Reference-quality bat genomes illuminate adaptations to viral tolerance and disease resistance

Ariadna Morales
Senckenberg Research Institute
Yue Dong
Edinburgh Medical School: Biomedical Sciences, College of Medicine and Veterinary Medicine, The University of Edinburgh, Edinburgh EH8 9AG, UK
Thomas Brown
Max Planck Institute of Molecular Cell Biology and Genetics https://orcid.org/0000-0001-8293-4816
Kaushal Baid
Vaccine and Infectious Disease Organization, University of Saskatchewan, Saskatoon, SK S7N 5E3, Canada
Dimitrios Kontopoulos
LOEWE Centre for Translational Biodiversity Genomics
Victoria Gonzalez
Vaccine and Infectious Disease Organization, University of Saskatchewan, Saskatoon, SK S7N 5E3, Canada https://orcid.org/0000-0002-6788-2939
Zixia Huang
School of Biology and Environmental Science, University College Dublin, Dublin, Ireland
Alexis Ahmed
LOEWE Centre for Translational Biodiversity Genomics
Leon Hilgers
LOEWE Centre for Translational Biodiversity Genomics
Sylke Winkler
MPI CBG https://orcid.org/0000-0002-0915-3316
Graham Hughes
Laboratory of Molecular Evolution and Mammalian Phylogenetics, University College Dublin
https://orcid.org/0000-0003-3088-345X
Xiaomeng Li
Zhejiang University-University of Edinburgh Institute, Zhejiang University School of Medicine, Zhejiang University, Haining 314400, China.
Bogdan Kirilenko
LOEWE Centre for Translational Biodiversity Genomics

Paolo Devanna
Max Planck Institute for Psycholinguistics
Tanya Lama
University of Massachusetts Cooperative Fish and Wildlife Research Unit
Yomiran Nissan
School of Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University, 6997801, Tel Aviv, Israel
Martin Pippel
Max Planck Institute of Molecular Cell Biology and Genetics https://orcid.org/0000-0002-8134-5929
Liliana Dávalos
Stony Brook University https://orcid.org/0000-0002-4327-7697
Sonja Vernes
Neurogenetics of Vocal Communication Group, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
Sebastien Puechmaille
Université de Montpellier
Stephen Rossiter
Queen Mary University of London
Yossi Yovel
Tel Aviv University https://orcid.org/0000-0001-5429-9245
Joe Prescott
Centre for Biological Threats and Special Pathogens, Robert Koch Institute, Nordufer 20, 13353 Berlin, Germany
Andreas Kurth
Robert Koch Institute
David Ray
Texas Tech University
Burton Lim
Royal Ontario Museum https://orcid.org/0000-0002-0884-0421
Eugene Myers
Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstraße 108, 01307 Dresden, Germany
Emma Teeling
University College Dublin, Ireland https://orcid.org/0000-0002-3309-1346
Arinjay Banerjee
Vaccine and Infectious Disease Organization, University of Saskatchewan https://orcid.org/0000-0002-2821-8357
Aaron Irving
Zhejiang University https://orcid.org/0000-0002-0196-1570

Michael Hiller

michael.hiller@senckenberg.de

LOEWE Centre for Translational Biodiversity Genomics

Biological Sciences - Article

Keywords: genome sequencing and assembly, comparative genomics, immunity, viral tolerance, coronaviruses, SARS-CoV-2

Posted Date: February 14th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2557682/v1

Additional Declarations: There is **NO** Competing Interest.

Reference-quality bat genomes illuminate adaptations to viral tolerance and disease resistance

Ariadna E. Morales [1,2,3,*], Yue Dong [4,5,*], 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], Leon Hilgers [1,2,3], Sylke Winkler [6,7], Graham Hughes [10], Xiaomeng Li [5,4], Bogdan M. Kirilenko [1,2,3], Paolo Devanna [11], Tanya M. Lama [12,], 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], Yovel Yossi [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 [26,5,4,#], Michael Hiller [1,2,3,#]

[1] LOEWE Centre for Translational Biodiversity Genomics, Senckenberganlage 25, 60325 Frankfurt, Germany
[2] Senckenberg Research Institute, Senckenberganlage 25, 60325 Frankfurt, Germany
[3] Goethe-University, Faculty of Biosciences, Max-von-Laue-Str. 9, 60438 Frankfurt, Germany
[4] Edinburgh Medical School: Biomedical Sciences, College of Medicine and Veterinary Medicine, The University of Edinburgh , Edinburgh EH8 9AG, UK
[5] Zhejiang University-University of Edinburgh Institute, Zhejiang University School of Medicine, Zhejiang University, Haining 314400, China.
[6] Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstraße 108, 01307 Dresden, Germany
[7] DRESDEN concept Genome Center, Dresden, Germany
[8] Vaccine and Infectious Disease Organization, University of Saskatchewan, Saskatoon, SK S7N 5E3, Canada
[9] Department of Veterinary Microbiology, University of Saskatchewan, Saskatoon, SK S7N 5B4, 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, New York 11794, USA
[13] School of Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University, 6997801, Tel Aviv, Israel
[14] Sagol School of Neuroscience, Tel Aviv University, 6997801, Tel Aviv, Israel
[15] Consortium for Inter-Disciplinary Environmental Research, SUNY Stony Brook, Stony Brook, New York 11794, 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

[20] Centre for Biological Threats and Special Pathogens, Robert Koch Institute, Nordufer 20, 13353 Berlin, Germany
[21] Department of Biological Sciences, Texas Tech University, Lubbock TX 79409 USA
[22] Department of Natural History, Royal Ontario Museum, 100 Queen's Park, Toronto, Ontario M5S 2C6, Canada
[23] Department of Biology, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[24] Department of Laboratory Medicine and Pathobiology, Temerty Faculty of Medicine, University of Toronto, Toronto, ON M5S 1A8, Canada
[25] Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of British Columbia, Vancouver, BC V6T 1Z3, Canada
[26] Department of Clinical Laboratory, Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310058, China.

* These authors contributed equally
# Corresponding authors: aaronirving@intl.zju.edu.cn, michael.hiller@senckenberg.de

Running title: Viral tolerance and disease resistance in bats.

Keywords: bats, genome sequencing and assembly, comparative genomics, immunity, viral tolerance, coronaviruses, SARS-CoV-2

Abstract

Bats carry viruses that can cause severe disease in other mammals. Asymptomatic infections in bats suggest 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. A systematic analysis showed that signatures of selection in immune genes are more prevalent in bats compared with other mammals. We found an excess of immune gene adaptations in the ancestral Chiroptera and many descending bat lineages, highlighting viral entry and detection factors, and regulators of antiviral and inflammatory responses. ISG15, an antiviral gene contributing to hyperinflammation during COVID-19, exhibits a deletion of a cysteine, required for homodimer formation, in rhinolophid and hipposiderid bats. Cellular infection experiments showed enhanced intracellular protein conjugation of bat ISG15 and lack of secretion into extracellular space, where human ISG15 stimulates inflammation. Our work highlights molecular mechanisms contributing to viral tolerance and disease resistance in bats.

Introduction

Bats are recognized as natural reservoirs for a large diversity of viruses, some of which can cross species barriers and cause zoonotic disease in humans and other animals [1,2]. To date, viruses from 31 families have been found in bats, including paramyxoviruses (e.g. Hendra, Nipah, Mumps), filoviruses (e.g. Marburg, Bombali ebolavirus), rhabdoviruses (e.g. Rabies) and coronaviruses [3–6]. Among the coronaviruses, close relatives of the betacoronaviruses MERS-CoV, SARS-CoV and SARS-CoV-2 are found in bats. These viruses can cause diseases in humans such as COVID-19, driving fever, cough, pneumonia, acute respiratory distress, and sometimes leading to death [7]. Although transmission of MERS-CoV and SARS-CoV to humans likely occurred via intermediate mammalian hosts (e.g. civets, camels), accumulating data suggest that all three coronaviruses originated from bats, with Asian horseshoe bats (family Rhinolophidae) as the likely ancestral source of SARS-CoV and SARS-CoV-2 and the Egyptian tomb bat (family Emballonuridae) as the likely source of MERS-CoV [8–10].

Coronaviruses are especially widely distributed in bats and have been detected in species from 15 of the 21 bat families [3,11]. A survey of >19,000 mammalian individuals across African, Latin American and Asian countries detected coronaviruses in 8.6% of bat individuals but only 0.2% of non-bat individuals, and showed that bat species diversity correlates with coronavirus diversity [12]. Consistently, metagenomic screens revealed that 1.4% of the viruses detected in rodents are coronaviruses, while 17.8% of viruses detected in bats belong to *Coronaviridae* (Fig. 1A, Supplementary Table 1). In particular, coronaviruses appear to be more frequently detected in horseshoe bats (family Rhinolophidae) and roundleaf bats (family Hipposideridae) than in other bat families examined (Fig. 1B).

While corona- and other zoonotic viruses can cause severe disease or death in humans and other mammals, viral infections in natural reservoir hosts are often asymptomatic. Indeed, experimentally inoculating bats with corona- or Marburg viruses showed productive viral infection and replication, but a lack of clinical signs of disease [13–17]. This suggests that bats have a higher viral tolerance and that they evolved a unique immune response to balance antiviral defense with disease resistance [18,19]. One strategy for resisting disease upon viral infection is to eliminate the virus from the body, while also controlling infection-induced inflammation, which can lead to cytotoxicity and collateral tissue damage [20,21]. Previous studies showed that bats mount effective antiviral responses, but limit the expression of inflammatory cytokines and dampen uncontrolled immune responses and thus reduce immunopathology [22–24]. For example, Egyptian fruit bats (*Rousettus aegyptiacus*) that are infected with Marburg virus upregulate antiviral genes such as *IRF7*, *RIG-I*, *ISG15*, *MX1*, *IFIT1/2/3*, and *STAT1*, yet do not strongly induce proinflammatory genes [16].

Genomic analyses revealed insights into immune system changes that may contribute to enhanced resistance to viral disease in bats, such as selection of viral entry factors and innate immune response genes [18,25–28], diversified activating and inhibiting natural killer cell receptors [18], selection and losses of pro-inflammatory genes that regulate canonical NF-κB

signaling [29], expansions and contractions of type 1 interferon genes [18,28,30], and the absence of PHYIN genes that activate inflammasomes [25,31]. Further studies have revealed a general dampening of the inflammasome system to multiple immune stimuli [32,33]. Since powered flight requires high metabolic rates and many by-products of rapid metabolism and cellular stress can activate the immune system, greater viral tolerance could have evolved as a byproduct of immune adaptations to counter flight-induced sterile inflammation [22,32]. Nevertheless, the molecular changes that underlie viral tolerance in bats are not yet fully uncovered.

