral Infection. OPEN ACCESS 27. .

Iwata-Yoshikawa, N., Okamura, T., Shimizu, Y., Hasegawa, H., Takeda, M., and Nagata, N. (2019). TMPRSS2 Contributes to Virus Spread and Immunopathology in the Airways of Murine Models after Coronavirus Infection. J. Virol. *93*, e01815-18. https://doi.org/10.1128/JVI.01815-18.

Jeon, S., Blazyte, A., Yoon, C., Ryu, H., Jeon, Y., Bhak, Y., Bolser, D., Manica, A., Shin, E.-S., Cho, Y.S., et al. (2021). Regional TMPRSS2 V197M Allele Frequencies Are Correlated with COVID-19 Case Fatality Rates. Mol. Cells *44*, 680–687. https://doi.org/10.14348/molcells.2021.2249.

Johansson, M.U., Zoete, V., Michielin, O., and Guex, N. (2012). Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. BMC Bioinformatics *13*, 173. https://doi.org/10.1186/1471-2105-13-173.

Kilkenny, M.L., Veale, C.E., Guppy, A., Hardwick, S.W., Chirgadze, D.Y., Rzechorzek, N.J., Maman, J.D., and Pellegrini, L. (2022). Structural basis for the interaction of SARS-CoV-2 virulence factor nsp1 with DNA polymerase α–primase. Protein Sci. *31*, 333–344. https://doi.org/10.1002/pro.4220.

Kirchdoerfer, R.N., and Ward, A.B. (2019). Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. Nat. Commun. *10*, 2342. https://doi.org/10.1038/s41467-019-10280-3.

Konkolova, E., Klima, M., Nencka, R., and Boura, E. (2020). Structural analysis of the putative SARS-CoV-2 primase complex. J. Struct. Biol. *211*, 107548. https://doi.org/10.1016/j.jsb.2020.107548.

Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H., and Frost, S.D.W. (2006). GARD: a genetic algorithm for recombination detection. Bioinformatics *22*, 3096–3098. https://doi.org/10.1093/bioinformatics/btl474.

Li, K., Meyerholz, D.K., Bartlett, J.A., and McCray, P.B. (2021). The TMPRSS2 Inhibitor Nafamostat Reduces SARS-CoV-2 Pulmonary Infection in Mouse Models of COVID-19. MBio *12*, e0097021. https://doi.org/10.1128/mBio.00970-21.

Limburg, H., Harbig, A., Bestle, D., Stein, D.A., Moulton, H.M., Jaeger, J., Janga, H., Hardes, K., Koepke, J., Schulte, L., 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*, e00649-19. https://doi.org/10.1128/JVI.00649-19.

Liu, X., Li, Y., Peng, S., Yu, X., Li, W., Shi, F., Luo, X., Tang, M., Tan, Z., Bode, A.M., et al. (2018). Epstein-Barr virus encoded latent membrane protein 1 suppresses necroptosis through targeting RIPK1/3 ubiquitination. Cell Death Dis. *9*, 53. https://doi.org/10.1038/s41419-017-0081-9.

Löytynoja, A., and Goldman, N. (2008). Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. Science *320*, 1632–1635. https://doi.org/10.1126/science.1158395.

Matsuyama, S., Nagata, N., Shirato, K., Kawase, M., Takeda, M., and Taguchi, F. (2010). Efficient Activation of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein by the Transmembrane Protease TMPRSS2. J. Virol. *84*, 12658–12664. https://doi.org/10.1128/JVI.01542-10.

Meng, B., Abdullahi, A., Ferreira, I.A.T.M., Goonawardane, N., Saito, A., Kimura, I., Yamasoba, D., Gerber, P.P., Fatihi, S., Rathore, S., et al. (2022). Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity. Nature 1–9. https://doi.org/10.1038/s41586-022-04474-x.

Mitchell, P.S., Emerman, M., and Malik, H.S. (2013). An evolutionary perspective on the broad antiviral specificity of MxA. Curr. Opin. Microbiol. *16*, 493–499. https://doi.org/10.1016/j.mib.2013.04.005.

Mompeán, M., Li, W., Li, J., Laage, S., Siemer, A.B., Bozkurt, G., Wu, H., and McDermott, A.E. (2018). The Structure of the Necrosome RIPK1-RIPK3 Core, a Human Hetero-Amyloid Signaling Complex. Cell *173*, 1244-1253.e10. https://doi.org/10.1016/j.cell.2018.03.032.

Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K., and Kosakovsky Pond, S.L. (2012). Detecting individual sites subject to episodic diversifying selection. PLoS Genet. *8*, e1002764. https://doi.org/10.1371/journal.pgen.1002764.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S.L., and Scheffler, K. (2013). FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. Mol. Biol. Evol. *30*, 1196–1205. https://doi.org/10.1093/molbev/mst030.

Murrell, B., Weaver, S., Smith, M.D., Wertheim, J.O., Murrell, S., Aylward, A., Eren, K., Pollner, T., Martin, D.P., Smith, D.M., et al. (2015). Gene-wide identification of episodic selection. Mol. Biol. Evol. *32*, 1365–1371. https://doi.org/10.1093/molbev/msv035.

Narayanan, K., Huang, C., Lokugamage, K., Kamitani, W., Ikegami, T., Tseng, C.-T.K., and Makino, S. (2008). Severe Acute Respiratory Syndrome Coronavirus nsp1 Suppresses Host Gene Expression, Including That of Type I Interferon, in Infected Cells. J. Virol. *82*, 4471–4479. https://doi.org/10.1128/JVI.02472-07.

Nieto-Torres, J.L., Shanahan, S.-L., Chassefeyre, R., Chaiamarit, T., Zaretski, S., Landeras-Bueno, S., Verhelle, A., Encalada, S.E., and Hansen, M. (2021). LC3B phosphorylation regulates FYCO1 binding and directional transport of autophagosomes. Curr. Biol. *31*, 3440-3449.e7. https://doi.org/10.1016/j.cub.2021.05.052.

Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A.D., Rawlik, K., Pasko, D., Walker, S., Parkinson, N., Fourman, M.H., Russell, C.D., et al. (2021). Genetic mechanisms of critical illness in COVID-19. Nature *591*, 92–98. https://doi.org/10.1038/s41586-020-03065-y.

Parkinson, N., Rodgers, N., Head Fourman, M., Wang, B., Zechner, M., Swets, M.C., Millar, J.E., Law, A., Russell, C.D., Baillie, J.K., et al. (2020). Dynamic data-driven meta-analysis for prioritisation of host genes implicated in COVID-19. Sci. Rep. *10*, 22303. https://doi.org/10.1038/s41598-020-79033-3.

Pavlovich, S.S., Lovett, S.P., Koroleva, G., Guito, J.C., Arnold, C.E., Nagle, E.R., Kulcsar, K., Lee, A., Thibaud-Nissen, F., Hume, A.J., et al. (2018). The Egyptian Rousette Genome Reveals Unexpected Features of Bat Antiviral Immunity. Cell *173*, 1098-1110.e18. https://doi.org/10.1016/j.cell.2018.03.070.

Peacock, T.P., Brown, J.C., Zhou, J., Thakur, N., Newman, J., Kugathasan, R., Sukhova, K., Kaforou, M., Bailey, D., and Barclay, W.S. (2022). The SARS-CoV-2 variant, Omicron, shows rapid replication in human primary nasal epithelial cultures and efficiently uses the endosomal route of entry. 2021.12.31.474653. https://doi.org/10.1101/2021.12.31.474653.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612. https://doi.org/10.1002/jcc.20084.

Pia, L., and Rowland-Jones, S. (2022). Omicron entry route. Nat. Rev. Immunol. *22*, 144–144. https://doi.org/10.1038/s41577-022-00681-9.

Picard, L., Ganivet, Q., Allatif, O., Cimarelli, A., Guéguen, L., and Etienne, L. (2020). DGINN, an automated and highly-flexible pipeline for the detection of genetic innovations on protein-coding genes. Nucleic Acids Res. *48*, e103. https://doi.org/10.1093/nar/gkaa680.

Pond, S.L.K., Frost, S.D.W., and Muse, S.V. (2005). HyPhy: hypothesis testing using phylogenies. Bioinformatics *21*, 676–679. https://doi.org/10.1093/bioinformatics/bti079.

Prescott, J., Guito, J.C., Spengler, J.R., Arnold, C.E., Schuh, A.J., Amman, B.R., Sealy, T.K., Guerrero, L.W., Palacios, G.F., Sanchez-Lockhart, M., et al. Rousette Bat Dendritic Cells Overcome Marburg Virus-Mediated Antiviral Responses by Upregulation of Interferon-Related Genes While Downregulating Proinflammatory Disease Mediators. MSphere *4*, e00728-19. https://doi.org/10.1128/mSphere.00728-19.

Rangel, H.R., Ortega, J.T., Maksoud, S., Pujol, F.H., and Serrano, M.L. (2020). SARS-CoV-2 host tropism: An in silico analysis of the main cellular factors. Virus Res. *289*, 198154. https://doi.org/10.1016/j.virusres.2020.198154.

Riebeling, T., Jamal, K., Wilson, R., Kolbrink, B., von Samson-Himmelstjerna, F.A., Moerke, C., Ramos Garcia, L., Dahlke, E., Michels, F., Lühder, F., et al. (2021). Primidone blocks RIPK1-driven cell death and inflammation. Cell Death Differ. *28*, 1610–1626. https://doi.org/10.1038/s41418-020-00690-y.

Satoh, K., Takemura, Y., Satoh, M., Ozaki, K., and Kubota, S. (2021). Loss of FYCO1 leads to cataract formation. Sci. Rep. *11*, 13771. https://doi.org/10.1038/s41598-021-93110-1.

Sawyer, S.L., Wu, L.I., Emerman, M., and Malik, H.S. (2005). Positive selection of primate TRIM5α identifies a critical species-specific retroviral restriction domain. Proc. Natl. Acad. Sci. *102*, 2832–2837. https://doi.org/10.1073/pnas.0409853102.

Schoggins, J. (2021). Defective viral RNA sensing linked to severe COVID-19. Science *374*, 535–536. https://doi.org/10.1126/science.abm3921.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Schubert, K., Karousis, E.D., Jomaa, A., Scaiola, A., Echeverria, B., Gurzeler, L.-A., Leibundgut, M., Thiel, V., Mühlemann, O., and Ban, N. (2020). SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. Nat. Struct. Mol. Biol. *27*, 959–966. https://doi.org/10.1038/s41594-020-0511-8.

SENAPATI, S., BANERJEE, P., BHAGAVATULA, S., KUSHWAHA, P.P., and KUMAR, S. (2021). Contributions of human ACE2 and TMPRSS2 in determining host–pathogen interaction of COVID-19. J. Genet. *100*, 12. https://doi.org/10.1007/s12041-021-01262-w.

Shapira, T., Monreal, I.A., Dion, S.P., Buchholz, D.W., Imbiakha, B., Olmstead, A.D., Jager, M., Désilets, A., Gao, G., Martins, M., et al. (2022). A TMPRSS2 inhibitor acts as a pan-SARS-CoV-2 prophylactic and therapeutic. Nature 1–13. https://doi.org/10.1038/s41586-022-04661-w.

Sigrist, C.J., Bridge, A., and Le Mercier, P. (2020). A potential role for integrins in host cell entry by SARS-CoV-2. Antiviral Res. *177*, 104759. https://doi.org/10.1016/j.antiviral.2020.104759.

Simpson, D.S., Gabrielyan, A., and Feltham, R. (2021). RIPK1 ubiquitination: Evidence, correlations and the undefined. Semin. Cell Dev. Biol. *109*, 76–85. https://doi.org/10.1016/j.semcdb.2020.08.008.

Sironi, M., Cagliani, R., Forni, D., and Clerici, M. (2015). Evolutionary insights into host–pathogen interactions from mammalian sequence data. Nat. Rev. Genet. *16*, 224–236. https://doi.org/10.1038/nrg3905.

Smith, M.D., Wertheim, J.O., Weaver, S., Murrell, B., Scheffler, K., and Kosakovsky Pond, S.L. (2015). Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. Mol. Biol. Evol. *32*, 1342–1353. https://doi.org/10.1093/molbev/msv022.

Souilmi, Y., Lauterbur, M.E., Tobler, R., Huber, C.D., Johar, A.S., Moradi, S.V., Johnston, W.A., Krogan, N.J., Alexandrov, K., and Enard, D. (2021). An ancient viral epidemic involving host coronavirus interacting genes more than 20,000 years ago in East Asia. Curr. Biol. *31*, 3504-3514.e9. https://doi.org/10.1016/j.cub.2021.05.067.

Soveg, F.W., Schwerk, J., Gokhale, N.S., Cerosaletti, K., Smith, J.R., Pairo-Castineira, E., Kell, A.M., Forero, A., Zaver, S.A., Esser-Nobis, K., et al. (2021). Endomembrane targeting of human OAS1 p46 augments antiviral activity. ELife *10*, e71047. https://doi.org/10.7554/eLife.71047.

Starokadomskyy, P., Gemelli, T., Rios, J.J., Xing, C., Wang, R.C., Li, H., Pokatayev, V., Dozmorov, I., Khan, S., Miyata, N., et al. (2016). DNA polymerase-α regulates type I interferon activation through cytosolic RNA:DNA synthesis. Nat. Immunol. *17*, 495–504. https://doi.org/10.1038/ni.3409.

Starokadomskyy, P., Wilton, K.M., Krzewski, K., Lopez, A., Sifuentes-Dominguez, L., Overlee, B., Chen, Q., Ray, A., Gil-Krzewska, A., Peterson, M., et al. (2019). NK cell defects in X-linked pigmentary reticulate disorder. JCI Insight *4*. https://doi.org/10.1172/jci.insight.125688.

