Article

Bat genomes illuminate adaptations to viral tolerance and disease resistance

https://doi.org/10.1038/s41586-024-08471-0

Received: 6 February 2023

Accepted: 28 November 2024

Published online: 29 January 2025

Open access

Check for updates

Ariadna E. Morales[1,2,3,30], Yue Dong[4,5,30], Thomas Brown[6,7], Kaushal Baid[8], Dimitrios - Georgios Kontopoulos[1,2,3], Victoria Gonzalez[8,9], Zixia Huang[10], Alexis-Walid Ahmed[1,2,3], Arkadeb Bhuinya[8,9], Leon Hilgers[1,2,3], Sylke Winkler[6,7], Graham Hughes[10], Xiaomeng Li[4,5], Ping Lu[5], Yixin Yang[5], Bogdan M. Kirilenko[1,2,3], Paolo Devanna[11], Tanya M. Lama[12,29], Yomiran Nissan[13,14], Martin Pippel[6,7], Liliana M. Dávalos[12,15], Sonja C. Vernes[11,16], Sebastien J. Puechmaille[17,18], Stephen J. Rossiter[19], Yossi Yovel[14,15], Joseph B. Prescott[20], Andreas Kurth[20], David A. Ray[21], Burton K. Lim[22], Eugene Myers[6,7], Emma C. Teeling[10], Arinjay Banerjee[8,9,23,24,25], Aaron T. Irving[26,27,28] ✉ & Michael Hiller[1,2,3] ✉

Zoonoses are infectious diseases transmitted from animals to humans. Bats have been suggested to harbour more zoonotic viruses than any other mammalian order[1]. Infections in bats are largely asymptomatic[2,3], indicating limited tissue-damaging inflammation and immunopathology. To investigate the genomic basis of disease resistance, the Bat1K project generated reference-quality genomes of ten bat species, including potential viral reservoirs. Here we describe a systematic analysis covering 115 mammalian genomes that revealed that signatures of selection in immune genes are more prevalent in bats than in other mammalian orders. We found an excess of immune gene adaptations in the ancestral chiropteran branch and in many descending bat lineages, highlighting viral entry and detection factors, and regulators of antiviral and inflammatory responses. *ISG15*, which is an antiviral gene contributing to hyperinflammation during COVID-19 (refs. 4,5), exhibits key residue changes in rhinolophid and hipposiderid bats. Cellular infection experiments show species-specific antiviral differences and an essential role of protein conjugation in antiviral function of bat *ISG15*, separate from its role in secretion and inflammation in humans. Furthermore, in contrast to humans, *ISG15* in most rhinolophid and hipposiderid bats has strong anti-SARS-CoV-2 activity. Our work reveals molecular mechanisms that contribute to viral tolerance and disease resistance in bats.

Bats (Chiroptera) are the only mammals capable of powered flight and have remarkable adaptations, such as echolocation and exceptional longevity. Bats are recognized as natural reservoirs for many viruses, some of which can cross species barriers and cause zoonotic diseases[1]. Viruses from 31 families have been found in bats, including paramyxoviruses (such as Hendra virus, Nipah virus and mumps virus), filoviruses (Marburg virus and Bombali virus), rhabdoviruses (rabies virus) and coronaviruses (close relatives of MERS-CoV, SARS-CoV and SARS-CoV-2)[6].

Coronaviruses are especially widely distributed in bats and have been detected in species from 15 of the 21 bat families[6,7]. In contrast to humans, in whom coronaviruses and other zoonotic viruses can cause hyperinflammation, respiratory insufficiency and multi-organ failure[8], infections in bats are largely asymptomatic. Indeed, experimentally

[1]LOEWE Centre for Translational Biodiversity Genomics, Frankfurt, Germany. [2]Senckenberg Research Institute, Frankfurt, Germany. [3]Faculty of Biosciences, Goethe-University, Frankfurt, Germany. [4]Deanery of Biomedical Sciences, College of Medicine and Veterinary Medicine, The University of Edinburgh, Edinburgh, UK. [5]Zhejiang University–University of Edinburgh Institute, Zhejiang University School of Medicine, Zhejiang University, Haining, China. [6]Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany. [7]DRESDEN concept Genome Center, Dresden, Germany. [8]Vaccine and Infectious Disease Organization, University of Saskatchewan, Saskatoon, Saskatchewan, Canada. [9]Department of Veterinary Microbiology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada. [10]School of Biology and Environmental Science, University College Dublin, Dublin, Ireland. [11]Neurogenetics of Vocal Communication Group, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. [12]Department of Ecology and Evolution, SUNY Stony Brook, Stony Brook, NY, USA. [13]School of Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel. [14]Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel. [15]Consortium for Inter-Disciplinary Environmental Research, SUNY Stony Brook, Stony Brook, NY, USA. [16]School of Biology, University of St Andrews, St Andrews, UK. [17]Institut Universitaire de France, Paris, France. [18]ISEM, University of Montpellier, CNRS, IRD, Montpellier, France. [19]School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK. [20]Centre for Biological Threats and Special Pathogens, Robert Koch Institute, Berlin, Germany. [21]Department of Biological Sciences, Texas Tech University, Lubbock, TX, USA. [22]Department of Natural History, Royal Ontario Museum, Toronto, Ontario, Canada. [23]Department of Biology, University of Waterloo, Waterloo, Ontario, Canada. [24]Department of Laboratory Medicine and Pathobiology, Temerty Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada. [25]Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada. [26]Department of Infectious Diseases, Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China. [27]Center for Infection, Immunity and Cancer, Zhejiang University–University of Edinburgh Institute, Zhejiang University School of Medicine, Zhejiang University, Haining, China. [28]Department of Biomedical Sciences, College of Medicine and Veterinary Medicine, The University of Edinburgh, Edinburgh, UK. [29]Present address: Department of Biological Sciences, Smith College, Northampton, MA, USA. [30]These authors contributed equally: Ariadna E. Morales, Yue Dong. ✉e-mail: aaronirving@intl.zju.edu.cn; michael.hiller@senckenberg.de

inoculating bat species with SARS-like or Marburg viruses showed productive viral infection and replication, but no clinical signs of disease[2,3], indicating that bats have evolved differential immune responses[9,10]. Infection-induced inflammation is necessary to mount effective antiviral responses, but inflammation can lead to collateral tissue damage. Bats limit the expression of inflammatory cytokines and dampen uncontrolled immune responses, reducing immunopathology[11–13]. For example, Egyptian fruit bats infected with Marburg virus upregulate antiviral genes but do not strongly induce pro-inflammatory genes[3].

Analyses of bat genomes revealed insights into immune system changes that may contribute to enhanced resistance to viral disease, such as selection of viral entry and innate immune factors, differences in natural killer cell receptors and type 1 interferon genes, selection and losses of pro-inflammatory regulators of canonical NF-κB signalling, and the absence of inflammasome-activating PYHIN genes[9,14–17]. Further studies revealed a general dampening of the inflammasome system to multiple immune stimuli[17,18]. Because powered flight requires high metabolic rates, and by-products of rapid metabolism activate the immune system, greater viral tolerance could have evolved as a by-product of immune adaptations to counter flight-induced sterile inflammation[11,18].

Here we integrated sequencing and assembly of new reference-quality bat genomes with a comprehensive mammalian genome analysis and experiments to better understand the genomic basis of disease resistance in bats.

Frequency of coronaviruses in bats

To compare how often coronaviruses were detected in comparison with other viruses in bats and rodents (another order that harbours a rich viral diversity), we used data from the ZOVER database[6], considering only metagenomic studies. This analysis revealed that 17.8% of the viral sequences detected in bats are from coronaviruses, compared with only 1.4% for rodents (two-tailed Fisher's exact test, $P = 5 \times 10^{-50}$; Fig. 1a and Supplementary Table 1). Among bat families, 41.6% and 31.7% of the viruses detected in individuals from the families Rhinolophidae and Hipposideridae, respectively, are coronaviruses, which is significantly higher than for other bat families (Fig. 1b).

Ten new reference-quality bat genomes

The Bat1K project used long-read and long-range sequencing technologies to generate ten new reference-quality genomes (Supplementary Table 2). Specifically, we sequenced four rhinolophid (*Rhinolophus yonghoiseni*, *Rhinolophus perniger lanosus*, *Rhinolophus affinis* and *Rhinolophus trifoliatus*) and three hipposiderid (*Hipposideros larvatus*, *Aselliscus stoliczkanus* and *Doryrhina cyclops*) species. Most of these species live in Southeast Asia, which is a hotspot for zoonotic diseases. *Rhinopoma microphyllum* (family Rhinopomatidae) and *Megaderma spasma* (family Megadermatidae) were sequenced to represent close sister families of the Rhinolophidae–Hipposideridae clade. Finally, we sequenced *Mops condylurus* (family Molossidae), which is implicated as a natural reservoir for Bombali virus (genus *Orthoebolavirus*)[19]. For nine species, we sequenced 27–42× PacBio circular consensus (HiFi) reads for contig assembly, and roughly 60× Hi-C data for scaffolding. The *M. spasma* genome was assembled from Oxford Nanopore long reads (81× coverage), Bionano optical maps, 10× Genomics linked reads and HiC data.

All ten assemblies far exceed the minimum Bat1K standards[20] (Fig. 1c,d and Supplementary Figs. 1 and 2). At least 50% of all assemblies are in contiguous sequences (contig N50) ranging from 12.5 Mb to 72.2 Mb. Contig N90 values range from 3.8 MB to 32.6 Mb. Chromosome-level scaffolds comprise 91.8–99.7% of our assemblies, and estimated base accuracies (consensus quality, QV = 61.8–69.7, indicating less than 1 error per megabase) are very high for all HiFi-based assemblies. To assess gene completeness, we used TOGA, a method that integrates reference-based gene annotation, inferring orthologous genes, and gene classification[21]. Compared with assemblies based on short reads, long-read assemblies consistently exhibit more genes with intact and complete reading frames (Fig. 1e, Supplementary Figs. 3 and 4, and Supplementary Table 3). We show that high assembly completeness and quality facilitates comprehensive screens for positive selection in genes (Extended Data Fig. 1). Together, the combination of HiFi and HiC data is a powerful strategy to generate reference-quality assemblies for bats and other species. Similar to other bats, the new genomes exhibited an accumulation of recent DNA transposon insertions (Supplementary Fig. 5).

To generate a genome-based bat phylogeny that includes the newly sequenced species, we used both multi-species coalescence and concatenated alignments of 1:1 orthologous genes, representing 30,354,372 base pairs (bp), which consistently inferred the same tree topology (Fig. 1f), in agreement with previous phylogenies[22]. Support values are high for all nodes. Finally, we used 17 fossil calibration points to infer a time-calibrated tree (Fig. 1f, Supplementary Fig. 6 and Supplementary Table 4), which estimated the Rhinolophidae–Hipposideridae common ancestor to be around 35 million years old.

Immune gene selection in mammals

The ability of bats to limit disease during viral infections is probably the result of immune system adaptations, which may be detectable as signatures of episodic positive selection in genes. To test how prevalent positive selection is among different mammalian orders and different functional groups of genes, we devised the following genome-wide screen. First, we used TOGA[21] to obtain orthologous genes across 20 bat genomes and 95 non-chiropteran mammals that represent 9 other orders, including per order up to 20 species selected for having 16,000 or more intact orthologues (Fig. 2a, Supplementary Fig. 7 and Supplementary Table 5). Second, we used the sensitive branch-site model aBSREL[23] to test all 228 branches in the 115-species tree for positive selection. Instead of testing predefined hypotheses, our exploratory screen also allows for recurrent and convergent positive selection. Considering 17,130 genes, we found 8,608 genes that show selection in at least one branch (Supplementary Table 6). Third, for each of the ten mammalian orders, we determined functional enrichments of positively selected genes (Fig. 2, Extended Data Fig. 2 and Supplementary Table 7).

Considering all high-level biological processes, as defined by top-level Gene Ontology (GO) terms, we found that bats have the most-significant enrichment for 'immune system process' (a corrected P value (P_{cor}) of 2.02×10^{-16}), followed by rodents ($P_{cor} = 6.75 \times 10^{-11}$; Fig. 2a). This pattern is not driven by unequal taxonomic representation or substantial genome quality differences, because four other orders (Primates, Rodentia, Artiodactyla and Carnivora) are also represented by 20 species with comparable assembly qualities (Supplementary Fig. 7). This pattern is also robust to subsampling 10 species from the 20-species orders and using the alternative instead of the primary bat assemblies (Extended Data Fig. 2 and Supplementary Figs. 8 and 9). Moreover, immune-related genes positively selected in bats are not more prone to undergoing structural rearrangements in bats than not-selected immune genes (Supplementary Fig. 10). The number of positively selected immune-related genes per bat branch did not correlate with viral diversity, longevity quotient or hibernation (Supplementary Fig. 11).

We observed that branch length measured in millions of years and substitutions per site in both neutrally evolving and coding regions correlate significantly with the number of positively selected immune genes (Fig. 2b and Extended Data Fig. 3). Using the best-fitting regression model (Supplementary Tables 8 and 9), we coloured all branches according to the difference between observed and expected number of positively selected immune genes, which highlights many bat lineages as outliers (Fig. 2c). Consistent with an excess of immune gene

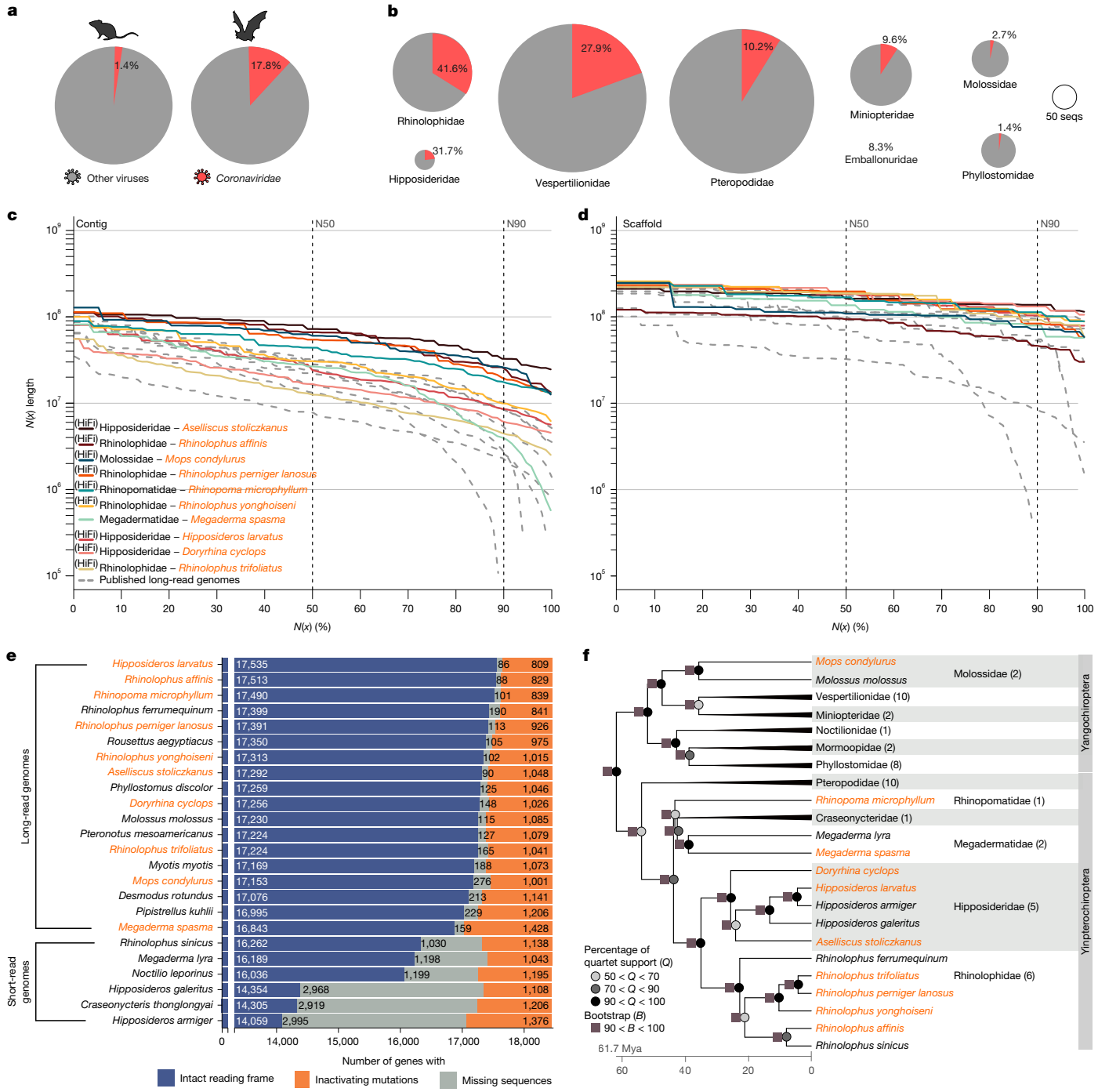


Fig. 1 | High-quality chromosome-level genome assemblies of ten bat species. **a**,**b**, Percentage of viruses from the family *Coronaviridae* detected in rodents (left) and bats (right) (**a**) and in different bat families (**b**). Pie size is proportional to the number of viral sequences (seqs). Pies $n < 50$ not shown (Emballonuridae $n = 12$). **c**,**d**, Assembly contiguity visualized as $N(x)$ graphs showing contig (**c**) or scaffold (**d**) size for which x% of the assembly consists of contigs or scaffolds of at least that size. New bat assemblies (solid coloured lines) are sorted by contig N50 in the inset legend. **e**, Status of 18,430 ancestral placental mammalian genes per assembly inferred by TOGA. An excess of genes with missing sequences (grey) indicates lower assembly completeness; an excess of genes with inactivating mutations (premature stop codons, frameshifts, splice-site disruptions and exon or gene deletions; orange) indicates lower base accuracy. **f**, Phylogenetic placement of newly sequenced species. The timetree was reconstructed for 50 bat species covering 12 families using exon-by-exon alignments of 16,860 orthologous genes. New Bat1K genomes are in orange text. Mya, million years ago. The rodent silhouette was obtained from PhyloPic (phylopic.org). *Mus musculus*, created by Daniel Jaron under a CC0 1.0 Universal Public Domain licence. The bat silhouette was vectorized by A.E.M.

selection in bats, models with two intercepts reveal a significantly higher intercept for bats than for non-flying mammals (Extended Data Fig. 3). Notably, the ancestral chiropteran branch has almost twice as many immune genes under selection than expected (42 observed versus 22 expected), thus exhibiting a higher excess of immune gene selection than ancestral branches of all other orders (for example, 42 versus 25 for Afrotheria and 20 versus 21 for Rodentia) (Fig. 2c and Supplementary Table 8). A similar pattern is observed for GO 'immune response' and SARS-CoV-2 relevant immune pathways (Extended Data Fig. 4 and Supplementary Fig. 12). By contrast, genes that are potentially relevant for other bat adaptations (such as longevity and echolocation) show no excess of selection (Supplementary Fig. 13). Together, these

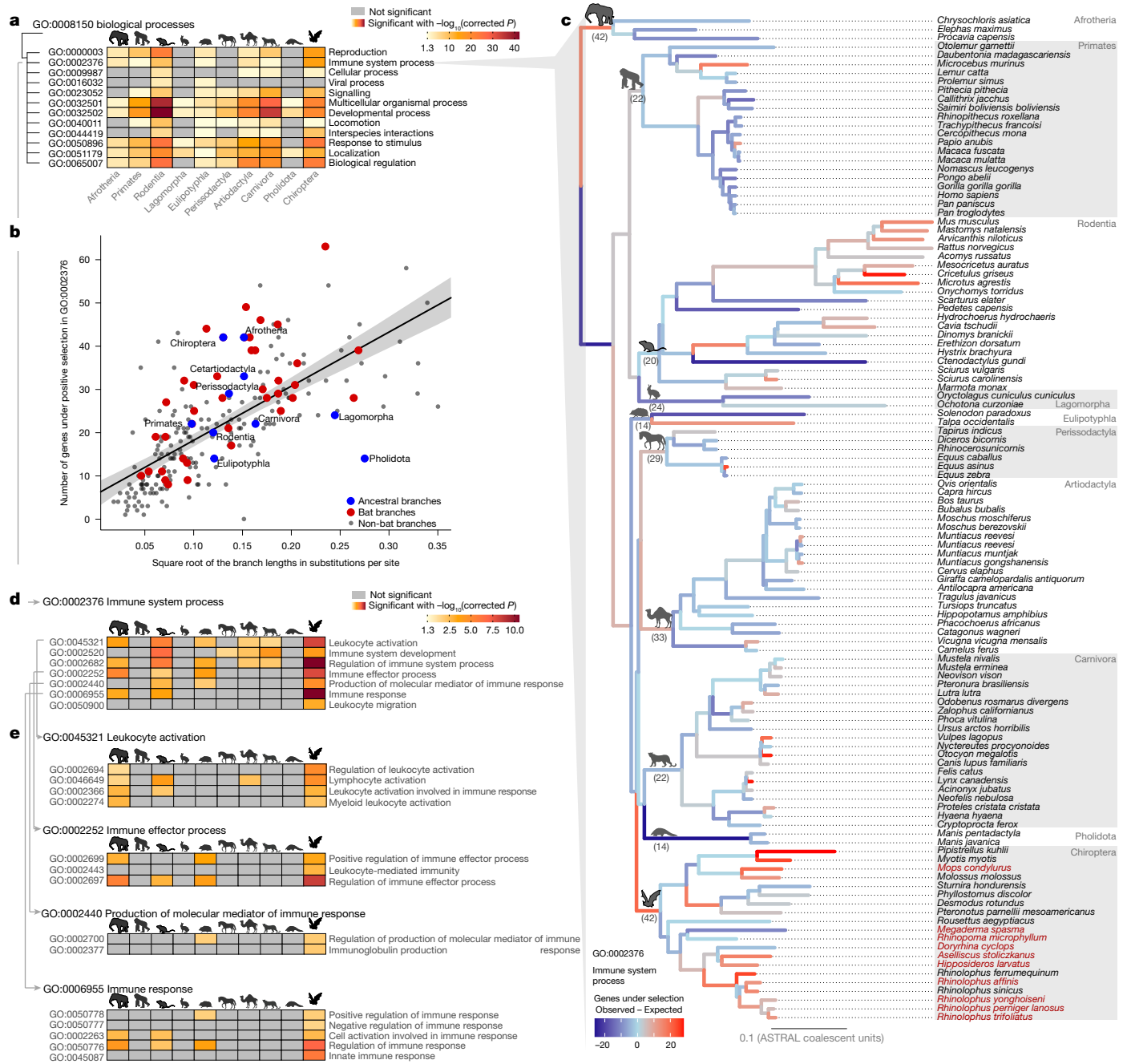


Fig. 2 | Selection of immune-related genes is most prevalent in bats. **a**, Functional enrichments of genes under positive selection in mammalian orders (columns). All high-level GO biological processes were tested (rows), but only terms significantly enriched in at least one order are shown. Coloured cells indicate significant enrichments after multiple test corrections. **b**, Linear regression model showing a significant correlation between the number of genes positively selected on a branch and the square root of the branch length. The grey shaded area indicates the 95% confidence interval. **c**, Per-branch signal of immune selection for the 115-mammal phylogeny. Branches are colour-coded according to the difference between the observed and expected number of positively selected 'immune system process' genes, calculated from the regression model. New Bat1K genomes are shown in red text. Observed numbers for ancestral branches of mammalian orders are in parentheses. **d**,**e**, Enrichments of direct child terms of GO 'immune system process' (**d**) and child terms thereof (**e**). All animal silhouettes apart from the bat were obtained from PhyloPic (phylopic.org). WT, wild type; C78A, Cys78 to Ala mutation; ΔC78, Cys78 deletion; S77C, Ser77 to Cys mutation. *Mus musculus*, by Daniel Jaron under a CC0 1.0 Universal Public Domain licence; *Gorilla gorilla gorilla* by T. Michael Keesey (after Colin M. L. Burnett) under a CC0 1.0 Universal Public Domain licence; *Leporidae*, by Sarah Werning under a CC BY 3.0 licence; *Erinaceus europaeus* by Roberto Díaz Sibaja under a CC BY 3.0 licence; *Equus ferus przewalskii*, by Mercedes Yrayzoz (vectorized by T. Michael Keesey) under a CC BY 3.0 licence; *Camelus dromedarius*, by Steven Traver under a CC0 1.0 Universal Public Domain licence; *Panthera pardus*, by Margot Michaud under a CC0 1.0 Universal Public Domain licence; *Manis culionensis*, by Steven Traver under a CC0 1.0 Universal Public Domain licence. The bat silhouette was vectorized by A.E.M., the Afrotheria (Thai Elephant) silhouette was obtained from OpenClipArt (https://openclipart.org/) under a CC0 1.0 Universal Public Domain licence.

data indicate that chiropteran immune system changes are distinctive adaptations that originated early in the chiropteran lineage at the branch in which powered flight also evolved.

We next analysed the child terms of 'immune system process' (Fig. 2d). Compared with other mammalian orders, genes positively selected in Chiroptera have the most significant enrichments for

'immune response' ($P_{cor} = 1.44 \times 10^{-10}$), 'regulation of immune system process' ($P_{cor} = 2.12 \times 10^{-11}$), 'immune effector process' ($P_{cor} = 5.66 \times 10^{-8}$) and 'leukocyte activation' ($P_{cor} = 9.31 \times 10^{-9}$), which were also robustly detected in our subsampling analysis (Extended Data Fig. 2). More specific GO terms further highlight chiropteran enrichments related to innate and adaptive immune systems (Fig. 2e).

Immune-related changes in bats

To gain insights into the immunomodulatory pathways that can contribute to bat tolerance of coronaviruses and other viruses, we explored more specific enrichments and integrated knowledge from the literature, focusing on genes under selection in the ancestral branches of Chiroptera, Rhinolophidae–Hipposideridae and Rhinolophidae (Extended Data Fig. 5 and Supplementary Fig. 14).

To enter host cells, viruses use cell-surface receptors. We identified positive selection on ANPEP (a receptor used by human coronavirus 229E for entry[24], selected in Rhinolophidae), the cofactor SCARB1 (which facilitates SARS-CoV-2 entry by enhancing cell-surface attachment[25], selected in Rhinolophidae–Hipposideridae) and the endosomal protease CTSB (which mediates entry of ebolaviruses and reoviruses[26], selected in Chiroptera and in Rhinolophidae) (Extended Data Fig. 5b).

Viral infections are detected by pattern-recognition receptors, including toll-like receptors (TLRs) and RIG-I-like receptors (RLRs), which induce innate immune responses and pro-inflammatory cytokine release. Interestingly, our analysis showed that 'innate immune response' ($P_{cor} = 1.37 \times 10^{-5}$) is enriched only for genes positively selected in bats (Fig. 2e). Furthermore, 'inflammatory response' ($P_{cor} = 2.51 \times 10^{-14}$) and 'regulation of cytokine production' ($P_{cor} = 1.58 \times 10^{-7}$) have the most significant enrichments for bats (Supplementary Table 7). These GO terms include several viral sensors and inflammatory immune response regulators (Extended Data Fig. 5b). Among them, *TLR8* (repeatedly selected in Rhinolophidae–Hipposideridae and Rhinolophidae) induces pro-inflammatory cytokines after detecting single-stranded RNA of endocytosed viruses such as SARS-CoV-2 (ref. 27). *TRIM38* (selected in Rhinolophidae) encodes a ubiquitin/SUMO ligase that has multiple roles in immunity. During early infection, TRIM38 enhances innate immunity by preventing the degradation of viral RNA and DNA sensors (RIG-I, MDA5, cGAS and STING)[28,29]. During late infection, TRIM38 is upregulated by interferons and inhibits TNF- and IL-1β–mediated NF-κB activation to suppress inflammation[30] (Extended Data Fig. 5c). Thus, TRIM38 contributes to strong early innate immune responses and dampens inflammation at later stages, processes that are intensified in bats[12,13]. *BTK* (selected in Chiroptera) encodes a tyrosine kinase that contributes to TLR-induced production of anti-inflammatory IL-10 (ref. 31) (Extended Data Fig. 5b,c). Furthermore, BTK interacts with NLRP3 inflammasome components and is essential for inflammasome activation[32]. Consistently, BTK inhibitors reduce hyperinflammation in severe COVID-19 (ref. 33). Hence, changes in BTK function could be yet another mechanism for inflammasome dampening in bats[17]. *TNFAIP2* (selected in Rhinolophidae–Hipposideridae) is a TNF-inducible gene that is involved in negative feedback regulation of NF-κB signalling[34]. HP (haptoglobin, repeatedly selected for in Chiroptera and Rhinolophidae–Hipposideridae) inhibits T-cell proliferation and pro-inflammatory cytokine secretion from various immune cell types[35], and directly activates TLR4 signalling, which stimulates the secretion of IFNβ (encoded by *IFNB1*, selected in Chiroptera)[36]. Finally, *IL36A*, a pro-inflammatory IL-1 superfamily member that stimulates NF-κB signalling[37], was lost in the common ancestor of Rhinolophidae, Hipposideridae, Megadermatidae and Rhinopomatidae (Supplementary Fig. 15A). Another pro-inflammatory IL-36 family member, *IL36G*, was lost in the rhinolophid ancestor (Supplementary Fig. 15B–D).