To elucidate the genomic basis of disease resistance, we generated new reference-quality genomes for ten bat species. We selected four rhinolophid (*R. yonghoiseni*, *R. lanosus*, *R. affinis*, *R. trifoliatus*) and three hipposiderid species (*H. larvatus*, *Aselliscus stoliczkanus*, *Doryrhina cyclops*), mostly from SouthEast Asia. These species represent divergent clades within these families. Several of these species are known to harbor coronaviruses [34–36] and SouthEast Asian Rhinolophidae are considered as natural reservoirs to the ancestors of SARS-CoV and SARS-CoV-2 [10]. To include representatives from the sister families Rhinopomatidae and Megadermatidae, we sequenced *Rhinopoma microphyllum* and *Megaderma spasma*. Finally, we sequenced *Mops condylurus* (family Molossidae), which is implicated as a natural Bombali ebola virus reservoir [37]. A systematic analysis across 115 mammalian genomes revealed that signatures of adaptive evolution in immune genes are most prevalent in bats, providing genomic evidence for special immune system adaptations in this mammalian order. Our screen identified bat changes in antiviral effector genes and regulators of inflammatory immune responses that may be relevant for human disease. Comparative experiments with *ISG15*, an antiviral factor that exhibits a key cysteine deletion in rhinolophid and hipposiderid bats, revealed fundamental differences between human and bats in the extracellular pro-inflammatory role of ISG15. Together, our study provides insights into the genomic underpinnings of bats' resistance to viral diseases, and implicates functional changes in bat ISG15 in limited infection-induced inflammation.

Results

Ten new reference-quality chromosome-level genomes of bats

We used long-read and long-range sequencing technologies to generate highly contiguous and complete reference genome assemblies that meet the standards of the Bat1K project[38]. We used samples from museum collections as a source of genetic material for eight of the 10 bat species (Supplementary Table 2), highlighting the importance of museum collections for biodiversity genomics [39]. Nine of these 10 species were sequenced to 27-42X coverage of PacBio circular consensus (HiFi) reads, providing long and accurate reads for contig assembly, and ~60X coverage of chromosome conformation capture (Hi-C) Illumina read data for scaffolding. The *Megaderma spasma* genome –generated from a 25-year old tissue sample– was assembled from Oxford Nanopore long reads (81X coverage), Bionano

optical maps, and Hi-C read pairs, and short read (10X Genomics Illumina) data was used to correct base errors.

All ten assemblies far exceed the minimum Bat1K standards [38]. Five of the ten new genomes showed improved contiguity over the best bat assemblies available to date, with contig N50 values ranging from 12.5-72.2 Mb (Fig. 1C, Supplementary Figures 1-2). At least 90% of all assemblies are contained in contiguous sequences that span several megabases, as indicated by the contig N90 values ranging from 3.8-32.6 Mb. Furthermore, 91.8-99.7% of our assemblies are contained in chromosome-level scaffolds, with scaffold N90 values ranging from 45.4-137.8 Mb (Fig. 1D, Supplementary Figure 3). Consistent with a previous HiFi-based bat assembly [40], we estimate a very high base accuracy (QV=61.8-69.7, indicating <1 error per megabase) for our nine HiFi-based assemblies. These base accuracy values are two orders of magnitude higher compared with previous PacBio CLR or Nanopore-based assemblies [28,29]. We then compared the status of 18,430 ancestral placental mammal coding genes per assembly, as inferred by TOGA (Tool to infer Orthologs from Genome Alignments), a method that integrates comparative gene annotation, inferring orthologous genes, and gene classification[41]. Compared to short-read assemblies, new and previous long-read based assemblies consistently exhibit more genes that have intact reading frames and lack missing sequences caused by assembly incompleteness or fragmentation (Fig. 1E, Supplementary Figure 4, Supplementary Table 3). This supports a high assembly completeness and quality, which is also a prerequisite for a comprehensive annotation of transposable elements. Compared to typical mammals, the newly sequenced bat genomes exhibited an accumulation of recent DNA transposon insertions (Supplementary Figure 5), similar to observations for other bats [42].

For comparative analysis, we placed the new assemblies in the context of the bat phylogeny. Using both multi-species coalescence and concatenated alignments of 16,860 1:1 orthologous genes (representing 30,354,372 bp), we consistently inferred the same tree topology (Fig. 1F), in agreement with previous phylogenies inferred from sparser data [43]. Support values are high for all nodes. Finally, we used a penalized likelihood method [44,45] and 17 fossil calibration points [46,47] to infer a time-calibrated tree (Fig. 1F, Supplementary Figure 6, Supplementary Table 4), which estimated the divergence of Rhinolophidae and Hipposideridae to be ~35 million years (My) old.

Selection in immune genes is most prevalent in bats

The ability of bats to limit disease upon viral infections is likely the result of a long history of coevolution between hosts and viruses that shaped immune system adaptations. Some of these adaptations may be detectable as signatures of episodic positive selection in genes. We thus devised a genome-wide screen to test how prevalent positive selection is among different orders of mammals and different functional groups of genes. To this end, we first used TOGA to obtain orthologous genes across the newly- and previously-sequenced bat genomes (N=20 species) and 95 additional non-chiropteran species that represent ten mammalian orders (Fig. 2A, Supplementary Table 5). When possible given the available genomes, we included up to

20 species per order, selecting assemblies with at least 16,000 intact orthologs (Supplementary Figure 7). We then used the sensitive branch-site model implemented in aBSREL [48] to test for positive selection on each branch in the 115-species tree. Instead of testing pre-defined hypotheses, this approach allows for an exploratory screen, as selection on a gene can occur on multiple individual branches and recurrent or convergent selection patterns may be detected. Considering 17,130 genes, we found 8,608 genes that show selection in at least one of the 228 branches in the 115-species tree (Supplementary Table 6).

For each of the ten included mammalian orders, we then determined functional enrichments of genes under selection. Considering all high-level biological processes, as defined by top-level Gene Ontology (GO) terms, we found that bats have the strongest enrichment for “immune system process”, followed by rodents (an order also known to harbor diverse viruses) and Afrotheria (Fig. 2A, Supplementary Figure 8, Supplementary Table 7). This pattern is not driven by unequal taxonomic representation or substantial genome quality differences, since four other mammalian orders (Primates, Rodentia, Cetartiodactyla, Carnivora) are also represented by 20 species and have a comparable genome assembly quality spectrum (Supplementary Figure 7). Furthermore, applying the selection screen to four subsamples, obtained by randomly selecting ten species from the 20-species orders Chiroptera, Primates, Rodentia, Cetartiodactyla, and Carnivora, robustly detects the strongest “immune system process” enrichment for bats (Supplementary Figure 9), validating that this result is not driven by a few individual bat species but representative for the order Chiroptera.

We observed that branch length, measured in My and substitutions per site in both neutral and coding regions, significantly correlates with the number of selected immune genes (Fig. 2B, Supplementary Figure 10). Consistent with previous simulations on few taxa [48], this likely reflects a higher incidence for episodic positive selection to occur over longer periods and increased power to detect it on longer branches. We therefore used a regression model to compute the expected number of selected “immune system process” genes for each branch (Supplementary Tables 8 and 9). Labeling branches by the difference between observed and expected selected immune genes highlights many bat lineages and species as outliers, indicating higher-than-expected immune-related selection in Chiroptera (Fig. 2C). Furthermore, fitting models with two intercepts, one for bats and another for the remaining mammals, shows that the intercept for bats is significantly higher, consistent with more immune gene selection in bats (Supplementary Figure 11). Notably, with 42 observed vs. 21 expected selected immune-related genes, the ancestral Chiroptera branch has a larger relative number of selected genes than ancestral branches of all other orders (e.g. 42 observed vs. 24 expected in Afrotheria, and 20 observed vs. 21 expected in Rodentia) (Fig. 2C, Supplementary Table 8). A similar pattern is observed for the individual GO term “immune response”, and the WikiPathways “network map of SARS-CoV-2” and “SARS-CoV-2 innate immunity evasion and cell-specific immune response” (Supplementary Figures 12-15). Together, this suggests that immune system changes originated early in the chiropteran lineage and coincided with the evolution of powered flight.

We next analyzed the child terms of "immune system process" across the 115 mammals, showing that, compared to other orders, genes under selection in Chiroptera are most enriched in "immune response" (which also contributes to the enrichment of the high-level process "response to stimulus" shown in Fig. 2C), "regulation of immune system process", "immune effector process" and "leukocyte activation" (Fig. 2D). These categories are also robustly and most strongly enriched in Chiroptera in our subsampling analysis (Supplementary Figure 16). More specific GO terms further highlight Chiropteran enrichments related to both the innate and adaptive immune systems (Fig. 2E).

Relevant immune-related changes in bats

In contrast to humans, for whom SARS-CoV-2, MERS-CoV and other coronaviruses can cause, in severe cases, hyperinflammation, respiratory insufficiency, and multi-organ failure [7], infections in rhinolophid/hipposiderid and other bats appear to be largely asymptomatic, a phenotype also observed when some bat species are experimentally infected with MERS-CoV or SARS-CoV-2 [13,17,49]. To gain insights into genes that are likely involved in viral tolerance in bats, we intersected knowledge about genes involved in immune responses to corona- and other viruses with positively selected genes, focusing on selection in the ancestral Chiroptera branch (C), the common stem branch of Rhinolophidae and Hipposideridae, and the Rhinolophidae stem branch (R). This revealed selected genes linked to viral entry and detection, regulation of inflammation and antiviral mechanisms, activation of the complement system, and B cell signaling (Fig. 3A, Supplementary Figure 17).

To enter host cells, (corona)viruses use cell-surface receptors. These receptors are often subject to evolutionary arms races between the host and the virus. Similar to the SARS-CoV1/2 receptor *ACE2* [50–52], we identified selection on *ANPEP* (selected in R), which encodes a receptor used by human coronavirus 229E for entry [53,54]. Additionally, the cofactor *SCARB1* (selected in RH) that facilitates SARS-CoV-2 entry by enhancing cell-surface attachment [55] and the endosomal protease *CTSB* (selected in C and in R) that mediates entry of ebolaviruses and reoviruses [56,57] are under selection in bats (Figure 3B).