Starokadomskyy, P., Wilton, K.M., Krzewski, K., Lopez, A., Sifuentes-Dominguez, L., Overlee, B., Chen, Q., Ray, A., Gil-Krzewska, A., Peterson, M., et al. NK cell defects in X-linked pigmentary reticulate disorder. JCI Insight *4*, e125688. https://doi.org/10.1172/jci.insight.125688.

Tao, R.-H., Kawate, H., Ohnaka, K., Ishizuka, M., Hagiwara, H., and Takayanagi, R. (2006). Opposite effects of alternative TZF spliced variants on androgen receptor. Biochem. Biophys. Res. Commun. *341*, 515–521. https://doi.org/10.1016/j.bbrc.2005.12.213.

Temmam, S., Vongphayloth, K., Salazar, E.B., Munier, S., Bonomi, M., Regnault, B., Douangboubpha, B., Karami, Y., Chrétien, D., Sanamxay, D., et al. (2022). Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. Nature 1–10. https://doi.org/10.1038/s41586-022-04532-4.

The COVID-19 Host Genetics Initiative (2020). 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. https://doi.org/10.1038/s41431-020-0636-6.

The Severe Covid-19 GWAS Group (2020). Genomewide Association Study of Severe Covid-19 with Respiratory Failure. N. Engl. J. Med. *383*, 1522–1534. https://doi.org/10.1056/NEJMoa2020283.

Wang, L.-F., and Anderson, D.E. (2019). Viruses in bats and potential spillover to animals and humans. Curr. Opin. Virol. *34*, 79–89. https://doi.org/10.1016/j.coviro.2018.12.007.

Wang, S., Sun, S., Li, Z., Zhang, R., and Xu, J. (2017). Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model. PLOS Comput. Biol. *13*, e1005324. https://doi.org/10.1371/journal.pcbi.1005324.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Wettstein, L., Weil, T., Conzelmann, C., Müller, J.A., Groß, R., Hirschenberger, M., Seidel, A., Klute, S., Zech, F., Prelli Bozzo, C., et al. (2021). Alpha-1 antitrypsin inhibits TMPRSS2 protease activity and SARS-CoV-2 infection. Nat. Commun. *12*, 1726. https://doi.org/10.1038/s41467-021-21972-0.

Wickenhagen, A., Sugrue, E., Lytras, S., Kuchi, S., Noerenberg, M., Turnbull, M.L., Loney, C., Herder, V., Allan, J., Jarmson, I., et al. (2021). A prenylated dsRNA sensor protects against severe COVID-19. Science *374*, eabj3624. https://doi.org/10.1126/science.abj3624.

Willett, B.J., Grove, J., MacLean, O.A., Wilkie, C., Logan, N., Lorenzo, G.D., Furnon, W., Scott, S., Manali, M., Szemiel, A., et al. (2022). The hyper-transmissible SARS-CoV-2 Omicron variant exhibits significant antigenic change, vaccine escape and a switch in cell entry mechanism. 2022.01.03.21268111. https://doi.org/10.1101/2022.01.03.21268111.

Xie, J., Li, Y., Shen, X., Goh, G., Zhu, Y., Cui, J., Wang, L.-F., Shi, Z.-L., and Zhou, P. (2018). Dampened STING-Dependent Interferon Activation in Bats. Cell Host Microbe *23*, 297-301.e4. https://doi.org/10.1016/j.chom.2018.01.006.

Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. *24*, 1586–1591. https://doi.org/10.1093/molbev/msm088.

Yang, J., and Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res. *43*, W174–W181. https://doi.org/10.1093/nar/gkv342.

Yedavalli, V.S., and Jeang, K.-T. (2011). Matrin 3 is a co-factor for HIV-1 Rev in regulating post-transcriptional viral gene expression. Retrovirology *8*, 61. https://doi.org/10.1186/1742-4690-8-61.

Zeberg, H., and Pääbo, S. (2020). The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. Nature *587*, 610–612. https://doi.org/10.1038/s41586-020-2818-3.

Zhai, Y., Sun, F., Li, X., Pang, H., Xu, X., Bartlam, M., and Rao, Z. (2005). Insights into SARS-CoV transcription and replication from the structure of the nsp7–nsp8 hexadecamer. Nat. Struct. Mol. Biol. *12*, 980–986. https://doi.org/10.1038/nsmb999.

Zhang, G., Cowled, C., Shi, Z., Huang, Z., Bishop-Lilly, K.A., Fang, X., Wynne, J.W., Xiong, Z., Baker, M.L., Zhao, W., et al. (2013). Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. Science *339*, 456–460. https://doi.org/10.1126/science.1230835.

Zhang, Q., Bastard, P., Liu, Z., Le Pen, J., Moncada-Velez, M., Chen, J., Ogishi, M., Sabli, I.K.D., Hodeib, S., Korol, C., et al. (2020). Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. Science *370*, eabd4570. https://doi.org/10.1126/science.abd4570.

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure Legends

Figure 1. Identification of the SARS-CoV-2 interactome with signatures of positive selection in bats and primates. A, Overview of the DGINN pipeline to detect adaptive evolution in SARS-CoV-2 VIPs. 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 bpp 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 positive selection (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 (positive selection) or PS, within each host – 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 (see Figure S4 for details). Of note, seven primate genes are "false positive": EMC1 (ER membrane protein complex subunit 1), MOV10 (Mov10 RISC complex RNA helicase), POR (cytochrome p450 oxidoreductase), PITRM1 (pitrilysin metallopeptidase 1), RAB14, RAB2A, and TIMM8B (translocase of inner mitochondrial membrane 8 homolog B).

Figure 2. SARS-CoV-2 VIPs under positive selection are interacting proteins of other coronaviruses, as well as other viral families. Virus-host protein-protein interactions' network of VIP genes under positive selection 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 2-6 different viral families). Node sizes at the virus families are proportional to the number of edges. The VIPs not interconnected are shown in Table S1.

Figure 3. TMPRSS2 has evolved under strong positive selection in primates, but not in bats. A, Role of TMPRSS2 in SARS-CoV-2 entry. B, Diagram of TMPRSS2 predicted domains, with sites under positive selection in primates represented by triangles (Table 1). Codon numbering based on *Homo sapiens* TMPRSS2. C, 3D-structure modeling of human TMPRSS2 (amino acids 1-492) with the positively selected sites (red), the SARS-CoV-2 predicted interface (light blue), the catalytic site (dark blue). D, The positively selected sites identified in primate

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 exhibits different patterns of variability in other mammals: 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 QSSKS motif in *Mustela putoris* was found in eleven rodent species). The corresponding motif in species/cells susceptible or permissive to coronaviruses is shown in Figure S7.

Figure 4. Domains of FYCO1 that are associated with severe COVID-19 in human have also evolved under significant positive selection in primates, but not in bats.
A, Known cellular role of FYCO1. B, Diagram of FYCO1 predicted domains, with sites under positive selection in primates represented by triangles (Table 1). Codon numbering 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 Panel E. Amino-acid color-coding, RasMol properties (Geneious, Biomatters). D, Sites identified in the coding sequence of FYCO1 as under positive selection (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.

Figure 5. The multi-functional 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 pro-survival signals that notably lead to NFkB 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 positive selection in bats (black triangles) with the corresponding positions 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 positive selection are in red and numbered is according to their position in bat RIPK1. D, The positively selected

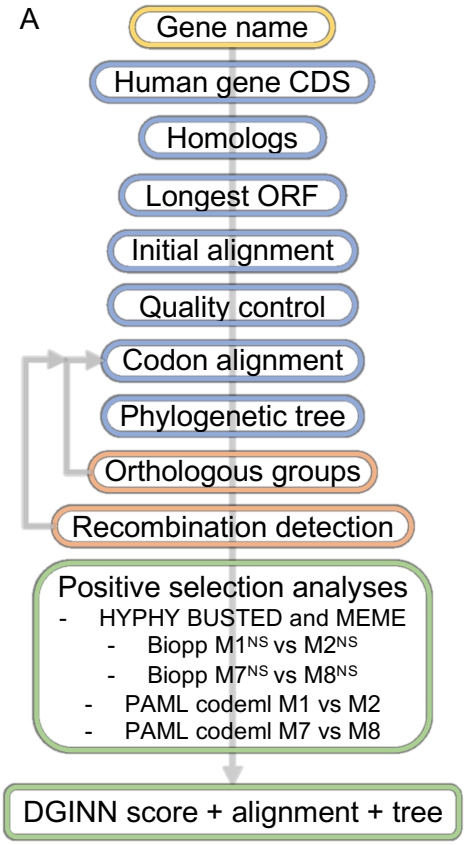
bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (grey) to human numbering (black) is shown at the top.

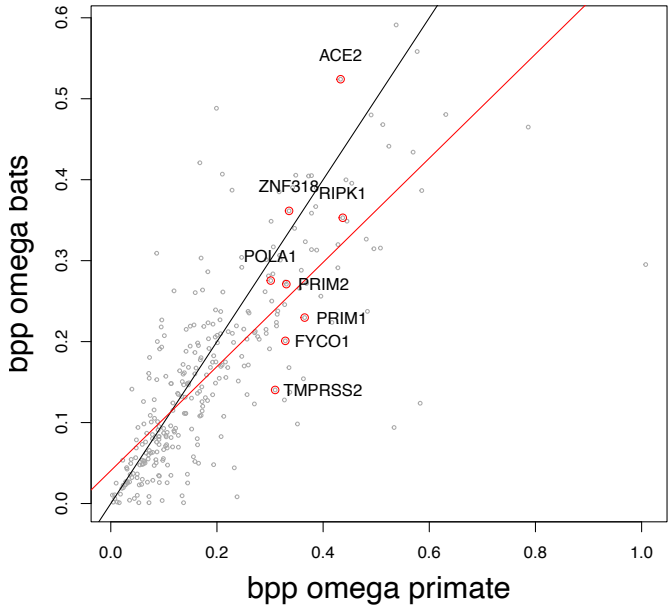
bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1.

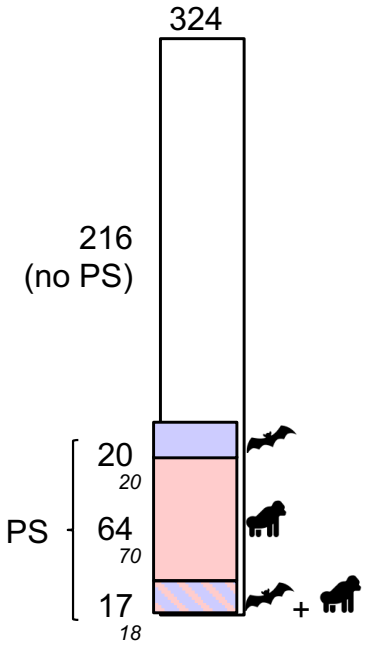
A



B



C



D

Number of methods detecting positive selection in bats (x) and primates (y)

Primates \ Bats	0	1	2	3	4	5
5	ZNF318 REEP6 PRIM2 NDUFAF2	NUP58 DNMT1 CEP68	SEPSECS PVR MYCBP2 HECTD1 FYCO1 EMC1	**CDK5RAP2**	ITGB1 ACADM	VPS39 TOR1AIP1 GOLGA7 GGH ACE2
4	SLC25A21 UBAP2 NDUFB9 LARP4B GHITM TUBGCP2 GCC2 ERO1B DPH5 CEP135 TMPRSS2	SIRT5 SAAL1 PRKAR2A MPHOSPH10 MARK1 LARP1 GLA GIGYF2 UGGT2	TMEM39B RIPK1 PCNT CLIP4 BRD4	**MIPOL1**	POLA1 LMAN2 IDE	SLC27A2
3	NAT14 TRIM59 LARP7 TIMM8B HS6ST2 RBX1 HDAC2 RAB2A GORASP1 RAB1A EXOSC2 RAB14 BZW2 PTBP2 ATE1 NPC2	RPL36 USP54 PLOD2 PLAT PITRM1 TRMT1 NGLY1 EIF4E2 CNTRL AP2A2 STOM	POR PMPCB MRPS5 MDN1 GOLGB1	**SCCPDH** **MOV10** **ADAM9**		RAP1GDS1 EDEM3
2	n= 15	n= 15	n= 10	PRIM1 NUP214 COL6A1	AKAP9	TBK1
1	n= 37	n= 27	n= 13	PUSL1 GRIPAP1 AGPS ACAD9	IMPDH2 CISD3 ATP6AP1	RAB18
0	n= 49	n= 28	n= 22	SLC9A3R1 SLC30A9 MARK2 GRPEL1	ERGIC1	SLC44A2 INHBE

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.07.487460; this version posted April 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2.