The GO term 'defence response to virus' ($P_{cor} = 4.61 \times 10^{-5}$) is uniquely enriched for genes positively selected in bats (Supplementary Table 7). Crucial for viral defence is a potent type I interferon (IFN-I) response. Although delayed or ineffective interferon responses are linked to severe COVID-19 (ref. 38), interferons need to be downregulated after infection to prevent immunopathology and autoimmune disease. We found selection not only on *IFNB1*, but also on several other genes that regulate or are regulated by interferon signalling (Extended Data Fig. 5d). IFNB1 suppresses the secretion of the potent pro-inflammatory cytokine IL-17A (selected in Chiroptera), which can trigger cytokine storms in severe COVID-19 by inhibiting the differentiation of IL-17-producing T helper cells[39]. IFNB1 and other type I interferons activate the JAK–STAT cascade, which induces, among other genes, *ISG15* (an antiviral protein investigated in detail below), *IFIT2* and *IFIT3*. Both *IFIT2* (selected in Chiroptera) and *IFIT3* (selected in Rhinolophidae) encode cytoplasmic proteins that restrict the replication of coronaviruses and other viruses by inhibiting viral RNA translation[40]. During viral infections, *LRRC25* (under selection in Chiroptera) downregulates IFN signalling by mediating the autophagic degradation of ISG15-associated RIG-I, thus suppressing RIG-I-mediated expression of IFNβ- and IFN-I-induced genes such as *IFIT1* and *IFIT2* (ref. 41). Similarly, when located intracellularly, N-myc and STAT interactor (NMI, under selection in Rhinolophidae–Hipposideridae) promotes the proteasomal degradation of IRF7 to inhibit the expression of *IFIT1* and *IFIT3*, *ISG15* and other IFN-stimulated genes[42]. By contrast, extracellular NMI induces the release of pro-inflammatory TNFs and IL-6 by activating canonical NF-κB[43]. Thus, depending on the location, NMI has anti- or pro-inflammatory roles.

Functional enrichments of positively selected genes related to leukocyte migration (*CCR2* and *CCR5*), complement system (*C7* and *C1S*) and B cell signalling (*CD79A* and *BTK*) are detailed in Extended Data Fig. 5e,f. In summary, our analysis revealed multiple promising targets for future experimental validation.

Conserved Cys-residue deletion in ISG15

ISG15 is a ubiquitin-like protein that is strongly induced by IFN-I during viral infections and upregulated in patients with COVID-19 (refs. 4,44). It has multiple functions, both intracellular and extracellular. Intracellularly, free ISG15 can be conjugated to hundreds of newly synthesized host and viral proteins (termed ISGylation), helping to restrict virus replication. Notably, genes that encode two protein ligases required for ISG15 conjugation, *UBA7* (encoding UBE1L) and *HERC5*, are positively selected in *Rhinolophus sinicus* and the ancestor of *R. trifoliatus*/*R. perniger lanosus* (*HERC5*) (Supplementary Table 6). Furthermore, the gene encoding the above-mentioned ISG15/RIG-I-binding LRRC25 is positively selected in the ancestral Chiroptera branch. ISGylation can be antagonized by viral immune-evasion proteins, exemplified by SARS-CoV-2-encoded papain-like protease (PLpro), which de-ISGylates IRF3 and MDA5 to suppress antiviral responses[45,46]. Extracellular ISG15 can enhance secretion of pro-inflammatory cytokines and chemokines in a paracrine manner, a process implicated in hyperinflammation during severe COVID-19 (refs. 4,5). In particular, SARS-CoV-2 PLpro-mediated de-ISGylation increases the pool of free ISG15, enhancing its extracellular secretion and production of pro-inflammatory factors.

ISG15 exists both as a monomer and a dimer. Although only the monomer can be used for intracellular ISGylation, the formation of homodimers, stabilized by disulfide bonds with a conserved cysteine (Cys78 in humans), is thought to be required for the extracellular cytokine function of ISG15 (refs. 47,48).

We discovered that Cys78 is deleted in all rhinolophid and hipposiderid bats (Fig. 3a). Mutating Cys78 prevents ISG15 homodimerization, altering extracellular cytokine activity and possibly enhancing ISGylation[47,48] (dimerized ISG15 cannot be ISGylated), so we tested the hypothesis that Cys78 deletion alters the function of rhinolophid and hipposiderid ISG15.

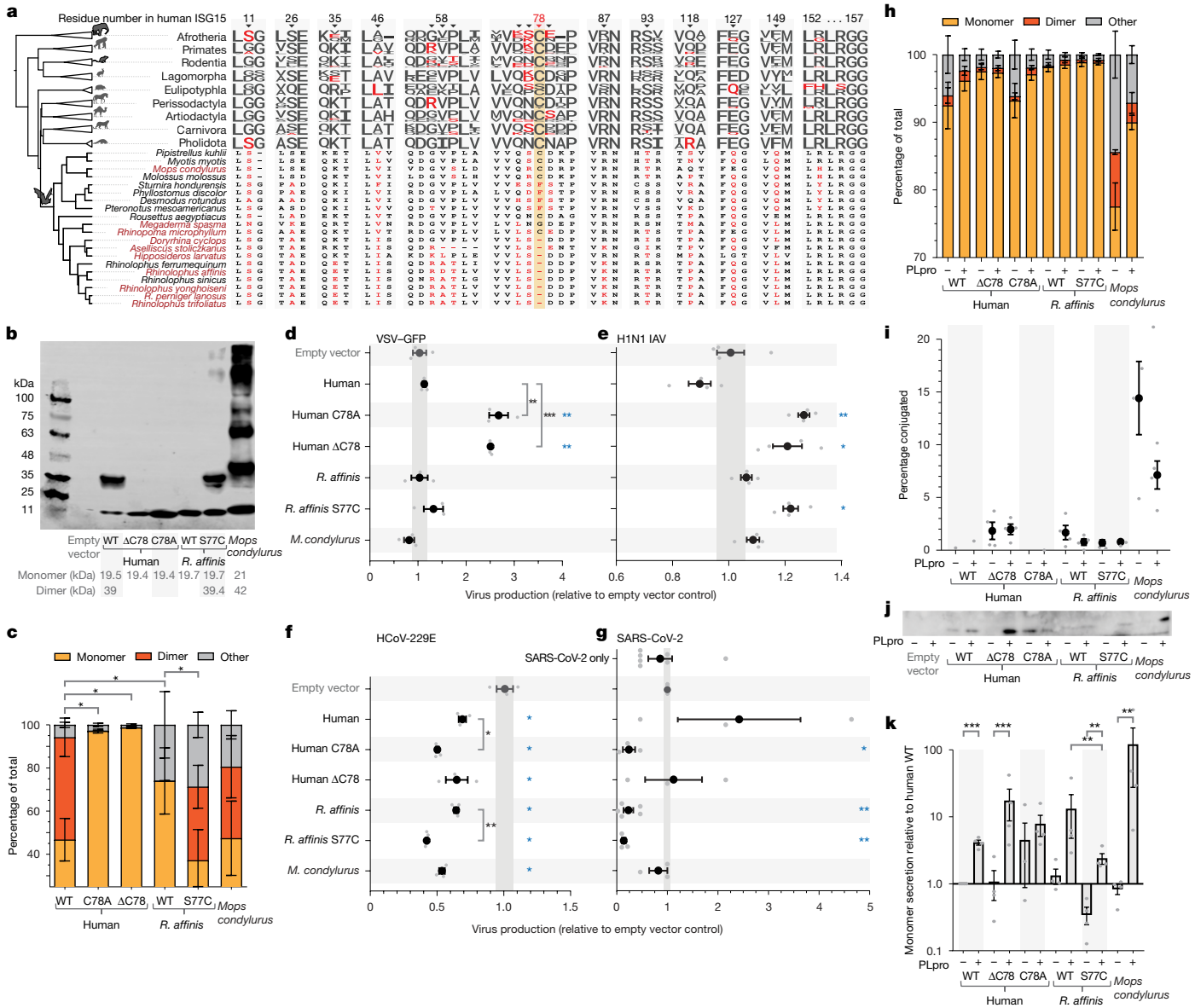


Fig. 3 | Altered dimerization and antiviral capacity of Cys78-mutated ISG15. **a**, Alignment of functionally important ISG15 residues. The sequence logos show the protein conservation of mammalian orders. **b**, Representative western blot of HEK293-subclone-ΔISG15 cells overexpressing Myc-tagged wild-type or mutated *ISG15*. Lysates were treated with 1,6-bis(maleimido)hexane (BMH) to cross-link disulfide bonds. **c**, Quantification of ISG15 western blots. Two-tailed *t*-tests test for differences in dimer formation; $n = 4$. **d**–**g**, Viral output, measured by $TCID_{50}$ (**f** and **g**) or a $TCID_{50}$ surrogate (Crystal Violet absorbance at 450 nm; **d** and **e**) in a HEK293-subclone (**d**), Vero-E6 (**e**), stable ANPEP-expressing HEK293 cells (**f**) and stable ACE2-expressing A549 cells (**g**) that express ISG15 constructs. **d**, VSV–GFP infection (multiplicity of infection (MOI) of 0.1 for 24 h); $n = 3$. **e**, H1N1 PR8 IAV infection (MOI of 0.1 for 48 h); $n = 4$. **f**, HCoV-229E infection (MOI of 0.1 for 72 h); $n = 3$. **g**, SARS-CoV-2 infection (MOI of 0.1 for 48 h); $n \geq 3$ for ISG15, $n = 7$ for SARS-CoV-2 control. **h**, Quantification of western blots (Extended Data Fig. 10e) of ISG15-transfected HEK293-subclone-ΔISG15 cells, co-transfected with SARS-CoV-2 PLpro (NSPC3C/L); $n = 4$. **i**, Quantification of ISG15 conjugated in cells after PLpro cleavage (matched to **h**). Decrease in conjugation with PLpro indicates de-ISGylation; $n = 4$. **j**, ISG15 released into the supernatant after PLpro cleavage (matched to **h**). In supernatants, ISG15 is not detected before infection; $n = 4$. **k**, Quantification of western blots for ISG15 cleavage and release into the supernatant (example in **j**), on a logarithmic scale; $n = 4$. Data are presented as mean (solid oval) and standard error of the mean (s.e.m.) (bars), showing individual data points of biological replicates as grey circles (**c**–**g**,**h**,**i** and **k**). All quantifications are from independent experiments. Significant differences to vector control (blue asterisks) or wild-type ISG15 (black asterisks) were determined by two-tailed *t*-tests (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$). Uncropped images, *P* values and all data are in Supplementary Data and Supplementary Tables 10, 11, 15, 16 and 18–21. WT, wild type; C78A, Cys78 to Ala mutation; ΔC78, Cys78 deletion; S77C, Ser77 to Cys mutation. All animal silhouettes apart from the bat were obtained from PhyloPic (phylopic.org). *Mus musculus*, by Daniel Jaron under a CC0 1.0 Universal Public Domain licence; *Gorilla gorilla gorilla* by T. Michael Keesey (after Colin M. L. Burnett) under a CC0 1.0 Universal Public Domain licence; *Leporidae*, by Sarah Werning under a CC BY 3.0 licence; *Erinaceus europaeus* by Roberto Díaz Sibaja under a CC BY 3.0 licence; *Equus ferus przewalskii*, by Mercedes Yrayzoz (vectorized by T. Michael Keesey) under a CC BY 3.0 licence; *Camelus dromedarius*, by Steven Traver under a CC0 1.0 Universal Public Domain licence; *Panthera pardus*, by Margot Michaud under a CC0 1.0 Universal Public Domain licence; *Manis culionensis*, by Steven Traver under a CC0 1.0 Universal Public Domain licence. The bat silhouette was vectorized by A.E.M., the Afrotheria (Thai Elephant) silhouette was obtained from OpenClipArt (https://openclipart.org/) under a CC0 1.0 Universal Public Domain licence.

Cys78 deletion prevents dimer formation

To investigate whether Cys78 deletion prevents stable homodimerization of bat ISG15, we first used structural modelling and molecular-dynamics simulations. Although Cys78-containing human ISG15 did indeed show stable homodimers, the Cys78-lacking ISG15 dimer of two rhinolophid and hipposiderid bats seemed more unstable, adopting topologically distinct conformations that are not stabilized by non-covalent interactions (Extended Data Figs. 6 and 7 and Supplementary Figs. 16–19).

To confirm this experimentally, we transiently transfected HEK293-subclone-ΔISG15 cells (human immortalized embryonic kidney cells with endogenous *ISG15* knockout that naturally express the ISGylation machinery) with constructs expressing wild-type Myc-tagged Cys78-containing ISG15 of human and *M. condylurus* (an outgroup bat to the Rhinolophidae–Hipposideridae) and Myc-tagged naturally Cys78-lacking ISG15 of *R. affinis* (Fig. 3b and Supplementary Figs. 20 and 21). Using a highly sensitive Myc antibody to exclusively detect transfected ISG15 and avoid cross-species antibody-detection problems, we found that human and *M. condylurus*, but not *R. affinis*, ISG15 forms stable dimers (Fig. 3b,c). Mutating Cys78 to alanine (C78A) or deleting Cys78 in human ISG15 prevented stable dimers and increased the monomer percentage, whereas restoring Cys78 (S77C mutation) in *R. affinis* ISG15 enabled dimerization (Fig. 3b,c). Furthermore, removing Cys78 from human ISG15 increased the percentage of other ISGylated proteins (Fig. 3c). Together, these data demonstrate that Cys-mediated disulfide bonds are required for the formation of stable ISG15 dimers.

Effect of ISG15 against vesicular stomatitis virus

To explore whether Cys78 deletion affects the antiviral function of ISG15, we first transiently transfected HEK293-subclone-ΔISG15 cells with the above-mentioned ISG15 constructs, infected them with GFP-tagged vesicular stomatitis virus (VSV, a representative of the *Rhabdoviridae* family that is common in bats), and measured viral production as VSV–GFP release into the supernatant. Although we observed no antiviral activity of any wild-type ISG15, mutating or deleting Cys78 from human ISG15 increased VSV–GFP production (pro-viral), indicating altered antiviral capacity of Cys78-deficient human ISG15 (Fig. 3d).

We next examined how human and Cys78-lacking ISG15 from six rhinolophid and three hipposiderid bats affects viral entry and replication by quantifying the percentage of VSV–GFP positive cells. ISG15 from humans and all the bats except *R. trifoliatus* reduced the viral load (Extended Data Fig. 8a), implicating ISGylated proteins in blocking viral entry. Notably, we observed significant differences among rhinolophid species that share the ISG15 Cys78 deletion, indicating that other amino acid mutations alter ISG15 function in a species-specific manner. Next, we measured GFP intensity during early infection stages as a proxy of viral replication, which revealed a conserved role for ISG15 in reducing viral load (Extended Data Fig. 8b). Measuring Ki67 levels showed that human ISG15 had a positive effect on cell proliferation during infection, but no effect was detectable without viral infection (Extended Data Fig. 8c). Surprisingly, ISG15 from six bats increased cell proliferation even in the absence of infection. Together, our VSV infection experiments reveal species-specific antiviral differences of rhinolophid–hipposiderid ISG15, not explained by their shared Cys78 deletion, and indicate further roles for ISG15 beyond antiviral defence, which is relevant because several bats constitutively express ISG15 before IFN induction[49,50].

Effect of ISG15 against influenza A virus and coronaviruses

We next examined how human, *R. affinis* (with and without Cys78) and *M. condylurus* ISG15 affects the production of three other viruses: influenza A virus (IAV), human coronavirus 229E (HCoV-229E) and SARS-CoV-2. First, we tested IAV as a member of the *Orthomyxoviridae* family that is commonly identified in bats. We stably transduced Vero-E6 cells with the ISG15 constructs, infected them directly with IAV (H1N1/PR8 strain) and examined viral production by median tissue-culture infectious dose assay ($TCID_{50}$). Whereas human ISG15 reduced viral output in a small but non-significant manner, no difference compared with the empty vector control was observed for ISG15 of *R. affinis* and *M. condylurus* (Fig. 3e). Interestingly, mutating or deleting Cys78 from human ISG15 and restoring Cys78 in *R. affinis* ISG15 significantly increased viral production (pro-viral), which is not caused by decreased IRF3 activation or interferon-stimulated gene expression, and does not depend on the extracellular cytokine function of ISG15 (Extended Data Fig. 9). Rather, this indicates that conformational changes induced by the cysteine mutations interfere with the intracellular function of ISG15.

Second, to compare ISG15 activity against coronaviruses, we measured the production of HCoV-229E (a virus, probably originating in bats, that can cause the common cold[51]) in HEK293-subclone cells that stably express the HCoV-229E receptor ANPEP (CD13) and ISG15 constructs. All ISG15 constructs significantly reduced HCoV-229E production (Fig. 3f). Interestingly, human ISG15-C78A significantly enhanced antiviral efficacy compared with wild-type ISG15, but no difference was observed when deleting Cys78 (Fig. 3f). Restoring Cys78 in *R. affinis* ISG15 also enhanced the antiviral effect of ISG15. This corroborates the idea that ISG15 antiviral activity is influenced by epistatic interactions between Cys78 and other residues.

Because PLpro of sarbecoviruses, including HCoV-229E, can de-ISGylate human ISG15 (ref. 52), thus increasing the pool of free ISG15, we investigated whether HCoV-229E infection affects ISG15 secretion. Wild-type human and *R. affinis* ISG15 was poorly secreted from HEK293 cells, and no increase in secretion was observed during HCoV-229E infection (Extended Data Fig. 10). By contrast, *M. condylurus* ISG15 was strongly secreted already in the absence of HCoV-229E, and secretion was further enhanced during viral infection. Mutating or deleting Cys78 from human ISG15 increased monomer secretion in the absence of virus, indicating that the presence of Cys78 affects the basal secretion of human ISG15. During HCoV-229E infection, secretion of human Cys-lacking ISG15 was reduced (Extended Data Fig. 10), probably because HCoV-229E causes an upregulation of the ISGylation machinery and conjugation[53], resulting in a smaller pool of ISG15 available for secretion. Overall, these data indicate species-specific differences in ISG15 secretion during HCoV-229E infection, and that Cys78 affects both conjugation and secretion, separately to dimer stabilization.

Third, we tested ISG15 against the original SARS-CoV-2 strain that started the pandemic, using A549 cells stably expressing the SARS-CoV-2 receptor ACE2. Whereas human and *M. condylurus* ISG15 had no effect on SARS-CoV-2 virus production, which could be explained by Molossidae and humans not being known as sarbecovirus reservoirs (Fig. 1b), *R. affinis* ISG15 strongly reduced viral production (Fig. 3g), independently of cytokine enhancement (Extended Data Fig. 9). Intriguingly, the C78A mutation, but not the Cys78 deletion, enhanced the anti-SARS-CoV-2 function of human ISG15 compared with the wild type, indicating that Cys78 may have other roles besides affecting secretion and cytokine signalling activity, because both mutations abolish stable dimer formation.

Given that SARS-CoV-2 PLpro evades the antiviral function of human ISG15 by efficiently de-ISGylating it[54], we examined the effect of Cys78 on PLpro cleavage and ISGylation. Overexpressing SARS-CoV-2 PLpro in HEK293-subclone-ΔISG15 cells transfected with ISG15 constructs showed that PLpro did indeed decrease ISGylation (the percentage of proteins conjugated to ISG15) and increase the amount of monomer in cell lysates for all three wild-type constructs (Fig. 3h,i). PLpro led to a strong increase in the secretion of human and *M. condylurus* ISG15 into the supernatant, but only a smaller non-significant increase

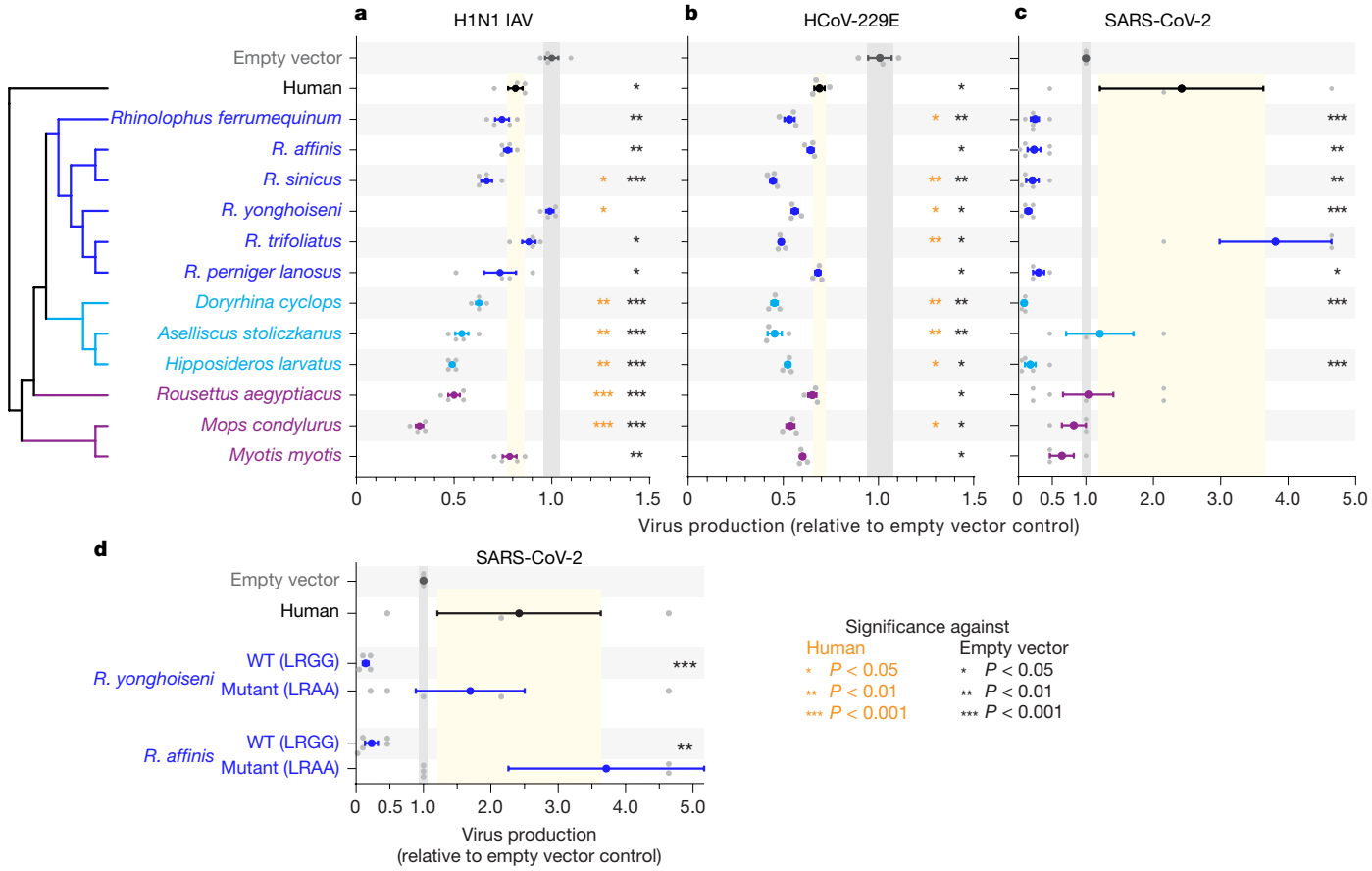


Fig. 4 | Species-specific ISG15 conjugation and antiviral function. a–c, Viral output, expressed relative to the empty vector (IRES–mCherry) control, in A549 (**a**), stable ANPEP-expressing HEK293 (**b**) and stable ACE2-expressing A549 (**c**) cells that express ISG15 constructs. **a**, H1N1 PR8 IAV infection (as per Fig. 3e). A549 cell supernatants titrated on Vero-E6 cells by plaque assays; $n = 4$. **b**, HCoV-229E infection (as per Fig. 3f). Viral output was measured by $TCID_{50}$ in Huh7 cells; $n = 3$. **c**, SARS-CoV-2 infection (as per Fig. 3g). Viral output from A549 cells was measured by $TCID_{50}$ in Vero76 cells; $n \geq 4$ for ISG15, $n = 7$ for SARS-CoV-2 control. **d**, As **c** except using wild-type (WT) and LRGG>LRAA ISG15 mutants that remove the ISG15 conjugation motif; $n \geq 5$ for LRGG mutations, $n = 7$ for SARS-CoV-2 control. Graphs represent mean and s.e.m. with individual data points in grey; *n* refers to independent biological replicates. Significant differences to vector control (black asterisks) or human ISG15 (orange asterisks) were determined by two-tailed *t*-tests (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$). All data are given in Supplementary Tables 22–25.

for *R. affinis*, again indicating that *R. affinis* is a poorer substrate for SARS-CoV-2 PLpro (Fig. 3j,k and Extended Data Fig. 10). Cys78 deletion, but not C78A, in human ISG15 increased PLpro-mediated secretion. Notably, restoring Cys78 in *R. affinis* ISG15 also increased PLpro-mediated secretion, highlighting the fact that Cys78 has different effects in *R. affinis* and human ISG15, and indicating potential structural changes, rather than the ability to form a disulfide bond, being important for evading SARS-CoV-2 PLpro. Our experiments also show that some epithelial cells can secrete ISG15 when high PLpro amounts are present (previously only shown in immune cells), indicating a potential role for ISG15 trans-signalling to immune cells in the surrounding environment during infection.

Bat ISG15s differ in antiviral activity

The presence or absence of Cys78 has no consistent effect on the antiviral activity of ISG15, and epistatic effects with other ISG15 residues may have a role, so results for *R. affinis* and *M. condylurus* may not generalize to other bats. We therefore compared the antiviral function of 12 bats, including Cys78-lacking ISG15 from six Rhinolophidae and three Hipposideridae, and Cys78-containing ISG15 from *M. condylurus*, a Pteropodidae (*Rousettus aegyptiacus*) and a Vespertilionidae (*Myotis myotis*).

For IAV, ISG15 of all bats except *R. yonghoiseni* showed antiviral activity compared with the vector control, and many have stronger antiviral activity than human ISG15; the presence of Cys78 did not correlate with antiviral function (Fig. 4a). For HCoV-229E, the ISG15 of all bats decreased viral production and, with the exception of *R. perniger lanosus* and *R. affinis* ISG15, the antiviral effect was stronger than for human ISG15 (Fig. 4b). Notably, for SARS-CoV-2, ISG15 of seven rhinolophid and hipposiderid bats strongly suppressed SARS-CoV-2 production by 80–90%, whereas the ISG15 of human, *M. condylurus*, *R. aegyptiacus* and *M. myotis* failed to reduce SARS-CoV-2 production (Fig. 4c). This shows that ISG15 of most rhinolophid and hipposiderid bats, for which coronaviruses are most commonly identified (Fig. 1b), can efficiently restrict cellular SARS-CoV-2 production.

To confirm that the enhanced intracellular anti-SARS-CoV-2 function of bat ISG15 is mediated by ISGylation, we prevented ISGylation by mutating the conserved carboxy-terminal LRGG ISGylation motif (Fig. 4d). This mutation ablated the anti-SARS-CoV-2 function of ISG15 and caused a pro-viral outcome, similar to human ISG15. Consistent with conservation of the E1-E2-E3-ligase genes in bats, we observed ISGylation with both bat and human ISG15 in bat cells, indicating compatible ISGylation machinery (Supplementary Fig. 21). Taken together, these data indicate that ISGylation is the key mediator of the enhanced anti-SARS-CoV-2 activity of rhinolophid and hipposiderid bat ISG15.