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 the release of pro-inflammatory cytokines. The inflammatory process is crucial to control pathogens and then restore tissue homeostasis, but needs to be tightly regulated to limit nonspecific immune-mediated tissue damage, as exemplified in severe COVID-19 where hyperinflammation can cause lung tissue damage [7]. We found several genes under selection in bats that are involved in detecting pathogen-associated molecular patterns and regulating inflammatory immune responses. *TLR8* (repeatedly selected in RH and in R) induces pro-inflammatory cytokine production after detecting single-stranded RNA of endocytosed viruses such as SARS-CoV-2 [58]. *TRIM38* (selected in R) encodes an interferon-induced enzyme with E3 ubiquitin and SUMO ligase activities that has multiple roles in immunity. During early infection, TRIM38 enhances innate immune responses to RNA and DNA viruses by SUMOylating and thereby preventing degradation of the viral RNA sensors RIG-I and MDA5

and the viral DNA sensors cGAS and STING [59,60]. During late infection, *TRIM38* is upregulated by interferons and suppresses inflammatory responses in several ways. By promoting the degradation of the TLR3/4 adapter protein TRIF, TRIM38 inhibits canonical NF-κB and IRF3 activation [61]. By promoting the degradation of TAB2/3 in the TNF-α and IL-1β signaling pathway, TRIM38 further inhibits NF-κB activation and proinflammatory cytokine production [61,62]. Thus, *TRIM38* mediates a strong early innate immune response and contributes to dampening inflammation at later stages, processes that are intensified in bats. *BTK* (selected in C) encodes an intracellular tyrosine kinase that has several roles in adaptive (below) and innate immunity. By interacting with several components of TLR signaling pathways, such as TLR4/6/8/9, the adapter protein MyD88 and the kinase IRAK1, BTK contributes to TLR-induced production of the anti-inflammatory cytokine IL-10 [63,64]. By interacting with NLRP3 inflammasome components, BTK is essential for inflammasome activation [65]. In severe COVID-19 patients, BTK inhibitors can reduce hyperinflammation [66]. *TNFAIP2* (selected in RH) is a TNFα-inducible gene that is involved in negative feedback regulation of NF-κB signaling [67]. HP (haptoglobin; repeatedly selected in C and RH), an acute-phase protein with immunomodulatory functions, inhibits T cell proliferation and the secretion of proinflammatory cytokines (interleukins, TNF-α) from various immune cell types [68,69]. Furthermore, HP can directly bind to TLR4 and activates TLR4 signaling, which stimulates the secretion of IFN-β (encoded by *IFNB1*, selected in C) [70]. Finally, we found that *IL36A*, a pro-inflammatory interleukin of the IL-1 superfamily that stimulates NF-κB signaling [71], was lost in the common ancestor of Rhinolophidae, Hipposideridae, Megadermatidae, and Rhinopomatidae (Supplementary Figure 18A). Another pro-inflammatory IL36 family member, *IL36G*, was lost in the rhinolophid ancestor, but this gene retains an intact reading frame in all investigated hipposiderid bats (Supplementary Figure 18B-D).

The type I interferon (IFN-I) response counteracts viral infections prior to activation of the adaptive immune system. While a delayed or ineffective IFN response is linked to severe COVID-19 [72], IFNs need to be downregulated after infection as sustained IFN production leads to immunopathology and is a hallmark of autoimmune diseases (Figure 3D) [73]. We found selection not only on *IFNB1*, but also several other genes that regulate or are regulated by IFN signaling. IFNB1 suppresses the secretion of IL17A (selected in C), a potent pro-inflammatory cytokine that is involved in triggering cytokine storms in severe COVID-19, by inhibiting the differentiation of IL17-producing T helper cells [74]. Interestingly, IL17A is mimicked by the SARS-CoV-2 encoded secreted glycoprotein ORF8, which can also bind to the IL17 receptor to induce proinflammatory factors [75]. IFNB1 and other type I interferons activate the JAK-STAT cascade that induces interferon-stimulated genes such as *IFIT2* (also called ISG54), *IFIT3* (ISG60), and *ISG15*. *IFIT2* (selected in C) and *IFIT3* (selected in R) encode cytoplasmic proteins that restrict replication of corona- and many other viruses by sensing viral mRNAs and inhibiting their translation [76–79]. ISG15 is an antiviral protein that lacks a key Cysteine residue in rhinolophid and hipposiderid bats which is required for homodimer formation (investigated in detail below). Other genes related to fine-tuning of IFN regulation include *LRRC25* (under selection in C), which participates in a negative feedback loop to avoid prolonged immune

activation by promoting autophagic degradation of the cytosolic dsRNA sensor RIG-I. Specifically, after viral infection, LRRC25 binds to RIG-I in an ISG15-dependent manner, promotes RIG-I degradation, and thus negatively regulates RIG-I-mediated expression of IFN-β and IFN-I induced genes such as *IFIT1/2* [80]. In a similar manner, when located intracellularly, NMI (N-myc and STAT interactor; under selection in RH) inhibits the expression of IFN-I and IFN stimulated genes such as *IFIT1/3* and *ISG15* by promoting the proteasomal degradation of IRF7, a transcription factor that is important for IFN-I induction during the late infection phase [81,82]. In contrast, extracellular NMI, released by activated macrophages, induces the release of proinflammatory cytokines such as TNFs and IL-6 by binding to TLR4 and activating canonical NF-κB [83]. Thus, depending on the location, NMI has anti- or proinflammatory roles.

Inflammation triggers the cellular release of chemokines that direct the migration of leukocytes to sites of infection. We identified selection in two CC chemokine receptor genes, *CCR2* (repeatedly selected in C, RH and R) and *CCR5* (selected in RH). CCR2 is a major receptor that promotes the infiltration of proinflammatory cells such as Ly6C-high monocytes to the lung during viral or bacterial infections [84,85]. Similarly, CCR5 is expressed on subsets of macrophages, dendritic, natural killer and T cells and directs them to virus-infected tissues [86,87] (Figure 3E). Importantly, *CCR2* and *CCR5* are located in a major risk locus for severe COVID-19 that contains several SNPs linked to increased *CCR2/5* expression in monocytes and macrophages [88,89]. Consistent with CCR2/5 mediated hyperinflammation, CCR2/CCR5 receptor antagonists can reduce cytokine storms in patients and are investigated as a treatment for severe COVID-19 [90,91].

The complement system helps to phagocytose or lyse pathogens, stimulates adaptive immune responses and promotes inflammation [92]. However, excessive complement activation during COVID-19 can lead to hyperinflammation and thrombosis [93]. We detected two complement components, *C7* and *C1S*, that are both under selection in RH. C1S encodes a serine protease that is involved in early activation cascade of the classical pathway [94]. 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 [95] (Figure 3F).

Finally, two selected genes, *CD79A* (selected in RH) and the above mentioned tyrosine kinase *BTK*, are key factors for B cell signaling. CD79A is the signal transduction subunit of the B cell antigen receptor that upon phosphorylation mediates the phosphorylation of BTK [96,97]. Activated BTK promotes B cell receptor mediated survival of B cells by signaling through Akt, NF-κB and other signaling pathways [98] (Figure 3E).

In summary, multiple genes with functions involving viral host cell entry, innate immune response regulation, complement activation, and B-cell survival are under selection in bats and provide promising target genes for future research.

ISG15 of rhinolophid and hipposiderid bats vary in antiviral activity

ISG15 is an antiviral, ubiquitin-like protein that is strongly induced by IFN and plays an important role in hyperinflammation during COVID-19. ISG15 can be conjugated to hundreds

of newly-synthesized host and viral proteins (a process known as ISGylation) and this process is antagonized by viral immune-evasion proteins [99]. For example, ISGylation is required for IRF3- and MDA5-mediated antiviral responses, and the SARS-CoV-2 encoded papain-like protease (PLpro) suppresses these responses by de-ISGylating IRF3 and MDA [100–102]. Interestingly, *HERC5*, encoding the protein ligase that mediates ISG15 conjugation [103], is positively selected in the ancestor of *R. trifoliatus* and *R. lanosus* and in *R. sinicus* (Supplementary Table 6). In addition to intracellular conjugation-dependent roles, free ISG15 can be secreted into the extracellular space, where it functions as a cytokine that stimulates the secretion of proinflammatory cytokines and chemokines [99]. The de-ISGylating activity of SARS-CoV-2 PLpro increases the pool of free ISG15, resulting in enhanced ISG15 secretion and increased production of proinflammatory factors, which is consistent with the immunopathology in COVID-19 patients [104].

We found that a highly conserved cysteine residue in ISG15 (Cys78 in human ISG15) is deleted in all rhinolophid and hipposiderid bats (Fig. 4A). Importantly, this Cys residue is required for the formation of stable ISG15 homodimers and its extracellular cytokine function [105–107]. Furthermore, mutating Cys78 enhances ISGylation, likely because dimerized ISG15 is not usable for ISGylation [106]. To confirm that the Cys78 deletion also prevents the formation of stable homodimers of the bat ISG15, we performed structural modeling. To this end, we used AlphaFold2 [108] to infer the 3D structures of the putative homodimers of *Rhinolophus sinicus, Doryrhina cyclops* and, for comparison, human ISG15, and conducted molecular dynamics simulations for a total duration of 3 µs (about 550,000 CPU hours) per ISG15. While the Cys78-containing human ISG15 dimer was indeed stable over the course of the simulations, the Cys78-lacking dimer of both bats appeared more unstable as it adopted a range of topologically distinct conformations that strongly deviated from the initial AlphaFold2 structure (Supplementary Figures 19-22). Given the relevance of Cys78 for ISG15 function and the role of ISG15 in hyperinflammation during COVID-19, we investigated functional differences between bat and human ISG15.

To explore whether bat ISG15's have altered antiviral capacity against viruses and in particular coronaviruses, we synthesized ISG15 from six Rhinolophidae (*Rhinolophus affinis, R. lanosus, R. yonghoiseni, R. sinicus, R. trifoliatus, R. ferrumequinum*), three Hipposideridae bats (*Aselliscus stoliczkanus*, *H. larvatus*, *Doryrhina cyclops*) and human and compared their antiviral function against four different viruses (Vesicular Stomatitis virus, Influenza A virus, human coronavirus 229E, and SARS-CoV-2). First, we transiently transfected HEK293 (human immortalized embryonic kidney) cells with the various ISG15 constructs and infected them with GFP-tagged Vesicular Stomatitis virus (VSV), a representative of the Rhabdoviridae family that is common amongst bats. FACS analysis revealed a clear antiviral restriction capacity compared to the vector control for all ISG15 proteins, except *R. trifoliatus* ISG15 (Fig. 4B, Supplementary Table 10). Notably, several bat species (*R. lanosus*, *R. yonghoiseni* and *R. trifoliatus*, all belonging to the same clade) were significantly less efficient at preventing VSV infection compared to human ISG15, revealing species-specific variation in bat ISG15 function. Viral load, measured by GFP-intensity in infected HEK293 cells, was reduced for all species

compared to the vector control, but exhibited no significant variation between the tested species (Supplementary Figure 23, Supplementary Table 11).

Interestingly, we noticed that uninfected ISG15-expressing cells sometimes appeared to grow faster. To test this, we measured cell growth and ATP turnover in uninfected HEK293 cells that were stably transfected (pools) with ISG15 of human or one of the nine bats (Supplementary Figures 24-25, Supplementary Tables 12-13). Indeed, in contrast to human ISG15 or the vector control, uninfected HEK293 cells showed a significantly increased growth for six of the nine bat species. This indicates that ISG15 of some bats may have functions beyond antiviral defense, which may be relevant as some bats exhibit constitutive ISG15 expression levels [109,110].