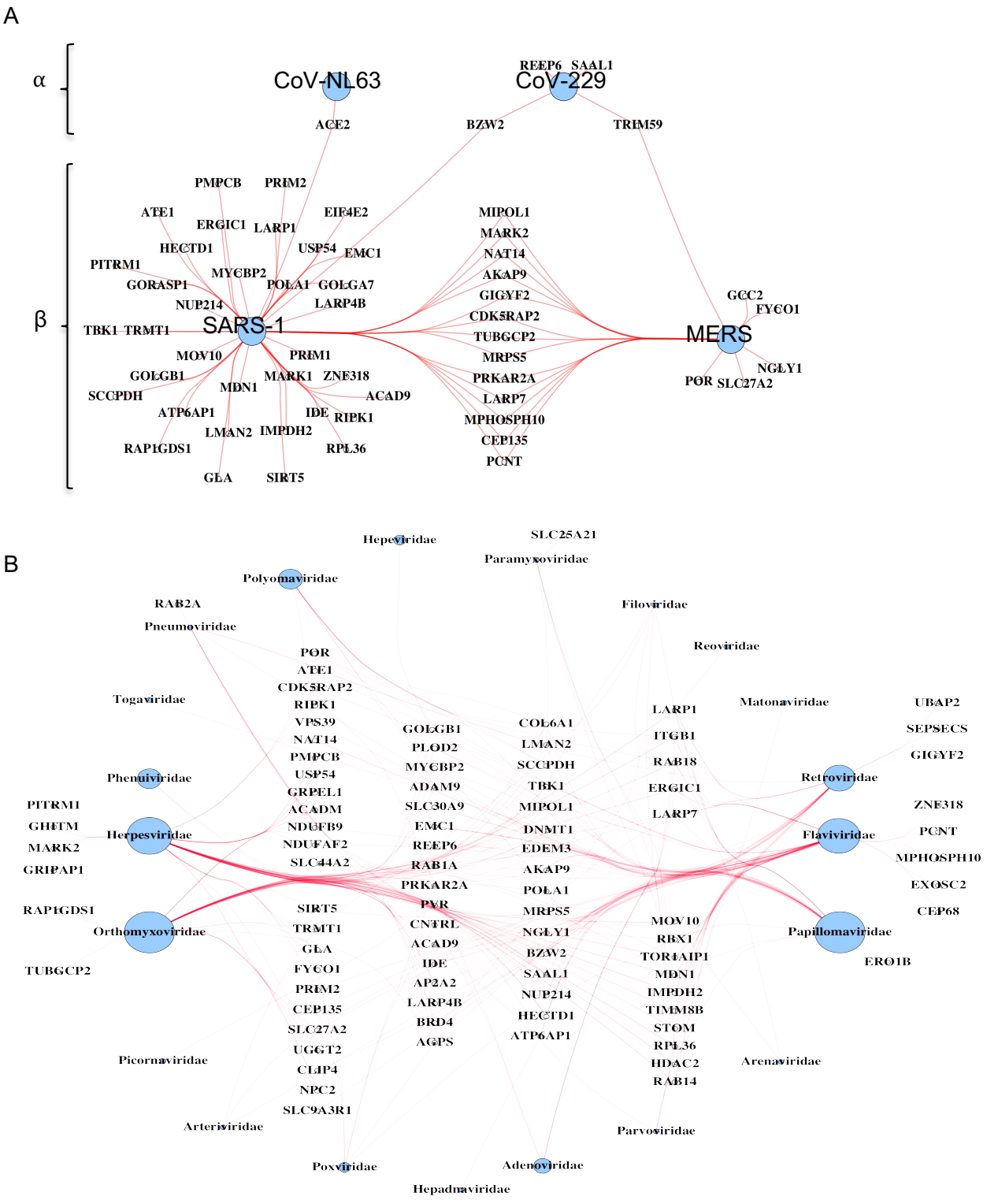
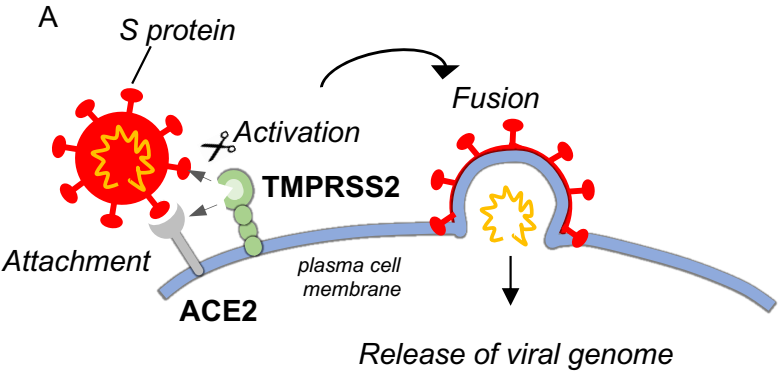
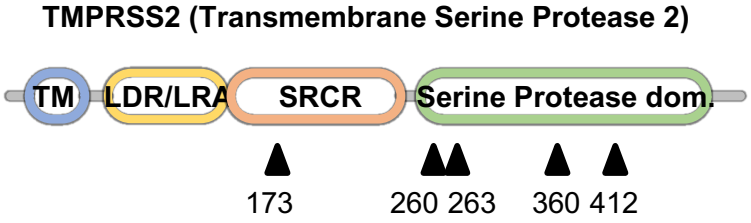
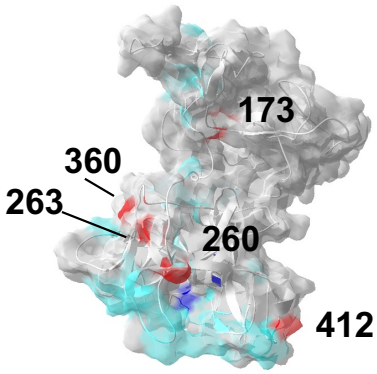
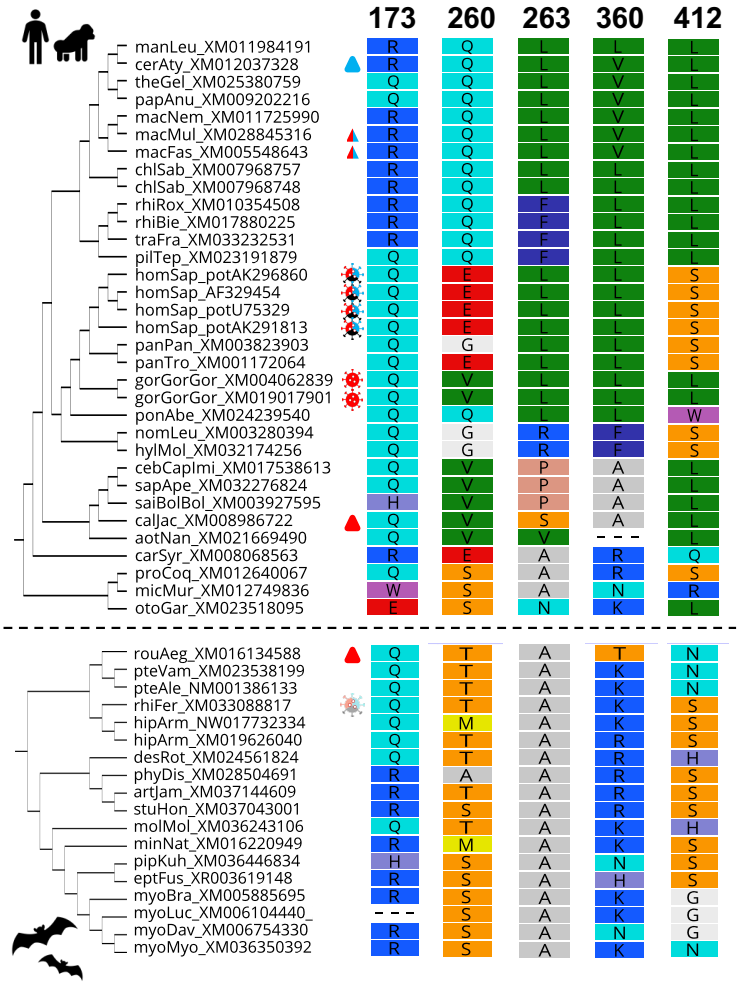


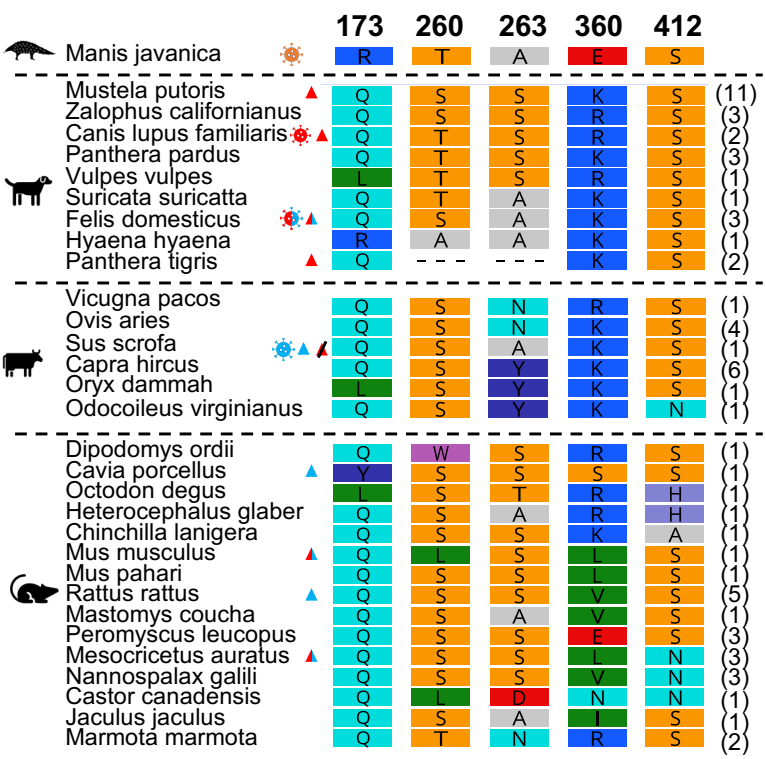
Figure 3.











Natural infection by:
- SARS-2
- SARS-1
- MERS
- Bat SARS-2-like
- Pangolin SARS-2-like
- Bat SARS-1-like
- MERS-like

Experimental infection by:
- SARS-2
- SARS-1

Figure 4.

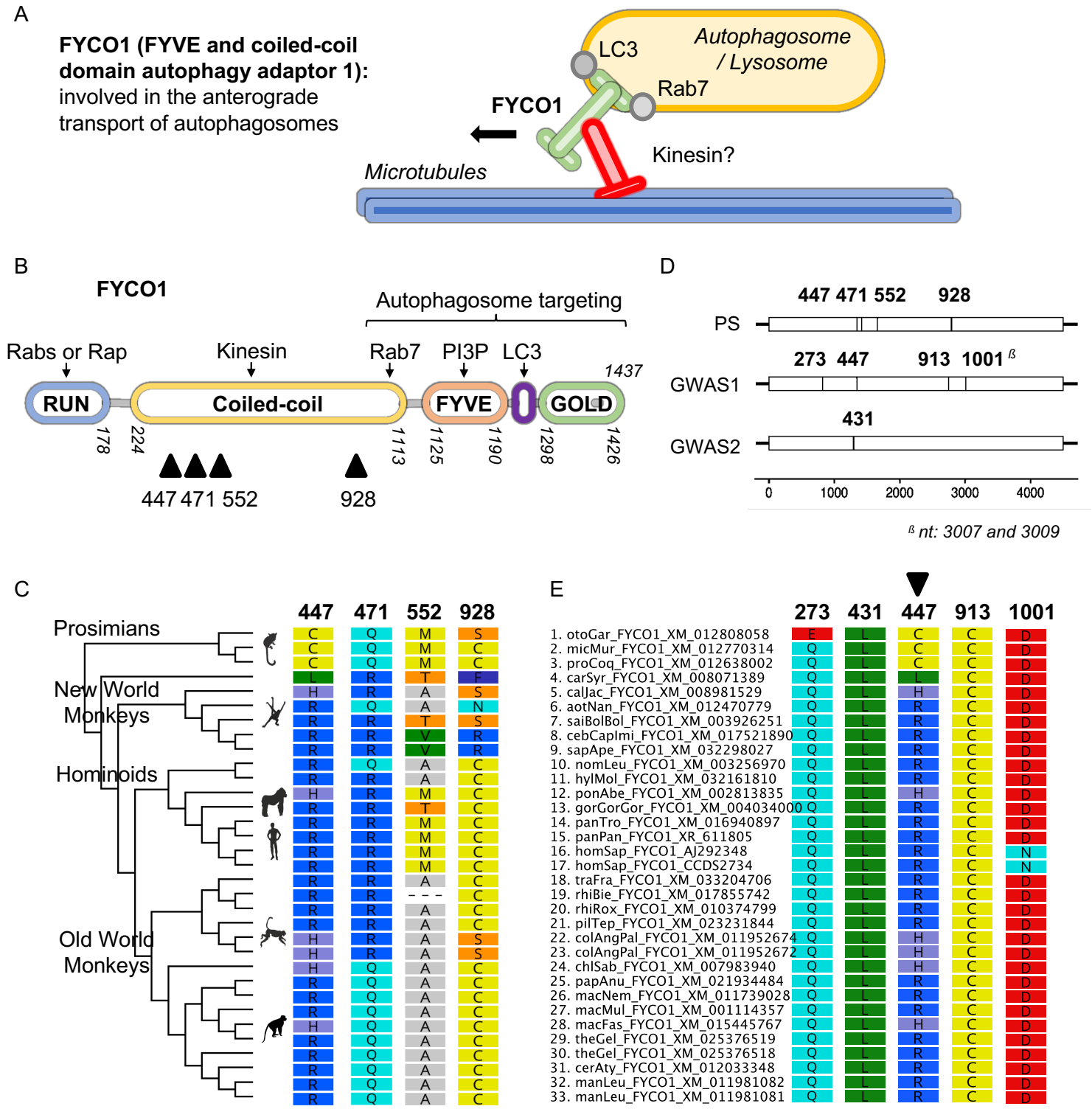
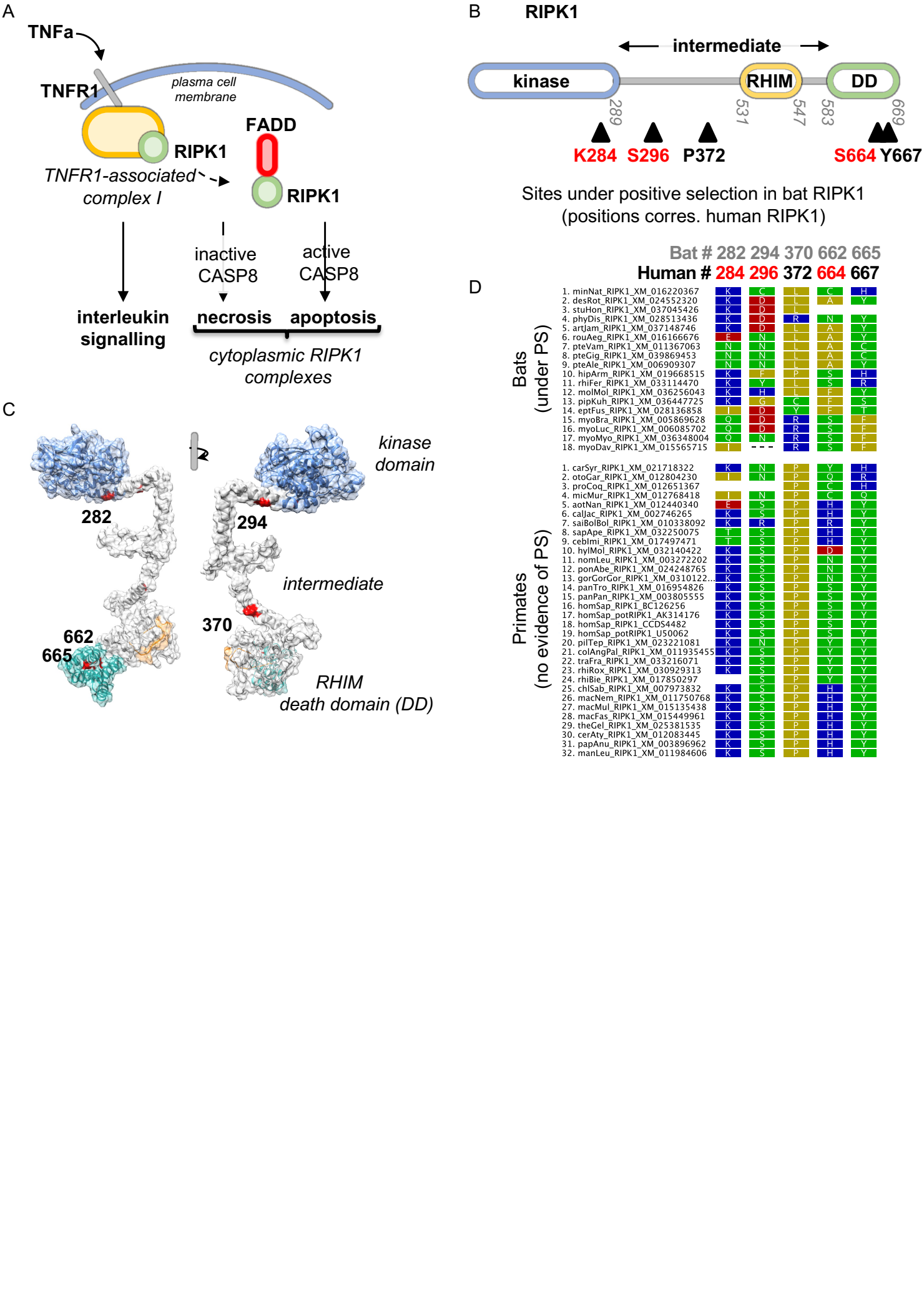