To further understand why *R. trifoliatus* is an outlier in the SARS-CoV-2 infection assays, we used the crystal structure of the SARS-CoV-2 PLpro-human ISG15 complex[54] and modelled human, *R. affinis*,

R. sinicus, *R. yonghoiseni* and *R. trifoliatus* ISG15 against PLpro. Compared with the human ISG15–PLpro model, we identified subtle changes in bat ISG15, for which three of six key contact residues differ from the human structure (N89E, T125S and F145L mutations) and contribute to altering the PLpro contact sites (Extended Data Fig. 11 and Supplementary Videos 10–19). Consistent with our experimental findings, *R. trifoliatus* ISG15, which has the weakest anti-SARS-CoV-2 activity among the tested horseshoe bats (Fig. 4c), had a contact region near the active site of PLpro that is very similar to that of human ISG15. By contrast, *R. yonghoiseni* ISG15, which has the strongest antiviral effect, had only two of the identified key contact residues still in contact with PLpro, but it also formed new contact surfaces away from the catalytic site associated with cleavage (Supplementary Video 15), potentially inhibiting the function of PLpro. This may be related to the evolution and diversification in function after the initial loss of Cys78, a scenario that requires further investigation in future studies.

Discussion

To understand the genomic basis of viral disease resistance in bats, we generated ten new reference-quality genomes, focusing on rhinolophid and hipposiderid bats that harbour close relatives of zoonotic coronaviruses, and performed a comprehensive exploratory selection screen covering 115 genomes. This screen revealed enrichments for immune gene selection in many mammalian orders, consistent with widespread selection exerted by pathogens driving rapid immune gene evolution[55]. However, compared with other orders, bats exhibit the most significant enrichments for immune gene selection, providing genomic evidence that bats have pronounced immune system adaptations. We uncovered an excess of immune gene selection in the ancestral chiropteran branch, raising the possibility that the evolution of flight is directly or indirectly linked to immune system changes. We highlight several regulators of inflammatory responses that inhibit pro-inflammatory cytokine production and participate in negative-feedback control of interferon signalling, indicating that these genes may contribute to preventing uncontrolled inflammation during viral infection in bats. Together, our findings provide promising experimental targets and a guide to unlocking the secrets behind bats' immune system adaptations.

Our ISG15 experiments have several implications. First, our mutation experiments show that Cys78 has other unpredictable effects on antiviral activity. Exemplified by IAV and HCoV-229E, we observed opposite effects when introducing an equivalent Cys mutation into human or *R. affinis* ISG15. Similarly, deleting or mutating Cys78 to alanine often had differential effects, exemplified by SARS-CoV-2, in which C78A but not Cys78 deletion conferred human ISG15 with an antiviral function. Hence, epistatic interactions with other residues also modulate the shape and function of ISG15. Second, although multiple studies suggested that ISG15 dimerization is required for secretion, we observed secretion of Cys-lacking ISG15 that cannot form stable dimers, in agreement with a study showing cytokine enhancement of C78S human ISG15 (ref. 56). As a result, further studies are required to fully understand the secretion mechanism and the functional roles of secreted ISG15. Third, by testing ISG15 from six rhinolophid and three hipposiderid bats, we unexpectedly observed ISG15-mediated antiviral differences not only between humans and bats, but also between closely related and co-roosting bat species (Fig. 4). Thus, one cannot necessarily generalize from one or a few tested species to all bats in the respective phylogenetic group. Fourth, we identified an effect of ISG15 of several bats (but not human) on cell proliferation, even in the absence of an infection. This reveals an unknown non-infection-related function of ISG15 that is probably relevant for bats that constitutively express ISG15. Fifth, whereas ISG15 of human and bats that are not frequently associated with coronaviruses lacked anti-SARS-CoV-2 activity, ISG15 of most rhinolophid and hipposiderid bats reduced SARS-CoV-2 production by 80–90%, indicating enhanced potency against SARS-CoV-2-related viruses. Sixth, previous work identified a key role of human ISG15 as a secreted pro-inflammatory cytokine, particularly when secretion is increased by virus-encoded PLpro-mediated cleavage during COVID-19. Our mutation experiments revealed that the anti SARS-CoV-2 activity of rhinolophid and hipposiderid ISG15 depends on the LRGG motif, indicating that intracellular ISGylation is key for the enhanced antiviral activity of bat ISG15. This indicates differences in intracellular ISGylation between human and bat ISG15, so 'ISGylome' comparisons should be the subject of future investigations. Together, our data indicate that *ISG15* is a fast-evolving antiviral effector gene that probably contributes to viral disease resistance in bats.

To fully elucidate the history of host–virus coevolution in bats, reference-quality genomes covering the diversity of bat families are needed; the ongoing phase 1 of the Bat1K consortium will soon provide these data[20]. Reference bat genomes, single-cell transcriptomics maps of their immune systems, breeding colonies, pluripotent stem cells and organoids[57–60] provide new tools to reveal the molecular adaptations behind bats' asymptomatic viral infections and other traits, which may have translational potential. Together, bats are an emergent model system for comparative mammalian biology, offering insights into immune system adaptations, healthy ageing and enhanced disease resistance.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-08471-0.

1. Olival, K. J. et al. Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650 (2017).
2. Schlottau, K. et al. SARS-CoV-2 in fruit bats, ferrets, pigs, and chickens: an experimental transmission study. *Lancet Microbe* **1**, e218–e225 (2020).
3. Guito, J. C. et al. Asymptomatic infection of Marburg virus reservoir bats is explained by a strategy of immunoprotective disease tolerance. *Curr. Biol.* **31**, 257–270 (2021).
4. Perng, Y.-C. & Lenschow, D. J. ISG15 in antiviral immunity and beyond. *Nat. Rev. Microbiol.* **16**, 423–439 (2018).
5. Munnur, D. et al. Altered ISGylation drives aberrant macrophage-dependent immune responses during SARS-CoV-2 infection. *Nat. Immunol.* **22**, 1416–1427 (2021).
6. Zhou, S. et al. ZOVER: the database of zoonotic and vector-borne viruses. *Nucleic Acids Res.* **50**, D943–D949 (2022).
7. Anthony, S. J. et al. Global patterns in coronavirus diversity. *Virus Evol.* **3**, vex012 (2017).
8. Lopes-Pacheco, M. et al. Pathogenesis of multiple organ injury in COVID-19 and potential therapeutic strategies. *Front. Physiol.* **12**, 593223 (2021).
9. Pavlovich, S. S. et al. The Egyptian rousette genome reveals unexpected features of bat antiviral immunity. *Cell* **173**, 1098–1110 (2018).
10. Letko, M., Seifert, S. N., Olival, K. J., Plowright, R. K. & Munster, V. J. Bat-borne virus diversity, spillover and emergence. *Nat. Rev. Microbiol.* **18**, 461–471 (2020).
11. Kacprzyk, J. et al. A potent anti-inflammatory response in bat macrophages may be linked to extended longevity and viral tolerance. *Acta Chiropt.* **19**, 219–228 (2017).
12. Banerjee, A. et al. Novel insights into immune systems of bats. *Front. Immunol.* **11**, 26 (2020).
13. Irving, A. T., Ahn, M., Goh, G., Anderson, D. E. & Wang, L.-F. Lessons from the host defences of bats, a unique viral reservoir. *Nature* **589**, 363–370 (2021).
14. Zhou, P. et al. Contraction of the type I IFN locus and unusual constitutive expression of IFN-α in bats. *Proc. Natl Acad. Sci. USA* **113**, 2696–2701 (2016).
15. Jebb, D. et al. Six reference-quality genomes reveal evolution of bat adaptations. *Nature* **583**, 578–584 (2020).
16. Scheben, A. et al. Long-read sequencing reveals rapid evolution of immunity- and cancer-related genes in bats. *Genome Biol. Evol.* **15**, evad148 (2023).
17. Ahn, M. et al. Bat ASC2 suppresses inflammasomes and ameliorates inflammatory diseases. *Cell* **186**, 2144–2159 (2023).
18. Ahn, M. et al. Dampened NLRP3-mediated inflammation in bats and implications for a special viral reservoir host. *Nat. Microbiol.* **4**, 789–799 (2019).
19. Bokelmann, M. et al. Tolerance and persistence of Ebola virus in primary cells from *Mops condylurus*, a potential Ebola virus reservoir. *Viruses* **13**, 2186 (2021).
20. Teeling, E. C. et al. Bat biology, genomes, and the Bat1K project: to generate chromosome-level genomes for all living bat species. *Annu. Rev. Anim. Biosci.* **6**, 23–46 (2018).
21. Kirilenko, B. M. et al. Integrating gene annotation with orthology inference at scale. *Science* **380**, eabn3107 (2023).
22. Doronina, L. et al. Contradictory phylogenetic signals in the Laurasiatheria anomaly zone. *Genes* **13**, 766 (2022).
23. Smith, M. D. et al. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* **32**, 1342–1353 (2015).

24. Yeager, C. L. et al. Human aminopeptidase N is a receptor for human coronavirus 229E. *Nature* **357**, 420–422 (1992).
25. Wei, C. et al. HDL-scavenger receptor B type 1 facilitates SARS-CoV-2 entry. *Nat. Metab.* **2**, 1391–1400 (2020).
26. Chandran, K., Sullivan, N. J., Felbor, U., Whelan, S. P. & Cunningham, J. M. Endosomal proteolysis of the Ebola virus glycoprotein is necessary for infection. *Science* **308**, 1643–1645 (2005).
27. Salvi, V. et al. SARS-CoV-2-associated ssRNAs activate inflammation and immunity via TLR7/8. *JCI Insight* **6**, e150542 (2021).
28. Hu, M.-M., Liao, C.-Y., Yang, Q., Xie, X.-Q. & Shu, H.-B. Innate immunity to RNA virus is regulated by temporal and reversible sumoylation of RIG-I and MDA5. *J. Exp. Med.* **214**, 973–989 (2017).
29. Hu, M.-M. et al. Sumoylation promotes the stability of the DNA sensor cGAS and the adaptor STING to regulate the kinetics of response to DNA virus. *Immunity* **45**, 555–569 (2016).
30. Hu, M.-M. et al. TRIM38 inhibits TNFα- and IL-1β-triggered NF-κB activation by mediating lysosome-dependent degradation of TAB2/3. *Proc. Natl Acad. Sci. USA* **111**, 1509–1514 (2014).
31. Schmidt, N. W., Thieu, V. T., Mann, B. A., Ahyi, A.-N. N. & Kaplan, M. H. Bruton's tyrosine kinase is required for TLR-induced IL-10 production. *J. Immunol.* **177**, 7203–7210 (2006).
32. Ito, M. et al. Bruton's tyrosine kinase is essential for NLRP3 inflammasome activation and contributes to ischaemic brain injury. *Nat. Commun.* **6**, 7360 (2015).
33. Roschewski, M. et al. Inhibition of Bruton tyrosine kinase in patients with severe COVID-19. *Sci. Immunol.* **5**, eabd0110 (2020).
34. Thair, S. A. et al. TNFAIP2 inhibits early TNFα-induced NF-x03BA;B signaling and decreases survival in septic shock patients. *J. Innate Immun.* **8**, 57–66 (2016).
35. Arredouani, M. et al. Haptoglobin directly affects T cells and suppresses T helper cell type 2 cytokine release. *Immunology* **108**, 144–151 (2003).
36. Kwon, J.-O. et al. Haptoglobin acts as a TLR4 ligand to suppress osteoclastogenesis via the TLR4-IFN-β axis. *J. Immunol.* **202**, 3359–3369 (2019).
37. Murrieta-Coxca, J. M. et al. IL-36 cytokines: regulators of inflammatory responses and their emerging role in immunology of reproduction. *Int. J. Mol. Sci.* **20**, 1649 (2019).
38. Christie, M. J. et al. Of bats and men: immunomodulatory treatment options for COVID-19 guided by the immunopathology of SARS-CoV-2 infection. *Sci. Immunol.* **6**, eabd0205 (2021).
39. Guo, B., Chang, E. Y. & Cheng, G. The type I IFN induction pathway constrains Th17-mediated autoimmune inflammation in mice. *J. Clin. Invest.* **118**, 1680–1690 (2008).
40. Diamond, M. S. & Farzan, M. The broad-spectrum antiviral functions of IFIT and IFITM proteins. *Nat. Rev. Immunol.* **13**, 46–57 (2013).
41. Du, Y. et al. LRRC25 inhibits type I IFN signaling by targeting ISG15-associated RIG-I for autophagic degradation. *EMBO J.* **37**, 351–366 (2018).
42. Ouyang, W. et al. NMI facilitates influenza A virus infection by promoting degradation of IRF7 through TRIM21. *Am. J. Respir. Cell Mol. Biol.* **65**, 30–40 (2021).
43. Xiahou, Z. et al. NMI and IFP35 serve as proinflammatory DAMPs during cellular infection and injury. *Nat. Commun.* **8**, 950 (2017).
44. Lowery, S. A., Sariol, A. & Perlman, S. Innate immune and inflammatory responses to SARS-CoV-2: implications for COVID-19. *Cell Host Microbe* **29**, 1052–1062 (2021).
45. Shin, D. et al. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **587**, 657–662 (2020).
46. Liu, G. et al. ISG15-dependent activation of the sensor MDA5 is antagonized by the SARS-CoV-2 papain-like protease to evade host innate immunity. *Nat. Microbiol.* **6**, 467–478 (2021).
47. Okumura, F., Lenschow, D. J. & Zhang, D.-E. Nitrosylation of ISG15 prevents the disulfide bond-mediated dimerization of ISG15 and contributes to effective ISGylation. *J. Biol. Chem.* **283**, 24484–24488 (2008).
48. Napolitano, A. et al. Cysteine-reactive free ISG15 generates IL-1β-producing $CD8\alpha^+$ dendritic cells at the site of infection. *J. Immunol.* **201**, 604–614 (2018).
49. Irving, A. T. et al. Interferon regulatory factors IRF1 and IRF7 directly regulate gene expression in bats in response to viral infection. *Cell Rep.* **33**, 108345 (2020).
50. Shaw, A. E. et al. Fundamental properties of the mammalian innate immune system revealed by multispecies comparison of type I interferon responses. *PLoS Biol.* **15**, e2004086 (2017).
51. Corman, V. M. et al. Evidence for an ancestral association of human coronavirus 229E with bats. *J. Virol.* **89**, 11858–11870 (2015).
52. Xiong, Y. et al. The substrate selectivity of papain-like proteases from human-infecting coronaviruses correlates with innate immune suppression. *Sci. Signal.* **16**, eade1985 (2023).
53. Okuno, S. et al. SARS-CoV-2 spike receptor-binding domain is internalized and promotes protein ISGylation in human induced pluripotent stem cell-derived cardiomyocytes. *Sci. Rep.* **13**, 21397 (2023).
54. Wydorski, P. M. et al. Dual domain recognition determines SARS-CoV-2 PLpro selectivity for human ISG15 and K48-linked di-ubiquitin. *Nat. Commun.* **14**, 2366 (2023).
55. Shultz, A. J. & Sackton, T. B. Immune genes are hotspots of shared positive selection across birds and mammals. *eLife* **8**, e41815 (2019).
56. Swaim, C. D. et al. Modulation of extracellular ISG15 signaling by pathogens and viral effector proteins. *Cell Rep.* **31**, 107772 (2020).
57. Wang, L.-F., Gamage, A. M., Chan, W. O. Y., Hiller, M. & Teeling, E. C. Decoding bat immunity: the need for a coordinated research approach. *Nat. Rev. Immunol.* **21**, 269–271 (2021).
58. Gamage, A. M. et al. Single-cell transcriptome analysis of the in vivo response to viral infection in the cave nectar bat *Eonycteris spelaea*. *Immunity* **55**, 2187–2205 (2022).
59. Déjosez, M. et al. Bat pluripotent stem cells reveal unusual entanglement between host and viruses. *Cell* https://doi.org/10.1016/j.cell.2023.01.011 (2023).
60. Chan, L. L. Y. et al. Generation of self-replicating airway organoids from the cave nectar bat *Eonycteris spelaea* as a model system for studying host–pathogen interactions in the bat airway epithelium. *Emerg. Microbes Infect.* **12**, e2148561 (2023).

Methods

Frequency of coronaviruses in bats and rodents

To assess how often coronaviruses were detected in bats and rodents, and in species belonging to different bat families, we used data from ZOVER[6] (last access, 25 January 2023), considering only metagenomic studies (Supplementary Table 1). We curated the bat taxonomic classification used in ZOVER to match the one described in https://batnames.org/. For example, species now recognized as part of the family Hipposideridae were assigned in ZOVER to the former classification in the family Rhinolophidae. Likewise, the family Miniopteridae, representing a separate bat family, was classified as Vespertilionidae. We then built contingency tables and applied a two-tailed Fisher's exact test to assess statistically significant differences in the prevalence of coronaviruses detected in bats versus rodents, in bats belonging to the family Rhinolophidae versus any other bats, and in bats belonging to the family Hipposideridae versus any other bats. Summary statistics and Fisher's exact tests were performed in R stats (v.4.1.1)[61].

Ethical statements and samples and collecting permits

For eight of the ten sequenced bat species, we used tissue samples from the Royal Ontario Museum mammal collection that had been kept frozen since their collection between 1992 and 2007 (Supplementary Table 2), highlighting the importance of frozen collections for biodiversity genomics[62]. The eight species are *Aselliscus stoliczkanus* (from Shuipu Village in China), *Doryrhina cyclops* (Parc National de Taï, Côte d'Ivoire), *Hipposideros larvatus* (Shuipu Village, China), *Rhinolophus affinis* (Shiwandashan National Reserve, China), *Rhinolophus perniger lanosus* (Shuipu Village, China), *Rhinolophus yonghoiseni* (Endau Rompin National Park, Malaysia), *Rhinolophus trifoliatus* (Endau Rompin National Park, Malaysia), and *Megaderma spasma* (Cat Tien National Park, Viet Nam) (Supplementary Table 2). Samples of the remaining two species, *Rhinopoma microphyllum* (northern Israel) and *Mops condylurus* (Bregbo, Côte d'Ivoire), were acquired from field expeditions. Samples were flash frozen in liquid nitrogen after being collected and stored at −80 °C until further processing.

The ethical statements of collecting and processing tissue samples for each species followed the procedures required by the following permits:

- *Doryrhina cyclops* (ROM-M100513); permit number 81 DPN from Direction de la Protection de la Nature, Côte d'Ivoire;
- *Rhinolophus yonghoiseni* (ROM-M113050); *Rhinolophus trifoliatus* (ROM-M113012); reference number PTN(J) 3/8 from Perbadanan Taman Negara (National Parks Corporation), Johor, Malaysia;
- *Aselliscus stoliczkanus* (ROM-M118506), *Hipposideros larvatus* (ROM-M118627), *Rhinolophus affinis* (ROM-M116429), *Rhinolophus perniger lanosus* (ROM-M118548); certificate numbers 2007/CN/ES133-137/KM from the Endangered Species Import and Export Management Office of the People's Republic of China;
- *Megaderma spasma* (ROM-M110751); Number 138/STTN from the Institute of Ecology and Biological Resources, National Center for Science and Technology, Viet Nam;
- *Mops condylurus* (ID: 03#106); capture of bats and animal work were performed with the permission of the Laboratoire Central Veterinaire, Laboratoire National d'Appui au Développement Agricole (LANADA), Bingerville, Côte d'Ivoire (no. 05/virology/2016) and the Ministère des Eaux et Forêts (No. 0474/MINEF/DGFF/FRC-aska);
- *Rhinopoma microphyllum*; National Parks Authority, permit 2013/04169, IACUC 04-20-019.

Extraction of long genomic DNA

Ultralong and long genomic DNA from various tissues (Supplementary Table 2) was isolated with the Nanobind Tissue Big DNA Kit from Circulomics (part number NB-900-701-01, protocol version Nanobind Tissue Big DNA Kit Handbook v.1.0 11/19) following the manufacturer's instructions (https://www.circulomics.com/nanobind). In brief, 25–40 mg of liver, spleen or heart tissue was minced to small slices on a clean and cold surface. Tissues were homogenized with the TissueRuptor II device (Qiagen), making use of its maximum settings. After complete tissue lysis, remaining cell debris was removed and the genomic DNA (gDNA) was bound to Circulomics Nanobind discs in the presence of isopropanol. High-molecular-weight (HMW) gDNA was eluted from the nanobind disks in elution buffer. The integrity of the HMW gDNA was determined by pulse field gel electrophoresis using the Pippin Pulse device (SAGE Science). Most of the gDNA was between 10 and 500 kilobases (kb) in length. All pipetting steps of ultralong and long gDNA were done with wide-bore pipette tips. Supplementary Fig. 22 shows an example gel image of the HMW gDNA extraction of *Rhinopoma microphyllum*.

HiFi library preparation and sequencing

Long insert libraries were prepared as recommended by Pacific Biosciences according to the guidelines for preparing HiFi SMRTbell libraries using the SMRTbell Express Template Prep Kit 2.0 (PN 101-853-100, v.03) for *Aselliscus stoliczkanus*, *Hipposideros cyclops*, *Hipposideros larvatus*, *Mops condylurus*, *Rhinolophus affinis*, *Rhinolophus perniger lanosus*, *Rhinolophus yonghoiseni*, *Rhinolophus trifoliatus* and *Rhinopoma microphyllum*. In brief, HMW gDNA was sheared to 20-kb fragments with the MegaRuptor device (Diagenode) and 10 µg sheared gDNA was used for library preparation. The SMRTbell library was size-selected for fragments of 9–13 kb with the BluePippin device according to the manufacturer's instructions. The size-selected libraries were run on Sequel II SMRT cells with the SEQUEL II sequencing kit 2.0 for 30 h on a SEQUEL II. Circular consensus sequences were called, making use of the default SMRTLink tools. For each species, a total of 65 Gb to 92 Gb of HiFi reads were generated, representing between 27× and 42× effective genome coverage.

Because Pacific Biosciences HiFi sequencing for *Megaderma spasma*, which was 25 years old, produced very little output despite good DNA and library quality, we used Oxford Nanopore Technologies for this species. Two Oxford Nanopore Technology ligation sequencing libraries were prepared following the manufacturer's instructions (article number SQK-LSK110, protocol version GDE_9108_v110_revH_10Nov2020). Input gDNA was either unsheared or sheared gDNA (50 kb), using the Diagenode MegaRuptor device as described for Pacific Biosciences HiFi sequencing. After repair of the sheared and unsheared gDNA, Oxford Nanopore Technologies sequencing adapters were ligated to the gDNA fragments and the resulting libraries were enriched for fragments larger than 3 kb. Both libraries were loaded on a Promethion device using R9.4.1 flow cells, generating 173 Gb of reads representing 81× effective genome coverage (12,173,706 total reads, 173,867,637,954 bp total yield, 14,282 bp mean read length, 10,552 bp median read length and 22,685 bp N50 read length).

ARIMA HiC

Chromatin conformation capture was done using the ARIMA-HiC (material number A510008) and the HiC+ kit (material number A410110) and following the user guide for animal tissues (ARIMA-HiC kit, document A160132 v.01 and ARIMA-HiC 2.0 kit, document number A160162 v.00). In brief, about 50 mg of flash-frozen powdered tissue was cross-linked chemically. The cross-linked gDNA was digested with a restriction-enzyme cocktail consisting of two and four restriction enzymes, respectively. The 5′-overhangs were filled in and labelled with biotin. Spatially proximal digested DNA ends were ligated. The ligated biotin-containing fragments were enriched and used for Illumina library preparation, which followed the ARIMA user guide for library preparation using the Kapa Hyper Prep kit (ARIMA document part number A160139 v.00). The barcoded HiC libraries were run on an S4 flow cell of a NovaSeq6000 with 300 cycles. Supplementary Table 2 shows an overview of species and the HiC protocol applied to each.

10× Genomics linked reads

To scaffold and correct base errors in the *Megaderma spasma* contig assembly, we generated linked Illumina reads with the 10× Genomics Chromium genome application, following the genome reagent kit protocol v.2 (document CG00043, rev. B, 10× Genomics). In brief, 1 ng of long or megabase-size gDNA was partitioned across 1 million gel bead-in-emulsions using the Chromium device. Individual gDNA molecules were amplified in these individual gel bead-in-emulsions in an isothermal incubation using primers that contain a specific 16-bp 10× barcode and the Illumina R1 sequence. After breaking the emulsions, pooled amplified barcoded fragments were purified, enriched and used for Illumina sequencing library preparation, as described in the protocol. Pooled Illumina libraries were sequenced to around 40× genome coverage on an S4 flow cell of a NovaSeq6000 with 300 cycles.

Genome assembly

Contig assembly. Pacific Biosciences CCS (HiFi) reads were generated from the subreads.bam files using the ccs command from the Pacific Biosciences pipeline v.4.2.0 (https://github.com/PacificBiosciences/ccs). For six species (*Aselliscus stoliczkanus*, *Hipposideros larvatus*, *Rhinolophus affinis*, *R. perniger lanosus*, *R. yonghoiseni* and *R. trifoliatus*), we created contig assemblies using hifiasm v.0.13 (ref. 63) with the argument -l0. The primary assembly was created by using purge_dups v.1.2.3 (ref. 64) on the p_ctg.fa output file. The alternative assembly was created by combining the haplotype-purged output from the p_ctg contigs with the a_ctg.fasta created by hifiasm. We then ran purge_dups on this combined alternative assembly to create the final alternative assembly for each species.

For *Rhinopoma microphyllum*, we assembled the contigs using hifiasm v.0.15.5-r352 with purging argument l2. For *Mops condylurus*, we used hifiasm v.0.15.4-r432. For both assemblies, we created the primary and alt contigs sets using purge-dups v.1.2.3 as above.

For *Doryrhina cyclops*, hifiasm created a large number of misassemblies joining regions from distinct chromosomes, which could not be reasonably corrected by hand. We therefore ran HiCanu v.2.1 (ref. 65) to create the initial contigs. Because this resulted in an assembly twice the expected size of the genome, we ran purge_dups on the contig assembly using custom cut-offs based on a haploid coverage of 13×: 8, 1, 1, 20, 2, 60 as in ref. 65. The purged output from purge_dups was taken as the primary contig assembly and the haplotype-purge output as the alternative assembly.

For *Megaderma spasma*, we ran Canu v.2.2 in -nanopore mode and created the primary contig sets using purge-dups as above. We used the basecaller Guppy v.5.0.12 for the libraries PAG52988 and PAG53246, and Guppy v.5.0.13 for the library PAH65658, both in high-accuracy mode.

Scaffolding of *Megaderma spasma*. We first scaffolded the contigs created by Oxford Nanopore Technologies reads using the 10× Genomics data. To this end, we mapped the 10× Genomics reads using longranger v.2.2.2 and scaffolded using Scaff10× v.4.2 and Break10× v.3.1. Next, we used bionano optical maps to further scaffold the assembly after 10× scaffolding. We created an optical map de novo assembly and then created the scaffold using Bionano Hybrid Scaffold using tools in Bionano Solve v.1.6.1. The resulting assembly was further scaffolded with HiC data, as described below.

HiC scaffolding. To scaffold contigs into chromosome-level scaffolds, we first mapped the Arima V2 HiC data to the genome assemblies using bwa-mem v.0.7.17-r1188 (ref. 66) and filtered reads on the basis of mapping quality and proper-paired alignments following the Arima mapping pipeline from the VGP (https://github.com/VGP/vgp-assembly/blob/master/pipeline/salsa/arima_mapping_pipeline.sh). We then scaffolded using salsa2 v.2.2 (ref. 67) with arguments: -m yes -p yes.

Manual curation. To join the contigs missed by salsa2 and break the joins that were spuriously created, we curated the scaffolds manually. In a few cases, hifiasm created false joins between two different chromosomes in one contig. To break these contigs, we mapped the CCS data to the contigs and found regions of the genome at these spurious joins. Then we identified either regions of low coverage (below 5, often 1 or 2 reads) or highly repetitive regions, in which repetitive tips of contigs from different chromosomes were falsely joined. In these cases, these ambiguous regions were removed from the genome and separate regions of the contigs were rescaffolded.