Next, we examined the antiviral function of ISG15 against Influenza A virus (IAV), a member of the *Orthomyxoviridae* family that is commonly identified in bats. We stably transduced A549 (adenocarcinomic human alveolar basal epithelial) cells with the ISG15 constructs, infected each line directly with IAV (H1N1/PR8 strain) and performed plaque assays (Fig. 4C, Supplementary Figure 26, Supplementary Table 14). While human ISG15 has antiviral efficacy against IAV, confirming previous results [111], it was significantly less antiviral than ISG15 of five of the six rhinolophid bats. Only *R. sinicus* ISG15 did not have obvious antiviral activity. Among the three hipposiderid bats, only *A. stoliczkanus* ISG15 had antiviral capacity. Compared to human ISG15, IAV-infected A549 cells expressing ISG15 of bats generally had higher levels of MX1, a known IAV antiviral restriction factor, whereas MX1 was not detectable in infected cells in the absence of ISG15 (Supplementary Figure 27).

To compare activity of ISG15 against coronaviruses, we generated HEK293 cells stably-expressing the human coronavirus (HCoV) 229E receptor ANPEP (CD13) and bat/human ISG15 constructs, and infected them with HCoV-229E. While human ISG15 had a clear antiviral effect, for the bats, only *R. ferrumequinum* ISG15 was significantly antiviral (Fig. 4D, Supplementary Table 15). Consistently, viral load, measured by viral N protein staining, was also substantially reduced by human and *R. ferrumequinum* ISG15 (Supplementary Figure 28, Supplementary Table 16). In contrast, *R. sinicus* ISG15 increased the amount of intracellular N protein in infected cells, suggesting increased viral replication. Whether this correlates with an increased release of infectious particles remains to be determined.

Finally, we tested the effect of ISG15 on SARS-CoV-2 infection. A549 cells that stably express the SARS-CoV-2 receptor ACE2 were transfected with bat/human ISG15 constructs and infected with SARS-CoV-2. While human ISG15 failed to decrease SARS-CoV-2 production compared to the vector control, ISG15 of five of nine bats significantly reduced viral release, as measured by $TCID_{50}$ assays (Fig. 4E, Supplementary Table 17).

In comparison to human ISG15 that has the Cys78 residue, ISG15 of rhinolophid and hipposiderid bats lacking Cys78 did not reveal a consistent antiviral difference in our experiments. Therefore, we directly investigated the cysteine by mutating Cys78 in human ISG15 and restoring it in *R. affinis* ISG15. For HCoV-229E, mutating Cys78 or restoring the cysteine resulted in a small but significant antiviral difference compared to the wild-type ISG15 (Fig. 4F, Supplementary Table 15). Furthermore, mutating Cys78 in human ISG15 conferred

a considerable and significant ability to reduce SARS-CoV-2 viral production (Fig. 4G). Restoring the cysteine in *R. affinis* ISG15 maintained a significant antiviral activity against SARS-CoV-2 (Fig. 4G, Supplementary Table 17).

In summary, our experiments show that ISG15 of Rhinolophidae and Hipposideridae largely maintains its antiviral effector function, but there are species- and virus-specific differences not explained by the shared Cys78 deletion, indicating that additional mutations in individual bats (Fig. 4A) also affect ISG15's antiviral activity. Further investigation of species-specific effects of ISG15 may provide additional insight into human ISG15 antiviral capacity and function.

ISG15 of rhinolophid and hipposiderid bats remains intracellular and shows increased ISGylation

Cys78 is required for homodimer formation and the extracellular cytokine function of ISG15 [106] Therefore, we measured ISG15 in the supernatant of HCoV-229E-infected cells to compare the level of free ISG15 that is present in the extracellular space. In contrast to human ISG15, which was readily detectable extracellularly, extracellular ISG15 of all tested rhinolophid and hipposiderid bats was not or barely detectable (Fig. 4H, Supplementary Figures 29-34, Supplementary Table 18). Inside the cells, free ISG15 of all tested bats was reduced and ISG15 conjugated to proteins was increased in comparison to human ISG15. Unexpectedly, mutating Cys78 in human ISG15 did not drastically reduce secretion into the extracellular space (Fig. 4H). However, mutating human Cys78, particularly to serine, decreased free ISG15 and increased ISG15 conjugation intracellularly (Fig. 4H). Restoring Cys78 in *R. affinis* ISG15 increased secretion into the extracellular space. Together, this suggests ISG15's extracellular pro-inflammatory function, which is reliant on Cys78, is minimized in rhinolophid and hipposiderid bats.

Discussion

To shed light on the genomic basis of viral disease resistance in bats, we sequenced ten new reference-quality genomes, putting a focus on rhinolophid and hipposiderid bats that harbor several zoonotic coronaviruses. We found that bat assemblies based on PacBio HiFi reads often have superior contiguity and base accuracy, which corroborates results from the Vertebrate Genome Project, Darwin Tree of Life and Earth BioGenome Project [112–114] and supports that the combination of "HiFi+HiC" data is a powerful strategy to generate reference-quality genomes of bats.

Our exploratory selection screen revealed that many mammalian orders show significant enrichments for selection in immune-related genes, which is consistent with selection pressure exerted by pathogens driving a rapid evolution of immune genes [115]. However, compared to other mammalian orders, bats exhibit the strongest enrichments for immune gene selection, providing genomic evidence that bats possess unique immune system

adaptations. Using our comprehensive dataset comprising 115 mammalian species and 228 phylogenetic branches, we observed that branch length is significantly correlated with the number of selected genes. A regression model to estimate the expected number of selected genes showed that the ancestral Chiroptera branch exhibits more immune gene selection than expected. This supports that the evolution of powered flight and immune system adaptations are linked [22,116,117]; however, it remains to be determined whether flight is directly or indirectly linked to immune system changes. While we applied the regression model specifically to immune 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.

Immune genes that are positively selected in ancestral Chiroptera or the Rhinolophidae/Hipposideridae lineages are involved in viral entry, virus detection, and antiviral responses by the innate, adaptive, and complement systems. Furthermore, several selected genes regulate inflammatory responses by inhibiting the production of pro-inflammatory cytokines and participating in negative feedback control of interferon signaling, indicating that these genes may contribute to preventing uncontrolled inflammation during viral infection in bats. Together, our findings provide promising target genes for experimental exploration and a guide to unlock the secrets behind the adaptations that make bats' immune system unique.

ISG15 is one of these target genes and it exhibits a key difference (Cys78 deletion) shared among rhinolophid and hipposiderid bats. Our experiments revealed virus- and species-specific differences in antiviral efficacy between human and bat ISG15 as well as between ISG15 of different bats, which could potentially be due to differences in the repertoire of ISGylation target proteins. Although the Cys78 deletion is not associated with a consistent viral restriction pattern of Rhinolophidae/Hipposideridae ISG15, indicating that mutations at other residues also modulate ISG15 function, it is noteworthy that mutating Cys78 confers human ISG15 the ability to restrict SARS-CoV-2. Consistent with Cys78 being required for homodimer formation and ISG15's role as an extracellular cytokine [105–107], we found that human ISG15 is secreted into the extracellular space. In contrast, bat ISG15 is not secreted but showed enhanced intracellular ISGylation. Furthermore, secretion and ISGylation can be partially altered by mutating Cys78 in human ISG15 or restoring it in the *R. affinis* protein. Although further experiments are required, these findings suggest that Cys78 deletion (in conjunction with other mutations in bats) may have two effects on ISG15 function. First, by preventing dimer formation and leading to enhanced ISGylation, Cys78 deletion may counteract the de-ISGylating activity of viral evasion proteins such as SARS-CoV-2 PLpro, which could contribute to maintaining ISG15's intracellular antiviral activity. Second, by preventing secretion, Cys78 deletion may reduce ISG15's extracellular pro-inflammatory function. Thus, ISG15 could be one of the factors that contribute to the ability of rhinolophid and hipposiderid bats to launch effective antiviral responses without triggering excessive inflammation.

To fully elucidate the history of host-viral 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 [38]. Reference genomes, single cell transcriptomics maps of their

514 immune systems, breeding colonies, bat cell lines, organoids and the ability to generate
515 induced pluripotent stem cells [118–121] provide new tools to elucidate the molecular adaptations
516 that enable viral tolerance with asymptomatic infections. Together, this makes bats an
517 emergent model system for comparative mammalian biology, offering insights not only into
518 special immune system adaptations, but also healthy aging, enhanced disease resistance, and
519 other remarkable chiropteran traits.
520
521

Materials and Methods

Ethical statements and samples and collecting permits

Ten tissue samples were acquired from the Royal Ontario Museum mammal collection or field expeditions for the species *Aselliscus stoliczkanus* (China, Shuipu Village), *Doryrhina cyclops* (Ivory Coast – Parc National de TaÏ), *Hipposideros larvatus* (China, Shuipu Village), *Rhinolophus affinis* (China, Shiwandashan National Reserve), *Rhinolophus lanosus* (China, Shuipu Village), *Rhinolophus yonghoiseni* (Malaysia, Endau Rompin National Park), *Rhinolophus trifoliatus* (Malaysia, Endau Rompin National Park), *Rhinopoma microphyllum* (Northern Israel), *Megaderma spasma* (Vietnam, Cat Tien National Park), and *Mops condylurus* (Côte d'Ivoire, Bregbo Village) (Supplementary Table 2). Samples were flash frozen in liquid nitrogen after being collected and stored at -80C 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, 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 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 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 v1.0 11/19) following the manufacturer's instructions (https://www.circulomics.com/nanobind). In brief, 25-40 mg of liver, spleen, or heart tissue were minced to small slices on a clean and cold surface. Tissues were homogenized with the TissueRuptor II device (Qiagen) making use of its maximal settings. After complete tissue lysis, remaining cell debris was removed, and the gDNA was bound to Circulomics Nanobind discs in the presence of isopropanol. High molecular weight (HMW)

gDNA was eluted from the nanobind discs in elution buffer (EB). The integrity of the HMW gDNA was determined by pulse field gel electrophoresis using the Pippin PulseTM device (SAGE Science). The majority of the gDNA was between 10 and 500 kb in length. All pipetting steps of ultra-long and long gDNA were done carefully with wide-bore pipette tips.

PacBio 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, version 03) for *Aselliscus stoliczkanus*, *Hipposideros cyclops*, *Hipposideros larvatus*, *Mops condylurus*, *Rhinolophus affinis*, *Rhinolophus lanosus*, *Rhinolophus yonghoiseni*, *Rhinolophus trifoliatus*, and *Rhinopoma microphyllum*. Briefly, HMW gDNA was sheared to 20 kb fragments with the MegaRuptor™ device (Diagenode) and 10µg sheared gDNA was used for library preparation. The PacBio SMRTbell™ library was size-selected for fragments between 9 to 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 hours on the SEQUEL II. Circular consensus sequences were called, making use of the default SMRTLink tools. For each species, a total of 65 to 92 Gb of HiFi reads were generated, representing between 27X and 42X effective genome coverage.