A
FYCO1 (FYVE and coiled-coil domain autophagy adaptor 1): involved in the anterograde transport of autophagosomes
LC3
Autophagosome / Lysosome
Rab7
FYCO1
Kinesin?
Microtubules
B
FYCO1
Autophagosome targeting
Rabs or Rap
Kinesin
Rab7
PI3P
LC3
RUN
Coiled-coil
FYVE
GOLD
1437
178
224
1113
1125
1190
1298
1426
447 471 552
928
D
447 471 552
928
PS
273 447
913 1001 ß
GWAS1
431
GWAS2
0
1000
2000
3000
4000
ß nt: 3007 and 3009
C
447 471 552 928
Prosimians
New World Monkeys
Hominoids
Old World Monkeys
E
273 431 447 913 1001
1. otoGar_FYCO1_XM_012808058
2. micMur_FYCO1_XM_012770314
3. proCoq_FYCO1_XM_012638002
4. carSyr_FYCO1_XM_008071389
5. calJac_FYCO1_XM_008981529
6. aotNan_FYCO1_XM_012470779
7. saiBolBol_FYCO1_XM_003926251
8. cebCapImi_FYCO1_XM_017521890
9. sapApe_FYCO1_XM_032298027
10. nomLeu_FYCO1_XM_003256970
11. hylMol_FYCO1_XM_032161810
12. ponAbe_FYCO1_XM_002813835
13. gorGorGor_FYCO1_XM_004034000
14. panTro_FYCO1_XM_016940897
15. panPan_FYCO1_XR_611805
16. homSap_FYCO1_AJ292348
17. homSap_FYCO1_CCDS2734
18. traFra_FYCO1_XM_033204706
19. rhiBie_FYCO1_XM_017855742
20. rhiRox_FYCO1_XM_010374799
21. pilTep_FYCO1_XM_023231844
22. colAngPal_FYCO1_XM_011952674
23. colAngPal_FYCO1_XM_011952672
24. chlSab_FYCO1_XM_007983940
25. papAnu_FYCO1_XM_021934484
26. macNem_FYCO1_XM_011739028
27. macMul_FYCO1_XM_001114357
28. macFas_FYCO1_XM_015445767
29. theGel_FYCO1_XM_025376519
30. theGel_FYCO1_XM_025376518
31. cerAty_FYCO1_XM_012033348
32. manLeu_FYCO1_XM_011981082
33. manLeu_FYCO1_XM_011981081

Figure 5.


A
TNFa
TNFR1
plasma cell membrane
FADD
RIPK1
TNFR1-associated complex I
RIPK1
inactive CASP8
active CASP8
interleukin signalling
necrosis
apoptosis
cytoplasmic RIPK1 complexes
B
RIPK1
intermediate
kinase
RHIM
DD
289
531
547
583
669
K284
S296
P372
S664
Y667
Sites under positive selection in bat RIPK1 (positions corres. human RIPK1)
C
kinase domain
282
294
intermediate
370
662
665
RHIM
death domain (DD)
D
Bat # 282 294 370 662 665
Human # 284 296 372 664 667
Bats (under PS)
1. minNat_RIPK1_XM_016220367
2. desRot_RIPK1_XM_024552320
3. stuHon_RIPK1_XM_037045426
4. phyDis_RIPK1_XM_028513436
5. artJam_RIPK1_XM_037148746
6. rouAeg_RIPK1_XM_016166676
7. pteVam_RIPK1_XM_011367063
8. pteGig_RIPK1_XM_039869453
9. pteAle_RIPK1_XM_006909307
10. hipArm_RIPK1_XM_019668515
11. rhiFer_RIPK1_XM_033114470
12. molMol_RIPK1_XM_036256043
13. pipKuh_RIPK1_XM_036447725
14. eptFus_RIPK1_XM_028136858
15. myoBra_RIPK1_XM_005869628
16. myoLuc_RIPK1_XM_006085702
17. myoMyo_RIPK1_XM_036348004
18. myoDav_RIPK1_XM_015565715
Primates (no evidence of PS)
1. carSyr_RIPK1_XM_021718322
2. otoGar_RIPK1_XM_012804230
3. proCoq_RIPK1_XM_012651367
4. micMur_RIPK1_XM_012768418
5. aotNan_RIPK1_XM_012440340
6. calJac_RIPK1_XM_002746265
7. saiBolBol_RIPK1_XM_010338092
8. sapApe_RIPK1_XM_032250075
9. ceblmi_RIPK1_XM_017497471
10. hylMol_RIPK1_XM_032140422
11. nomLeu_RIPK1_XM_003272202
12. ponAbe_RIPK1_XM_024248765
13. gorGorGor_RIPK1_XM_0310122...
14. panTro_RIPK1_XM_016954826
15. panPan_RIPK1_XM_003805555
16. homSap_RIPK1_BC126256
17. homSap_potRIPK1_AK314176
18. homSap_RIPK1_CCDS4482
19. homSap_potRIPK1_U50062
20. pilTep_RIPK1_XM_023221081
21. colAngPal_RIPK1_XM_011935455
22. traFra_RIPK1_XM_033216071
23. rhiRox_RIPK1_XM_030929313
24. rhiBie_RIPK1_XM_017850297
25. chlSab_RIPK1_XM_007973832
26. macNem_RIPK1_XM_011750768
27. macMul_RIPK1_XM_015135438
28. macFas_RIPK1_XM_015449961
29. theGel_RIPK1_XM_025381535
30. cerAty_RIPK1_XM_012083445
31. papAnu_RIPK1_XM_003896962
32. manLeu_RIPK1_XM_011984606

Tables

Table 1. Results from the comprehensive positive selection analyses of the genes of interest.

For each gene, are presented the results of the comprehensive phylogenetic and positive selection analyses: BUSTED, MEME, FUBAR, aBSREL from HYPHY/Datamonkey.com, M1vsM2, M7vsM8, M8avsM8 from Bpp, and M1vsM2, M7vsM8, M8avsM8 from PAML Codeml. The genes identified under positive selection are highlighted in grey. The sites considered under positive selection 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/ .

Legend details: Size, length of the codon alignment; n. sp., number of species included in the alignment; PS?, if the gene is under positive selection: Y, yes, N, no; p value, supporting a model under positive selection; PSS, positive selection sites; the cutoff for each method is given in the table; omega (PS), corresponds to the omega value in the positive selection class (dN/dS>1). ZNF318 and the proteins from the Primase complex are in Supplementary Information, and in Figures S9 and S10, respectively. *, for bpp M8 PSS analyses there were dozens of sites under positive selection due to the low omega value in the class w >1. For aBSREL, the branch identified under positive selection is given by the DGINN nomenclature (three letters from the genus and three from the species). na, not available.

Sequence alignment info				BUSTED		MEME (p<0.05)	FUBAR (PP>0.9)	Bpp M1vsM2				Bpp M7vsM8				M8a vs M8		Codeml M1vsM2		Codeml M7vsM8		M8a vs M8		aBSREL	PSS aln	PSS in human ref
Gene	Order	Size	n sp.	PS?	p value	PSS	PSS	PS?	p value	ω (PS)	PSS	PS?	p value	ω (PS)	PSS	PS?	p value	PS?	p value	PS?	p value	PSS	PS?	p value	(p<0.1)		
FYCO1	bats	1481	18	N	0.22	375, 504, 566, 688, 790, 1059, 1092	-	N	0.9861	-	-	N	0.4786	-	-	N	0.9240	N	1	-	Y	6.56E-08	607	N	0.2281	-		
FYCO1	**primates**	1500	29	N	0.30	355, 416, **472**, 484, 601, 629, 728, 869, 919, 1102, 1218, 1219, 1242, 1259, 1267, 1407	**472**	Y	0.0058	4.42	**448, 553**	Y	4.77E-05	1.73	***n=206****	Y	0.0136	Y	0.0103		Y	3.97E-08	**448, 930**	Y	0.0032	rhiBie	**448, 472, 553, 930**	**R447, R471, M552, C928**
POLA1	**bats**	1484	17	Y	0.00	147, **246, 249**, 317, 1314, 1449	109, 196, **246, 249, 250,** 276, **284, 315, 1080,** 1270	Y	4.37E-07	59.91		Y	5.76E-07	2.05	*n=186**	Y	0.0016	Y	2.17E-08	**249, 296, 1080**	Y	1.94E-10	**246, 249, 250, 284, 296, 315, 1080**	Y	7.01E-08	phyDis, pteAle	**246, 249, 250, 284, 296, 315, 1080**	**gap, E239, E240, W273, Q285, V304, V1069**
POLA1	**primates**	1451	25	N	0.26	590, 707, 718, 1082	**817**	N	0.4902	-	-	Y	0.0332	14.36	**221**	Y	0.0343	Y	0.0100		Y	0.0060	**221**, 226, 227, **817**, 1058	Y	0.0102	N	**221, 817**	**D232, N828**
PRIM1	**bats**	421	17	N	0.50	-	257, 258, 275, 278, **289, 291, 292**	N	0.4226	-	-	N	0.0693	-	-	N	0.4204	Y	0.0116		Y	2.40E-05	**289, 291, 292**	Y	0.0060	N	**289, 291, 292**	**Y288, P290, W291**
PRIM1	primates	420	26	N	0.48	277, 361	-	N	0.4672	-	-	N	0.2021	-	-	N	0.6951	N	0.1427	-	N	0.0619	-	N	0.0619	-		
PRIM2	**bats**	517	16	N	0.50	5, **11**, 20, 71, **168, 176**	**11**, 69, 81, **168, 176**, 476	Y	0.0005	23.14	**186**	Y	4.20E-06	12.34	**186**, 187	Y	0.0009	Y	5.78E-05	**186**	Y	8.09E-05	**186**	Y	0.0135	na	**11, 168, 176, 186**	**L11, V167, L173, K177**
PRIM2	primates	513	20	N	0.50	5, 8, 73, 278	-	N	0.1629	-	-	N	0.0632	-	-	N	0.2565	Y	5.88E-07	8, 10	Y	2.75E-07	8, 10, 76, 173, 176, 178	Y	5.72E-07	-		
RIPK1	**bats**	669	18	Y	0.01	127, 230, **282**, 289, 370, 400, 480, 497, 668	**294, 370**	Y	0.0008	54.40		Y	0.0055	1.61	*n=87**	N	0.1172	Y	0.0138		Y	9.37E-05	**294, 370, 662, 665**	Y	0.0009	myoDav	**294, 370, 662, 665**	**K284, S296, P372, S664, Y667**
RIPK1	**bats**	669	15	Y	0.05	127, 230, **282, 370,** 400, 480, 668	16, 277, **282,** 294, **370**	Y	0.0009	70.02		Y	0.0002	10.77	**370, 665**	Y	0.0033	Y	0.0192		Y	0.0005	**370, 665**	Y	0.0023	myoDav	**282, 370, 665**	
RIPK1	primates	671	29	N	0.50	491, 493, 592	-	N	0.3312	-	-	N	0.0855	-	-	N	0.4059	N	0.0907	-	Y	0.0013	664	N	0.0848	-		
TMPRSS2	bats	496	17	N	0.43	19, 136, 249, 364, 412, 416	-	N	1	-	-	Y	0.0133	1.29	49, 216, 364, 413, 416, 435	N	0.9939	N	1	-	Y	0.0417		N	0.9990	-		
TMPRSS2	**primates**	563	28	N	0.14	168, 232, 299, **312,** 316, 424, 467, **500**	117, **224, 312, 412, 500**	Y	0.0095	5.70	**224, 412**	Y	2.95E-05	5.19	**224, 312, 315, 412, 500**	na	na	Y	4.27E-06	**224, 412**	Y	8.43E-09	**224, 312, 315, 412, 500**	Y	3.09E-07	N	**224, 312, 315, 412, 500**	**Q173, E260, L263, L360, S412**
ZNF318	bats	2142	16	N	0.50	127, 586, 851, 1647, 1665, 2015, 2016	-	N	0.7917	-	-	na	na	na	na	na	na	N	0.4266	-	Y	0.0037		N	0.1361	-		
ZNF318	bats	2395	11	Y	0.02	78, 370, 830, 1107, 1588, 2153, 2267	-	N	0.9461	-	-	N	0.0740	-	-	N	0.0757	N	0.6505	-	N	0.1220	-	N	0.6408	-		
ZNF318	**primates**	2108	29	N	0.50	93, 843, **1302,** 1436, 1449, 1884, 2095	**1302,** 1878	N	0.9957	-	-	N	0.7354	-	-	N	0.8185	Y	0.0017	**1302**	Y	5.14E-06	**1302,** 1589, 1733	Y	0.0005	N	**1302**	**V1481**
ZNF318	**primates**	2496	24	Y	0.00	38, 75, 104, 174, 1372, **1636,** 1814, **1929,** 2166	**1636, 1817, 2098, 2243**	Y	3.59E-09	501.94		Y	0.0278	1.45	*n=68**	N	0.3927	Y	7.81E-05	**1636, 2098**	Y	2.08E-06	**1636, 1817, 1929,** 1954, **2098, 2243**	Y	3.17E-05	rhiBie, micMur, homSap	**1636, 1817, 1929, 2098, 2243**	