Polishing assemblies. To polish the final HiFi-based genomes and remove the remaining base errors, we used the CCS reads. To perform a polishing round, we mapped all CCS reads to the scaffolded, gap-closed assemblies using pbmm2 (https://github.com/PacificBiosciences/pbmm2) with arguments: --preset CCS -N 1 and called variants using DeepVariant (https://github.com/google/deepvariant/). We then filtered for sites with genotype 1/1 and a PASS filter value, meaning that all or nearly all the reads support an alternative sequence at this position and passed DeepVariant's internal filters. With this method, we did not polish any heterozygous or polymorphic regions of the genome, but only those that were incorrect and not supported by any CCS reads. We then corrected base errors using bcftools consensus v.1.12 (ref. 68).

For *Megaderma spasma*, we first mapped the 10× Genomics linked reads to the assembly using Longranger v.2.2.2. We then called variants using DeepVariant v.1.2.0, filtered the vcf file using Merfin v.1.1-development r197 (ref. 69) and determined the consensus using bcftools consensus v.1.12. We performed two rounds of polishing.

QV metrics. We used Merqury v.1.3 (ref. 70) to estimate QV values. For *Megaderma spasma*, which was based on Oxford Nanopore Technology reads, we used the 10× Genomics Illumina linked reads. For the other nine HiFi-based assemblies, we used the HiFi reads. For the HiFi-based assemblies, the base accuracy values were two orders of magnitude higher than with previous Pacific Biosciences CLR or Nanopore-based assemblies[15,16], although we note that these values are an upper bound because the reads used for assembly were used to estimate the QV, rather than an independent Illumina read dataset. Therefore, it is possible that any remaining errors in the assembly are also in the HiFi reads, which are known to have issues in homopolymer and other simple repeat regions[63,65].

Assembly completeness. We used two metrics. First, we used BUSCO v.5.1.1 (Benchmarking Universal Single-Copy Orthologs; odb10 dataset) with arguments: --mode 'genome' to compare the percentage of completely detected near-universally conserved mammalian genes across different assemblies[71]. Second, to further assess assembly completeness and base accuracy, we used an alternative benchmark metric provided by TOGA 1.0. We thus considered a set of 18,430 genes that probably existed in the placental mammal ancestor, defined as human genes that have an intact reading frame in at least one afrotherian and at least one xenarthran genome[21]. For each assembly, we determined how many ancestral genes have the following: an intact reading frame (TOGA classification intact, stating that the middle 80% of the coding sequence is present and lacks gene-inactivating mutations); inactivating mutations (TOGA classifications loss and uncertain loss); or missing sequence owing to assembly gaps or fragmentation (TOGA classifications partly intact and missing).

Annotation of transposable elements

To annotate transposable elements (TEs) in the newly sequenced bats, we first generated a de novo repeat library for each genome assembly using a novel pipeline consisting of RepeatModeler, RepeatClassifier,

custom scripts (https://github.com/davidaray/bioinfo_tools/blob/master/extract_align.py, RepeatAfterMe (RAM) (https://zenodo.org/record/7076442) and the TE-Aid package included in ref. 72.

In brief, each assembly was subjected to an initial RepeatModeler analysis. Because RepeatModeler will often produce incomplete putative consensus sequences, each putative consensus was subjected to an extension using RAM. These extended consensus sequences were then curated using a custom bash script (TEcurate.sh) that categorized each sequence into one of four categories (LINE, LTR, DNA or Unidentified) using RepeatClassifier, which is part of the RepeatModeler package. TEcurate.sh would then use the TE-Aid package to generate genome coverage plots, self-alignment dot-plots, structure and ORF plots, and copy number estimates.

For any elements clearly categorized as LINE, LTR or DNA, the identity provided by RepeatClassifier was used to generate a unique identifier that included species of origin, the RepeatModeler ID and TE class/family information. For example, hCyc.1.18-#LINE/L1 was discovered in *Doryrhina cyclops*. Its RepeatModeler ID was rnd-1_family-18 (1.18) and RepeatClassifer identified it as being a LINE1 element. LTR elements were further processed by hand to subdivide them into their LTR and internal segments, as in ref. 73. Consensus sequences with fewer than ten full-length copies were discarded.

TE-Aid plots of elements in the unidentified group were examined by eye to determine likely group membership using structural hallmarks, such as terminal inverted repeats (TIRs), long terminal repeats (LTRs) and sequence characteristics (such as repetitive tails, Helitron-specific CTAG motifs, SINE A-B boxes). Using these characteristics, putative consensus sequences were categorized as LINE, LTR, DNA, SINE, RC (rolling circle) or, when no clear hallmarks were identifiable, unknown.

After all the putative TE consensus sequences were classified and named, all consensus sequences were collapsed with previously known mammalian TEs per a variation of the 80–80–80 rule using USEARCH[74] with parameters -id 0.80 -minsl 0.95 -maxsl 1.05 -maxaccepts 32 -maxrejects 128 -userfields query+target+id+ql+tl and comparison to the mammalian TE library from ref. 75. All novel TE consensus sequences have been submitted to the Dfam TE database[76]. The resulting mammalian TE library was used to mask all assemblies with RepeatMasker. Output was processed to eliminate overlapping hits using RM2Bed.py, part of the RepeatMasker installation package to generate BED files for downstream analyses.

Annotation of microRNAs

Annotation of microRNA (miRNA) genes in the newly sequenced bats was done in a similar way to that in ref. 15. In brief, before miRNA prediction, repetitive and low-complexity regions in each bat genome were masked with the Dfam database (v.3.5; https://www.dfam.org/releases/Dfam_3.5/) using RepeatMasker (v.4.0.6; http://www.repeatmasker.org). For each masked genome, conserved miRNA genes were predicted using the Rfam database (v.14)[77] and Infernal (v.1.1.2)[78]. Infernal uses not only sequence similarity but also miRNA secondary structures for homology searches. We manually inspected 'spurious miRNAs' with multiple copies and determined the authenticity of these copies on the basis of the secondary structure we predicted with RNAfold (v.2.4.18)[79].

Repeat masking for pairwise genome alignments

To align newly sequenced genomes, we generated a de novo repeat library for each genome assembly using RepeatModeler (http://www.repeatmasker.org/, parameter -engine NCBI). The resulting library was then used to soft-mask the genome using RepeatMasker v.4.0.9 (parameters: -engine crossmatch -s).

Pairwise genome alignments

To infer orthologous genes for phylogenomics and selection screens, we used the human hg38 assembly as a reference species and generated pairwise genome alignments with bats and other mammals as query species. To this end, we used LASTZ 1.04.15 (ref. 80) to obtain local alignments. We used LASTZ parameters (K = 2,400, L = 3,000, Y = 9,400, H = 2,000, and the LASTZ default scoring matrix), which have a sufficiently high sensitivity to align orthologous exons between placental mammals[81]. Local alignments were chained using axtChain 1.0 (ref. 82) with default parameters except linearGap=loose. We used RepeatFiller 1.0 (ref. 83) (using the default parameters) to add missed repeat-overlapping local alignments to the alignment chains and chainCleaner 1.0 (ref. 84) (using default parameters except minBrokenChainScore = 75,000 and -doPairs) to improve alignment specificity.

Inferring and annotating orthologous genes

To infer orthologues for phylogenomic, selection and gene-loss analyses, we used TOGA 1.0 (ref. 21) (https://github.com/hillerlab/TOGA, commit v.c4bce48). In brief, TOGA uses pairwise genome alignment chains between a reference species (human hg38 assembly) and a query species (other mammals) to infer and annotate orthologous genes and to classify them as intact or lost. TOGA implements a novel paradigm to infer orthologous gene loci that relies largely on intronic and intergenic alignments and uses machine learning to accurately distinguish orthologous from paralogous or processed pseudogene loci. We used the human GENCODE V38 (Ensembl 104) annotation as input for TOGA, providing 39,664 transcripts of 19,456 coding genes.

Annotation completeness

We ran BUSCO with arguments: --mode 'protein' on the TOGA annotations to compare the percentage of completely detected almost universally conserved mammalian genes[71]. Note that applying BUSCO to a gene annotation (annotated proteins) produced by TOGA results in the detection of many duplicated genes because comprehensive annotations frequently include more than one transcript (splice variant) per gene. Because this does not indicate a problem, but rather a comprehensive transcript annotation, we reported only the number of completely detected BUSCO genes in TOGA annotations in Supplementary Tables 3 and 5.

Exon-by-exon alignments of orthologous genes

For phylogenomics and genome-wide selection screens, we used orthologues that are classified by TOGA as intact. TOGA is aware of orthology at the exon level, allowing the implementation of an exon-by-exon alignment to generate a comprehensive set of multiple codon alignments. For each human gene, we considered only the longest isoform. We included only 1:1 orthologues and excluded species for which no or multiple co-orthologues were inferred. Codons with frameshifting insertions or deletions and premature stop codons were masked with 'NNN' to maintain the reading frame. For each gene, every orthologous exon was aligned using MACSE v.2 (ref. 85), and all exons, together with codons split by introns, were concatenated into a multiple-codon alignment. Codon alignments were cleaned with HmmCleaner v.0.180750 (ref. 86) using default cost values to identify poorly aligned sequence segments and selectively remove them. From the multiple codon alignments of 19,288 genes, we used 17,130 (about 88%) that included at least 60% of the 115 mammals for phylogenetic inferences and selection screenings.

Phylogenetic and divergence-time estimation

To place the newly sequenced genomes into a bat phylogeny, we reconstructed phylogenetic relationships using whole-gene codon alignments, considering in total 50 bat species (Supplementary Table 3) and 16,860 genes. We also inferred a phylogenetic tree for all 115 mammals using 17,130 genes and used it as input for our selection screen and regression analysis (below).

To estimate a species tree, we followed both a coalescence-based approach, as implemented in ASTRAL v.5.5.9 (refs. 87,88), and a concatenated approach, as implemented in IQTREE v.2.1.3 (ref. 89). For the ASTRAL analysis, input trees were estimated in RAxML v.8.1.16

(ref. 90). Each gene was analysed using three independent replicates, a GTR + GAMMA model and a rapid-hill-climbing algorithm. Gene trees were used as input in ASTRAL with default parameters, and 100 bootstrap replicates were used to calculate the node support. Branch-support values were estimated using a transfer bootstrap expectation implemented in BOOSTER v.1 (ref. 91). For the IQTREE analysis, gene alignments were concatenated into a supermatrix and partitioned using best-fit models of sequence evolution for each gene, determined using ModelFinder (as implemented in IQTREE)[92]. A maximum-likelihood tree was inferred using IQTREE, with nodal support calculated using 1,000 bootstrap pseudo-replicates.

To estimate time-calibrated trees, we used a penalized likelihood method implemented in treePL[93,94] and 17 fossil calibration points[95,96]. First, one analysis was run to determine the best optimization parameters for treePL, and then a second analysis was run using the optimized values. Fossil calibrations[95] were applied to constrain maximum divergence times at relevant nodes (Supplementary Table 4). The time-calibrated phylogenies of bats and mammals are available at http://genome.senckenberg.de/download/Bat1KImmune/.

Selection of non-chiropteran genome assemblies

To obtain a broad genome representation of mammals for our selection screen, we included 95 other mammal and ten bat genomes, representing the main mammalian groups and bat families (Supplementary Table 5). We selected only assemblies for which at least 16,000 ancestral placental mammal genes have an intact reading frame, as determined by TOGA (detailed below) (Supplementary Fig. 7). For Chiroptera, we included the 10 new and 10 previously published bat assemblies; 18 of these 20 bats were assembled from long sequencing reads[15,16,97], and the remaining two genomes were assembled from Illumina short-read data[98,99] (Supplementary Table 3). For five mammalian orders (Primates, Rodentia, Artiodactyla, Carnivora and Chiroptera), we selected exactly 20 species. The other mammalian orders are represented by fewer species, because there were fewer sequenced genomes available that met our selection criteria. Details and sources of all 115 assemblies are provided in Supplementary Table 5.

Genome-wide unbiased selection screen

To identify genes under positive selection, we used aBSREL[23], an adaptive branch-site random-effects likelihood method implemented in HYPHY v.2.5.8 (ref. 100) that infers the strength of natural selection using the ratio of non-synonymous to synonymous substitution rates (the dN/dS metric). The values of dN/dS are interpreted as sites that experience purifying selection (dN/dS < 1), evolve neutrally (dN/dS ≈ 1) or experience positive diversifying selection (dN/dS > 1). For every phylogenetic branch, aBSREL tests if positive selection has occurred by determining whether more-complex models that allow a fraction of the codons to evolve with dN/dS > 1 fit the data best. The method uses the Akaike information criterion to assess the goodness of fit while penalizing increased model complexities (number of parameters). We ran aBSREL in exploratory mode to test all the branches and nodes in the phylogenetic tree for 115 mammals. For each gene, aBSREL corrected for multiple testing over all tested branches using the Benjamini–Hochberg procedure. In total, 17,130 genes were screened for selection using our ASTRAL topology as input. Alignments of genes of interest (Supplementary Fig. 14) were inspected by eye to rule out spurious signals resulting from misalignments.

To test whether enrichment results of genes under selection (see below) are representative of mammalian orders or driven by individual species, we did a subsampling analysis. We ran four more selection screens using the same dataset of 17,130 genes but subsampled the 5 mammalian orders with 20 species (Chiroptera, Carnivora, Artiodactyla, Rodentia and Primates) by randomly selecting only 10 species. Orders with less than 20 genomes were not subsampled, thus each subsampled dataset included 115 − 50 = 65 species. Subsamples 1–3 removed species at random, whereas subsample 4 included the 10 species in the five 20-species orders that were left out of subsample 1. For each subsampled set of species, codon alignments were generated and cleaned as for the full dataset, and the same input transcripts were screened for selection.

Gene-enrichment analyses

To explore whether genes under selection in different mammalian orders are enriched in specific functional groups, we performed gene-enrichment analyses, as implemented in gProfiler[101,102], using as background all the genes annotated for human reported in Ensembl v.106 (Ensembl Genomes 53, database built on 18 May 2022, e106_eg53_p16_65fcd97). As databases, we used Gene Ontology (http://geneontology.org/) and pathways from KEGG (https://www.genome.jp/kegg/), Reactome (https://reactome.org/) and WikiPathways (https://www.wikipathways.org/index.php/WikiPathways); miRNA targets from miRTarBase (http://mirtarbase.mbc.nctu.edu.tw/) and regulatory motif matches from TRANSFAC (http://genexplain.com/transfac/); tissue specificity from the Human Protein Atlas (https://www.proteinatlas.org/); protein complexes from CORUM (http://mips.helmholtz-muenchen.de/corum/); and human disease phenotypes from Human Phenotype Ontology (https://hpo.jax.org/app/).

It is expected that the more species are included in a mammalian order, the more genes are under selection in at least one of these branches. Therefore, we focused our comparison on the 5 orders that are represented by exactly 20 species. Because *P* values depend, among other factors, on sample size and the magnitude of the difference, differences in *P* values can be due to a larger sample size in one group (here, number of selected genes per order), despite having the same or a lower magnitude of the difference. gProfiler calculates the precision value, defined as the proportion of selected genes that are annotated with a particular term, which is listed in Supplementary Table 7 for all enriched terms. For all GO terms that are discussed in the main text as having the strongest enrichment for bats, we confirmed that among the five 20-species mammalian orders, the *P* value is the lowest and the proportion of selected genes with the GO term is the highest for Chiroptera. For example, immune system process (GO:0002376) has the lowest corrected *P* value (2.02×10^{-16} for bats, followed by 6.75×10^{-11} for rodents) and also the highest proportion of selected genes annotated with this term (0.184 for bats and 0.17 for rodents).

Correlation between branch length and number of genes under selection

We tested whether there is a significant correlation between branch lengths and the number of genes under selection (note that the aBSREL branch-site model tests for positive selection on every branch in the tree). For branch lengths, we used three independent estimations: first, millions of years from our time-calibrated phylogeny inferred using treePL[94] and fossil calibrations (Supplementary Table 4); second, the number of substitutions per neutral site estimated from 4D sites using phyloFit[103]; and third, the number of substitutions per site estimated from coding regions using IQTREE[89]. Normal probability plots indicate heavy tails (non-normality), which could be attributed to the unequal error variance of branch-length distribution. We then explored whether remedial measurements such as the Box–Cox approximation can be applied to find appropriate power transformations. In all cases, the likelihood function reaches its maximum when $\lambda \approx 0.05$ (where λ is the transformation parameter, ranging from −5 to 5); we therefore applied a square-root transformation. We fitted linear models with and without transformations and used Akaike's information criterion (AIC) to select the model(s) that best fit(s) the data given the model complexity[104]. AIC can be interpreted as a measure of lack of model fit, and to better interpret these relative values, Akaike weights (wAIC) are used to compare models. These weights are analogous to model probabilities because the sum of all wAIC values in a given set of models equals 1. The model

with the square root of substitutions per site estimated from coding regions fits the data best with a wAIC = 0.7025 (Supplementary Table 9) and supports a significant correlation with the number of selected genes (r^2 = 0.4917, F-statistic = 220.6, 1 and 226 d.f., $P < 2.2 \times 10^{-16}$). A significant correlation between branch length and the number of selected genes was also found for time and number of substitutions per neutral site (Extended Data Fig. 3). Using the best-fit model, we then considered specific immune gene sets and coloured the branches in the phylogenetic reconstruction by the observed number of selected genes minus the expected number based on the model.

To further test whether the number of immune genes under selection is higher in bats than in other mammals, we introduced a categorical taxonomy variable (bats and non-bats). First, we analysed the relationship between the number of immune genes under selection and branch lengths without accounting for different taxonomic groups, corresponding to one intercept and one slope. Second, we included the taxonomic group (bats or non-bats) as an independent correlate corresponding to different intercepts. Third, taxonomic group was included as a correlate, but interacting with the continuous branch-length variable, resulting in two models with two slopes for bats and non-bats, one with one intercept and another with two intercepts. This series was repeated with different branch-length estimates as a covariate. Based on previous analyses, branch-length variables were square-root transformed, with an untransformed analysis included for comparison. Finally, we compared the fit of a simpler frequency distribution than the negative binomial. A Bayesian approach was adopted to run these models, as a flexible way to both fit the model series and generate fit comparison statistics. A negative binomial frequency distribution was used to model the number of immune genes under selection, such that:

$$y_i \sim \text{negative binomial}\ (\lambda = \exp(l_i), pr)$$

where y_i is the count of positively selected genes for branch i, λ is the rate or mean of the Poisson distribution, exp is the inverse logarithmic link function, and $(1\text{-}pr)/pr$ defines a rate or shape parameter for the gamma distribution of a mixture of Poisson distributions, which relaxes the expectation of equality of mean and variance of the Poisson distribution. As a result, the negative binomial distribution is usually a better fit to biological data[105]. With a linear model applied to l:

$$l = \beta_0 + \beta_1 X$$

where β_0 represents the intercept, which is global for analyses with a single intercept or group-specific for testing bats versus non bats, β_1 is the coefficient on branch length, and X represents branch length. Both coefficients are normally distributed. To implement Bayesian sampling for these analyses, we used brms[106], a package that enables coding models in R for implementation in the stan statistical language[107]. For each model, we ran four separate Markov chain Monte Carlo chains using a Hamiltonian Monte Carlo approach. Compared with other Bayesian implementations, the Hamiltonian Monte Carlo approach saves time in sampling parameter spaces by generating efficient transitions spanning the posterior based on derivatives of the density function of the model. We estimated the R^2 of all models using the procedure outlined in ref. 108. To compare model fits, we used WAIC (the widely applicable information criterion), which weighs log pointwise predictive density against the expected effective number of parameters as defined in ref. 109 and provides estimates of the standard error of the difference between the best fit and other models.

ISG15 3D structure modelling

To test whether the Cys78 deletion in ISG15 of some bats affects the formation of stable ISG15 homodimers, we first used AlphaFold2 (ref. 110) through ColabFold v.1.3.0 (ref. 111) to infer the structure of the putative ISG15 homodimer of human, *Rhinolophus sinicus*, *Rhinolophus affinis*, *Rhinolophus yonghoiseni* and *Doryrhina cyclops*. Starting from each ISG15 sequence, ColabFold identified homologous sequences by running MMseqs2 (ref. 112) against the UniRef100 database[113] and against a set of environmental sequences[114]. Structural template information was obtained from the PDB70 database[115]. Next, the AlphaFold-multimer-v2 model[116] was used to infer five structural models of the dimer, with 12 rounds of recycling for model improvement. The resulting models were relaxed using the Amber force field[117] and were ranked according to their predicted template modelling score, which we used to identify the best model.

To further investigate the stability of the dimers inferred with AlphaFold2 (Supplementary Fig. 16), we conducted three replicate molecular dynamics simulations for ISG15 of human, *R. sinicus* and *D. cyclops* using GROMACS v.2022.1 (refs. 118,119) and the CHARMM36-Jul2021 force field[120]. More precisely, we prepared each dimer by treating termini as ionized (NH_3^+ and COO^-), assigning appropriate protonation states to amino acids (assuming pH = 7) as determined using PROPKA3 (refs. 121,122) and adding hydrogen atoms. Each dimer was subsequently placed in a periodic dodecahedral box, at a minimum distance of 2.5 nm from each box edge (Supplementary Fig. 17). The box was filled with TIP3P water molecules and with Na^+ and Cl^- ions as required to neutralize the system (referred to as low salt concentrations below). Following this, we performed energy minimization of the system, and examined the values of the potential energy and the maximum force to ensure that the system was sufficiently relaxed. Next, we applied position restraints on non-hydrogen protein atoms and equilibrated the system in two steps: first, under an NVT ensemble to stabilize the temperature (at 300 K); and second, under an NPT ensemble to stabilize the pressure (at 1 bar) and density of the system. For these, we used the velocity rescaling thermostat[123] and the Parrinello–Rahman barostat[124,125], set the integration time step to 2 fs and the duration of each equilibration step to 100 ps. As before, we manually examined the temperature, pressure and density to ensure that the system was successfully equilibrated (Supplementary Fig. 18). Finally, we removed the position restraints and conducted production simulations for 1 µs each, recording snapshots of the system every 100 ps. For each species, the three simulations ran for around 6 months on a compute node with 128 cores, summing to a total of about 550,000 CPU hours per species. To explore the effect of higher (physiological) salt concentrations, we conducted three more replicate molecular-dynamics simulations of the ISG15 dimer of *Rhinolophus sinicus* for 0.5 µs each using a physiological salt concentration (150 mM) (Extended Data Fig. 6).

To analyse the resulting molecular-dynamics trajectories, we combined the snapshots from the three replicate simulations per ISG15 dimer, removed water molecules and ions, and constructed a matrix of pairwise root mean square deviations using Carma v.2.01 (ref. 126). We then clustered the three (human and bats) root mean square deviation matrices according to the Partitioning Around Medoids algorithm[127] implemented in the cluster R package v.2.1.3 (https://CRAN.R-project.org/package=cluster). This allowed us to identify representative conformations, separately for each dimer. In particular, we set the number of clusters to all possible values between 2 and 10, and selected the clustering with the highest mean silhouette score[128]. Finally, we extracted the protein snapshots corresponding to the medoid of each cluster and compared them with the initial protein model obtained from AlphaFold (Extended Data Fig. 6). We also produced videos of the simulations using PyMOL v.2.5.0 for all nine molecular-dynamics simulations (Supplementary Videos 1–9).

For the two bat ISG15s, we also used the gmx hbond program with the -contact parameter to examine the contacts among ISG15 monomers (within 5 Å) that were present at each simulation snapshot. This allowed us to investigate how the fraction of native contacts (contacts in the original AlphaFold models that are hence present at the beginning of the simulations) changed across simulation time (Supplementary Fig. 19). As well as native contacts, we also constructed a presence/absence

matrix of all contacts between ISG15 monomers during the course of the simulations, which we analysed through PCA (Supplementary Fig. 19).

3D structure modelling of ISG15–PLpro complexes

We used Alphafold2-Multimer in ColabFold[111] and the recently identified crystal structure of SARS-CoV-2 PLpro in complex with human ISG15 (ref. 54) to model the complex of human, *R. affinis*, *R. sinicus*, *R. yonghoiseni* and *R. trifoliatus* ISG15 and PLpro (displayed as a monomer of each). The ISG15 chain of human from the crystal structure was replaced with full-length human ISG15 (including the remaining C-terminal residues after the cleavage site) then routinely swapped for each species of ISG15. The top model was chosen as previously and predicted template modelling displayed with scaling as per the predicted local distant difference test AlphaFold2 scale in ChimeraX. We used only the catalytic unit of the PLpro monomer (residues 819–2,763) that was previously shown to support independent modelling that reflected well the crystal structure. Supplementary Videos 10, 12, 14, 16 and 18 show the complete view of one molecule of PLpro (catalytic domain) with one molecule of ISG15. Supplementary Videos 11, 13, 15, 17 and 19 (human, *R. affinis, R. yonghoiseni, R. sinicus* and *R. trifoliatus*, respectively) represent zoomed views with a 360° rotational spin, highlighting the contact residues (green) calculated in ChimeraX (structure > contacts, VDW overlap, >-0.4, inter-model contacts) between PLpro and ISG15 and the key (homologous) amino acid residues (yellow) identified as contact points in the crystal structure between human ISG15 and PLpro. Snapshots of the entire complex and zoomed regions of the catalytic site with the C-terminal tail of ISG15 are shown in Extended Data Fig. 11.

Investigation of antiviral mechanisms of ISG15 in Rhinolophidae and Hipposideridae

Cell culture. Huh7, HEK293, A549, Vero-76 (CRL-1587, ATCC) and Vero-E6 cells were grown in typical DMEM (Gibco) supplemented with 10% FBS (ExcelBio) and 1% Pen/Strep (Sigma). Importantly, the HEK293 cells used in this study are a HEK293 subclone selected for high production of pseudoviruses. Compared with most HEK293 cells, these subclone cells express the basal machinery of IFN/ISGs and the ISGylation machinery (E1/E2/E3 ligases) is upregulated, which we validated by western blots for ISGylation and its machinery (Extended Data Fig. 10 and Supplementary Fig. 21). RsKT.01 cells (*Rhinolophus sinicus*) were a gift from Z. Shi and were grown in DMEM (Gibco) supplemented with 10% FBS (Biological Industries) and 1% Pen/Strep (Sigma). For HCoV experiments, an ANPEP/CD13-Flag construct (Sino Biological) was transfected into HEK293 cells (PEI, PolySciences), then selected (mixed pool) with hygromycin for 2 weeks to generate stable cell lines, validated by surface CD13 staining (Sino Biological, 1:2,000 dilution, no permeabilization), and used for consequent infection with HCoV-229E. For SARS-CoV2 experiments, A549-ACE2 cells (human lung adenocarcinoma-derived cells overexpressing human angiotensin-converting enzyme 2 (ACE2)) were provided by the Colpitts laboratory[129], and we used the clonal population A549-ACE2 B9. The A549-ACE2 cells were maintained in Ham's F-12K (Kaighn's) medium supplemented with 10% FBS, 10 μg ml^{-1} blasticidin and 1% Pen/Strep.

Huh7 cells were transfected with Lip2000 (Biosharp), HEK293 cells with polyethylenimine (Polysciences), A549 cells with Lipo6000, and Vero-E6 cells with Lipo8000 (Beyotime), each according to the manufacturer's instructions. All cell lines were tested for mycoplasma and were free of mycoplasma contamination.

Lentiviruses were generated using third-generation HIV-VSV.G lentiviruses with the psPax2 (Addgene plasmid 12260; http://n2t.net/addgene:12260; RRID:Addgene_12260) vector system in HEK293 cells. Geneblocks were synthesized (Beijing Tsingke Biotech or Genscript) according to TOGA annotations for *ISG15* (transcript ENST00000649529, aligned to human for validation) and cloned into the pLVX-IRES–mCherry vector under a CMV promoter for direct transfection or lentivirus generation. Lentiviral supernatants were prepared in low-FBS DMEM supplemented with 1% NEAA (Phygene), sodium pyruvate (Thermo Scientific, Gibco) and filtered through a 0.45 μm low-binding PES PVDF filter (Jet Biofil). Lentiviral transduction was done with 100 μl supernatant per well (six-well plates) of cells in 1% FBS with 4 μg ml^{-1} Polybrene (Biosharp) for 4–6 h; the medium was replaced with 10% serum, and 48–72 h later, cells were sorted for mCherry fluorescence and grown as stable (mix pooled) cell lines to minimize clonal variation. A549 and RsKT cells included a 'spinfection' step for centrifuging at 100*g* for 1 h at 37 °C in six-well plates with the lentivirus.