Since PacBio HiFi sequencing for *Megaderma spasma* produced very little output despite a good DNA and library quality, we used Oxford Nanopore Technologies (ONT) for this species. Two Oxford Nanopore 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), making use of the Diagenode MegaRuptor device as described for PacBio HiFi sequencing. After repair of the sheared and unsheared gDNA, ONT sequencing adapters were ligated to the gDNA fragments and the resulting libraries were enriched for fragments larger than 3 kb in size. Both libraries were loaded on a Promethion device using R9.4.1 flow cells, generating 173 Gb of reads representing 81X effective genome coverage.

ARIMA HiC

Chromatin conformation capture was done by making use of the ARIMA-HiC (Material Nr. A510008) and the HiC+ Kit (Material Nr. A410110) and following the user guide for animal tissues (ARIMA-HiC kit, Document A160132 v01 and ARIMA-HiC 2.0 kit Document Nr: A160162 v00). In brief, ~50 mg of flash-frozen powdered tissue was crosslinked chemically. The crosslinked gDNA was digested with a restriction enzyme cocktail consisting of two and four restriction enzymes, respectively. The 5'-overhangs were filled in and labeled with biotin. Spatially proximal digested DNA ends were ligated. The ligated biotin-containing fragments were enriched and went for Illumina library preparation, which followed the ARIMA user guide for Library preparation using the Kapa Hyper Prep kit (ARIMA Document Part Number A160139 v00). 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.

10x Genomics linked reads

To scaffold and correct base errors in the *Megaderma spasma* contig assembly, we generated linked Illumina reads with the 10x Genomics ChromiumTM genome application, following the Genome Reagent Kit Protocol v2 (Document CG00043, Rev B, 10x Genomics, Pleasonton, CA). In brief, 1 ng of long or megabase-size gDNA was partitioned across 1 million gel bead-in-emulsions (GEMS) using the ChromiumTM device. Individual gDNA molecules were amplified in these individual GEMS in an isothermal incubation using primers that contain a specific 16 bp 10x barcode and the Illumina® R1 sequence. After breaking the emulsions, pooled amplified barcoded fragments were purified, enriched, and went into Illumina sequencing library preparation as described in the protocol. Pooled Illumina libraries were sequenced to ~40X genome coverage on an S4 flow cell of a NovaSeq6000 with 300 cycles.

Genome assembly

Contig assembly – PacBio 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*, *Rhinolophus lanosus*, *Rhinolophus yonghoiseni* and *Rhinolophus trifoliatus*), we created contig assemblies using hifiasm v.0.13 [122] with the argument -l0. The primary assembly was created by using purge_dups v.1.2.3 [123] 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 v0.15.5-r352 with purging argument l2. For *Mops condylurus*, we used hifiasm v0.15.4-r432. For both assemblies, we created the primary and alt contigs sets using purge-dups v1.2.3 as above.
For *Doryrhina cyclops*, hifiasm created a large number of mis-assemblies joining regions from distinct chromosomes, which could not be reasonably corrected by hand. Therefore, we ran HiCanu v.2.1 [124] to create the initial contigs. Since this resulted in an assembly two times the expected size of the genome, we ran purge_dups on the contig assembly using custom cutoffs based on a haploid coverage of 13X: 8, 1, 1, 20, 2, 60 as in [124]. 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 v2.2 in -nanopore mode and created the primary contig sets using purge-dups as above.

Scaffolding of Megaderma spasma – We first scaffolded the contigs created by ONT reads using the 10X Genomics data. To this end, we mapped the 10X Genomics reads using longranger v2.2.2 and scaffolded using Scaff10X v4.2 and Break10X v3.1. Next, we used

Bionano optical maps to further scaffold the assembly after 10X scaffolding. We created an optical map de-novo assembly and then created the scaffold using Bionano Hybrid Scaffold using tools within Bionano Solve v1.6.1. 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 v0.7.17-r1188 [125] and filtered reads based on 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 v2.2 [126] with arguments: -m yes -p yes.

Manual curation – To join those contigs missed by salsa2 and break those joins which were spuriously created, we manually curated the scaffolds. 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, where 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 re-scaffolded.

Polishing assemblies – To polish the final HiFi-based genomes and remove unambiguous heterozygous sites, 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 (Poplin et al., 2018). We then filtered for sites with genotype 1/1 and a 'PASS' filter value, meaning that all or nearly all reads support an alternative sequence at this position and passed DeepVariant's internal filters. With this method, we do not polish any heterozygous or polymorphic regions of the genome, but only those that are incorrect and not supported by any CCS reads. We then corrected base errors using bcftools consensus v.1.12 [127].

For *Megaderma spasma,* we first mapped the 10X Genomics linked-reads to the assembly using Longranger v2.2.2. We then called variants using DeepVariant v1.2.0, filtered the vcf file using Merfin v1.1-development r197 [128] and determined the consensus using bcftools consensus v1.12. We performed two rounds of polishing.

Annotation of Transposable Elements

To annotate transposable elements (TEs) in the newly-sequenced bats, first we 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 [129].

Briefly, 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 RepeatClassifier identified it as being a LINE1 element. LTR elements were further processed by hand to subdivide them into their LTR and Internal segments, per [130]. 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 (i.e. Terminal Inverted Repeats (TIRs), Long Terminal Repeats (LTRs), etc.), sequence characteristics (repetitive tails, Helitron-specific CTAG motifs, SINE A-B boxes, etc.). 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 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 of Wicker et al. using USEARCH [131] 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 [42]. All novel TE consensus sequences have been submitted to the Dfam TE database [132]. 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 miRNAs

Annotation of miRNA genes in the newly-sequenced bats was performed similarly to reference [29]. Briefly, prior to miRNA prediction, repetitive and low-complexity regions in each bat genome were masked with the Dfam database (v3.5) (https://www.dfam.org/releases/Dfam_3.5/) using RepeatMasker (v4.0.6, http://www.repeatmasker.org). For each masked genome, conserved miRNA genes were predicted using the Rfam database (v14)[133] and Infernal (v1.1.2)[134]. 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 based on their secondary structure that we predicted with RNAfold (v2.4.18)[135].

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 [136] to obtain local alignments. We used LASTZ parameters (K = 2400, L = 3000, Y = 9400, H = 2000, and the LASTZ default scoring matrix) that have a sufficiently high sensitivity to align orthologous exons between placental mammals [137]. Local alignments were chained using axtChain [138] with default parameters except linearGap=loose. We used RepeatFiller [139] (default parameters) to add missed repeat-overlapping local alignments to the alignment chains and chainCleaner [140] (default parameters except minBrokenChainScore = 75,000 and -doPairs) to improve alignment specificity.

Inferring and annotating orthologous genes

To assess gene completeness in assemblies and infer orthologs for phylogenomic, selection and gene loss analyses, we used TOGA [41] (https://github.com/hillerlab/TOGA, commit v.c4bce48). Briefly, 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 largely relies 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.

To compare assembly completeness and base accuracy, we considered a set of 18,430 genes that likely 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 [41]. For each assembly, we determined how many ancestral genes have (i) an intact reading frame (TOGA classification intact, stating that the middle 80% of the coding sequence is present and lacks gene-inactivating mutations), (ii) inactivating mutations (TOGA classifications loss and uncertain loss), or (iii) missing sequence due to assembly gaps or fragmentation (TOGA classifications partially intact and missing). An excess of genes with missing sequences

indicates a lower assembly completeness and an excess of genes with inactivating mutations indicates a lower base accuracy.

Exon-by-exon alignments of orthologous genes

For phylogenomics and genome-wide selection screens, we used orthologs 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 only included 1:1 orthologs and excluded species for which no or multiple co-orthologs were inferred. Codons having 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 [141], and all exons, together with codons split by introns, were concatenated into a multiple codon alignment. Codon alignments were cleaned with HmmCleaner [142] 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 (~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 bats into a phylogeny, we reconstructed phylogenetic relationships using whole gene codon alignments, considering in total 50 bat species 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 coalescent-based approach as implemented in ASTRAL v.5.5.9 [143,144], and a concatenated approach as implemented in IQTREE [145]. For the ASTRAL analysis, input trees were estimated in RAxML v. 8.1.16 [146]. Each gene was analyzed with 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 node support. Branch support values were estimated using a transfer bootstrap expectation implemented in BOOSTER [147]. 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 [148]. A maximum likelihood tree was inferred using IQTREE, with nodal support calculated using 1000 bootstrap pseudo-replicates.

To estimate a time-calibrated tree, we followed a penalized likelihood approach as implemented in treePL [45]. 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 [46] were applied to constrain maximum divergence times at relevant nodes (Supplementary Table 4). The time-calibrated phylogenies of bats and mammals are available on 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 only selected assemblies for which at least 16,000 ancestral placental mammal genes have an intact reading frame, as determined by TOGA (detailed below) (Supplementary Figure 7). For Chiroptera, we included the 10 new and ten previously published bat assemblies. 18 of these 20 bats were assembled from long sequencing reads [28,29,40] while the remaining two genomes were assembled from Illumina short-read data [149,150] (Supplementary Table 3). For five mammalian orders (Primates, Rodentia, Cetartiodactyla, Carnivora, Chiroptera), we selected exactly 20 species. The other mammalian orders are represented by fewer species, as 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 [48], an adaptive branch-site random effects likelihood method implemented in HYPHY [151]. aBSREL was run in exploratory mode to test all branches and nodes within the phylogenetic tree. For each gene, multiple test corrections over all tested branches were applied 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 were inspected by eye to rule-out spurious signals due to misalignments.

To test whether enrichment results of genes under selection (see below) are representative for mammalian orders or driven by individual species, we performed a subsampling analysis. We ran four additional selection screens using the same dataset of 17,130 genes, but subsampled the five groups having 20 species (Chiroptera, Carnivora, Cetartiodactyla, Rodentia, and Primates) by randomly selecting only ten species. Groups 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 ten species in the five 20-species orders that were left out in 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 if genes under selection in different mammalian orders are enriched in specific functional groups, we performed gene enrichment analyses as implemented in gProfiler [152,153] (last access on May 10, 2022) using all annotated human genes as the background. 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 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/).