SUPPLEMENTARY

Supplementary Information – Results and Discussion

ZNF318 has undergone positive selection at three sites in primates, but not in bats

ZNF318 (NP_055160, also known as TZF, testicular zinc finger protein) is a putative RNA binding Matrin-type zinc finger protein. Matrin-type zinc fingers are RNA interacting domains found most notably in matrins and U1 small nuclear ribonucleoprotein C (InterPro: IPR003604). Aside from two central matrin-type ZNF domains, ZNF318 interacts with the androgen receptor and forms homodimers through its N-terminal tail (Figure S9A) (Tao et al., 2006). Finally, despite accounting for about half the protein, the Proline-rich C-terminal domain of ZNF318 has currently no ascribed function (Figure S9A). While the precise molecular function of ZNF318 remains unknown, it was shown to interact with the HUSH chromatin silencing complex and contributes to splicing-coupled transposon, and potentially latent HIV, silencing (Chougui et al., 2018; Douse et al., 2020). Antiviral functions have also been described for some human matrin genes (Yedavalli and Jeang, 2011).

Our DGINN screens identified signatures of positive selection in ZNF318 from primates, but not in bat species. While we confirmed the presence of ZNF318 orthologs in 29 primate and 16 bat species, many were 5' truncated. Thus, we used 5' trimmed CDS for in-depth selection analyses (Table 1). Unlike the majority of codons that are highly conserved (mean dN/dS of ~0.3), the C-terminal portion of ZNF318 displays many non-synonymous substitutions between orthologs. Statistical analyses using CodeML identified three sites under positive selection in this domain (M7 vs M8, BEB > 0.95): residues 1481, 1756 and 1908 in the corresponding human full-length protein (NP_055160, Figure S9, Table 1). There does not seem to be strong co-evolutionary signatures between these three residues and the combination of sites found in human ("V", "R" and "P") arose once in the last common ancestor with chimps and bonobos (Figure S9B). Of the three sites, non-synonymous substitutions altering the proline residue 1908

might have strongest structural impact. Interestingly, these residues show some level of variation in bats, with bias for Threonines and Alanines at orthologous positions of sites 1481 and 1908.

While the function of ZNF318 during SARS-CoV2 infection remains to be determined, our finding of positive selection in primates suggest that it contributed to host adaptation in this lineage. Moreover, our analysis identified ZNF318 Proline-Rich domain as a putative functional interface with viruses. Hence, it would be interesting to investigate if this domain carries functions relating only to viral infections, including with non-coronaviruses (Figure 2), or if it is also involved in ZNF318 other cellular functions.

Rapid evolution of the Prim-Pol primase complex (POLA1, PRIM1, PRIM2) in primates and bats

Initiation of DNA replication in eukaryotes is dependent on the multisubunit primase-polymerase alpha (Prim-Pol α) complex that is responsible for the *de novo* synthesis of RNA/DNA primers on both the leading and lagging strands. Primase consists of the small catalytic subunit PRIM1 (p49) and the large regulatory subunit PRIM2 (p58), while polymerase consists of the catalytic subunit POLA1 (p180) and the accessory subunit POLA2 (p70). During primer synthesis, the primase subunits generate short RNA oligos, which are then subsequently extended with DNA by Pol α to be further elongated by replicative DNA polymerases. While many viruses encode their own Prim-Pol (Guilliam et al., 2015), including a putative CoV primase consisting of NSP7 and NSP8, SARS-CoV-2 NSP1 interacted with all four subunits of Prim-Pol α (Gordon et al., 2020).

From our initial DGINN analysis, we identified positive selection on three out of the four Prim-Pol α subunits: POLA1 (four tests for primates and bats), PRIM1 (three tests for primates and two tests for bats), and PRIM2 (five tests for primates only). POLA2 was not identified under positive selection in primates or bats (one test for bats and zero tests for primates).

To validate the finding and to precisely characterize gene evolution in primates and bats for POLA1, PRIM1, and PRIM2, we retrieved additional sequences for these orders and generated new high-quality codon alignments. POLA1 still showed significant signatures of positive selection across both primate and bat lineages (Table 1). PRIM1 maintained signatures of positive selection in bats (Table 1). PRIM2 maintained signatures of positive selection in primates for some tests, and the additional sequences identified positive selection in bats not seen with our initial pipeline (Table 1). Together this confirmed our initial analysis that POLA1, PRIM1, and PRIM2 are rapidly evolving in primate and bat lineages, with bats having more robust positive selection than primates.

Our codon-specific positive selection analysis identified multiple residues that have rapidly evolved in primates and bats. We specifically focused on sites that were found with more than one test of selection (Table 1). To determine if these sites of positive selection were found at the interface between subunits of the Prim-Pol α complex or at surface exposed sites, we mapped PS sites onto the human Prim-Pol α crystal structure (PDB: 5EXR) (Baranovskiy et al., 2016). None of the sites were found at Prim-Pol α complex interfaces, and while many were clustered close together, all were surface exposed or found in putative unstructured regions that were not present in the crystal structures (Figure S10). Together, this indicates that the positive selection identified for this complex is not being driven by complex formation and intra-complex coevolution, or DNA replication.

POLA1 is of particular interest as it has been modeled to dock with SARS-CoV-2 NSP1 (PDB 7OPL)(Kilkenny et al., 2022) (Chaudhuri, 2021). However, none of the sites identified in our positive selection analysis were at the predicted POLA1-NSP1 interface (residues 615-629; Figure S10A-B), suggesting that NSP1 is not driving positive selection on POLA1. Instead, 7/9 PS sites in bats and primates were found in unstructured regions that were not present in the crystal structure, with four sites (V235, D232, E239, and E240) all falling in a predicted disordered region of the protein (amino acids 232-251). That this unstructured region showed

strong positive selection in both primates and bats could suggest that primate and bat POLA1 rapid evolution is being driven by similar unknown pressures in these distinct families.

Altogether, our positive selection analysis has identified that the Prim-Pol α complex is under strong positive selection in bats and primates, however this is not driven by Prim-Pol α complex formation or NSP1. Thus, it remains unclear what is driving PS on the Prim-Pol α complex, whether SARS-CoV interact with the Prim-Pol α complex as a whole or individual proteins from this complex, and why SARS-CoV-2 directly interact with nuclear host DNA replication machinery.

So why does SARS-CoV-2 recruit the Prim-Pol α complex? One possibility is that Prim-Pol α may have a role in the innate immune response to viral infection, and thus viruses directly antagonize components of this complex. Prim-Pol α was identified to interact specifically with NSP1 (Gordon et al., 2020), which functions to inhibit host translation and innate immunity (Narayanan et al., 2008; Schubert et al., 2020). Pola1 is found in both the nucleus and cytoplasm, where it generates RNA-DNA hybrids that may be important for innate immune sensing (Starokadomskyy et al., 2016). Loss of function mutations in POLA1 lead to increased pathogen infection, innate immune activation, and decreased number and effectiveness of NK cells (Starokadomskyy et al., 2016)(Starokadomskyy et al., 2019). While this supports a role of POLA1 in innate immunity against pathogens, roles for PRIM1 and PRIM2 have yet to be investigated.

It is also possible that SARS-CoV are usurping the host primase complex (or components of this complex) to enhance genome replication. SARS-CoV NSP7 and NSP8 are proposed to function as primase important for initiation of genome replication (Konkolova et al., 2020; Zhai et al., 2005). However, recent cryoEM structures of NSP7-8-12 complex suggest that NSP7 and NSP8 are too far from the RdRP NSP12 active site to act as primase (Kirchdoerfer and Ward, 2019). Thus, it is possible that the host Prim-Pol α is recruited by NSP1 to further help initiation (and/or elongation) of CoV genome replication.

Supplementary Figures

Figure S1. Identification of the SARS-CoV-2 interactome with signatures of positive selection in bats. A, Overview of the key steps of the bat VIP DGINN screen workflow. Details in Table S1. B, VIP-encoding genes identified under significant positive selection by at least three methods in DGINN (embedded legend). The percentage of positively selected sites in each VIP is shown on the right panel.

Figure S2. Identification of SARS-CoV-2 interactome with signatures of positive selection in primates. A, Overview of the key steps of the bat VIP DGINN screen workflow. Details in Table S1. B, VIP-encoding genes identified under significant positive selection by at least three methods in DGINN (embedded legend). The percentage of positively selected sites in each VIP is shown on the right panel. Of note, seven genes are false positives due to erroneous sequences or alignments: EMC1 (ER membrane protein complex subunit 1), MOV10 (Mov10 RISC complex RNA helicase), POR (cytochrome p450 oxidoreductase), PITRM1 (pitrilysin metallopeptidase 1), RAB14, RAB2A, and TIMM8B (translocase of inner mitochondrial membrane 8 homolog B).

Figure S3. Comparison of primate positive selection analyses between this study and Gordon et al. A, Comparison of the omega (dN/dS) values in PAML Codeml M0 model of the primate VIP genes calculated using the automated DGINN pipeline (y axis) and from Gordon et al study (x axis). Raw data were kindly provided by Janet Young, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. In black, the bisector. In red, the linear regression. B, Comparative analysis of the number of VIPs “under positive selection” in primates. In dark grey, the “strong and weak positively selected genes with benjamini-hochberg correction” from Gordon et al study (Codeml M8 vs M8a, p-value < 0.10); in medium grey, same from “uncorrected p-values”; in light grey, the genes identified by at least three methods in the primate DGINN screen. A total of 322 genes were in common between the two studies. C, VIPs under positive selection in Gordon et al with a DGINN score below three.

Figure S4. Comparative analyses of adaptive signatures in SARS-CoV-2 interactome in primates and in bats. A, Full table (associated with Figure 1D) showing the genes identified by x,y DGINN methods in bats and primates, respectively. B, Tanglegram of genes under positive selection in bats and primates. At the top, genes with the highest DGINN score (DGINN scores: purple, 5; blue, 4; teal, 3; green, 2; brown, 1; red, 0).

Figure S5. Meta-Analysis by Information Content (MAIC) scores of the VIP genes under positive selection. A-B, MAIC rank of VIPs identified under positive selection (by at least three methods in DGINN; DGINN scores of 3-5) in bats (A) and primates (B). The ACE2 and TMPRSS2 genes are highlighted in red as references. C-D, MAIC rank for all VIPs with, or without, evidence of positive selection (DGINN scores > or = 3, or < 3, respectively) in bats (C) and primates (D).