Myc-tagged ISG15 constructs were synthesized as above to ensure antibody compatibility (Sangon Biotech) (Supplementary Table 27). HEK293-ΔISG15 cells were co-transfected with 200 ng pCDNA3.1-NSP3C (PLpro), 200 ng pCDNA3.1-NSP3L and 300 ng ISG15 Myc-tagged constructs, then cells were induced with 1 μg ml^{-1} of poly-I:C 24 h after transfection, cell lysates were treated with DSP or BMH (see below) and collected 48 h after transfection. Myc constructs were validated to have minimal impact compared with non-tagged ISG15 constructs (Fig. 3d and Supplementary Fig. 23). HEK293-ΔISG15 cells were transfected with 50 ng UBE1L (Addgene, 12438) and UBE2L6 (Addgene, 98380), respectively, and cell lysates were collected 24 h after transfection.

Fluorescence-activated cell sorting. To generate ISG15 stable cell lines, lentiviral transduced Huh7, HEK293, A549, RsKT and Vero-E6 cells were sorted by fluorescence-activated cell sorting (FACS) using the BD influx system for mCherry-positive cells, normalized against autofluorescence in the respective parental cell line. VSV–GFP load was measured directly by GFP fluorescence intensity. For HCoV-229E, CD13 stable HEK293 cells were stained with 229E N protein (Sino biological, 1:2,000) and Ki67 (Beyotime, 1:500) for 30 min in FACS buffer containing 1× PBS (Gibco), 1% FBS and 1% Pen/Strep, after permeabilization with 0.05% TX-100 in TBS and blocking in 5% BSA in TBS-T. Cells were subsequently rinsed, stained with anti-mouse/rabbit CF-488/568/647 secondary antibody for 15 min (Biotium, dilution 1:10,000), rinsed three times and run on the ACEA Novocyte flow system.

Cell-viability assays. To infer metabolic activity by turnover of ATP, viability assays were performed by adding 10 μl CCK-8 (Transgen Biotech) directly to cells grown in DMEM, 1%FBS, Pen/Strep, incubated for 4 h, then measured over time in a Tecan Spark microplate reader at Abs 450 nm. The background was subtracted and normalized against the control.

Protein conjugation. To detect protein–protein conjugations (cross-linking of amide bonds), di(*N*-succinimidyl)3,3′-dithiodipropionate (DSP) (Aladdin, PubChem CID: 93313) was used. To detect disulfide-bond-linked proteins, 1,6-bis(maleimido)hexane (BMH) (Aladdin, PubChem CID: 20992), a homobifunctional (sulfide-to-sulfide) sulfhydryl-reactive reagent was used, which facilitated dimer detection. DSP and BMH were dissolved in DMSO (high-purity, Sigma) and diluted in PBS before use at final concentrations of 0.2 mM and 0.1 mM, respectively. After transfection, cross-linkers were added to cells and incubated for 15 min and 25 min, respectively, at 37 °C, before proceeding to the western blot.

Western blots. Cell supernatants and cell lysates were collected before infection or 48 h after infection (HCoV-229E) or 24 h after infection (H1N1 IAV and VSV–GFP). Cells were lysed in Buffer 1 lysis buffer[18], supplemented with a phosphatase-inhibitor cocktail (Phygene, PH0321) and protease-inhibitor cocktail (Phygene, PH0320). Collected cell lysates and cell supernatants were mixed with 5× SDS–PAGE loading buffer (Phygene, PH0333) and boiled for 5 min. The protein ladder used was 11–180 kDa Colormixed Protein Marker (Solarbioc) or 15–130 kDa two-colour prestained ladder (Biomed 168), or 10–250 kDa SmartBuffers prestained ladders (N6619); slight size differences were observed

between 10% Bis-Tris gels and 4–20% gradient gels (run in Tris-MOPS). Subsequently, cell lysates and supernatants were separated by 10% SDS–PAGE gel, transferred to PVDF membranes (Millipore, 0.45 µm) and blocked with 5% skimmed milk in TBS. For ISGylation detection, protein samples were mixed with 4× NuPAG LDS sample buffer (Invitrogen, NP0007), separated by NuPAGE 4–12% Bis-Tris gels in running buffer (Invitrogen, NP0001) for 70 min under 120 V, and transferred (Invitrogen, NP000061) for 90 min under 100 V.

The following antibodies were used for detection: rabbit anti-MX1 polyclonal antibody (clone N2C2, Genetex, GTX110256, dilution 1:1,000); rabbit anti-ISG15 polyclonal antibody (middle region, Aviva Systems Biology, ARP59386_P050, dilution 1:1,000); rabbit anti-GAPDH monoclonal antibody (clone 14C10, Cell Signaling, 2118, dilution 1:2,000); rabbit anti-CD13 polyclonal antibody (Sino Biological, 10051-T60, dilution 1:2,000); rabbit HCoV-229E nucleocapsid polyclonal antibody (Sino Biological, 40640-T62, dilution 1:2,000); mouse anti-MYC monoclonal antibody (Sino Biological, 100029-MM08, dilution 1:2,000 for cell lysate and 1:1,000 for cell supernatants); rabbit anti-UBE1L monoclonal antibody (Huabio, HA721228, dilution:1:500); rabbit polyclonal anti-UBE2L6 antibody (Abclonal, A13670); rabbit polyclonal anti-HERC5 antibody (Abclonal, A14889); and HRP-conjugated goat anti-rabbit IgG (Transgen Biotech, HS101-01, dilution 1:5,000).

Chemiluminescence was detected using the Enhanced ECL chemiluminescence detection kit (Vazyme) according to the manufacturer's instructions, and subsequently imaged by the LI-COR ODYSSEY FC imaging system (LI-COR Biosciences). Uncropped western blot images are included as Supplementary Data. Densitometry measurements were calculated from FiJI ImageJ software based on equal-size rectangular ROIs (multi-measure) of greyscale TIFF raw files (inverted) for GAPDH (with subtraction of background), ISG15 cell lysates and ISG15 supernatant images. Counts were normalized to GAPDH levels and expressed relative to human cell lysate ISG15 signal (graph for $n = 3$ independent blots).

Knockout cell-line generation. Single guide RNAs (sgRNAs) of *ISG15* were designed following a published protocol[130], and subsequently each sgRNA sequence was cloned into plentiCRISPRv2 vector (Addgene plasmid 52961). Scramble non-targeting gRNAs were also designed to serve as a negative control. The gRNA sequences are shown in Supplementary Table 27. Lentiviruses were generated using the method described above. HEK293 cells were first transduced and selected with puromycin (1 µg ml^{-1}) for two rounds of 10 days, followed by western blot of ISG15 for knockout validation.

Virus infections. The HCoV-229E clinical isolate was a gift from J. Zhao (Guangzhou Medical University), and IAV H1N1 PR8 and VSV–GFP (Indiana) were gifts from L. Lu (Zhejiang University). A clinical isolate of SARS-CoV-2 (SARS-CoV-2/SB3-TYAGNC) was used for infection studies following sequence validation using next-generation sequencing[131]. HCoV-229E was cultured in Huh7 cells or MRC-5 cells. IAV was propagated in A549 or Vero-E6 cells, VSV–GFP in HEK293 cells and SARS-CoV-2 in Vero-76 cells using a previously published protocol[131]. All stocks were prepared in low serum, filtered for cell debris, aliquoted and titrated in the respective cell line. Virus stocks were thawed once and used for an experiment. A fresh vial was used for each experiment to avoid repeated freeze–thaws. Virus infections were done in 1% FBS at low MOI (0.1) for fluorescent reporter or HCoV-229E $TCID_{50}$ assays. HCoV-229E assays were run in ten-fold dilutions in low-serum media. H1N1 PR8 IAV infection was done at an MOI of 0.1 in Vero-E6 or A549 cells in 1% FBS for 2 h before removal and replacement with growth media. Supernatant and cell lysates were collected 24 h after infection. Supernatant collection was similarly done after VSV–GFP assays were rinsed after 4–6 h infection, replaced with growth media and followed over time until about 70–80% GFP-positive (overnight).

For SARS-CoV-2 infections, A549-ACE2 cells were seeded at a density of 1.5×10^5 cells per well in a 12-well plate for 24 h. Then the cells were transfected with 200 ng of plasmids encoding bat ISG15 (see above) or vector control for 24 h, followed by infection with ancestral SARS-CoV-2 (SARS-CoV-2/SB3-TYAGNC isolate) at an MOI of 0.01 for 48 h. Control cells were sham infected. Infected or sham-infected cells were incubated at 37 °C for 1 h with gentle rocking every 15 min. After 1 h, virus inoculum was removed, cells were washed with PBS and supplemented with growth medium. Bulk cellular RNA and media from infected and sham-infected cells were collected 48 h after infection using a previously published protocol[132]. Cells transfected with mCherry_pcDNA3.1(+)-P2A plasmid and infected with SARS-CoV-2 served as a control for plasmid DNA transfection-mediated impact on SARS-CoV-2 replication. All work with infectious SARS-CoV-2 was done in a containment level 3 laboratory at the Vaccine and Infectious Disease Organization, University of Saskatchewan using approved protocols.

Plaque assay. To test direct antiviral function in cells stably expressing *ISG15*, plaque assays with IAV H1N1 PR8 were done in A549-stable cell lines (as above) by the addition of 50 µl virus to 500 µl low-FBS media (in triplicate) and serial 10-fold dilutions were performed (×8) in 24-well plates. Cells were incubated with virus for 4 h before rinsing and replaced with 2% methyl-cellulose 4000 cP direct overlays (Beyotime) supplemented with 1% FBS and Pen/Strep for 3–4 days.

$TCID_{50}$ assay. The supernatants from SARS-CoV-2-infected cells were titrated in triplicates on Vero-76 cells using $TCID_{50}$ assay[133]. In brief, 1.5×10^4 cells were seeded in each well of a 96-well plate. The plates were incubated overnight to obtain a confluent layer of Vero-76 cells. The next day, medium was taken off the cells and 50 µl 1:10 serially diluted virus-containing supernatant was added to the plates. The plates were incubated at 37 °C for 1 h. After incubation, the virus-containing supernatant was discarded and 100 µl complete media with 2% FBS was added to the plates. The plates were incubated at 37 °C for three and five days, respectively, and cytopathic effect was observed using a light microscope. $TCID_{50}$ per ml was calculated using the Spearman and Karber algorithm[134,135].

Free ISG15 and point mutations. Based on sequences of ISG15 (ENST00000649529), residues corresponding to Cys78 in human ISG15 were swapped with the codon for alanine (which changes polarity and removes the cysteine disulfide bond) or serine (which has a similar shape and charge but lacks a disulfide bond), whichever required the fewest nucleotide changes. Similarly, Ser77 in *R. affinis* was changed to a cysteine or a combination mutant replacing the absent lysine at position 77 and swapping the serine for a cysteine residue (at human Cys78). These geneblocks were generated in the same IRES–mCherry backbone. Supernatants from Huh7 cells after transfection or transduction were collected after 24 h or 48 h, respectively, pelleted for cell-debris removal and added directly to SDS–PAGE loading dye for a western blot. Similarly, supernatants were collected after 48 h of HCoV-229E infection. Cell lysates were collected as described previously.

Site-directed mutagenesis. *Rhinolophus affinis* and *R. yonghoiseni* ISG15 (LRGG to LRAA) mutants were generated using a QuikChange II site-directed mutagenesis kit (Agilent) using a previous modification[136]; see Supplementary Table 27 for primer sequences. Mutations were confirmed by Sanger sequencing at the National Research Council, Canada.

Animal silhouettes

The animal silhouettes used in the figures were downloaded from PhyloPic (https://www.phylopic.org) except for the bat silhouette, which was vectorized by A.E.M.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Genome assemblies and sequencing data are available at NCBI for *Aselliscus stoliczkanus* (PRJNA949177), *Doryrhina cyclops* (PRJNA938463), *Hipposideros larvatus* (PRJNA938461), *Rhinolophus affinis* (PRJNA938462), *Rhinolophus perniger lanosus* (PRJNA955779), *Rhinolophus yonghoiseni* (PRJNA938455), *Rhinolophus trifoliatus* (PRJNA939732), *Rhinopoma microphyllum* (PRJNA971926), *Megaderma spasma* (PRJNA940731) and *Mops condylurus* (PRJNA949178). TOGA, transposable element and miRNA annotations of newly sequenced bats and alignments of positively selected genes are available at http://genome.senckenberg.de/download/Bat1KImmune/. Accession codes and identifiers of genomic data are listed in Supplementary Tables 3 and 5. The gRNAs, primers and geneblock sequences are listed in Supplementary Table 27. Other databases used are: Dfam database v.3.5 (https://www.dfam.org/releases/Dfam_3.5/), Infernal v.1.1.2, GENCODE V38 (Ensembl 104), Gene Ontology http://geneontology.org/, KEGG https://www.genome.jp/kegg/, WikiPathways https://www.wikipathways.org/index.php/WikiPathways, miRTarBase http://mirtarbase.mbc.nctu.edu.tw, TRANSFAC http://genexplain.com/transfac/, Human Protein Atlas https://www.proteinatlas.org/, CORUM http://mips.helmholtz847, muenchen.de/corum/, Human Phenotype Ontology https://hpo.jax.org/app/ and uniref100.

Code availability

Custom scripts used for data analysis are available at GitHub (https://github.com/ariadnamorales/2023_Bat1Kimmunity).

61. R Core Team. *R: A Language and Environment for Statistical Computing*. *R Foundation for Statistical Computing, Vienna, Austria*. https://www.R-project.org/ (2023).
62. Card, D. C., Shapiro, B., Giribet, G., Moritz, C. & Edwards, S. V. Museum genomics. *Annu. Rev. Genet.* **55**, 633–659 (2021).
63. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
64. Guan, D. et al. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898 (2020).
65. Nurk, S. et al. HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res.* **30**, 1291–1305 https://doi.org/10.1101/gr.263566.120 (2020).
66. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://arxiv.org/abs/1303.3997 (2013).
67. Ghurye, J. et al. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Comput. Biol.* **15**, e1007273 (2019).
68. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
69. Formenti, G. et al. Merfin: improved variant filtering, assembly evaluation and polishing via *k*-mer validation. *Nat. Methods* **19**, 696–704 (2022).
70. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 (2020).
71. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
72. Goubert, C. et al. A beginner's guide to manual curation of transposable elements. *Mob. DNA* **13**, 7 (2022).
73. Storer, J. M., Hubley, R., Rosen, J. & Smit, A. F. A. Curation guidelines for de novo generated transposable element families. *Curr. Protoc.* **1**, e154 (2021).
74. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
75. Osmanski, A. B. et al. Insights into mammalian TE diversity through the curation of 248 genome assemblies. *Science* **380**, eabn1430 (2023).
76. Storer, J., Hubley, R., Rosen, J., Wheeler, T. J. & Smit, A. F. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mob. DNA* **12**, 2 (2021).
77. Kalvari, I. et al. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* **49**, D192–D200 (2021).
78. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
79. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The Vienna RNA websuite. *Nucleic Acids Res.* **36**, W70–W74 (2008).
80. Harris, R. S. *Improved Pairwise Alignment of Genomic DNA*. PhD thesis, Pennsylvania State Univ. (2007).
81. Sharma, V. & Hiller, M. Increased alignment sensitivity improves the usage of genome alignments for comparative gene annotation. *Nucleic Acids Res.* **45**, 8369–8377 (2017).
82. Kent, W. J., Baertsch, R., Hinrichs, A., Miller, W. & Haussler, D. Evolution's cauldron: Duplication, deletion and rearrangement in the mouse and human genomes. *Proc. Natl Acad. Sci. USA* **100**, 11484–11489 (2003).
83. Osipova, E., Hecker, N. & Hiller, M. RepeatFiller newly identifies megabases of aligning repetitive sequences and improves annotations of conserved non-exonic elements. *Gigascience* **8**, giz132 (2019).
84. Suarez, H. G., Langer, B. E., Ladde, P. & Hiller, M. chainCleaner improves genome alignment specificity and sensitivity. *Bioinformatics* **33**, 1596–1603 (2017).
85. Ranwez, V., Douzery, E. J. P., Cambon, C., Chantret, N. & Delsuc, F. MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. *Mol. Biol. Evol.* **35**, 2582–2584 (2018).
86. Di Franco, A., Poujol, R., Baurain, D. & Philippe, H. Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evol. Biol.* **19**, 21 (2019).
87. Mirarab, S. & Warnow, T. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* **31**, i44–i52 (2015).
88. Mirarab, S. et al. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* **30**, i541–i548 (2014).
89. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
90. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
91. Lemoine, F. et al. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* **556**, 452–456 (2018).
92. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
93. Sanderson, M. J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302 (2003).
94. Smith, S. A. & O'Meara, B. C. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**, 2689–2690 (2012).
95. Foley, N. M., Springer, M. S. & Teeling, E. C. Mammal madness: is the mammal tree of life not yet resolved? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **371**, 20150140 (2016).
96. Teeling, E. C. et al. A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* **307**, 580–584 (2005).
97. Blumer, M. et al. Gene losses in the common vampire bat illuminate molecular adaptations to blood feeding. *Sci. Adv.* **8**, eabm6494 (2022).
98. Wang, K. et al. Molecular adaptation and convergent evolution of frugivory in Old World and neotropical fruit bats. *Mol. Ecol.* **29**, 4366–4381 (2020).
99. Dong, D. et al. The genomes of two bat species with long constant frequency echolocation calls. *Mol. Biol. Evol.* **34**, 20–34 (2017).
100. Kosakovsky Pond, S. L., Frost, S. D. W. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2005).
101. Raudvere, U. et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).
102. Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* **35**, W193–W200 (2007).
103. Siepel, A. & Haussler, D. Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Mol. Biol. Evol.* **21**, 468–488 (2004).
104. Burnham, K. P. & Anderson, D. R. in *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* 49–97 (Springer, 2002).
105. O'Hara, R. & Kotze, J. Do not log-transform count data. *Nat. Prec.* https://doi.org/10.1038/npre.2010.4136.1 (2010).
106. Bürkner, P.-C. brms: an R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1–28 (2017).
107. Carpenter, B. et al. Stan: a probabilistic programming language. *J. Stat. Softw.* **76**, 1 (2017).
108. Gelman, A., Goodrich, B., Gabry, J. & Vehtari, A. R-squared for Bayesian regression models. *Am. Stat.* **73**, 307–309 (2019).
109. Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2017).
110. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
111. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
112. Mirdita, M., Steinegger, M. & Söding, J. MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics* **35**, 2856–2858 (2019).
113. Mirdita, M. et al. Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res.* **45**, D170–D176 (2016).
114. Hildebrandt, P., Garda, H., Stier, A., Stockburger, M. & Van Dyke, R. A. Resonance Raman study of the cytochrome P-450 LM2-halothane intermediate complex. *FEBS Lett.* **237**, 15–20 (1988).
115. Steinegger, M. et al. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 473 (2019).
116. Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. Preprint at *bioRxiv* https://doi.org/10.1101/2021.10.04.463034 (2022).
117. Eastman, P. et al. OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **13**, e1005659 (2017).
118. Abraham, M. J. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

119. Bauer, P., Hess, B. & Lindahl, E. GROMACS 2022.1 Source Code. *Zenodo* https://doi.org/10.5281/zenodo.6451564 (2022).
120. Best, R. B. et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ(1) and χ(2) dihedral angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
121. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).
122. Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *J. Chem. Theory Comput.* **7**, 2284–2295 (2011).
123. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
124. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
125. Nosé, S. & Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **50**, 1055–1076 (1983).
126. Glykos, N. M. Software news and updates. Carma: a molecular dynamics analysis program. *J. Comput. Chem.* **27**, 1765–1768 (2006).
127. Reynolds, A. P., Richards, G., de la Iglesia, B. & Rayward-Smith, V. J. Clustering rules: a comparison of partitioning and hierarchical clustering algorithms. *J. Math. Model. Algor.* **5**, 475–504 (2006).
128. Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).
129. LeBlanc, E. V. & Colpitts, C. C. The green tea catechin EGCG provides proof-of-concept for a pan-coronavirus attachment inhibitor. *Sci. Rep.* **12**, 12899 (2022).
130. Concordet, J.-P. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* **46**, W242–W245 (2018).
131. Banerjee, A. et al. Isolation, sequence, infectivity, and replication kinetics of severe acute respiratory syndrome coronavirus 2. *Emerg. Infect. Dis.* **26**, 2054–2063 (2020).
132. Banerjee, A. et al. Experimental and natural evidence of SARS-CoV-2-infection-induced activation of type I interferon responses. *iScience* **24**, 102477 (2021).
133. Banerjee, A. et al. Selection of viral variants during persistent infection of insectivorous bat cells with Middle East respiratory syndrome coronavirus. *Sci. Rep.* **10**, 7257 (2020).
134. Spearman, C. The method of 'right and wrong cases' ('constant stimuli') without Gauss's formulae. *Br. J. Psychol.* **2**, 227–242 (1908).
135. Kärber, G. Beitrag zur kollektiven Behandlung pharmakologischer Reihenversuche. *Naunyn Schmiedebergs Arch. Exp. Pathol. Pharmakol.* **162**, 480–483 (1931).
136. Wang, W. & Malcolm, B. A. Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange Site-Directed Mutagenesis. *Biotechniques* **26**, 680–682 (1999).
137. Parker, J. et al. Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* **502**, 228–231 (2013).
138. Zhu, M. et al. Negative regulation of T cell activation and autoimmunity by the transmembrane adaptor protein LAB. *Immunity* **25**, 757–768 (2006).
139. Ni, G. et al. PPP6C negatively regulates STING-dependent innate immune responses. *mBio* **11**, e01728–20 (2020).
140. Fontenot, J. D. et al. Regulatory T cell lineage specification by the forkhead transcription factor foxp3. *Immunity* **22**, 329–341 (2005).
141. McArdel, S. L., Terhorst, C. & Sharpe, A. H. Roles of CD48 in regulating immunity and tolerance. *Clin. Immunol.* **164**, 10–20 (2016).
142. Haller, O., Staeheli, P. & Kochs, G. Interferon-induced Mx proteins in antiviral host defense. *Biochimie* **89**, 812–818 (2007).
143. Kohlmeier, J. E. et al. The chemokine receptor CCR5 plays a key role in the early memory $CD8^+$ T cell response to respiratory virus infections. *Immunity* **29**, 101–113 (2008).
144. Vanderheiden, A. et al. CCR2 signaling restricts SARS-CoV-2 infection. *mBio* **12**, e0274921 (2021).
145. Stikker, B. S. et al. Severe COVID-19-associated variants linked to chemokine receptor gene control in monocytes and macrophages. *Genome Biol.* **23**, 96 (2022).
146. Files, D. C. et al. Rationale of using the dual chemokine receptor CCR2/CCR5 inhibitor cenicriviroc for the treatment of COVID-19. *PLoS Pathog.* **18**, e1010547 (2022).
147. Huse, K. et al. Mechanism of CD79A and CD79B support for IgM^+ B cell fitness through B cell receptor surface expression. *J. Immunol.* **209**, 2042–2053 (2022).
148. Pal Singh, S., Dammeijer, F. & Hendriks, R. W. Role of Bruton's tyrosine kinase in B cells and malignancies. *Mol. Cancer* **17**, 57 (2018).
149. Afzali, B., Noris, M., Lambrecht, B. N. & Kemper, C. The state of complement in COVID-19. *Nat. Rev. Immunol.* **22**, 77–84 (2022).
150. Merle, N. S., Church, S. E., Fremeaux-Bacchi, V. & Roumenina, L. T. Complement system part I - molecular mechanisms of activation and regulation. *Front. Immunol.* **6**, 262 (2015).
151. Xie, C. B., Jane-Wit, D. & Pober, J. S. Complement membrane attack complex: new roles, mechanisms of action, and therapeutic targets. *Am. J. Pathol.* **190**, 1138–1150 (2020).
152. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
153. The PyMOL Molecular Graphics System, version 3.0 (Schrödinger, LLC).
154. Suzuki, T. et al. Infection with flaviviruses requires BCLXL for cell survival. *PLoS Pathog.* **14**, e1007299 (2018).
155. Shi, H.-X. et al. Positive regulation of interferon regulatory factor 3 activation by Herc5 via ISG15 modification. *Mol. Cell. Biol.* **30**, 2424–2436 (2010).
156. Swaim, C. D., Scott, A. F., Canadeo, L. A. & Huibregtse, J. M. Extracellular ISG15 signals cytokine secretion through the LFA-1 integrin receptor. *Mol. Cell* **68**, 581–590 (2017).

Acknowledgements M.H. was supported by the German Research Foundation (HI1423/5-1) and the LOEWE-Centre for Translational Biodiversity Genomics, funded by the Hessen State Ministry of Higher Education, Research and the Arts (LOEWE/1/10/519/03/03.001(0014)/52); T.B. by the German Research Foundation (INST 269/768-1); L.M.D. by the National Science Foundation (NSF-IOS 2032063 and 2031906, and NSF-DEB 1838273); T.M.L. by the National Science Foundation (award ID 2010853); A.T.I. by a Key grant from the National Science Foundation of Zhejiang Province (Z23C010003) and a National Science Foundation Research Fund for International Excellent Young Scientists (RFIS-II, 82350610279); B.K.L. by the Royal Ontario Museum Governors, the National Science Foundation (DEB-0344430) and the Toronto Zoological Society; A. Banerjee by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2022-03010) and the Canadian Institutes of Health Research (CIHR) Center for Research on Pandemic Preparedness and Health Emergencies, Early Career Investigator grant (PEE-183995) and CIHR Infection and Immunity, Early Career Research Support (PTT-192089); V.G. by a PGS-D scholarship funded by the Natural Sciences and Engineering Research Council of Canada (569587-2022); S.C.V. by a UKRI Future Leaders fellowship (MR/T021985/1), an ERC Consolidator grant (101001702; BATSPEAK) and a Max Planck Research Group award from the Max Planck Society; S.J.P. by a Junior Chair grant from the Institut Universitaire de France; E.M. by the Max Planck Society; A.K. and J.B.P. by the Robert Koch Institute; E.C.T. by Irish Research Council Laureate Award IRCLA/2017/58 and Science Foundation Ireland Future Frontiers 19/FFP/6790; D. - G.K. by an EMBO postdoctoral fellowship (ALTF 1089-2021). The Vaccine and Infectious Disease Organization receives operational funding from the government of Saskatchewan through Innovation Saskatchewan and the Ministry of Agriculture, and funding for its Containment Level 3 facility (InterVac) from the Canada Foundation for Innovation through the Major Science Initiatives. We thank the Long Read Team of the DRESDEN-concept Genome Center, the DFG NGS Competence Center, which is part of the Center for Molecular and Cellular Bioengineering, Technische Universität Dresden, and of the MPI-CBG. We also thank the core facility of the Zhejiang University–University of Edinburgh Institute. Unpublished genome assemblies of six bats and seven other mammalian species were used with permission from the DNA Zoo Consortium (dnazoo.org).

Author contributions Sample providers, species identification and taxonomic advice: B.K.L., S.J.R., S.J.P., A.K., J.B.P., T.M.L., L.M.D., Y.N. Genome sequencing and assembly: T.B., M.P., S.W. TOGA and pairwise genome alignments: A.E.M., B.M.K. Phylogenetic analyses: A.E.M., G.H. Codon alignments, selection and enrichment analyses: A.E.M. Regression analysis: A.E.M., L.M.D. Gene loss: A.-W.A., A.E.M. Immune-related genes investigation: A.E.M, M.H., A.-W.A., L.H., A.T.I. TE analysis: D.A.R. MicroRNAs: Z.H., P.D., S.C.V. ISG15 3D modelling: D. - G.K., A.T.I. ISG15 experiments, cell-line generation, virus infection and laboratory infrastructure: Y.D., X.L., K.B., V.G., P.L., Y. Yang, A. Bhuinya, A. Banerjee, A.T.I. Species selection, sample acquisition, sequencing, funding and infrastructure: Y. Yovel, S.C.V., E.M., L.M.D., E.C.T., M.H. Conception of study: A.E.M., E.C.T., A.T.I., M.H.