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. For branch lengths, we used three independent estimations: (i) millions of years from our time-calibrated phylogeny inferred using treePL [45] and fossil calibrations (Supplementary Table 4), (ii) number of substitutions per neutral site estimated from 4D sites using phyloFit [154], and (iii) number of substitution per site estimated from coding regions using IQTREE [145]. Normal probability plots suggest heavy tails (non-normality), which could be attributed to the unequal error variance of branch length distribution. We then explored if 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 $\lambda \sim 0.05$; therefore, we 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 [155]. 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-squared=0.4917, F-statistic = 220.6, 1 and 226 d.f., $p < 2.2e\text{-}16$). A significant correlation between branch length and the number of selected genes was also found for time and number of substitutions per neutral site (Supplementary Figure 10). Using the best-fit model, we then considered specific immune gene sets and colored 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 analyzed 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, 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, one with one intercept and another with two intercepts, and two slopes for bats/non-bats. This series was repeated with different branch length estimates as a covariate. Based on prior analyses, branch length variables were square root transformed, with an

884 untransformed analysis included for comparison. Lastly, we compared the fit of a simpler
885 frequency distribution than the negative binomial. A Bayesian approach was adopted to run
886 these models, as a flexible way to both fit the model series and generate fit comparison
887 statistics. A negative binomial frequency distribution was used to model the number of immune
888 genes under selection, such that:
889 $$yi \sim negative\ binomial\ (\lambda = exp(l_i),\ pr)$$
890 where λ is the rate or mean of the Poisson distribution, exp is the inverse logarithmic
891 link function, and (1-*pr*)/*pr* defines a rate or shape parameter for the gamma distribution of a
892 mixture of Poisson distributions, which relaxes the expectation of equality of mean and
893 variance of the Poisson distribution. As a result, the negative binomial distribution is usually a
894 better fit to biological data [156]. With a linear model applied to *l*:
895 $$l = \beta 0 + \beta 1 X$$
896 where $\beta 0$ represents the intercept, which is global for analyses with a single intercept
897 or group-specific for testing bats vs. non bats, $\beta 1$ is the coefficient on branch length, and X
898 represents branch length. Both coefficients are normally distributed. To implement Bayesian
899 sampling for these analyses, we used brms [157], a package that enables coding models in R for
900 implementation in the stan statistical language [158]. For each model, we ran four separate
901 Markov chain Monte Carlo chains using a Hamiltonian Monte Carlo approach. Compared to
902 other Bayesian implementations, the Hamiltonian Monte Carlo approach saves time in
903 sampling parameter spaces by generating efficient transitions spanning the posterior based on
904 derivatives of the density function of the model. We estimated the R^2 of all models using the
905 procedure outlined by [159]. To compare model fits, we used WAIC (widely applicable information
906 criterion), which weighs log pointwise predictive density against the expected effective number
907 of parameters as defined by [160] and provides estimates of the standard error of the difference
908 between the best fit and other models.
909
910 **ISG15 3D structure modeling**
911 To explore the impact that bat-specific residues at key sites in ISG15 have on the
912 overall 3D structure relative to their analogous human sites, the protein structure for human
913 ISG15 was modeled using *in silico* methods. The protein structure was predicted using the
914 Iterative Threading ASSEmbly Refinement (I- TASSER) [161,162] server. Given a query amino
915 acid sequence, I-TASSER finds PDB reference templates showing similar secondary
916 structures using the Local Meta-Threading Server (LOMETS) [163]. This template then serves as
917 a basis for downstream homology modeling, with loop regions being inferred using *ab initio*
918 modeling. The model with the highest confidence score (C-score = 0.04) was used for all
919 subsequent analyses.
920 Nine key residue changes showing functional importance in human (A11S [164], S26A
921 [165], K35E [166], R99A [167], T103K [167]), bat-specific changes (T6K, S21N, A46I, V117T/M), or a
922 combination of both were analyzed to infer how bat/human variants affect the overall protein
923 stability relative to the human wild-type. This was done by calculating the predicted changes
924 in Gibbs free energy (folding energy, DDG) for both wild type and mutant variants in ISG15

using Dynamut [168] with the I-TASSER predicted 3D structure. By replacing human residues with bat-specific or human residues at that specific loci and calculating the predicted increases or decreases in DDG, we determined whether human residues at these key sites were stabilizing or destabilizing. A total of five mutations were stabilizing (T6K, S21N, K35E, A46I, T103K), suggesting that the variant amino acid is a more stable residue compared to the human wild-type residue. In contrast, four mutations were destabilizing, with the wild-type human amino acid representing a more stable residue compared to *Rhinolophus ferrumequinum* for A11S, S26A and V117T, and the R99A mutation shown to abolish ISG15 signaling via ITGAL. These results suggest that 5 human wild-type amino acids represent the most stable residues for human ISG15, while 4 bat-specific mutations would result in an overall increase in stability of protein structure. This overall increase in stability was further confirmed by calculating the net DDG (0.55) when looking at all mutations at once using DynaMut2 [169].

To test whether the Cys78 deletion in ISG15 of certain bats affects the formation of stable ISG15 homodimers, we used AlphaFold2 [108] through ColabFold [170] to infer the structure of the putative ISG15 homodimer of human, of the Chinese rufous horseshoe bat (*Rhinolophus sinicus*), and of the cyclops roundleaf bat (*Doryrhina cyclops*). Starting from each ISG15 sequence, ColabFold identified homologous sequences by running MMseqs2 [171] against the UniRef100 database [172] and against a set of environmental sequences [173]. Structural template information was obtained from the PDB70 database [174]. Next, the AlphaFold-multimer-v2 model [175] was employed 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 [176] and were ranked according to their pTM score, which we used to identify the best model.

To further investigate the stability of the dimers inferred with AlphaFold, we conducted three replicate molecular dynamics simulations per ISG15 using GROMACS v.2022.1 [177,178] and the CHARMM36-Jul2021 force field [179]. More precisely, we prepared each dimer by treating termini as ionized (i.e., NH_3^+ and COO^-), assigning appropriate protonation states to amino acids (assuming pH = 7) as determined using PROPKA3 [180,181], 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. The box was filled with TIP3P water molecules and with Na^+ and Cl^- ions as required for neutralizing the system. 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: (i) under an NVT ensemble to stabilize the temperature (at 300 K), and (ii) under an NPT ensemble to stabilize the pressure (at 1 bar) and the density of the system. For these, we used the velocity rescaling thermostat [182] and the Parrinello-Rahman barostat [183,184], 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. Finally, we removed the position restraints and conducted production simulations for 1 μs each, recording snapshots of the system every 100 ps. These simulations ran for ~6 months on a compute node with 128 cores, summing to a total of ~550,000 CPU hours for each species.

To analyze 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 (RMSD) using Carma v.2.01 [185]. We then clustered the three (human and bats) RMSD matrices according to the Partitioning Around Medoids algorithm [186] 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 [187]. Finally, we extracted the protein snapshots corresponding to the medoid of each cluster and compared them with the initial protein model obtained from AlphaFold (Supplementary Figures 19-21).

Experimental investigation of antiviral mechanisms of ISG15 in Rhinolophidae and Hipposideridae

Cell cultures:

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 (Gibco). For HCoV experiments, an ANPEP/CD13-Flag construct (Sino Biological) was transfected into HEK293 cells (PEI), then selected (mixed pool) with hygromycin for 2 weeks to generate stable cell lines, validated by surface CD13-staining (Sino Biological, 1:2000 dilution, no permeabilization), and used for consequent infection with HCoV. For SARS-CoV2 experiments, A549-ACE2 cells (human lung adenocarcinoma derived cells overexpressing human angiotensin-converting enzyme 2; ACE2) were provided by Dr. Colpitt's laboratory [188], and we used the clonal population A549-ACE2 B9. A549-ACE2 cells were maintained in Ham's F-12K (Kaighn's) medium supplemented with 10% FBS, 10 ug/ml 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), according to the manufacturer's instructions. All cell lines were tested for mycoplasma and were free of mycoplasma contamination.

Lentiviruses were generated using 3rd-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 (Tsingke) according to TOGA annotations for *ISG15* (transcript ENST00000649529, aligned to human for validation) and cloned into the pLVX-IRES-mCherry vector under the CMV promoter for direct transfection or lentivirus generation. Lentiviral supernatants were prepared in low-FBS DMEM supplemented with 1% NEAA (Phygene), sodium pyruvate (Gibco) and filtered through a 0.45-μm low-PES PVDF filter (Jet Biofil). Lentiviral transduction was performed with 100μl supernatant/well (6-well plate) of cells in 1% FBS with 4μg/ml Polybrene (Biosharp) for 4-6 hrs, media was replaced with 10% serum and 48-72 hrs later cells were sorted for mCherry fluorescence and grown as stable (mix pooled) cell lines to minimize clonal variation.

Fluorescence-activated Cell Sorting (FACS):

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.

Cell viability assays:

To infer metabolic activity via turnover of ATP, viability assays were performed by addition of 10ul of CCK-8 (Transgen) directly to cells grown in DMEM, 1%FBS, P/S, incubated for 4 hours, then measured across time in the Tecan Spark microplate reader at Abs 450 nm. The background was subtracted and normalized against control.

Western Blot:

To test ISG15 protein expression levels and detect free ISG15s, cell supernatants were collected before infecting HEK293 cells with HCoV-229E virus, then cell supernatants and cell lysates were collected 48h post-infection. Cell lysates of ISG15-expressing A549 cells were similarly collected at 24 h post-infection of H1N1 IAV. Cells were lysed in 'Buffer 1' lysis buffer [32], supplemented with phosphatase inhibitor cocktail (Phygene, PH0321) and protease inhibitor cocktail (Phygene, PH0320). Collected cell lysates and cell supernatants were mixed with 5x SDS-Page loading buffer (Phygene, PH0333) and boiled for 5 min. Subsequently, cell lysates and supernatants were separated by 10% SDS-PAGE gel, transferred to PVDF membranes (Millipore, 0.45μm) and blocked with 5% skimmed milk in TBS.

The following antibodies were used for detection: rabbit anti-MX1 polyclonal antibody (clone N2C2, Genetex, GTX110256, dilution 1:1000), rabbit anti-ISG15 polyclonal antibody (middle region, Aviva Systems Biology, ARP59386_P050, dilution 1:1000), rabbit anti-GAPDH monoclonal antibody (clone 14C10, Cell Signaling, 2118, dilution 1:2000), rabbit anti-CD13 polyclonal antibody (Sino Biological, 10051-T60, dilution 1:2000), rabbit HCoV-229E nucleocapsid polyclonal antibody (Sino Biological, 40640-T62, dilution 1:2000) and HRP-conjugated goat anti-rabbit IgG (Transgen, HS101-01, dilution 1:5000).

Chemiluminescence was detected using the ECL chemiluminescence detection kit (Vazyme) according to the manufacturer's instructions, and subsequently imaged by LI-COR ODYSSEY® FC Imaging system (LI-COR Biosciences). Uncropped western blot images are shown in Supplementary Figures 27 and 29-33. Densitometry measurements were calculated from FiJI ImageJ software based on equal size rectangular ROIs (multi-measure) of grayscale

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 Homo sapiens cell lysate ISG15 signal (graph for n=3 independent blots).

Virus infections:

A HCoV-229E clinical isolate was a kind gift from Prof. Jincun Zhao (Guangzhou Medical University), IAV H1N1 PR8 and VSV-GFP (Indiana) were kind gifts from Prof. Linrong 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 [189]. 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 [189]. 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 performed in 1% FBS at low MOI (0.1) for fluorescent reporter or HCoV-229E TCID50 assays. HCoV-229E assays were run in 10-fold dilutions in low-serum media. VSV-GFP assays were rinsed after 4-6hrs infection, replaced with growth media and followed across time until ~70-80% GFP-positive (overnight).