Figure S6. Positively selected VIPs are involved in several biological processes. The graphs present the top 10 biological pathways retrieved after analysis on the Reactome database of bat (A) and primate (B) positively selected VIPs.

Figure S7. Genetic variation of mammalian TMPRSS2 at the corresponding residues under positive selection in primates. Top, sequence logo of the positively selected sites in mammals that are naturally and experimentally permissive to SARS and MERS coronaviruses. Logos generated using WebLogo based on the alignment of the positively selected sites identified in primates. Bottom, Alignment of the corresponding amino acids in the mammalian species.

Figure S8. RIPK1 interactome and 3D-RIPK1 structure prediction. A, RIPK1 was used to interrogate the Reactome database and to retrieve RIPK1 cellular interactors that were subsequently subdivided according to their involvement in the indicated pathways. B, Described RIPK1 microbial antagonists or interactors from viruses (top) and bacteria (bottom). C, Human and *Rhinolophus ferrumequinum* RIPK1 sequences were used to generate 3D-structure prediction models on RaptorX (grey and orange, respectively). Magnified views of the structural homologies between the indicated domains.

Figure S9. ZNF318 has undergone positive selection at three sites in primates, but not in bats. A, Schematic representation of human ZFN318 protein domains. Numbering of residues is relative to the human full-length protein sequence. The three residues subjected to positive selection in primates are shown with arrows. B, Amino-acids found at rapidly evolving sites across ZFN318 primate orthologs used in this study (top). Orthologs are organized according to the accepted species phylogeny. Bottom, same as top with the corresponding residues in bats. Dashes indicate an alignment gap due to an indel.

Figure S10. Rapid evolution of the Prim-Pol primase complex (POLA1, PRIM1, PRIM2) in primates and bats. A, Schematic representation of human DNA primase complex. B, D, & E, Diagrams of predicted domains for POLA1, PRIM1, and PRIM2, respectively. Sites under positive selection in primates are represented by black triangles and sites under positive selection in bats are represented by purple triangles (Table 1). Codon numbering based on Homo sapiens genes. C, POLA1 amino acid variation at the positively selected sites in primates (top) and bats (bottom). Codon numbering and coloring as in B.

Supplementary Tables

Table S1. Number of genes screened in the initial primate and bat automatic screens – associated to Figures S1-S2.

		Primates	**Bats**
Initial dataset (Gordon et al., 2020 + ACE2 + TMPRSS2)		334	334
Phylogenetic analyses	Failed runs	5	4
	Duplicated genes retrieved by DGINN	39	5
	Complete alignments and phylogenies	368	335
Positive selection analyses	Results from 0 to 4 (out of 5) DGINN methods	43	11
	Complete results (5/5 PS methods)	325	324

Table S2. SARS-CoV-2 interacting proteins with no other known virus interactors

HGNC names of the SARS-COV-2 VIPs with no other known virus interactors (*, except with other coronaviruses)
ACE2*
GOLGA7*
PRIM1*
EIF4E2*
GORASP1*
MARK1*

TRIM59*
CISD3
GGH
INHBE
PUSL1
DPH5
GCC2
HS6ST2
NUP58
PLAT
PTBP2
TMEM39B
TMPRSS2*

Figure S1. Identification of SARS-CoV-2 interactome with signatures of positive selection in bats

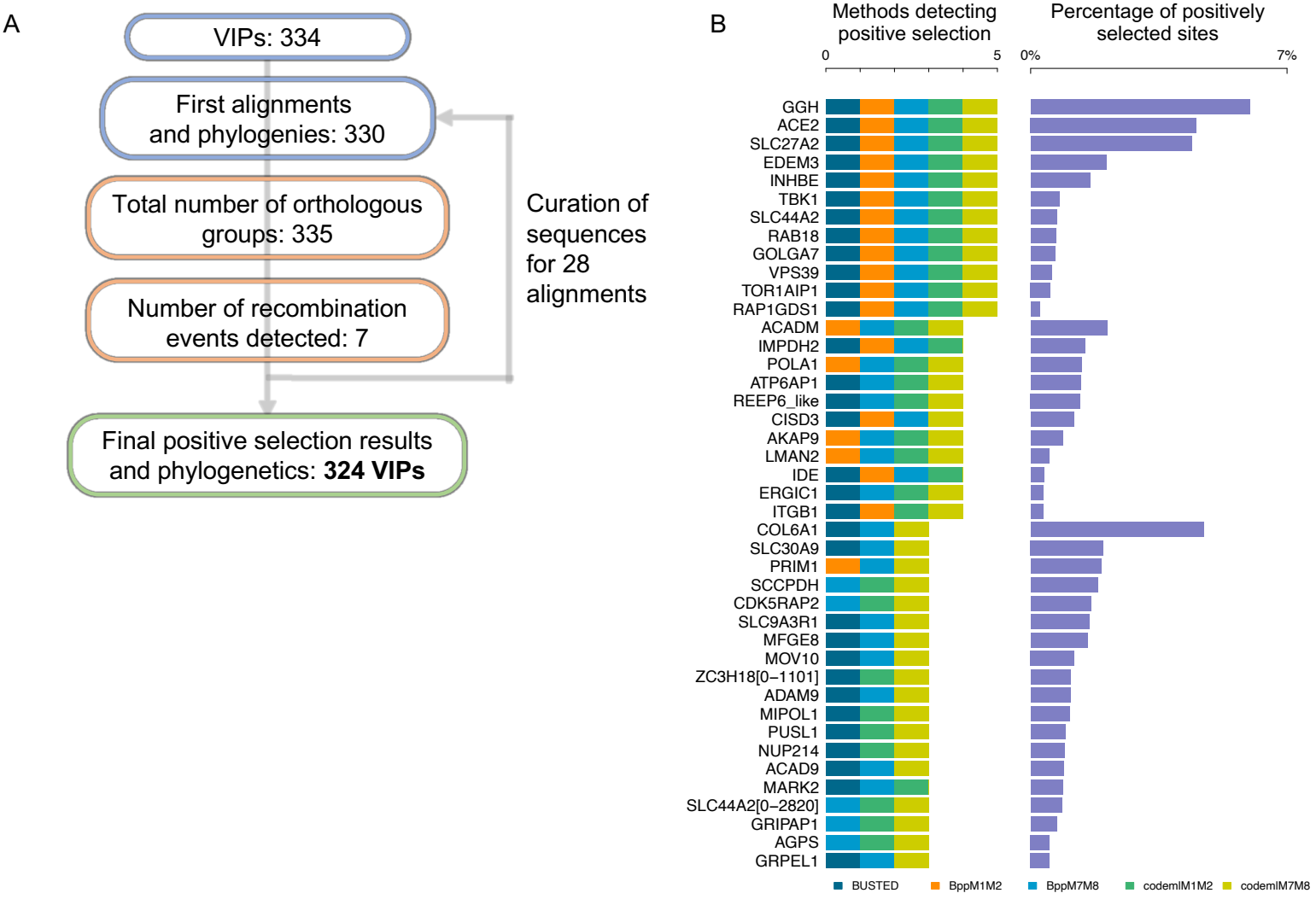


Figure S2. Identification of SARS-CoV-2 interactome with signatures of positive selection in primates

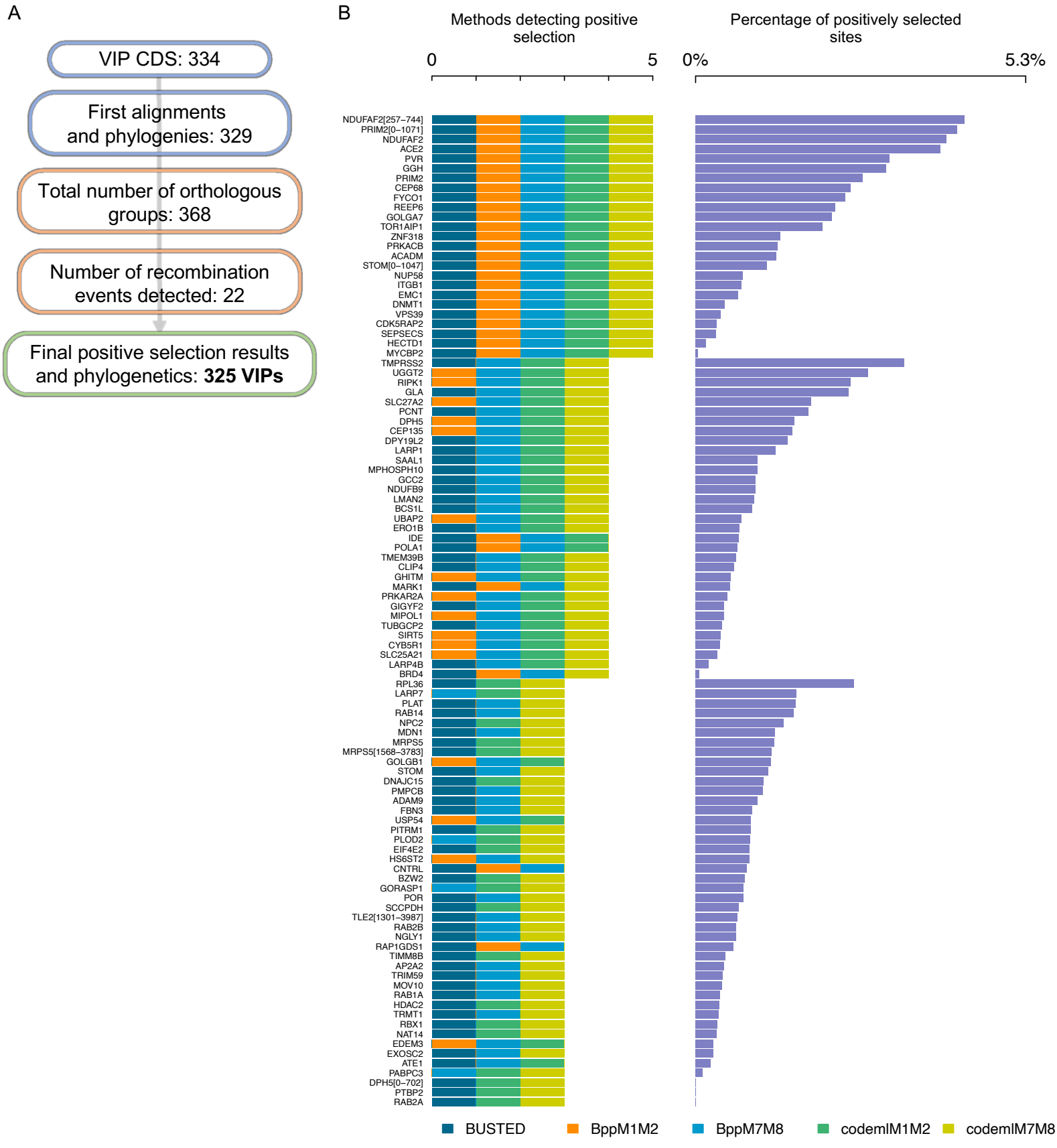
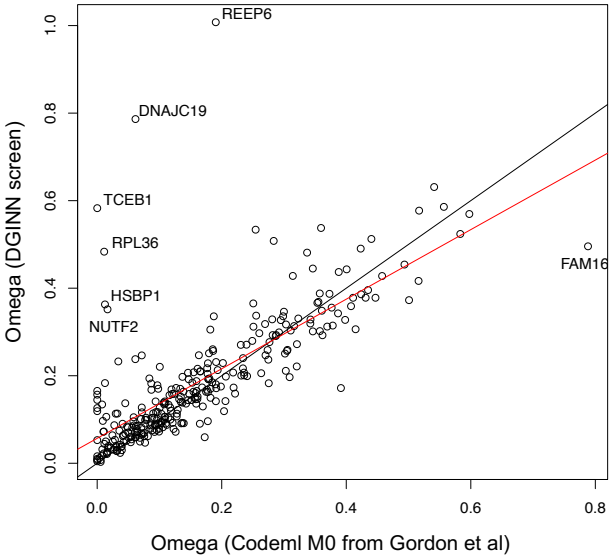
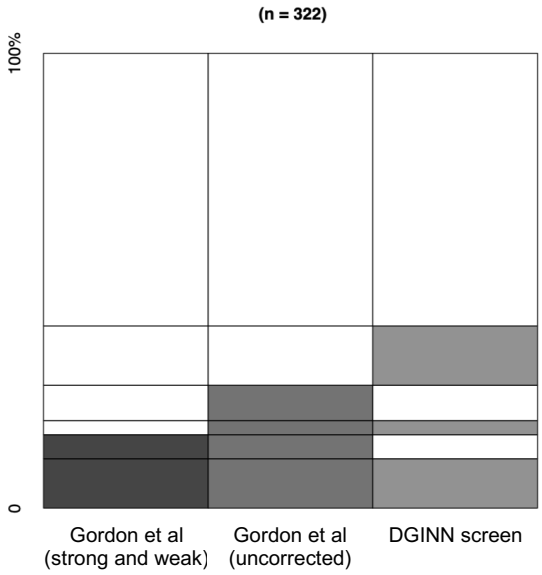


Figure S3. Comparison of primate positive selection analyses between this study and Gordon et al.