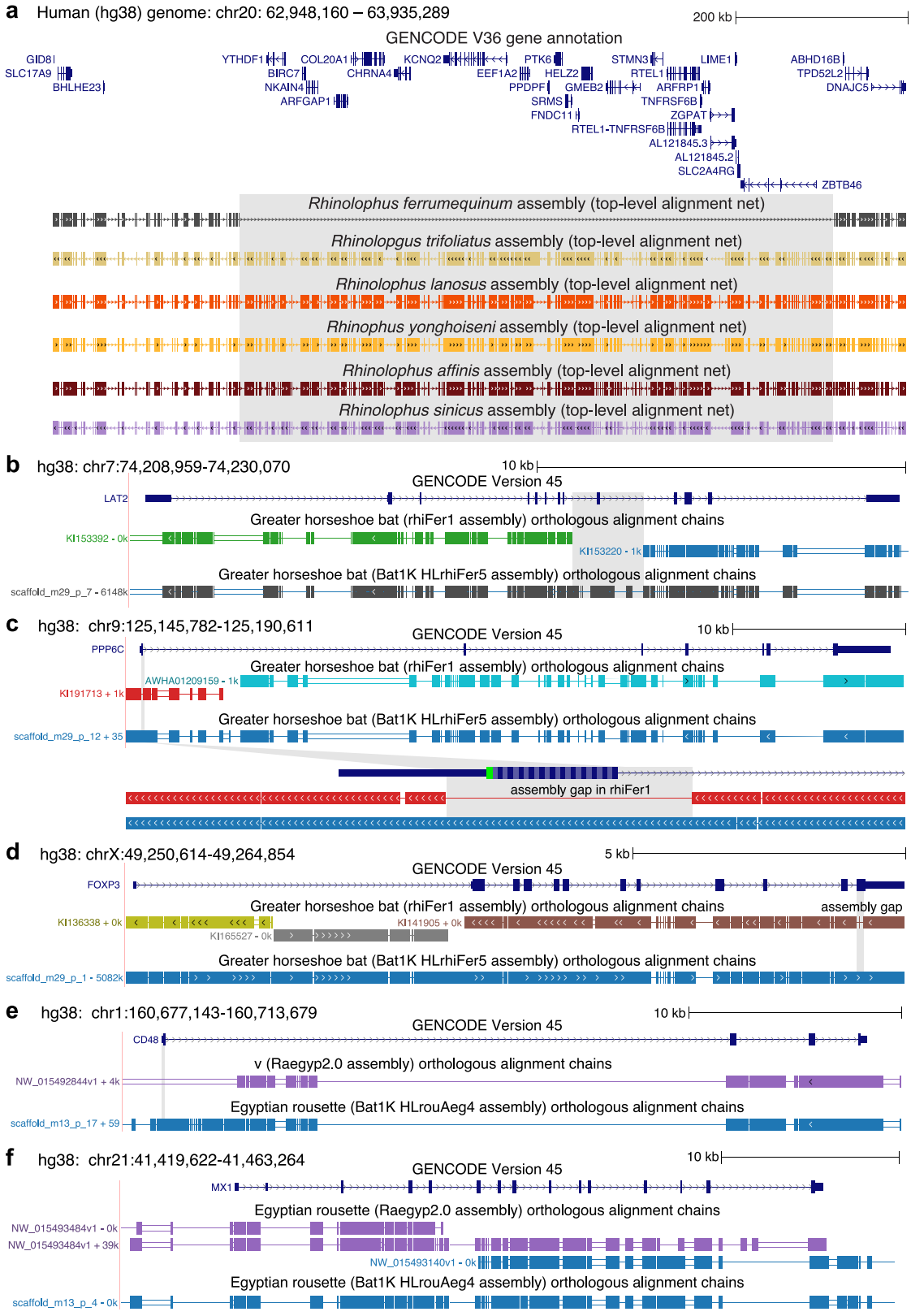
Competing interests The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-08471-0.

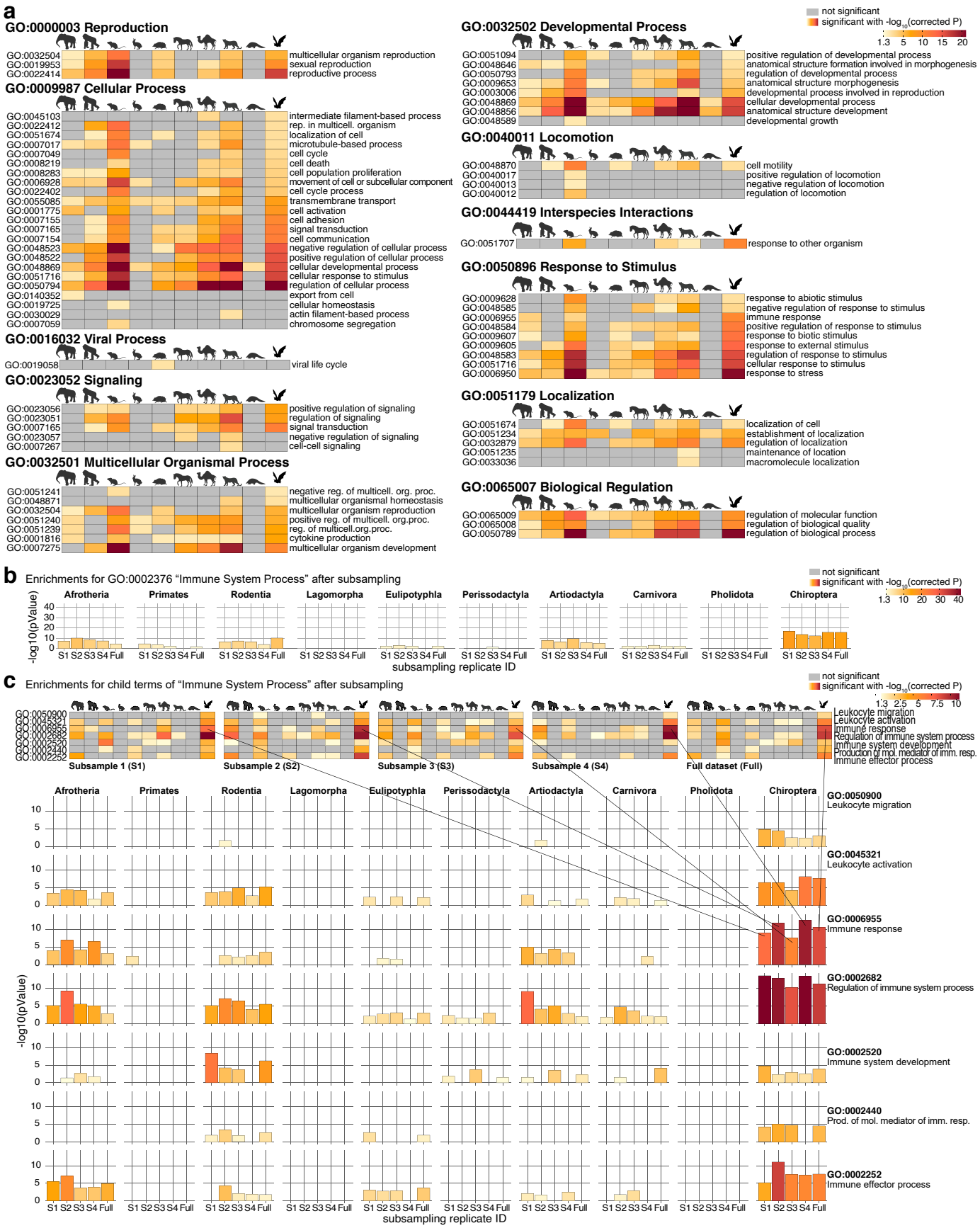
Correspondence and requests for materials should be addressed to Aaron T. Irving or Michael Hiller.

Peer review information *Nature* thanks Arang Rhie and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



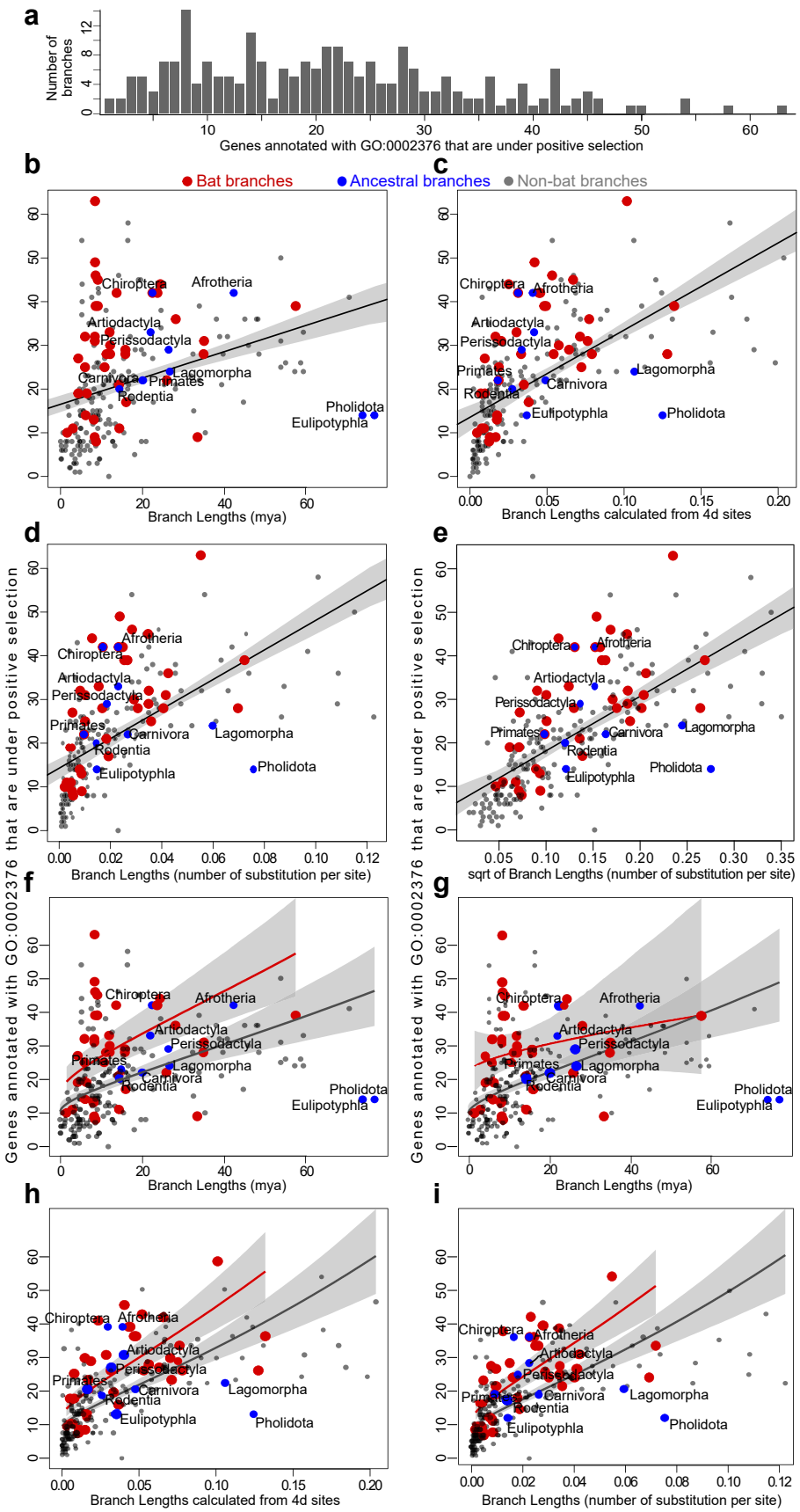
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Genome assembly quality and impact on selection screens. a, PacBio HiFi sequencing improves genome assembly contiguity. UCSC genome browser screenshot showing a 700 kb locus of the human chr20 and genome alignments to six rhinolophid bats (boxes in the alignment net represent aligning sequence and connecting lines deletions or unaligning sequence). The region highlighted in grey does not align between human and *Rhinolophus ferrumequinum*, because the *R. ferrumequinum* PacBio CLR-based assembly has a large 421,369 bp assembly gap in this locus. As a result, several genes contained in this locus are missing from this assembly, which contributes to the slightly higher number of missing genes for *R. ferrumequinum* vs. other rhinolophid bats (Fig. 1e). Consistent with PacBio HiFi-based assemblies often having higher contig N50 and N90 values (Fig. 1c), the other four HiFi-based rhinolophid assemblies have a contiguous sequence in this locus without an assembly gap. *Rhinolophus sinicus*, which was assembled from Illumina short reads, has 24 smaller assembly gaps with sizes 15–1058 bp in this locus. **b-f**, Lower assembly quality leads to missed signals of gene selection. To explore the effect of assembly quality on results from genome-wide selection screens, we replaced the high-quality Bat1K assembly (HLrhiFer5) of the Greater horseshoe bat (*Rhinolophus ferrumequinum*)[15] with a previous short read based assembly (rhiFer1) of the same species[137] that has a higher degree of incompleteness and fragmentation (panel b-d). We kept all other 114 mammalian species and tested if the genes under selection in the *R. ferrumequinum* branch with the HLrhiFer5 assembly are also under selection when using the rhiFer1 assembly. Similarly, we replaced the Bat1K assembly (HLrouAeg4) of the Egyptian rousette (*Rousettus aegyptiacus*) with a previous assembly (Raegyp2.0) that used long and short read data but has a smaller contig N50 and an excess of inactivating mutations (indicative of a higher base error rate; see Supplementary Fig. 3). We identified 272 vs. 133 genes under positive selection in HLrhiFer5 vs. rhiFer1, and 299 vs. 194 genes under positive selection in HLrouAeg4 vs. Raegyp2.0, indicating that lower assembly quality hampers the identification of selected genes. To illustrate this, the panels show UCSC genome browser screenshots of five immune-related genes, where we detected positive selection in the Bat1K but not the previous assembly. The first three examples (b-d) show cases where the previous assembly does not cover the gene on a single scaffold and exons are missing because of assembly gaps. The last two examples (e-f) show how assembly problems other than assembly gaps hamper selection screens. **b**, *LAT2* (linker for activation of T cells family member 2), a regulator of T cell activation[138], is split across two different scaffolds in rhiFer1, as shown by the two alignment chains between human (hg38 assembly) and *R. ferrumequinum* rhiFer1. Importantly, while other methods can capture only one of these gene fragments at best, TOGA recognizes both alignment chains as orthologous and joins both gene fragments, resulting in a more complete codon alignment. Nevertheless, coding exon 8 (blue highlight) is missing in rhiFer1, thus this exon is missing in the codon alignment. In HLrhiFer5, all exons align to a single scaffold. **c**, *PPP6C* (protein phosphatase 6 catalytic subunit), a factor that regulates STING phosphorylation and activation[139], is split across two scaffolds (alignment chains) in rhiFer1. While TOGA recognizes the red chain as an orthologous fragment of *PPP6C*, the inset shows that coding exon 1 overlaps an assembly gap in rhiFer1 (but not HLrhiFer5), thus this exon will be missed in the codon alignment. **d**, The gene locus of *FOXP3* (forkhead box P3), a master regulator involved in regulatory T-cell development and function[140], is split across several scaffolds in rhiFer1. Although the entire coding region is present on a single scaffold (brown chain), the last coding exon overlaps an assembly gap in rhiFer1 and thus will be missed in the codon alignment. **e**, *CD48* (CD48 antigen), a cell surface factor involved in adhesion and activation of adaptive immune cells[141], lacks an aligning exon 1 in the Raegyp2.0 assembly. Compared to the HLrouAeg4 assembly, Raegyp2.0 lacks ~22,800 bp of sequence and this sequence is also present in other Pteropodid assemblies. This indicates that this 'deletion' is likely an assembly error in Raegyp2.0. **f**, *MX1* (MX dynamin like GTPase 1), an interferon-induced antiviral gene[142], has two orthologous alignment chains that cover the gene. While this apparent 'duplication' is likely due to incomplete haplotype purging in Raegyp2.0, it leads TOGA to classify *MX1* as a 1:2 ortholog in this assembly and since our screen only considers 1:1 orthologs, this gene is missed in a screen including Raegyp2.0.



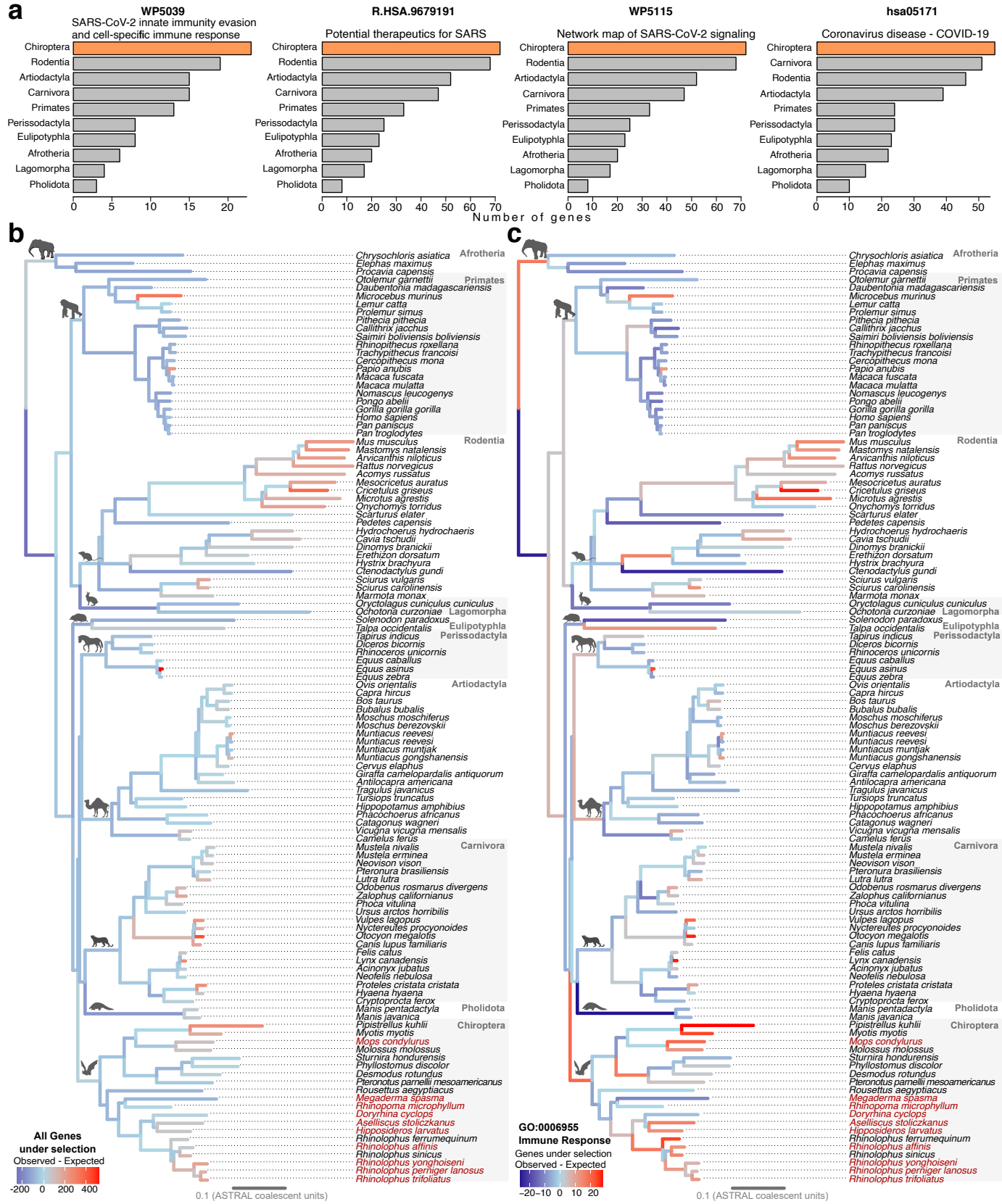
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Robustness of immune-related enrichments in bats. **a**, Heatmaps show enrichments for genes under positive selection for the child terms of high-level GO Biological Processes that are shown in Fig. 2a. Child terms of "immune system process" are shown in Fig. 2d. Enrichment tests for genes under selection in the different mammalian groups (columns) were performed in gProfiler. Only child terms with a significant enrichment in at least one mammalian order are shown. Cells with color indicate a significant enrichment after correcting for multiple testing. Supplementary Table 7 provides the full gProfiler output, including all significant p-values for all groups. **b-c**, Enrichments of genes under positive selection in four subsampled datasets. To explore whether functional enrichments of positively selected genes in mammalian clades are driven by individual species rather than being representative for the clade Chiroptera, we ran four additional for positive selection by subsampling 65 species from all 115 mammals, as explained in the Methods. In subsample 1–3 (s1-s3), we randomly selected ten of the 20 species in each clade that is represented by 20 species in the full dataset (Chiroptera, Primates, Rodentia, Artiodactyla, Carnivora). Subsample 4 (s4) includes all species left out in subsample 1 from the 20-species clades. Clades with less than 20 species were not subsampled. **b**, For each mammalian order, bar plots visualize the statistical significance ($-\log_{10}$ corrected p-value) of enrichments for GO:0002376 "Immune System Process" in the four subsamples and in comparison to the full dataset (fifth bar). Importantly, the chiropteran enrichments for "immune system process" are robustly observed across the subsamples, and no other mammalian group exhibits an enrichment as significant as in Chiroptera. **c**, Enrichments for child terms of "Immune System Process" in four subsampled datasets. Heatmaps at the top summarize enrichments with colored cells representing significant enrichments after correcting for multiple testing. Bar charts shown below visualize the statistical significance as in panel b. Genes positively selected in Chiroptera robustly show a strong enrichment in "Immune Response" and "Regulation of Immune System Process", where the enrichments are more significant than in any other mammalian group. All animal silhouettes apart from the bat were obtained from PhyloPic (phylopic.org). *Mus musculus*, by Daniel Jaron under a CC0 1.0 Universal Public Domain licence; *Gorilla gorilla gorilla* by T. Michael Keesey (after Colin M. L. Burnett) under a CC0 1.0 Universal Public Domain licence; *Leporidae*, by Sarah Werning under a CC BY 3.0 licence; *Erinaceus europaeus* by Roberto Díaz Sibaja under a CC BY 3.0 licence; *Equus ferus przewalskii*, by Mercedes Yrayzoz (vectorized by T. Michael Keesey) under a CC BY 3.0 licence; *Camelus dromedarius*, by Steven Traver under a CC0 1.0 Universal Public Domain licence; *Panthera pardus*, by Margot Michaud under a CC0 1.0 Universal Public Domain licence; *Manis culionensis*, by Steven Traver under a CC0 1.0 Universal Public Domain licence. The bat silhouette was vectorized by A.E.M., the Afrotheria (Thai Elephant) silhouette was obtained from OpenClipArt (https://openclipart.org/) under a CC0 1.0 Universal Public Domain licence.



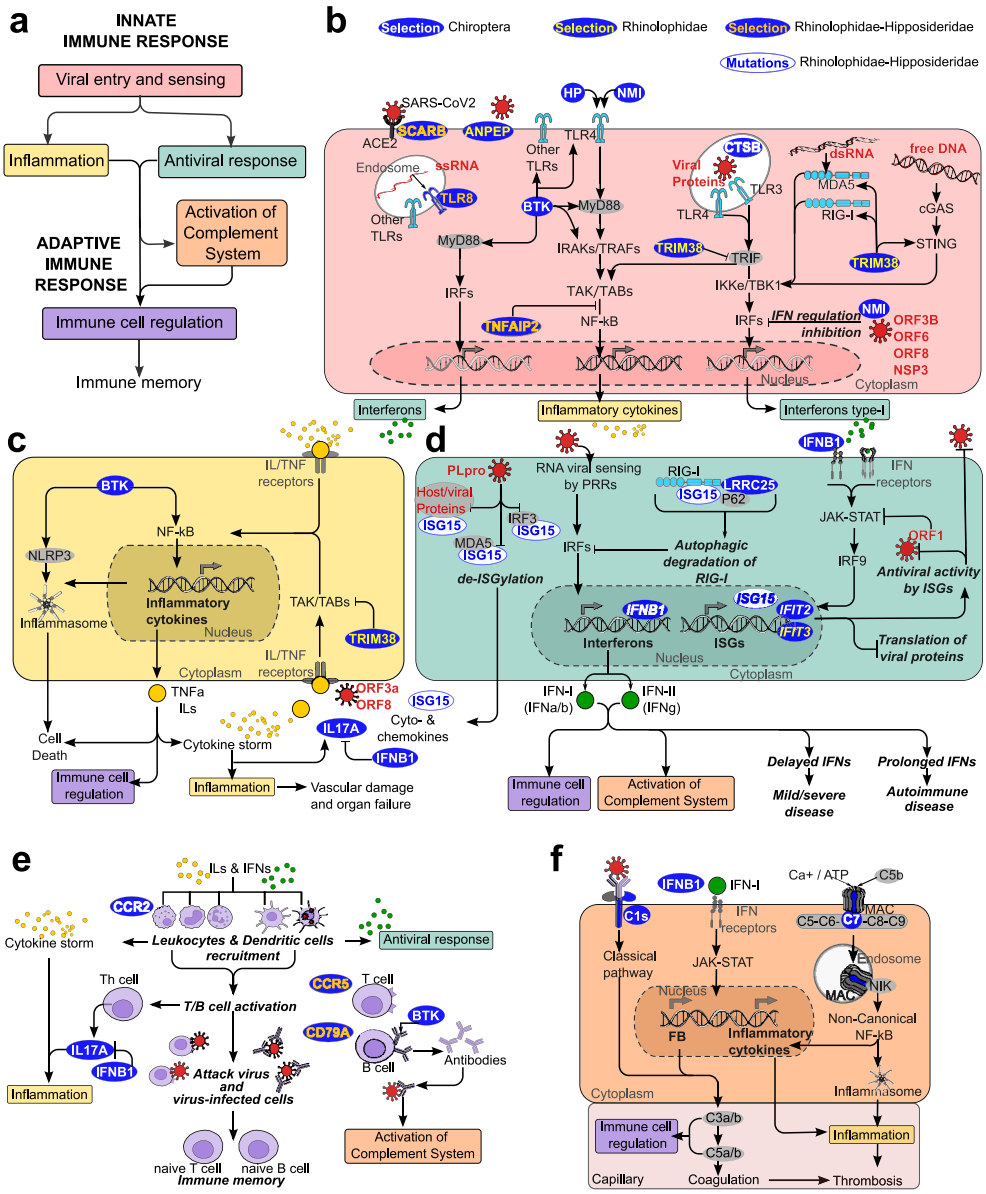
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Correlation between branch length and number of genes under positive selection. a, Distribution of the number of immune system genes that are under selection in a given number of branches, considering all 228 branches in the 115-species tree. **b-e**, Dispersion plots show a correlation between different measures of branch lengths and the number of "immune system process" (GO:0002376) genes under positive selection. For branch lengths, we considered **b**, millions of years taken from our time-calibrated phylogeny inferred using treePL and fossil calibrations (Supplementary Table 4); **c**, number of substitutions per neutral site estimated from 4D sites using phyloFit; **d**, number of substitutions per site estimated from coding regions using IQTREE; **e**, transformed branch lengths as the square-root of the number of substitutions per site estimated from coding regions using IQTREE. These tests robustly show that branch length is significantly correlated with the number of selected immune genes. This likely reflects a higher incidence for episodic positive selection to occur over longer evolutionary time periods and increased power to detect it on longer branches, consistent with previous simulations on few taxa[23]. Akaike's information criterion (AIC) indicates that the model in (E) fits the data best (Supplementary Table 9). **f-i**, Model selection with one or two slopes and/or intercepts. To determine whether the number of immune genes under selection is higher in bats than in other mammals, we introduced a categorical variable, corresponding to taxonomy (bats and non-bats) and fit a series of negative binomial regressions where two intercepts and two slopes would be allowed in the models. **f**, negative binomial, two-intercept model of square-root transformed millions of years from a time-calibrated phylogeny inferred using treePL and fossil calibrations (Supplementary Table 4). **g**, negative binomial, two-intercept and two slope model of square-root transformed millions of years from a time-calibrated phylogeny inferred using treePL and fossil calibrations (Supplementary Table 4). **h**, negative binomial, two-intercept model of number of substitutions per neutral site estimated from 4D sites using phyloFit. **i**, negative binomial, two-intercept model of number of substitution per site estimated from coding regions using IQTREE. The best-fit model shown in panel e has two intercepts corresponding to bats and non-bats, but a single slope. These results robustly show that bats have a higher number of immune-related genes under positive selection. While we applied the regression model specifically to immune-related genes, this approach could be a generally-applicable strategy to reveal lineages (branches) having an excess of selection on genes belonging to particular functional categories.