For SARS-CoV-2 infections, A549-ACE2 cells were seeded at a density of 1.5×10^5 cells/well in a 12-well plate for 24 hrs. After 24 hrs, cells were transfected with 200 ng of plasmids encoding bat ISG15 (see above) or vector control for 24 hrs, followed by infection with ancestral SARS-CoV-2 (SARS-CoV-2/SB3-TYAGNC isolate) at a multiplicity of infection (MOI) of 0.01 for 48 hrs. Control cells were sham infected. Infected or sham infected cells were incubated at 37°C for 1 hr with gentle rocking every 15 min. After 1 hr, virus inoculum was removed, cells were washed with phosphate buffered solution (PBS), and supplemented with growth medium. Bulk cellular RNA and media from infected and sham infected cells were harvested at 48 hrs post infection using a previously published protocol [190]. Cells transfected with mCherry_pcDNA3.1(+)-P2A plasmid and infected with SARS-CoV-2 served as control for plasmid DNA transfection-mediated impact on SARS-CoV-2 replication. All work with infectious SARS-CoV-2 was performed 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 performed in A549-stable cell lines (as above) by the addition of 50µl of virus to 500µl of low-FBS media (in triplicate) and serial 10-fold dilutions were performed (x8) in 24-well plates. Cells were incubated with virus for 4-6 hours prior to rinsing and replaced with 2% Methyl-cellulose 4000cP direct overlays (Beyotime) for 2-3 days.

Median Tissue Culture Infectious Dose assay ($TCID_{50}$):

The supernatants from SARS-CoV-2 infected cells were titrated in triplicates on Vero-76 cells using tissue culture infectious dose 50 ($TCID_{50}$) assay [191]. Briefly, 3 x 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 following day, media was taken off the cells and 50 µL of 1:10 serially diluted virus containing supernatant was added to the plates. The plates were incubated at 37 °C for 1 hr. After incubation, the virus containing supernatant was discarded and 100 µl of 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 under a light microscope. Tissue culture infectious dose 50/ml ($TCID_{50}$/ml) was calculated using the Spearman and Karber algorithm [192,193].

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 (changes polarity and removes the Cysteine disulfide bond) or Serine (similar shape/charge but no disulfide bond) that 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 post-transfection/transduction were collected after 24/48 hrs respectively, pelleted for cell debris removal and added directly to SDS-PAGE loading dye for western blot. Similarly, supernatants were collected after 48hrs of HCoV-229E infection. Cell lysates were collected as described previously.

Competing interests

The authors have no competing interests.

Acknowledgment

MH was supported by the German Research Foundation (HI1423/5-1) and the LOEWE-Centre for Translational Biodiversity Genomics (TBG) funded by the Hessen State Ministry of Higher Education, Research and the Arts (HMWK), TB by the German Research Foundation (INST 269/768-1), LMD by the National Science Foundation (NSF-IOS 2032063 and 2031906, NSF-DEB 1838273), TML by the National Science Foundation (Award ID 2010853), ATI by a Key grant from the National Science Foundation of Zhejiang Province (Z23C010003), BKL by the Royal Ontario Museum Governors, the National Science Foundation (DEB-0344430), and the Toronto Zoological Society, AB by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2022-03010) and the Canadian Institutes of Health Research (479539), VG by a PGS-D scholarship funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) (569587-2022), SCV by a UKRI Future Leaders Fellowship (MR/T021985/1), an ERC Consolidator Grant (101001702; BATSPEAK), and a Max Planck Research Group awarded by the Max Planck Society, SJP by a Junior Chair grant from the Institut Universitaire de France, GM by the Max Planck Society, AK and JP by the Robert Koch Institute, ECT by Irish Research Council Laureate Award IRCLA/2017/58 and Science Foundation Ireland Future Frontiers 19/FFP/6790. VIDO receives operational funding from the Government of Saskatchewan via Innovation Saskatchewan and the Ministry of Agriculture, and funding for its Containment Level 3 (CL3) 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, DFG NGS Competence Center, part of the Center for Molecular and Cellular Bioengineering (CMCB), Technische Universität Dresden and of the MPI-CBG. We also thank the core-facility of Zhejiang University - University of Edinburgh Institute.

Data availability

Genome assemblies and all sequencing data have been submitted to NCBI (accession number pending). 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 publicly available genomic data are listed in Supplementary Tables 3 and 5.

Code Availability

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

References

1. Hayman, D. T. S. Bats as Viral Reservoirs. *Annual review of virology* **3**, 77–99 (2016).
2. Olival, K. J. *et al.* Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650 (2017).
3. Zhou, S. *et al.* ZOVER: the database of zoonotic and vector-borne viruses. *Nucleic Acids Res.* **50**, (2022).
4. Halpin, K., Young, P. L., Field, H. E. & Mackenzie, J. S. Isolation of Hendra virus from pteropid bats: a natural reservoir of Hendra virus. *J. Gen. Virol.* **81**, 1927–1932 (2000).
5. Towner, J. S. *et al.* Isolation of genetically diverse Marburg viruses from Egyptian fruit bats. *PLoS Pathog.* **5**, e1000536 (2009).
6. Goldstein, T. *et al.* The discovery of Bombali virus adds further support for bats as hosts of ebolaviruses. *Nature microbiology* **3**, 1084–1089 (2018).
7. Lopes-Pacheco, M. *et al.* Pathogenesis of Multiple Organ Injury in COVID-19 and Potential Therapeutic Strategies. *Front. Physiol.* **12**, 593223 (2021).
8. Memish, Z. A. *et al.* Middle East respiratory syndrome coronavirus in bats, Saudi Arabia. *Emerg. Infect. Dis.* **19**, 1819–1823 (2013).
9. Ge, X. Y. *et al.* Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **503**, 535–538 (2013).
10. Temmam, S. *et al.* Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature* **604**, 330–336 (2022).
11. Li, W. *et al.* Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310**, 676–679 (2005).
12. Anthony, S. J. *et al.* Global patterns in coronavirus diversity. *Virus evolution* **3**, (2017).
13. Schlottau, K. *et al.* SARS-CoV-2 in fruit bats, ferrets, pigs, and chickens: an experimental transmission study. *Lancet Microbe* **1**, e218–e225 (2020).
14. Watanabe, S. *et al.* Bat coronaviruses and experimental infection of bats, the Philippines. *Emerg. Infect. Dis.* **16**, 1217–1223 (2010).
15. Munster, V. J. *et al.* Replication and shedding of MERS-CoV in Jamaican fruit bats (Artibeus jamaicensis). *Sci. Rep.* **6**, 21878 (2016).
16. 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.e5 (2021).
17. Bosco-Lauth, A. M. *et al.* Experimental Infection of Brazilian Free-Tailed Bats (Tadarida brasiliensis) with Two Strains of SARS-CoV-2. *Viruses* **14**, (2022).
18. Pavlovich, S. S. *et al.* The Egyptian Rousette Genome Reveals Unexpected Features of Bat Antiviral Immunity. *Cell* **173**, 1098–1110.e18 (2018).
19. 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).
20. Medzhitov, R., Schneider, D. S. & Soares, M. P. Disease tolerance as a defense strategy. *Science* **335**, 936–941 (2012).
21. Seal, S., Dharmarajan, G. & Khan, I. Evolution of pathogen tolerance and emerging infections: A missing experimental paradigm. *Elife* **10**, e68874 (2021).
22. 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).
23. Banerjee, A. *et al.* Novel Insights Into Immune Systems of Bats. *Front. Immunol.* **11**, 26 (2020).
24. 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).
25. Zhang, G. *et al.* Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. *Science* **339**, 456–460 (2013).
26. Ng, M. *et al.* Filovirus receptor NPC1 contributes to species-specific patterns of ebolavirus susceptibility in bats. *Elife* **4**, (2015).
27. Hawkins, J. A. *et al.* A metaanalysis of bat phylogenetics and positive selection based on genomes and transcriptomes from 18 species. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 11351–11360

(2019).
28. Scheben, A. *et al.* Long-read sequencing reveals rapid evolution of immunity- and cancer-related genes in bats. *bioRxiv* 2020.09.09.290502 (2021) doi:10.1101/2020.09.09.290502.
29. Jebb, D. *et al.* Six reference-quality genomes reveal evolution of bat adaptations. *Nature* **583**, 578–584 (2020).
30. Zhou, P. *et al.* Contraction of the type I IFN locus and unusual constitutive expression of IFN-α in bats. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 2696–2701 (2016).
31. Ahn, M., Cui, J., Irving, A. T. & Wang, L.-F. Unique Loss of the PYHIN Gene Family in Bats Amongst Mammals: Implications for Inflammasome Sensing. *Sci. Rep.* **6**, 21722 (2016).
32. Ahn, M. *et al.* Dampened NLRP3-mediated inflammation in bats and implications for a special viral reservoir host. *Nat Microbiol* **4**, 789–799 (2019).
33. Goh, G. *et al.* Complementary regulation of caspase-1 and IL-1β reveals additional mechanisms of dampened inflammation in bats. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 28939–28949 (2020).
34. Wu, Z. *et al.* A comprehensive survey of bat sarbecoviruses across China in relation to the origins of SARS-CoV and SARS-CoV-2. *Natl Sci Rev* nwac213 (2022) doi:10.1093/nsr/nwac213.
35. Gouilh, M. A. *et al.* SARS-Coronavirus ancestor's foot-prints in South-East Asian bat colonies and the refuge theory. *Infect. Genet. Evol.* **11**, 1690–1702 (2011).
36. Seltmann, A. *et al.* Seasonal Fluctuations of Astrovirus, But Not Coronavirus Shedding in Bats Inhabiting Human-Modified Tropical Forests. *Ecohealth* **14**, 272–284 (2017).
37. Bokelmann, M. *et al.* Tolerance and Persistence of Ebola Virus in Primary Cells from Mops condylurus, a Potential Ebola Virus Reservoir. *Viruses* **13**, (2021).
38. Teeling, E. C. *et al.* Bat Biology, Genomes, and the Bat1K Project: To Generate Chromosome-Level Genomes for All Living Bat Species. *Annual Review of Animal Biosciences* **6**, 23–46 (2018).
39. Card, D. C., Shapiro, B., Giribet, G., Moritz, C. & Edwards, S. V. Museum Genomics. *Annu. Rev. Genet.* **55**, 633–659 (2021).
40. Blumer, M. *et al.* Gene losses in the common vampire bat illuminate molecular adaptations to blood feeding. *Science Advances* **8**, eabm6494 (2022).
41. Kirilenko, B. M. *et al.* TOGA integrates gene annotation with orthology inference at scale. *bioRxiv* 2022.09.08.507143 (2022) doi:10.1101/2022.09.08.507143.
42. Osmanski, A. B. *et al.* Insights into mammalian TE diversity via the curation of 248 mammalian genome assemblies. *bioRxiv* 2022.12.28.522108 (2022) doi:10.1101/2022.12.28.522108.
43. Doronina, L. *et al.* Contradictory Phylogenetic Signals in the Laurasiatheria Anomaly Zone. *Genes* **13**, (2022).
44. 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).
45. Smith, S. A. & O'Meara, B. C. TreePL: Divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**, 2689–2690 (2012).
46. 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**, (2016).
47. Teeling, E. C. *et al.* A Molecular Phylogeny for Bats Illuminates Biogeography and the Fossil Record. *Science* **307**, 580–584 (2005).
48. 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).
49. Hall, J. S. *et al.* Experimental challenge of a North American bat species, big brown bat (Eptesicus fuscus), with SARS-CoV-2. *Transbound. Emerg. Dis.* **68**, 3443–3452 (2021).
50. Guo, H. *et al.* Evolutionary Arms Race between Virus and Host Drives Genetic Diversity in Bat Severe Acute Respiratory Syndrome-Related Coronavirus Spike Genes. *J. Virol.* **94**, (2020).
51. Damas, J. *et al.* Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 22311–22322 (2020).
52. Frank, H. K., Enard, D. & Boyd, S. D. Exceptional diversity and selection pressure on coronavirus host receptors in bats compared to other mammals. *Proc. Biol. Sci.* **289**, 20220193