A



B



C

VIPs under PS in Gordon *et al.* with a DGINN score < 3	
VIP name	DGINN score (identified under PS by x methods)
ABCC1	2
AKAP9	2
CENPF	2
CEP250	2
CEP350	2
ERLEC1	2
FASTKD5	0
MRPS27	2
NINL	2
PDE4DIP	1
PRRC2B	0
PUSL1	1
ALG11	0
ALG8	1
DCTPP1	0
GOLGA2	2
NUP210	2

Figure S4. Comparative analyses of adaptive signatures in SARS-CoV-2 interactome in primates and in bats

A

Genes supported by x,y methods in bats and primates

primates

	0	1	2	3	4	5
5	NDUFAF2 PRIM2 REEP6 ZNF318	CEP68 DNMT1 NUP58	EMC1 FYCO1 HECTD1 MYCBP2 PVR SEPSECS	CDK5RAP2	ACADM ITGB1	ACE2 GGH GOLGA7 TOR1AIP1 VPS39
4	CEP135 DPH5 ERO1B GCC2 GHITM GORASP1 HS6ST2 INHBE LARP4B NDUFB9 SLC25A21	TMPRSS2 UBAP2	BRD4 CLIP4 MDN1 PCNT RIPK1 TMEM39B	MIPOL1	IDE LMAN2 POLA1	SLC27A2
3	ATE1 BZW2 CNTRL EIF4E2 GIGYF2 GLA LARP1 MARK1 MPHOSPH10 PRKAR2A SAAL1 SIRT5	UGGT2 USP54	GOLGB1 MRPS5 PMPCB POR	ADAM9 MOV10 SCCPDH		EDEM3 RAP1GDS1
2	AP2A2 CEP68 CHMP2A COLGALT1 EXOSC2 FAR2 FBN1 FBN2 GDF15 HYOU1 MARK3 NSD2 POFUT1 RAB7A SBNO1 SCARB1	STOM TRMT1	ANO6 CHMP2A COLGALT1 EIF4H FAR2 FBN1 GOLGA2 NUP210 PSMD8 RETREG3 WFS1	COL6A1 NUP214 PRIM1	AKAP9 ERGIC1	TBK1 INHBE
1	37 values	27 values	ACAD9 AGPS ARL6IP1 C1QBP CHMP2A COLGALT1 CSNK2B DCAF7 GDF15 GFER GNB1 GNG5 GOLGA3 GPX1 HSBP1 HYOU1 MAP7D1 MRPS2 NDFIP2 NUP98 PDE4DIP PIGO PKP2 PLD3 PPIL3 PTGES2 RRP9 SELENOS SRP19 TARS2 TBCA TCF12 TLE1 TLE3 TMEM97 TOMM70 UBXN8 ZDHHC5 ZNF503	CISD3 GRIPAP1 IMPDH2 PUSL1 TUBGCP3	ATP6AP1 RAB18
0	49 values	28 values	22 values	ACAD9 MARK2 SLC30A9 SLC9A3R1	ERGIC1	INHBE SLC44A2

p=1/n=0 AATF ACSL3 AKAP8L AP2M1 BCKDK CLCC1 CSDE1 CSNK2B DCAF7 GFER GNB1 GNG5 GOLGA3 GPX1 HSBP1 MAP7D1 MRPS2 NDFIP2 NUP98 NUTF2 PDE4DIP PDZD11 PIGO PKP2 PLD3 PPIL3 RAB7A RETREG3 SELENOS SRP19 TARS2 TCF12 TLE3 TMEM97 TOMM70 UBXN8 ZNF503

p=1/n=1 ALG8 AP3B1 ATP6V1A CEP112 CHPF2 DDX21 ERC1 FAM8A1 FBLN5 G3BP1 HMOX1 KDELC1 LOX NARS2 NOL10 NUP62 QSOX2 RAE1 SLC30A7 SRP72 TIMM29 TM2D3 TOR1A UBAP2L VPS11 ZC3H18 ZC3H7A

p=0/n=0 AAR2 ADAMTS1 AKAP8 ARF6 BAG5 CCDC86 CSNK2A2 DCAKD DCTPP1 DNAJC19 DPY19L1 ECSIT EIF4H ETFA EXOSC5 EXOSC8 FAM162A FAM98A FASTKD5 FKBP15 GTF2F2 HOOK1 KDELC2 MARC1 MAT2B MOGS NGDN NUP54 NUP88 OS9 PABPC1 PIGS PLEKHA5 RAB10 RALA REEP5 RHOA RNF41 SDF2 SLC30A6 SNIP1 SRP54 STC2 STOML2 THTPA TIMM10B TIMM9 UPF1 YIF1A

p=0/n=2 AASS ATP1B1 COMT CUL2 DDX10 ELOC ERMP1 FOXRED2 GPAA1 HS2ST1 JAKMIP1 MEPCE MIB1 NEK9 NEU1 PCSK6 PLEKHF2 RBM28 RDX SUN2 TMED5 TYSND1

bats

B

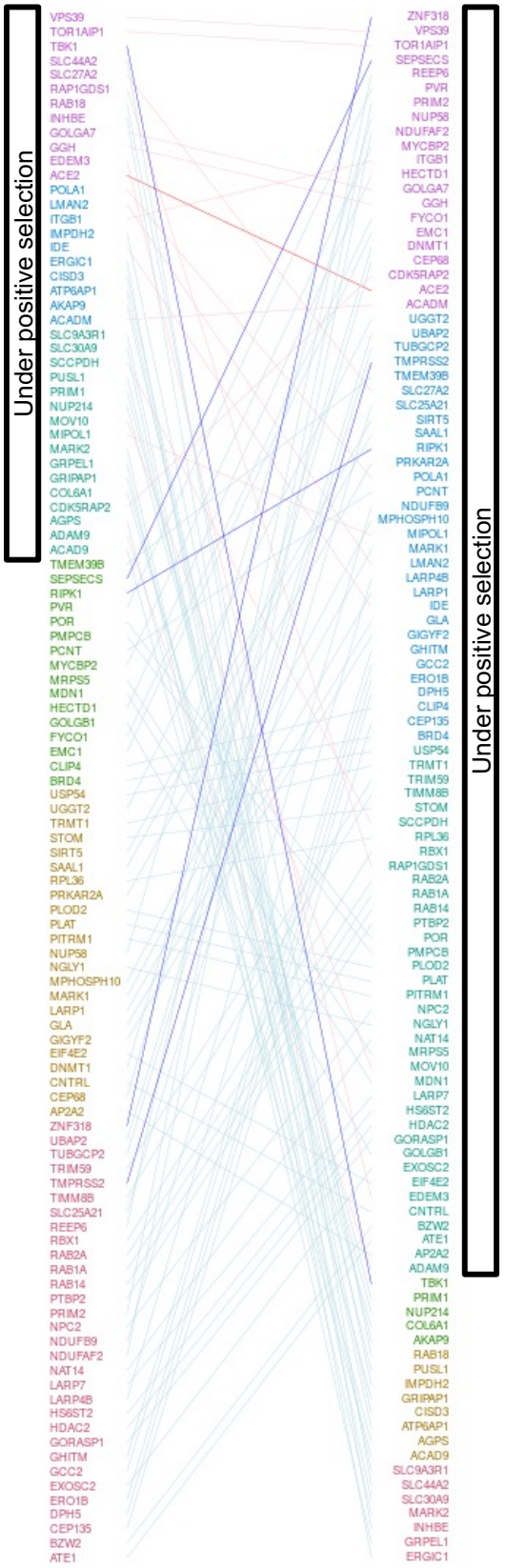


Figure S5. Meta-Analysis by Information Content (MAIC) scores of the VIP genes under positive selection.

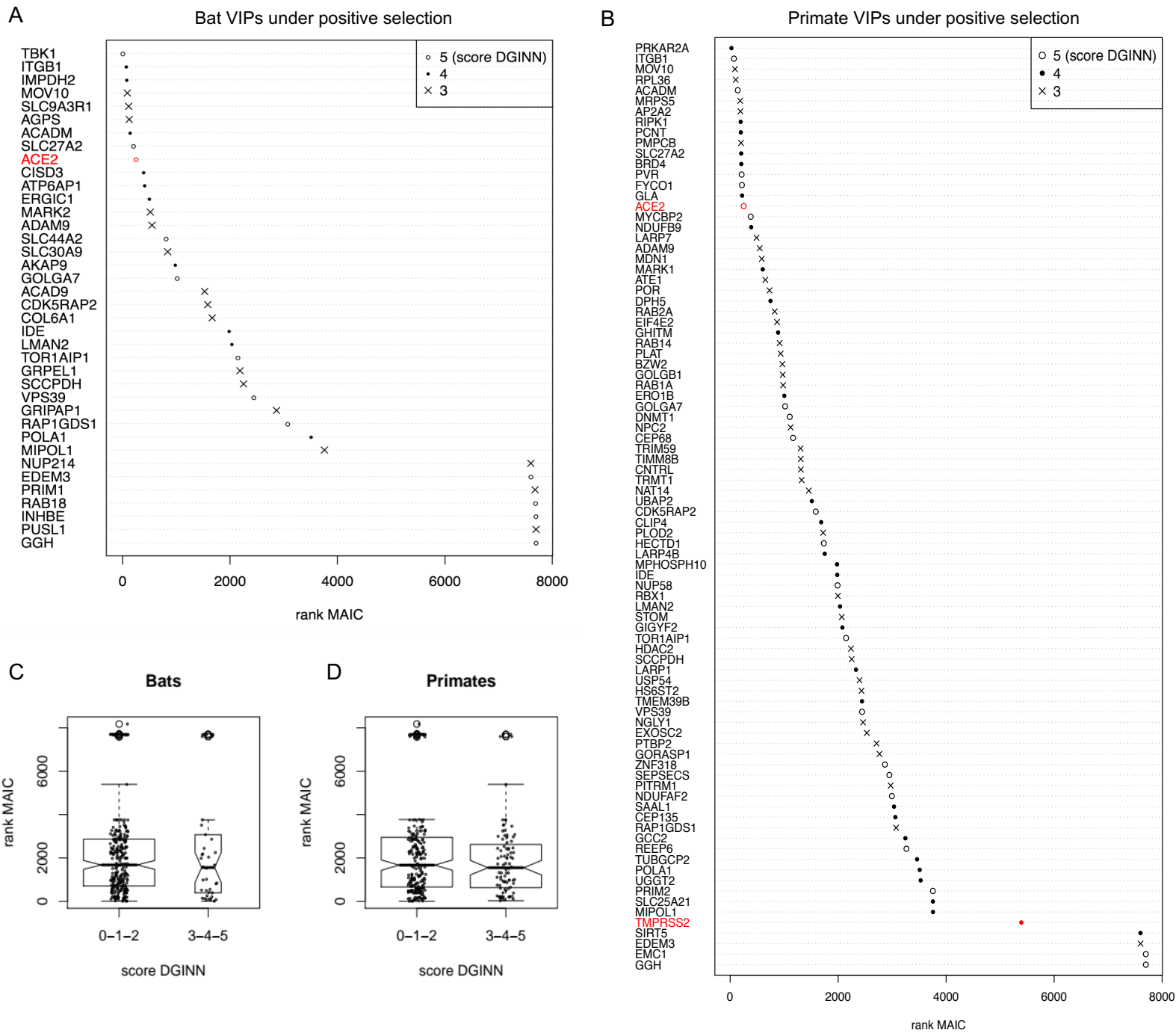
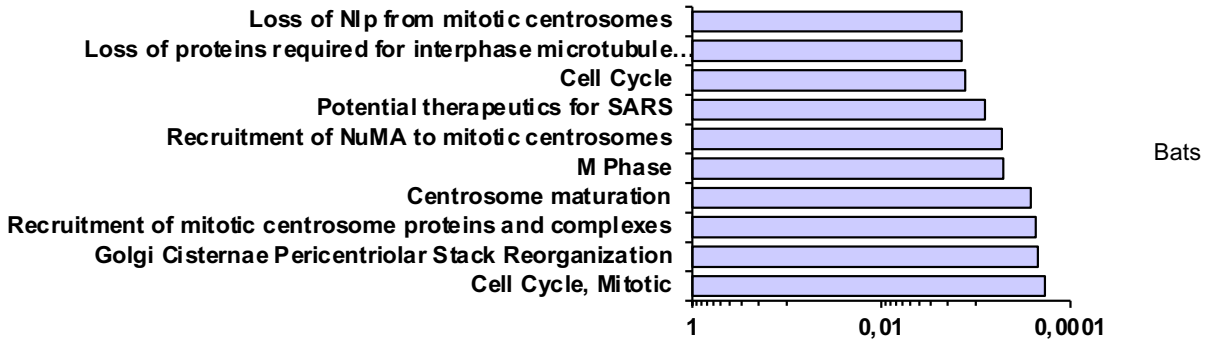


Figure S6. Positively selected VIPs are involved in several biological processes

A



B

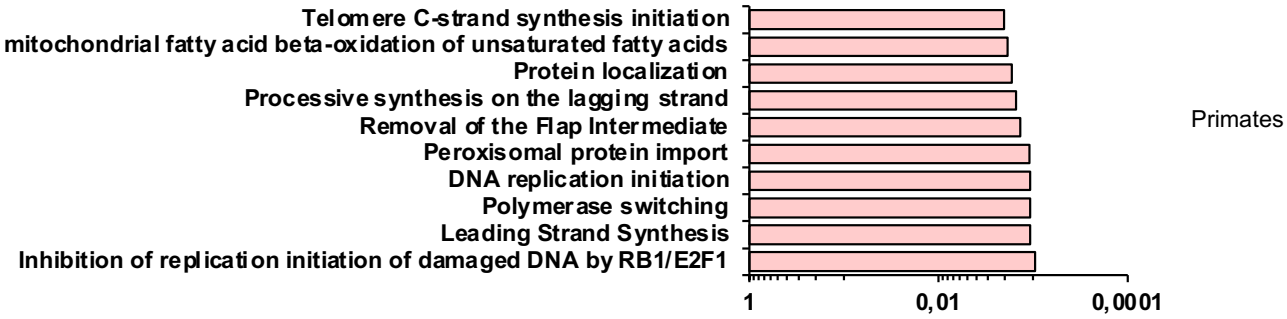
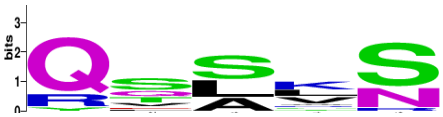