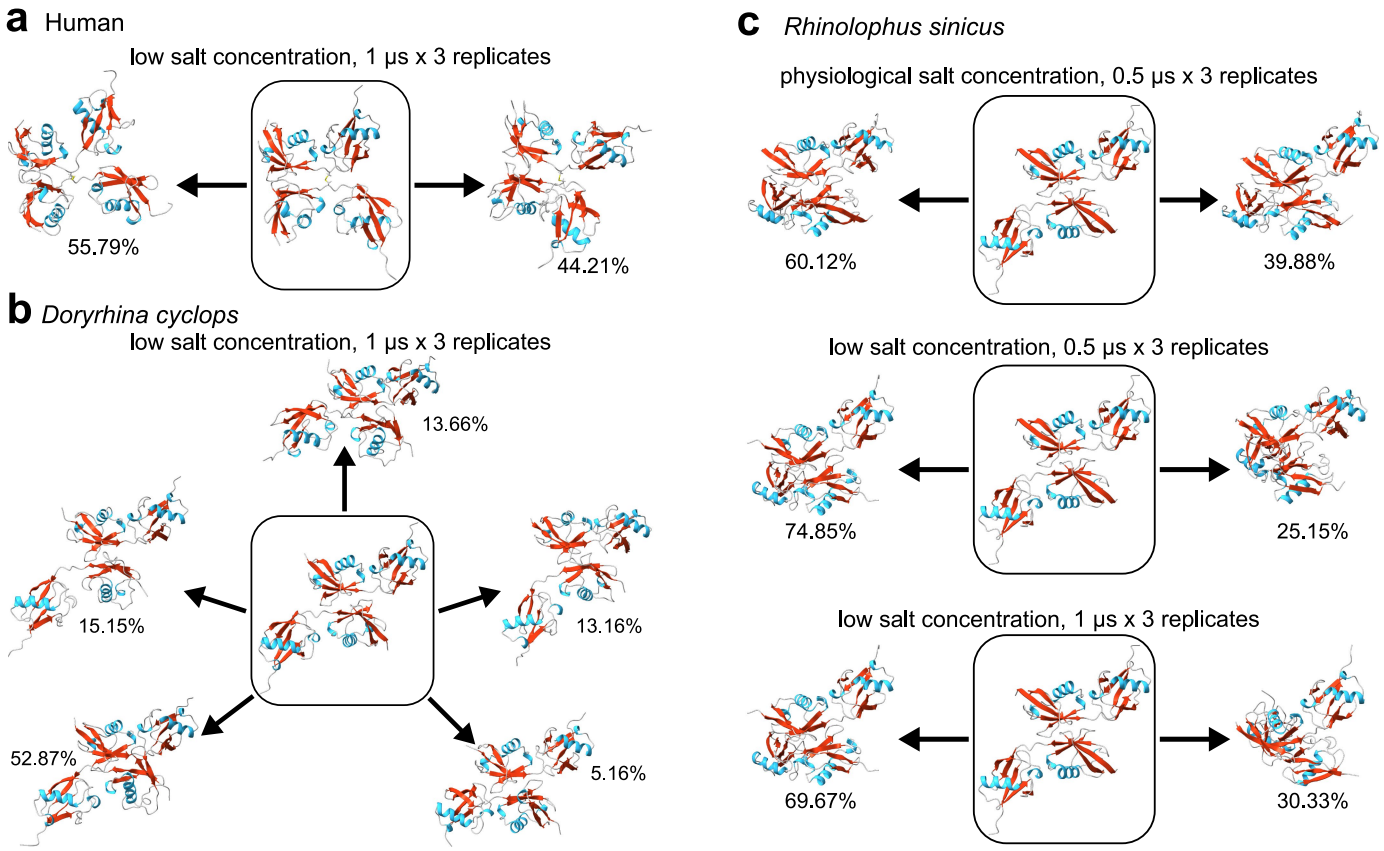


Extended Data Fig. 4 | See next page for caption.

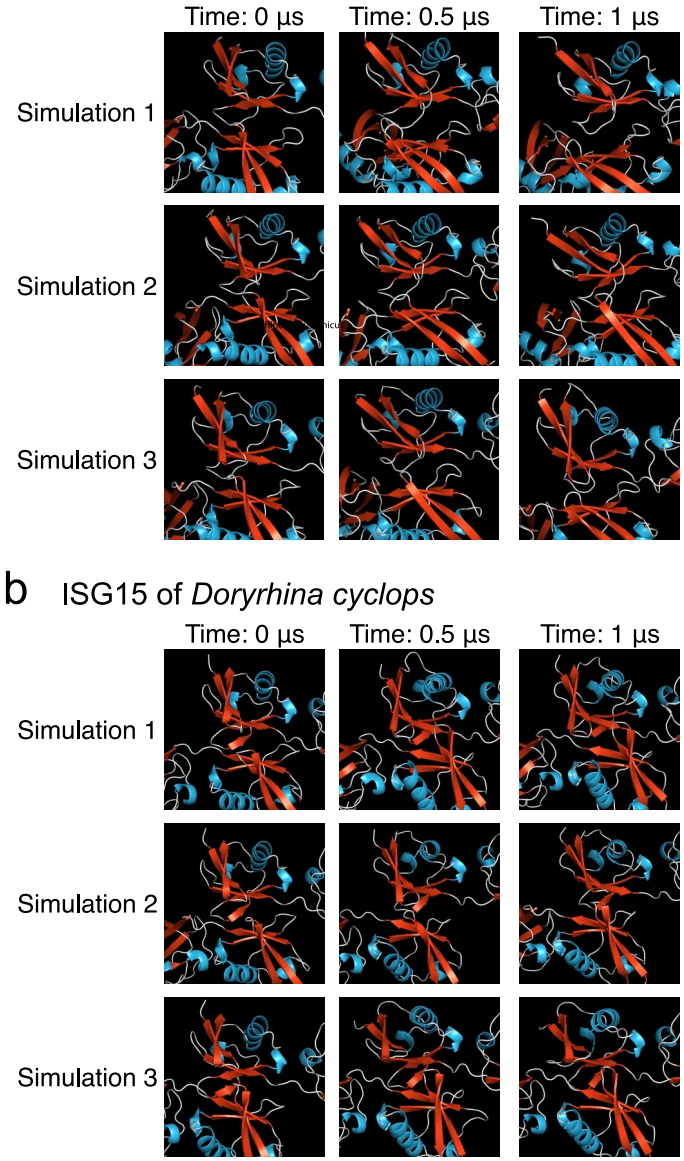
Extended Data Fig. 4 | Positive selection in SARS-CoV-2-related pathways and per-branch signal of selection for all and immune response genes. **a**, Genes under positive selection in SARS-CoV-2 related pathways. Absolute number of genes under positive selection in different mammalian orders (defined as at least one branch having a significant p-value) that are involved in pathways with relevance for SARS-CoV-2 and COVID-19. Pathway data was taken from WikiPathways (https://www.wikipathways.org/pathways/WP5039.html, https://www.wikipathways.org/pathways/WP5115.html), Reactome (https://www.reactome.org/content/detail/R-HSA-9679191), and KEGG pathways (https://www.genome.jp/pathway/hsa05171), last accessed on July 20th, 2022. Chiroptera consistently have the highest number of selected genes in these gene sets. We list all mammalian orders here for completeness, but it should be noted that five orders (Primates, Rodentia, Artiodactyla, Chiroptera, Carnivora) are represented by 20 species and thus have more branches than other orders. **b**-**c**, Per-branch signal of selection for all genes and genes annotated as "Immune Response" (GO:0006955) on the phylogeny for 115 mammals. Branches are color-coded based on the difference between the observed and expected number of selected genes in the gene set. The color scale indicates negative residual values (less genes under selection than expected) in blue and positive residual values (more genes than expected) in red per branch. Expected numbers were calculated from the regression models shown in Supplementary Tables 8-9. Ancestral branches for mammalian orders are indicated with icons. New Bat1K genomes are in red font. **b**, Considering all genes included in our screen without selecting a functional category shows no excess of selected genes in bats and on the ancestral Chiroptera branch. **c**, Considering genes annotated with the GO term "Immune Response" shows that the ancestral Chiroptera branch has the highest excess of selected "Immune Response" genes, substantiating the results for "Immune System Process" (Fig. 2c). All animal silhouettes apart from the bat were obtained from PhyloPic (phylopic.org). *Mus musculus*, by Daniel Jaron under a CC0 1.0 Universal Public Domain licence; *Gorilla gorilla gorilla* by T. Michael Keesey (after Colin M. L. Burnett) under a CC0 1.0 Universal Public Domain licence; *Leporidae*, by Sarah Werning under a CC BY 3.0 licence; *Erinaceus europaeus* by Roberto Díaz Sibaja under a CC BY 3.0 licence; *Equus ferus przewalskii*, by Mercedes Yrayzoz (vectorized by T. Michael Keesey) under a CC BY 3.0 licence; *Camelus dromedarius*, by Steven Traver under a CC0 1.0 Universal Public Domain licence; *Panthera pardus*, by Margot Michaud under a CC0 1.0 Universal Public Domain licence; *Manis culionensis*, by Steven Traver under a CC0 1.0 Universal Public Domain licence. The bat silhouette was vectorized by A.E.M., the Afrotheria (Thai Elephant) silhouette was obtained from OpenClipArt (https://openclipart.org/) under a CC0 1.0 Universal Public Domain licence.



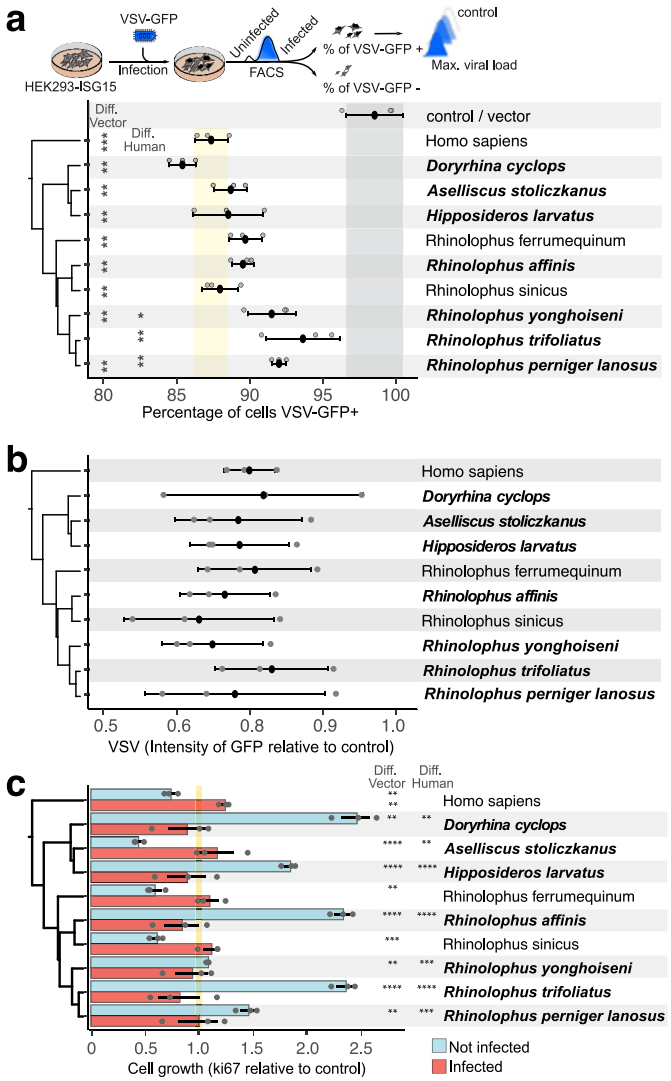
Extended Data Fig. 5 | Immune-related genes under positive selection in bats. a, Biological processes involved in immune responses triggered by viral infections. **b-f**, Schematic showing how ISG15 and genes selected in bats are involved in viral entry into cells and detecting viral patterns (b), regulating antiviral and inflammatory responses (c-d), B cell signaling (e), and activation of the complement system (f). Genes positively-selected in the ancestral branches of Chiroptera (white font), Rhinolophidae–Hipposideridae (yellow font), or Rhinolophidae (orange font) have a blue background. Genes highlighted in panels b-d are described in the main text. **e**, Inflammation triggers the release of chemokines that direct the migration of leukocytes to sites of infection. GO 'leukocyte migration' (GO:0050900, $P_{cor} = 0.0016$) is only enriched for genes selected in bats. We identified selection in two CC chemokine receptors, *CCR2* (repeatedly selected in Chiroptera, Rhinolophidae–Hipposideridae and Rhinolophidae) and *CCR5* (selected in Rhinolophidae–Hipposideridae) that promote the infiltration of pro-inflammatory cells[143,144]. Importantly, *CCR2* and *CCR5* are located in a major risk locus for severe COVID-19 that contains SNPs linked to increased *CCR2/5* expression[145]. Consistent with CCR2/5 mediated hyperinflammation, CCR2/CCR5 receptor antagonists can reduce cytokine storms in patients with severe COVID-19 (ref. 146). The GO term 'lymphocyte activation' (GO:0046649, $P_{cor} = 2.46 \times 10^{-5}$) has the most significant enrichment for genes under selection in bats, which includes two key factors for B cell signaling, *CD79A* (selected in Rhinolophidae–Hipposideridae) and *BTK* (described in the main text). CD79A is the signal transduction subunit of the B cell antigen receptor that upon phosphorylation mediates BTK phosphorylation[147,148]. Activated BTK promotes B cell receptor mediated survival of B cells. **f**, The complement system helps to phagocytose or lyse pathogens, stimulates adaptive immunity and inflammation, but excessive complement activation can lead to hyperinflammation and thrombosis during COVID-19[149]. GO 'Regulation of immune effector process' (GO:0002697, $P_{cor} = 1.55 \times 10^{-8}$; Fig. 2e), 'Complement system' (WikiPathways WP2806, $P_{cor} = 8.92 \times 10^{-8}$) and 'Complement and coagulation cascades' (KEGG:04610, $P_{cor} = 0.0003$; Supplementary Table 7) have the most significant enrichment for genes selected in bats. We detected two complement components, *C7* and *C1S* (both under selection in Rhinolophidae–Hipposideridae). C1S encodes a serine protease involved in early activation of the classical pathway[150]. C7 is a component of the membrane attack complex that forms membrane-disrupting pores and, when endocytosed, activates noncanonical NF-κB signaling and inflammasome assembly[151]. Pictograms of cellular elements were custom-drawn.



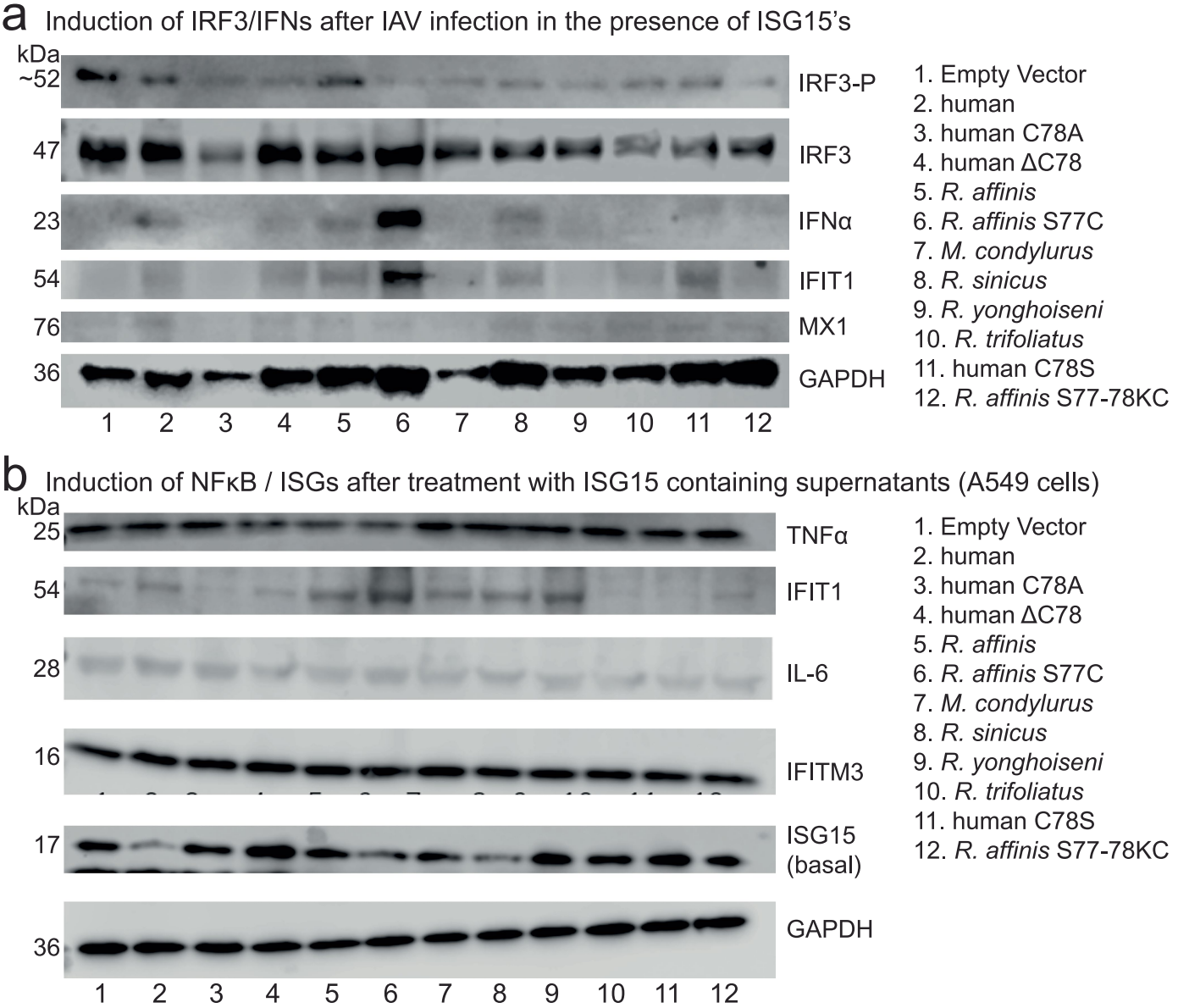
Extended Data Fig. 6 | Conformations of putative dimers of ISG15 for human, *Doryrhina cyclops*, and *Rhinolophus sinicus*. We used structural modeling to explore whether the Cys78 deletion affects homodimer formation of bat ISG15. We hypothesize two alternative outcomes. A stable homodimer will not be formed if we delete Cys78 due to the removal of the disulfide bond. Alternatively, stable homodimers might still form if the dimeric interface is maintained by compensating sets of non-covalent interactions (e.g., hydrogen bonds, hydrophobic interactions). To test these hypotheses, we initially performed 3D structural modeling using AlphaFold2[110] and compared the putative ISG15 homodimers of human and four bats lacking Cys78 (Supplementary Fig. 16). However, the AlphaFold models showed discrepancies between the dimeric interface confidence scores generated and previous experimental studies for human ISG15[47,48], as well as predicted variations in stability between the bat species (Supplementary Fig. 16); therefore we evaluated the stability of predicted dimers by molecular dynamics simulations shown in this figure. We conducted three replicate simulations of 1 µs for human ISG15 and for ISG15 of one representative rhinolophid and hipposiderid bat (total duration 3 µs, about 550,000 CPU hours for each species) (Supplementary Figs. 17-18 show the simulation boxes and equilibration phase). **a-c**, Conformations of putative dimers of human (**a**), *Doryrhina cyclops* (**b**), and *Rhinolophus sinicus* (**c**) ISG15. The dimer that was estimated with AlphaFold is circled, whereas the other dimers are major conformations, observed during molecular dynamics simulations (see Methods). Helices are shown in cyan, whereas β-strands in red. The percentages correspond to the proportion of each conformation across the simulations. The figure was rendered with UCSF ChimeraX v.1.2[152]. **a**, Both major conformations are very similar to the AlphaFold predicted conformation. Importantly, while the monomers slightly rotate around the disulfide bond, the Cys78-mediated disulfide bond is sufficient to maintain a dimeric interface with a highly similar conformation, indicative of a stable dimer (see also Supplementary Videos 1–3). The molecular dynamics simulation results of human ISG15 are consistent with experimental studies[47,48] and serve as a positive control as the disulfide bond cannot be broken in the simulation. **b**, If a stable dimer would exist, we would expect to observe a single dominant structural conformation, or several conformations that are highly similar, as observed for human ISG15. In contrast, for *D. cyclops* ISG15, the lack of the disulfide bond appears to result in an unstable dimer that adopts a range of conformations, showing remarkable differences in the spatial arrangement of the monomers (see also Supplementary Videos 4–6). **c**, As for *D. cyclops* ISG15, the lack of the Cys78-mediated disulfide bond appears to consistently result in an unstable dimer that adopts two major conformations that differ considerably in the spatial arrangement of the monomers, both from each other and from the starting AlphaFold structure (see also Supplementary Videos 7–9). This result was observed both at physiological (top) and at salt concentrations as low as required to neutralize the system (middle) for a comparable simulation time of up to 0.5 µs, indicating that different salt concentrations consistently result in unstable dimers. Comparing low salt conditions for a simulation time of 0.5 vs. 1 µs (middle vs. bottom), we found even larger structural differences between the major conformations, which is consistent with the expectation that as the simulation time increases, unstable dimers would increasingly deviate from the initial structure. Overall, these tests also indicate that our results are robust to variation in the salt concentration used in the simulations.



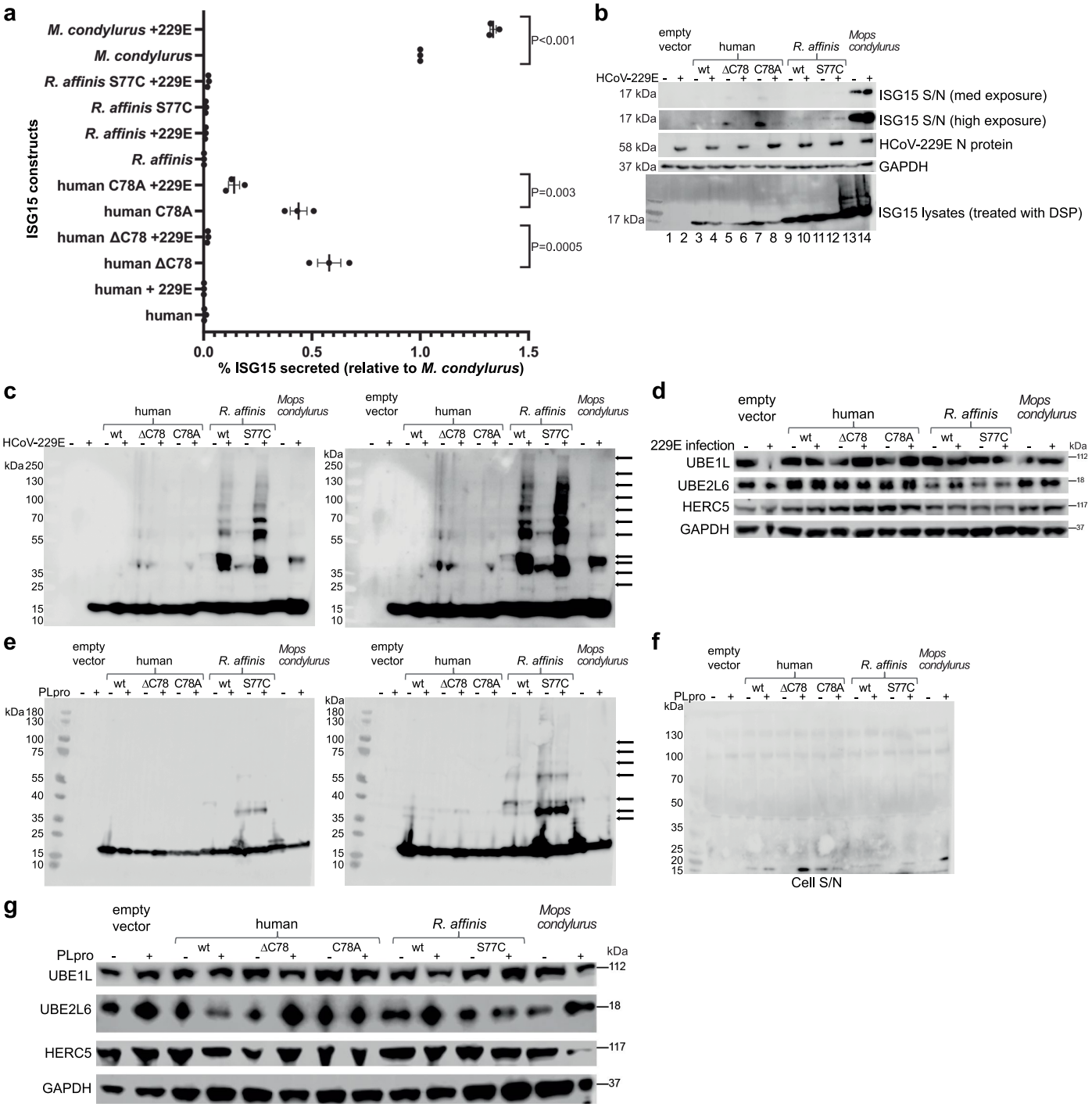
Extended Data Fig. 7 | Snapshots of the dimeric ISG15 interface during molecular dynamics simulations performed with a low salt concentration. **a**, *Rhinolophus sinicus* ISG15; **b**, *Doryrhina cyclops* ISG15. While the three simulations initially exhibit highly similar dimeric interfaces, because the starting point is the same AlphaFold-predicted dimer after energy minimization and equilibration, the interfaces increasingly deviate from each other as simulation time increases, indicative of an unstable dimer. See also Supplementary Videos 4–9. The figure was rendered with PyMOL v. 2.5.0[153].