Figure S7. Genetic variation of mammalian TMPRSS2 at the corresponding residues under positive selection in primates



Macaca mulatta	R	Q	L	V	N
Macaca fascicularis	R	Q	L	V	N
Cercocebus atys	R	Q	L	V	R
Callithrix jacchus	Q	V	S	A	S
Homo sapiens	Q	E	L	L	S
Gorilla gorilla	Q	V	L	L	S
Rousettus aegyptiacus	Q	T	A	T	N
Rinolophus ferrumequinum	Q	T	A	K	N
Canis lupus familiaris	Q	T	S	R	S
Ursus maritimus	Q	S	S	K	S
Felis catus	Q	S	A	K	S
Sus scrofa	Q	S	A	K	S
Panthera tigris	Q	- - -	- - -	K	S
Cavia porcellus	Y	S	S	S	S
Rattus rattus	Q	S	S	V	S
Mus musculus	Q	L	S	L	S
Cricetulus griseus	Q	S	S	L	N

Figure S8. RIPK1 interactome and 3D-RIPK1 structure prediction

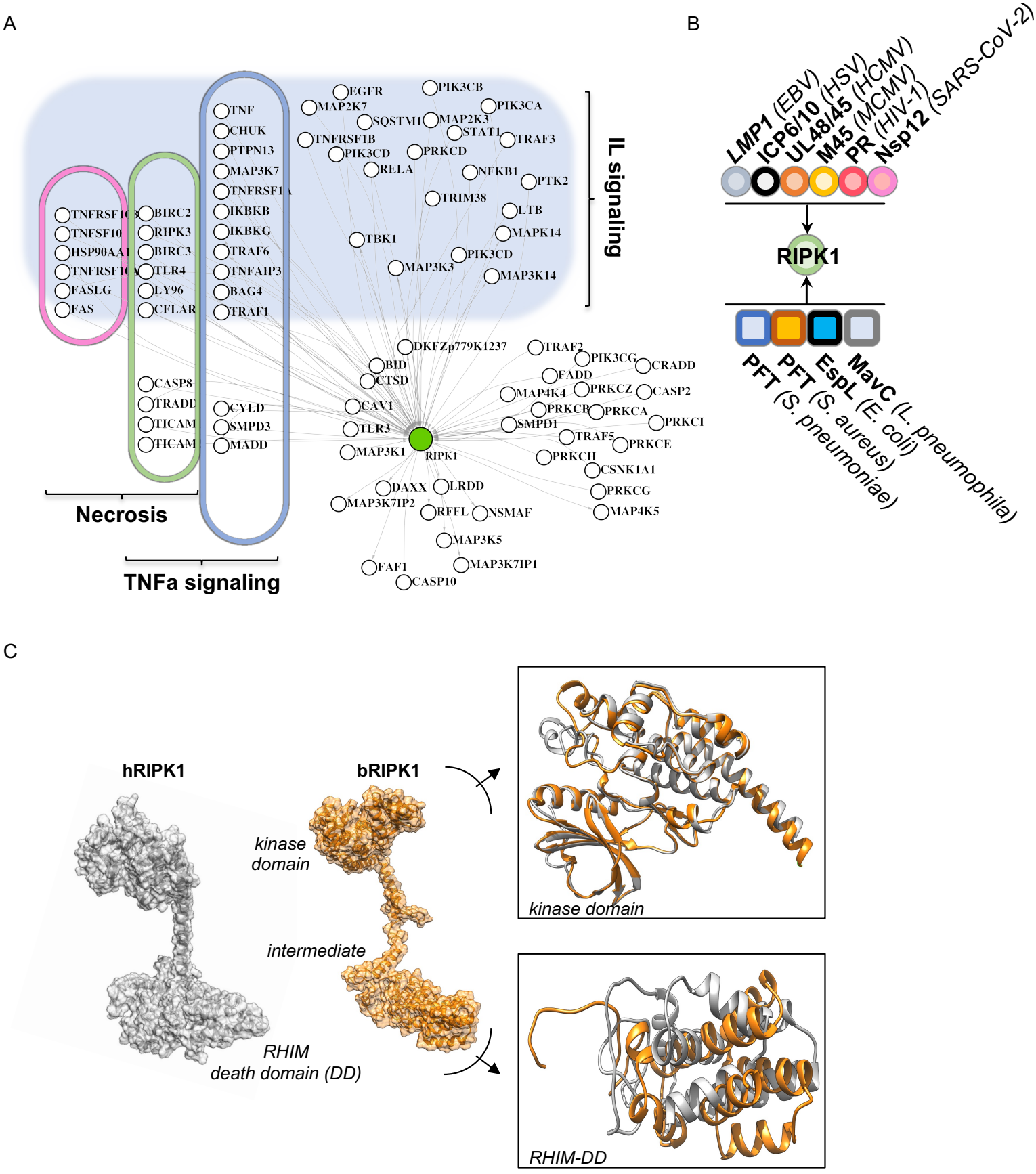


Figure S9. ZNF318 has undergone positive selection at three sites in primates, but not in bats

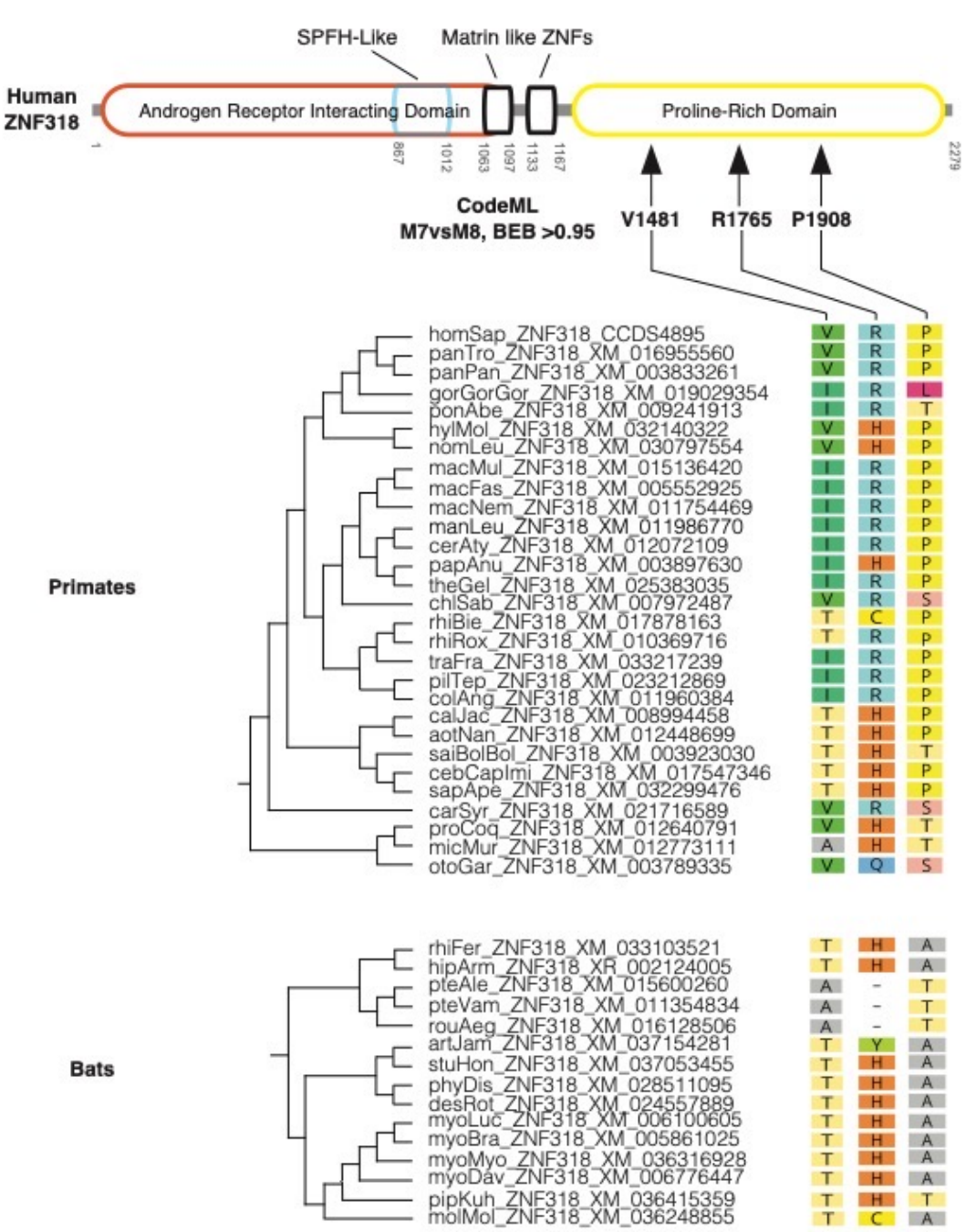
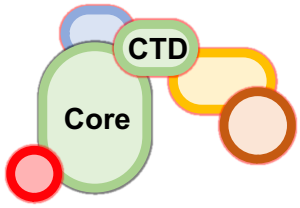


Figure S10. Rapid evolution of the Prim-Pol primase complex (POLA1, PRIM1, PRIM2) in primates and bats.

A



POLA1/p180
Polymerase activity (catalytic)

POLA2A/p70
B-subunit

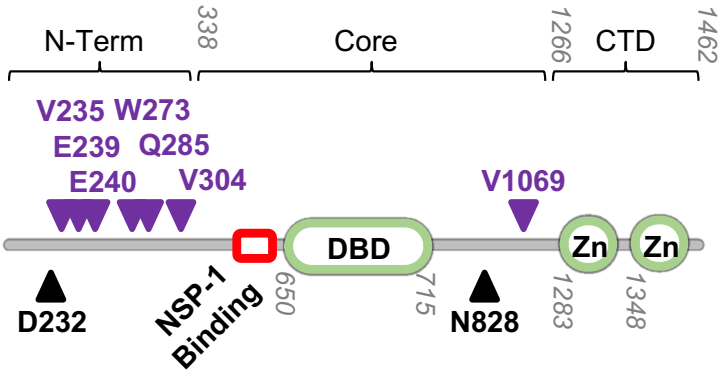
PRIM1/p49
Small primase (catalytic)

PRIM2/p58
Large primase

SARS-CoV-2 NSP1

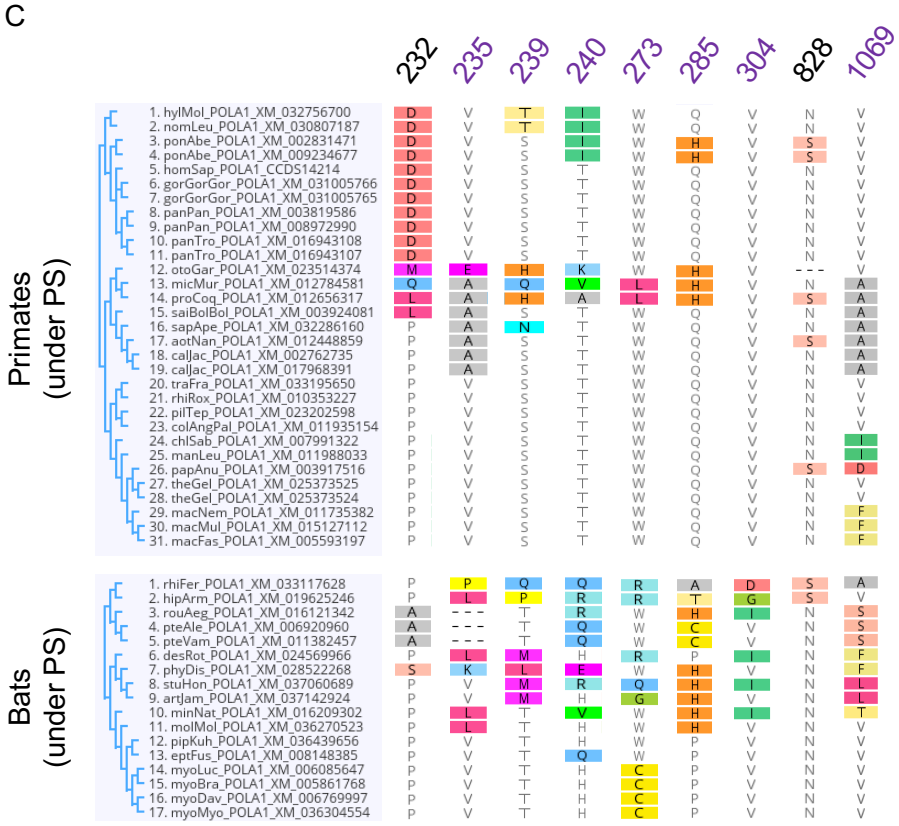
B

POLA1/p180



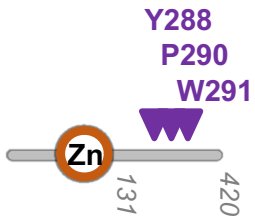
Sites under positive selection in **human** and **bat** genes (positions correspond to human ortholog)

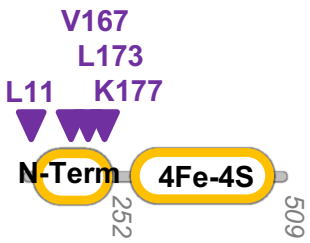
C



D

PRIM1/p49



E

PRIM2/p58