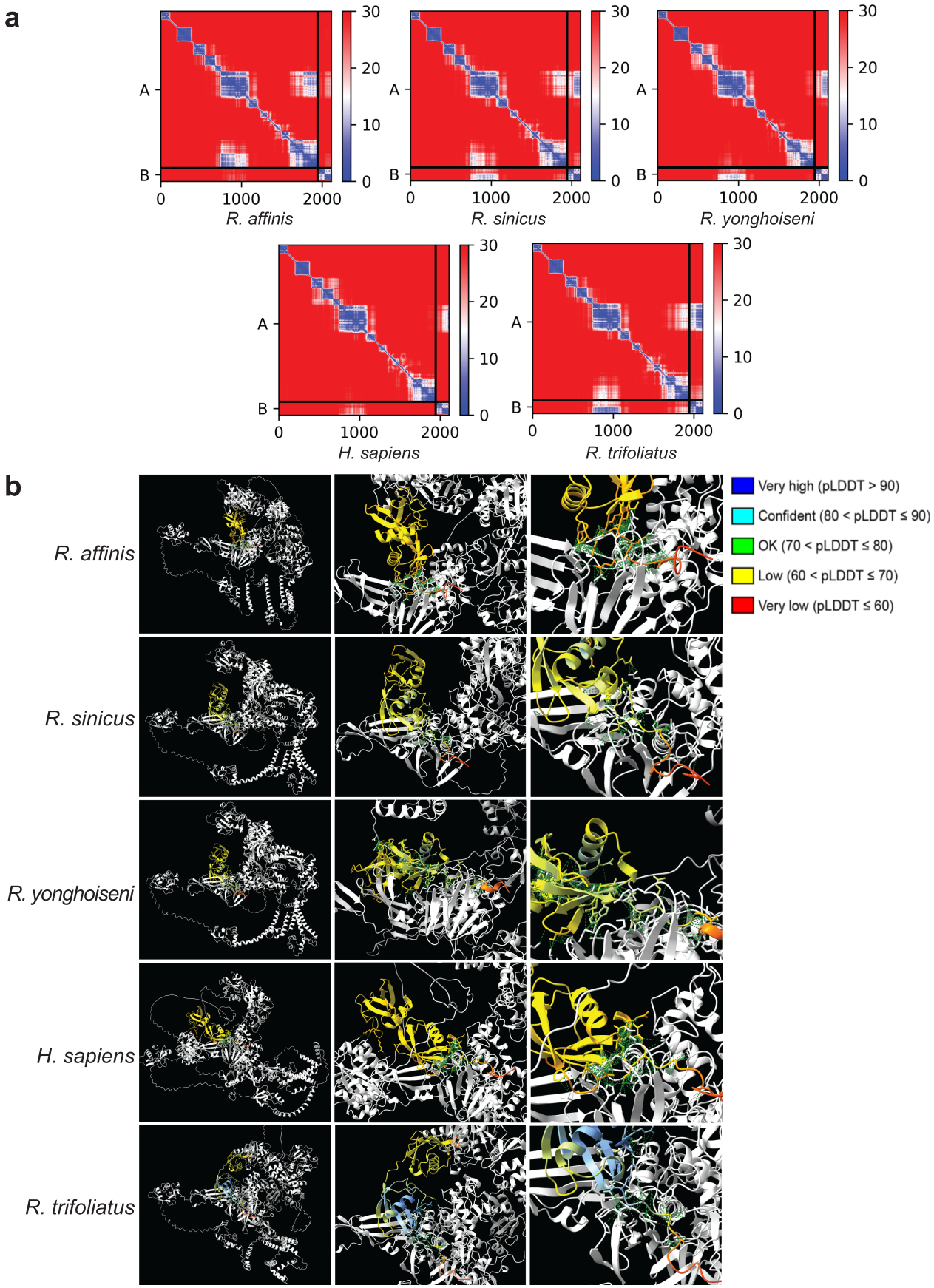
Extended Data Fig. 8 | ISG15 effects on VSV cellular entry and replication, and cell viability. To examine how human and various Cys78-lacking bat ISG15 constructs affect VSV entry and replication and cell viability, we synthesized in addition to *Rhinolophus affinis* ISG15 from five additional rhinolophid (*R. perniger lanosus, R. yonghoiseni, R. sinicus, R. trifoliatus, R. ferrumequinum*) and three hipposiderid bats (*Aselliscus stoliczkanus*, *H. larvatus*, *Doryrhina cyclops*). **a**, Viral entry, quantified as the percentage of VSV-GFP positive cells. HEK293 cells were transiently transfected with ISG15 constructs and infected with VSV-GFP. GFP-positive cells after 16 h infection with an MOI of 0.01 were measured by FACS, examining mCherry (or transfected) cells only. The percentage of GFP-positive cells is shown for three separate experiments (n > 10,000 cells each). Data is presented as mean (solid oval) and standard deviation (bars) with individual data points for the three biological replicates displayed as grey circles. Grey and light yellow backgrounds indicate empty vector and human, respectively. All ISG15 proteins reduce viral entry, with the exception of *R. trifoliatus* where the effect is not significant compared to the empty vector. Notably, viral entry differs substantially between ISG15 of different rhinolophid and hipposiderid bats. For example, ISG15 of *R. yonghoiseni*/*R. trifoliatus*/*R. perniger lanosus* (comprising a phylogenetic clade) had a significantly reduced ability to block viral entry compared to human ISG15. Because ISG15 of all rhinolophid and hipposiderid bats lacks Cys78 (Fig. 3a), this indicates that other amino acid mutations apart from the shared Cys78 deletion alter ISG15 function. P-values and raw data are in Supplementary Table 12. **b**, Viral replication, measured by GFP intensity during early infection stages. Viral load for the data shown in Fig. 3d, measured by mean fluorescence intensity (VSV-GFP) in HEK293 cells transiently transfected with ISG15 (IRES-mcherry) constructs, relative to the control vector (no ISG15). Data are presented as mean (solid oval) and standard deviation (bars) with individual data points of three biological replicates shown as grey circles. P-values and raw data are in Supplementary Table 13. **c**, ISG15 of some bats affects growth of uninfected HEK293 cells. Since we observed large differences between initial infection, intracellular viral load and final viral release into the supernatant, we additionally examined cell viability in the presence of ISG15. The panel shows FACS measurements of Ki-67, a cellular marker for proliferation, in HEK293 cells that were stably transfected with the various ISG15 constructs. Measurements were taken at 16 h post-infection and are normalized to the 0 h timepoint and the vector control. Thus, values > 1 indicate an increase in cell growth compared to the vector control. Three biological replicates are indicated by grey dots and black bars show standard error per treatment. P-values and raw data are in Supplementary Table 14. For uninfected cells (light blue bars), expression of human ISG15 and ISG15 of *A. stoliczkanus*, *R. ferrumequinum* and *R. sinicus* leads to a significant decrease in cell growth, compared to the vector control. In contrast, expression of ISG15 of the six other bats (*D. cyclops, H. larvatus, R. affinis, R. yonghoiseni, R. trifoliatus*, and *R. perniger lanosus*) results in a significantly increased growth of uninfected cells, showing that ISG15 of some bats positively affects cell growth in the absence of viral infection. Furthermore, ISG15 of six bats induces a significantly higher cell growth compared to human ISG15, while *A. stoliczkanus* ISG15 induces a significantly lower cell growth. For cells infected with GFP-VSV (red bars), none of the differences are significant when comparing bat to human ISG15. Human ISG15 was, however, still significantly different between infected and uninfected ($p = 0.0017$). Unlike human, several bats show large decreases in proliferation between uninfected and infected cells, indicating different dynamics inside the cell during virus replication. Furthermore, the basal increase in proliferation may have a substantial consequence on later virus production by allowing cells to continue producing higher amounts of virus prior to cell death[154], or by altering the basal immune state, possibly explaining a reduced effect of human or bat wildtype ISG15 in blocking VSV production. **a**-**c**, Significant differences to the vector control or to human ISG15 were determined with a two-tailed t-test and is indicated with * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. New Bat1K genomes are indicated in bold font.



Extended Data Fig. 9 | Effect of ISG15 Cys mutants on ISG expression and NFκB / IFN signaling. a, ISG15 Cys mutants do not alter ISG expression during H1N1 PR8 Influenza A virus (IAV) infection or ISG15 supernatant treatment in epithelial cells. To investigate why mutating or deleting Cys78 from human ISG15 and restoring Cys78 in *R. affinis* ISG15 significantly increased IAV production, we tested whether the ISG15 Cys mutants but not wildtype ISG15 impacted IRF3 activation (known to be modulated by ISG15[155]), IFN production, and induction of key antiviral interferon-stimulated genes (MX1 and IFIT1). IAV was infected at an MOI of 0.1 for 48 h after transfection with wildtype ISG15 constructs and ISG15 Cys78 mutants (indicated on right). The representative western blot shows that IAV infection causes some IRF3 activation and only minimal MX1 or IFIT1 protein induction in Vero-E6 cells when transfected with an empty vector or most ISG15 constructs. While *R. affinis* S77C induced more MX1 and IFIT1, this did not alter IAV production (Fig. 3e), indicating minimal effect on the final production of infectious IAV particles. This is consistent with VeroE6 cells having limited IFN signal amplification, and shows that while ISG15 modulates IFN signaling and some ISG-induction in Vero-E6 cells, other effects of ISG15 on viral production are more relevant. Importantly, there is no impact of Cys-removal from human ISG15 on MX1 or IFIT1 in infected cells, indicating that the observed pro-viral outcome of ISG15 Cys mutants is likely not due to downregulating interferon-stimulated gene expression. A representative western blot image is shown from three independent experiments. **b**, Extracellular ISG15 does not enhance NFκB and IFN signaling in A549 cells. Wildtype and Cys-mutant ISG15 showed limited anti-IAV activity. To rule out a reliance on extracellular cytokine enhancement from ISG15, we measured the effect of extracellular wildtype or mutant ISG15 on NFκB and IFN signaling. As this process requires the ISG15 receptor, LFA-1[156], which is not expressed in VeroE6 cells, we used A549 cells that are IFNɣ-competent and express a low-level of LFA-1. ISG15-containing supernatants from IAV-infected VeroE6 cells were UV-treated with 10000 gy of UV-C to deactivate IAV. A549 cells were then treated with these supernatants overnight. We then measured A549-cell lysates for TNFα and IL-6 to assess NFκB-signaling, and IFIT1, IFITM3, and endogenous ISG15 to assess IFN signaling. We observed no obvious induction of these pathways in these cells between wildtype and Cys-mutant ISG15, indicating the pro-viral effect of ISG15 mutants is likely not caused by impairing ISG15's cytokine enhancement function in these cells. A representative western blot image is shown from three independent experiments.



Extended Data Fig. 10 | ISG15 conjugation/secretion with PLpro or during HCoV-229E infection. a, HEK293-CD13-myc ISG15 stable cells were infected with HCoV-229E at an MOI of 0.01 for 72 h prior to collection of cell lysates (lysates) and supernatant (S/N) prior to western blot with anti-myc Ab (as above) to detect transfected ISG15. The level of ISG15 in the supernatant was normalized to cell lysates and loading control (GAPDH), and was expressed relative to the most abundant protein in the supernatant (*Mops condylurus*). Only ISG15 of *Mops condylurus* was significantly secreted into the supernatant and secretion further increased during HCoV-229E infection. Mean and standard error of the mean are displayed, including individual points (black) for n = 3 independent experiments. Significance was tested with a two-tailed t-test comparing infection to ISG15 alone. Raw data are provided in Supplementary Table 17. **b**, An example western blot image (as per panel a) showing high and low exposure of ISG15 supernatant (S/N), HCoV-229E N protein, GAPDH and ISG15 bands in the HEK293-subclone-ΔISG15 cell lysate with protein ladders. Lanes are as indicated. **c**, Example western blot image with low and high contrast showing ISG15 conjugation (α-myc) after transfection with ISG15's and HCoV-229E infection (−/+). Arrows indicate bands not seen in the empty vector control (conjugated proteins). **d**, Western blot of E1 ligase (UBE1L), E2 ligase (UBE2L6) and E3 ligase (HERC5) expression, together with GAPDH in the same samples shown in panel c, indicating sufficient expression of ISGylation machinery in the HEK293-subclone-ΔISG15 cell line. **e**, HEK293-subclone-ΔISG15 cells were transfected with ISG15 constructs as indicated and with/without NSP3C/L (SARS-CoV-2 PLpro). As per panel c, arrows indicate ISGylation bands, or ISG15 dimer/monomer. The amount of ISG15 present was calculated, normalized to GAPDH and quantified from three western blots. One of the three western blots is shown as an example. **f**, Representative western blot for the amount of ISG15 present in the supernatant (matched to e). **g**, ISGylation machinery expression, as per panel d, for NSP3C samples used in panel e and f and Fig. 3. Western blots in panels b-g are matched to the quantification in Fig. 3h–k.



Extended Data Fig. 11 | See next page for caption.

Extended Data Fig. 11 | Models of SARS-CoV-2 PLpro in complexes with ISG15 of bats and human. a, Predicted Aligned Error (PAE) plots obtained with Alphafold2-Multimer in ColabFold[111]. Alphafold2 modeled SARS-CoV-2 (wildtype, Wuhan-1) PLpro chain (amino acids 819–2763 of ORF1ab) monomer in complex with a monomer of ISG15 of the five indicated species. The PAE plots represent the error in the position of each amino acid. The crystal structure of PLpro in complex with ISG15 revealed it exists as a pentamer of five units of PLpro with five monomeric ISG15 units in a large complex[54]. This AlphaFold2 model represents the native monomer:monomer structure as folded in Alphafold2-Multimer (Mmseqs2 v 1.5.5) via ColabFold, due to computational limits. See also Supplementary Videos 10–19. **b**, Snapshots from Supplementary Videos 10–19 showing PLpro in a complex with ISG15 of bats and human. Images on the left show a whole protein view of the complex (as per panel a). The middle images show a zoom view of the active site and the right images display the flexible C-terminal tail of ISG15 adjacent to the PLpro binding groove with contacts visualized in green (Supplementary Videos 10–19 show the details). pLDDT scores and number of contacts for each of the five species are 64.84: 422, 65.77: 156, 66.04: 761, 60.44: 354, 65.79: 126, respectively. ISG15 is colored by pLDDT values. PLpro in white. The linear C-terminal tail of human ISG15 resembles that of *R. trifoliatus* ISG15, loading directly into the PLpro catalytic site, with additional contact sites in *R. trifoliatus* that seem to stabilize the interaction.

nature portfolio

Corresponding author(s): Aaron T. Irving, Michael Hiller

Last updated by author(s): Oct 30, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Custom scripts used for data analysis are available at GitHub (https://github.com/ariadnamorales/2023_Bat1Kimmunity).
Data analysis	Pacific Biosciences pipeline v.4.2.0 621 (https://github.com/PacificBiosciences/ccs), hifiasm v.0.13, purge_dups v.1.2.3, hifiasm v0.15.5-r352, hifiasm v0.15.4-r432 (see methods), HiCanu v.2.1, Canu v2.2, 10X Genomics longranger v2.2.2, Scaff10X v4.2, Break10X v3.1, Bionano Solve v1.6.1, g bwa-mem v0.7.17-r1188, Arima v2 (VGP) https://github.com/VGP/vgp-653, salsa2 v2.2, pbmm2 https://github.com/PacificBiosciences/pbmm2, DeepVariant v1.2.0, Merfin v1.1-development r197, bcftools consensus v1.12, RepeatClassifier https://github.com/davidaray/bioinfo_tools/blob/master/extract_align.py, RepeatAfterMe (RAM) https://zenodo.org/record/7076442, TE-Aid, USEARCH, RM2Bed.py, RepeatMasker v4.0.6, RNAfold v2.4.18, LASTZ v.1.04.15, axtChain v.1, ChainCleaner v.1, TOGA https://github.com/hillerlab/TOGA, commit v.c4bce48, MACSE v.2, HmmCleaner v.0.180750, ASTRAL v.5.5.9, IQTREE v.2.1.3, BOOSTER v.1, ModelFinder (as implemented in IQTREE), TreePL https://github.com/blackrim/treePL, gProfiler (database built on 2022-05-18 e106_eg53_p16_65fcd97), phyloFit v.1, brms v.2.20.4, AlphaFold2 Colabfold v.1.3.0 (release 13-45111), MMseqs2, Carma v.2.01 1, Partitioning Around Medoids algorithm implemented in Cluster v2.1.3, ggPlot2, FlowJo v10.0, Fiji IMageJ 2.9.0/1.53t, Graphpad Prism 9.4.1.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Genome assemblies and sequencing data is available at NCBI for Aselliscus stoliczkanus (PRJNA949177), Doryrhina cyclops (PRJNA938463), Hipposideros larvatus (PRJNA938461), Rhinolophus affinis (PRJNA938462), Rhinolophus perniger lanosus (PRJNA955779), Rhinolophus yonghoiseni (PRJNA938455), Rhinolophus trifoliatus (PRJNA939732), Rhinopoma microphyllum (PRJNA971926), Megaderma spasma (PRJNA940731), and Mops condylurus (PRJNA949178). TOGA, transposable element and miRNA annotations of newly-sequenced bats and alignments of selected genes are available for download at http://genome.senckenberg.de/download/Bat1KImmune/. Accession codes and identifiers of genomic data are listed in Supplementary Tables 3 and 5. gRNAs, primers and geneblock sequences are listed in Supplementary Table 27.

Other databases used are: Dfam database (v3.5) (https://www.dfam.org/releases/Dfam_3.5/, Infernal (v1.1.2), GENCODE V38 (Ensembl 104), Gene Ontology http://geneontology.org/, KEGG https://www.genome.jp/kegg/, WikiPathways https://www.wikipathways.org/index.php/WikiPathways, miRTarBase http://mirtarbase.mbc.nctu.edu.tw, TRANSFAC http://genexplain.com/transfac/, Human Protein Atlas https://www.proteinatlas.org/, CORUM http://mips.helmholtz847muenchen.de/corum/, Human Phenotype Ontology https://hpo.jax.org/app/, uniref100

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	not applicable
Population characteristics	not applicable
Recruitment	not applicable
Ethics oversight	not applicable

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For in vitro cellular work, where single cell quantification was required a minimum of 100 cells was utilized (microscopy) or 10,000 cells (FACS) based on minimal statistical requirements for each experiment (individual t test comparisons, one-way ANOVA etc). All other work was an average for each well of a plate as measured by the Plate Reader, counted for plaque assays or combined pool for qPCR/western blot and was performed in at least biological triplicate then replicated as below.
Data exclusions	No data was specifically excluded from the analysis.
Replication	All experiments were performed in triplicate and replicated (biologically) a minimum of 3 times.
Randomization	Randomization was generally not used due to blinding of genotypes and samples and internal normalization against controls (e.g. empty vector). Samples for genome extraction were randomly selected amongst those with high enough quality of HMW DNA. Randomization of genomes was performed for the phylogenetic downsampling by randomly selecting 10 species of the 20-species orders.
Blinding	Where possible- sample ID (genotype) was blinded and coded with a number. Viral infections, transfection experiments were coded/blinded without revealing the species. Quantification for microscopy and FACS was performed on blinded codes. WB samples were coded until after lysate collection then unblinded to ensure appropriate loading order was maintained. All data was unblinded at the time of generating graphs/visual representation.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Anti-CD13 Antibody Rabbit Polyclonal (Sinobiological, 10051-T60), Human coronavirus (HCoV-229E) Spike S1 Antibody, Rabbit PAb, Antigen Affinity Purified (Sinobiological, 40601-T62) , Ki67 Rabbit Monoclonal Antibody (Beyotime, AF1738), 20282 CF647 Goat Anti-Rabbit IgG (H+L), highly cross-adsorbed (biotium, 20282), 20281 CF647 Goat Anti-Mouse IgG (H+L), highly cross-adsorbed (biotium 20281), 20101 CF568 Goat Anti-Mouse IgG (H+L), highly cross-adsorbed (biotium 20181), Rabbit Anti-VSV-G tag antibody (Bioss, bs-2110R), ISG15 Antibody - middle region (avivasysbio, ARP59386_P050), Rabbit anti-IRF3 monoclonal antibody – Huabio (ET1612-14), Rabbit anti-Phospho IRF3 monoclonal antibody– Huabio (ET1608-22), MX1**抗体** [N2C2], Internal (Genetex, GTX110256), Mouse anti-MYC monoclonal antibody- Sino Biological (100029-MM08), Mouse anti-TNF alpha monoclonal antibody – Salarbio (K009343M), Mouse anti-IFIT1 monoclonal antibody – Sangon (D199761), GAPDH (14C10) Rabbit mAb (CST, #2118), ProteinFind ® Goat Anti-Mouse IgG (H+L), Anti-Rabbit Goat Anti-Rabbit IgG (H+L), Highly Cross-Adsorbed Secondary Antibody, Alexa Flour 488/568/647, HRP Conjugate (transgen, HS201-01), ProteinFind ® Goat Anti-Rabbit IgG (H+L), HRP Conjugate (transgen, HS101-01), **辣根**过氧化物酶标记的驴抗羊IgG Donkey Anti-Goat IgG/HRP (SolarBio, K0038D-HRP-100ul).rabbit anti-UBE1L monoclonal antibody (Huabio, HA721228, dilution:1:500), rabbit polyclonal anti-UBE2L6 antibody (Abclonal, A13670), rabbit polyclonal anti-HERC5 antibody (Abclonal, A14889)

Validation

Commercial companies performed specific validation via peptide binding to antibodies as per each spec sheet, additional validation was performed in the laboratory with non-transfected or non-infected controls showing no staining. Rabbit anti-ISG15 polyclonal antibody - Aviva Systems Biology (ARP59386_P050)
This antibody alone was validated to use for detecting human and mouse and predicted to have a broad-spectrum reactivity in following species: cow, dog, goat, horse, pig, rabbit, sheep by the manufacturer.
Myc-tag inserted into ISG15 plasmids was tested with anti-Myc antibody as the surrogate for detecting ISG15 thereafter.
Remaining antibodies were validated by the providers, concentration used for western blot and FACS were further however, optimized to ensure the results are accurate and repeatable.Rabbit anti-CD13 polyclonal antibody (Sinobiological, 10051-T60) was validated by the manufacturer by using U937 and THP-1 cell lines which naturally express ANPEP/CD13. Rabbit polyclonal anti-human coronavirus (HCoV-229E) nucleocapsid antibody (Sinobiological, 40640-T62) was validated by the manufacturer by using human coronavirus (HCoV-229E) nucleoprotein for western blot. Rabbit anti-Ki67 monoclonal antibody (Beyotime, AF1378) was validated by the manufacturer using HepG2 cell lines for western blot, human tonsil tissue for IHC and Hela cells stained with Ki67 antibody for flow cytometry analysis. Rabbit Anti-VSV-G tag antibody (Bioss, bs-2110R) was validated by the manufacturer using overexpressed E.coli as positive control for western blot. Rabbit anti-IRF3 monoclonal antibody (Huabio, ET1612-14) was validated by the manufacturer using 5 different cell lines including Hela, Jurkat and THP-1, along with more than 10 types of human and mouse tissue as positive control for western blot. Rabbit anti-Phospho IRF3 monoclonal antibody (Huabio, ET1608-22) was validated by the manufacturer using NIH/3T3 cell lysate and NIH/3T3 cells treated with 100nM Calyculin as positive control for western blot. Polyclonal rabbit anti-MX1 antibody (GeneTex, GTX110256) was validated by the manufacturer using MX1-overexpressed 293T cell lysate as positive control for western blot. Mouse anti-TNF alpha monoclonal antibody (Solarbio, K009343M) was validated by the manufacturer using rat plasma and rat serum as positive control for western blot. Mouse anti-IFIT1 monoclonal antibody (Sangon, D199761) was validated by the manufacturer using IFIT1-overexpressed cell lysates for western blot, validated for predicting human, mouse, rat and monkey IFIT1, this antibody was only used on human cell line in our research. Rabbit monoclonal anti-GAPDH antibody (CST, #2118) was validated using HELA cells as positive control for western blot and IH, this antibody was also validated in bat cell lines in our lab. Rabbit anti-UBE1L monoclonal antibody (Huabio, HA721228) was validated by the manufacturer using THP-1 cell lysates, rat kidney tissue, human testis tissue, mouse cerebellum tissue as positive control for western blot, validated for human, mouse and rat Ube1L. Rabbit polyclonal anti-UBE2L6 antibody (Abclonal, A13670) was validated by the manufacturer using mouse lung and rat kidney tissues as positive control for western blot. Rabbit polyclonal anti-HERC5 antibody (Abclonal, A14889) was validated by the manufacturer using U-87MG, OVCAR3 and HT-1080 cell lysates and mouse kidney and rat testis tissues as positive control for western blot.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)

Huh7/MRC-5 was supplied from Prof. Jincun Zhao (Guangzhou Medical University, original source ATCC), HEK293-subclone (original source: ATCC), A549 (CCL-185, ATCC), Vero-76 (CRL-1587, original source: ATCC), Vero-E6 cells sourced from Prof.

	Linrong Lu (Zhejiang University, original source: CTCCC repository). A549-ACE2 cells from Dr. Colpitt's laboratory (clone A549-ACE2 B9 as published). Hek293-CD13 cells and A549-ACE2-TMPRSS2 were generated for this project, as described. Hek293 subclone was sorted from parental Hek293 for high viral production. Cas9/crispr KO for ISG15 or scrambled guide were generated from this subclone. RsKT.01 (original cell line, Zhengli Shi, Wuhan Institute of Virology)
Authentication	No independent verification was performed by a commercial service however, Vero-E6 cells, RsKT, A549 and Hek293 cells transcriptome matches the appropriate species. ISGylation machinery was validated as shown in the supplemental figures.
Mycoplasma contamination	All cell lines were routinely tested for mycoplasma and found negative by pan-mycoplasma PCR.
Commonly misidentified lines (See ICLAC register)	None of the cell lines used belong to ICLAC

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	not applicable
Wild animals	Tissue samples were acquired from the Royal Ontario Museum mammal collection or field expeditions for: Aselliscus stoliczkanus China – Shuipu Village, Yuping Town M 3/28/2007 Doryrhina cyclops Ivory Coast – Parc National De Tai, Institute D'ecologie Tropicale F 2/18/1992 Hipposideros larvatus China – Shuipu Village, Yuping Town M 04/12/2007 Rhinolophus affinis China – Shiwandashan National Reserve M 4/25/2005 Rhinolophus perniger lanosus China – Shuipu Village, Yuping Town M 39174 Rhinolophus yonghoiseni Malaysia – Nature Education and Research Centre, Endau Rompin National Park F 8/14/2001 Rhinolophus trifoliatus Malaysia – Nature Education and Research Centre, Endau Rompin National Park M 08/10/2001 Rhinopoma microphyllum Israel – Northern Israel M 1/13/2013 Megaderma spasma Vietnam – Dong Nai, Cat Tien National Park Headquarters M 05/05/1998 Mops condylurus Côte d'Ivoire – Bregbo Village M 11/20/2019 All animals were euthanized and tissues were used for genome sequencing (all bat samples were re-purposed samples for genome sequencing with the exception of Mops condylurus which was collected specifically for generation of a genome).
Reporting on sex	Where possible Males were utilized for presence of both X and Y chromosomes, cyclops and yonghoiseni were F due to limited sample availability.
Field-collected samples	No housing, instant collection.
Ethics oversight	Doryrhina cyclops (ROM-M100513) – Permit number 81 DPN from Direction de la Protection de la Nature, République de Côte d'Ivoire Rhinolophus yonghoiseni (ROM-M113050), Rhinolophus trifoliatus (ROM-M113012) – Reference number PTN(J) 3/8 from Perbadanan Taman Negara (National Parks Corporation) Johor, Malaysia Aselliscus stoliczkanus (ROM-M118506), Hipposideros larvatus (ROM-M118627), Rhinolophus affinis (ROM-M116429), Rhinolophus perniger lanosus (ROM-M118548) – Certificate numbers 2007/CN/ES133-137/KM from The Endangered Species Import and Export Management Office of the People's Republic of China Megaderma spasma (ROM-M110751) – Number 138/STTN from Institute of Ecology and Biological Resources, National Center for Science and Technology, Vietnam Mops condylurus (ID: 03#106) – Capture of bats and animal work were performed with 547 the permission of the Laboratoire Central Veterinaire, Laboratoire National d'Appui au Développement Agricole (LANADA), Bingerville, Côte d'Ivoire (No. 05/virology/2016) and the Ministère des Eaux et Forêts (No. 0474/MINEF/DGFF/FRC-aska). Rhinopoma microphyllum – National Parks Authority, permit 2013/04169. IACUC 04-20-551 019. NB: All Ethics and guidelines for each specific country were valid at the time of collection.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	To generate ISG15 stable cell lines, lentiviral transduced Huh7, HEK293, A549 and Vero-E6 cells were sorted by fluorescence activated cell sorting (FACS) using the BD Influx System for mCherry-positive cells, normalized against autofluorescence in the

	respective parental cell line. VSV-GFP load was measured directly via GFP fluorescent intensity. For HCoV-229E, CD13 stable HEK293 cells were stained with 229E N protein (Sino biological, 1:2000) and Ki67 (Beyotime, 1:500) for 30 minutes in FACS buffer containing 1x PBS (Gibco), 1% FBS and 1% P/S, after permeabilization with 0.05% TX-100 in TBS and blocking in 5% BSA in TBS-T. Cells were subsequently rinsed, stained with anti-mouse / rabbit CF®-488/568/647 secondary antibody for 15 minutes (Biotium, dilution 1:10000), rinsed thrice and run on the ACEA Novocyte flow system. No primary controls, untransfected controls or uninfected controls and cells only for autofluorescence were used to generate compensation matrices.
Instrument	Acea Novocyte; BD Influx
Software	Acea Novocyte capture software, FlowJo v10.0
Cell population abundance	minimum of 10,000 events captured for the target population (transfected + infected).
Gating strategy	primary gates were used to separate live cells from dead based on SSC-H/FSC-H, then SS-H/SSC-A for single cells, then separate gates for mcherry positive or negative based on untransfected controls, followed by differential gates for infected or non-infected, gates drawn on uninfected control cells (stained with same antibodies).

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.