(2022).
53. Sungnak, W. *et al.* SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat. Med.* **26**, 681–687 (2020).
54. Yeager, C. L. *et al.* Human aminopeptidase N is a receptor for human coronavirus 229E. *Nature* **357**, 420–422 (1992).
55. Wei, C. *et al.* HDL-scavenger receptor B type 1 facilitates SARS-CoV-2 entry. *Nat Metab* **2**, 1391–1400 (2020).
56. Ebert, D. H., Deussing, J., Peters, C. & Dermody, T. S. Cathepsin L and cathepsin B mediate reovirus disassembly in murine fibroblast cells. *J. Biol. Chem.* **277**, 24609–24617 (2002).
57. 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).
58. Salvi, V. *et al.* SARS-CoV-2-associated ssRNAs activate inflammation and immunity via TLR7/8. *JCI Insight* **6**, (2021).
59. 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).
60. 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).
61. Hu, M.-M. *et al.* TRIM38 Negatively Regulates TLR3/4-Mediated Innate Immune and Inflammatory Responses by Two Sequential and Distinct Mechanisms. *J. Immunol.* **195**, 4415–4425 (2015).
62. 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. U. S. A.* **111**, 1509–1514 (2014).
63. Jefferies, C. A. *et al.* Bruton's Tyrosine Kinase Is a Toll/Interleukin-1 Receptor Domain-binding Protein That Participates in Nuclear Factor κB Activation by Toll-like Receptor 4*. *J. Biol. Chem.* **278**, 26258–26264 (2003).
64. 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).
65. 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).
66. Roschewski, M. *et al.* Inhibition of Bruton tyrosine kinase in patients with severe COVID-19. *Sci Immunol* **5**, (2020).
67. 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).
68. Arredouani, M. *et al.* Haptoglobin directly affects T cells and suppresses T helper cell type 2 cytokine release. *Immunology* **108**, 144–151 (2003).
69. Shen, H. *et al.* Haptoglobin activates innate immunity to enhance acute transplant rejection in mice. *J. Clin. Invest.* **122**, 383–387 (2012).
70. 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).
71. 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**, (2019).
72. 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).
73. Crow, M. K., Olferiev, M. & Kirou, K. A. Type I Interferons in Autoimmune Disease. *Annu. Rev. Pathol.* **14**, 369–393 (2019).
74. 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).
75. Wu, X. *et al.* Viral Mimicry of Interleukin-17A by SARS-CoV-2 ORF8. *MBio* **13**, e0040222 (2022).
76. Das Sarma, J. *et al.* Ifit2 deficiency restricts microglial activation and leukocyte migration

following murine coronavirus (m-CoV) CNS infection. *PLoS Pathog.* **16**, e1009034 (2020).
77. Schindewolf, C. *et al.* SARS-CoV-2 Uses Nonstructural Protein 16 to Evade Restriction by IFIT1 and IFIT3. *bioRxiv* (2022) doi:10.1101/2022.09.26.509529.
78. Zhang, W. *et al.* The emerging roles of IFIT3 in antiviral innate immunity and cellular biology. *J. Med. Virol.* e28259 (2022) doi:10.1002/jmv.28259.
79. Diamond, M. S. & Farzan, M. The broad-spectrum antiviral functions of IFIT and IFITM proteins. *Nat. Rev. Immunol.* **13**, 46–57 (2013).
80. 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).
81. Wang, J. *et al.* Negative regulation of Nmi on virus-triggered type I IFN production by targeting IRF7. *J. Immunol.* **191**, 3393–3399 (2013).
82. 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).
83. Xiahou, Z. *et al.* NMI and IFP35 serve as proinflammatory DAMPs during cellular infection and injury. *Nat. Commun.* **8**, 950 (2017).
84. Vanderheiden, A. *et al.* CCR2 Signaling Restricts SARS-CoV-2 Infection. *MBio* **12**, e0274921 (2021).
85. Lin, K. L., Suzuki, Y., Nakano, H., Ramsburg, E. & Gunn, M. D. CCR2+ monocyte-derived dendritic cells and exudate macrophages produce influenza-induced pulmonary immune pathology and mortality. *J. Immunol.* **180**, 2562–2572 (2008).
86. 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).
87. Glass, W. G. *et al.* Chemokine receptor CCR5 promotes leukocyte trafficking to the brain and survival in West Nile virus infection. *J. Exp. Med.* **202**, 1087–1098 (2005).
88. 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).
89. Pairo-Castineira, E. *et al.* Genetic mechanisms of critical illness in COVID-19. *Nature* **591**, 92–98 (2021).
90. Gupta, K., Kaur, G., Pathak, T. & Banerjee, I. Systematic review and meta-analysis of human genetic variants contributing to COVID-19 susceptibility and severity. *Gene* **844**, 146790 (2022).
91. 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).
92. Morgan, B. P. & Harris, C. L. Complement, a target for therapy in inflammatory and degenerative diseases. *Nat. Rev. Drug Discov.* **14**, 857–877 (2015).
93. Afzali, B., Noris, M., Lambrecht, B. N. & Kemper, C. The state of complement in COVID-19. *Nat. Rev. Immunol.* **22**, 77–84 (2022).
94. 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).
95. 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).
96. 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).
97. Pal Singh, S., Dammeijer, F. & Hendriks, R. W. Role of Bruton's tyrosine kinase in B cells and malignancies. *Mol. Cancer* **17**, 57 (2018).
98. Weber, A. N. R. *et al.* Bruton's Tyrosine Kinase: An Emerging Key Player in Innate Immunity. *Front. Immunol.* **8**, 1454 (2017).
99. Perng, Y.-C. & Lenschow, D. J. ISG15 in antiviral immunity and beyond. *Nat. Rev. Microbiol.* **16**, 423–439 (2018).
100. 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).
101. Shin, D. *et al.* Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **587**, 657–662 (2020).
102. 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. *Nature Microbiology 2021 6:4* **6**, 467–478 (2021).
103. Wong, J. J. Y., Pung, Y. F., Sze, N. S.-K. & Chin, K.-C. HERC5 is an IFN-induced HECT-type E3 protein ligase that mediates type I IFN-induced ISGylation of protein targets. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10735–10740 (2006).
104. Munnur, D. *et al.* Altered ISGylation drives aberrant macrophage-dependent immune responses during SARS-CoV-2 infection. *Nat. Immunol.* **22**, 1416–1427 (2021).
105. Narasimhan, J. *et al.* Crystal Structure of the Interferon-induced Ubiquitin-like Protein ISG15*. *J. Biol. Chem.* **280**, 27356–27365 (2005).
106. 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).
107. Napolitano, A. *et al.* Cysteine-Reactive Free ISG15 Generates IL-1β-Producing CD8α+ Dendritic Cells at the Site of Infection. *J. Immunol.* **201**, 604–614 (2018).
108. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
109. 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).
110. 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).
111. Zhao, C., Hsiang, T.-Y., Kuo, R.-L. & Krug, R. M. ISG15 conjugation system targets the viral NS1 protein in influenza A virus-infected cells. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2253–2258 (2010).
112. Rhie, A. *et al.* Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**, 737–746 (2021).
113. Null, N. *et al.* Sequence locally, think globally: The Darwin Tree of Life Project. *Proceedings of the National Academy of Sciences* **119**, e2115642118 (2022).
114. Lewin, H. A. *et al.* The Earth BioGenome Project 2020: Starting the clock. *Proc. Natl. Acad. Sci. U. S. A.* **119**, (2022).
115. Shultz, A. J. & Sackton, T. B. Immune genes are hotspots of shared positive selection across birds and mammals. *Elife* **8**, (2019).
116. Wang, L.-F., Walker, P. J. & Poon, L. L. M. Mass extinctions, biodiversity and mitochondrial function: are bats 'special' as reservoirs for emerging viruses? *Curr. Opin. Virol.* **1**, 649–657 (2011).
117. O'Shea, T. J. *et al.* Bat flight and zoonotic viruses. *Emerg. Infect. Dis.* **20**, 741–745 (2014).
118. 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).
119. 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.e5 (2022).
120. 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).
121. Déjosez, M. *et al.* Bat pluripotent stem cells reveal unique entanglement between host and viruses. *bioRxiv* 2022.09.23.509261 (2022) doi:10.1101/2022.09.23.509261.
122. 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).
123. Guan, D. *et al.* Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898 (2020).
124. Nurk, S. *et al.* HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res.* **30**, gr.263566.120 (2020).
125. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. (2013).
126. Ghurye, J. *et al.* Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Comput. Biol.* **15**, e1007273 (2019).

127. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
128. Formenti, G. *et al.* Merfin: improved variant filtering, assembly evaluation and polishing via k-mer validation. *Nat. Methods* **19**, 696–704 (2022).
129. Goubert, C. *et al.* A beginner's guide to manual curation of transposable elements. *Mob. DNA* **13**, 7 (2022).
130. 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).
131. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
132. 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).
133. Kalvari, I. *et al.* Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* **49**, D192–D200 (2021).
134. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
135. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The Vienna RNA websuite. *Nucleic Acids Res.* **36**, W70–4 (2008).
136. Harris, R. S. Improved pairwise alignment of genomic DNA. (Ph.D. Thesis, The Pennsylvania State University, 2007).
137. Sharma, V. & Hiller, M. Increased alignment sensitivity improves the usage of genome alignments for comparative gene annotation. *Nucleic Acids Res.* **45**, 8369–8377 (2017).
138. Kent, W. J., Baertsch, R., Hinrichs, A., Miller, W. & Haussler, D. Evolution's cauldron: Duplication, deletion, and rearrangement in the mouse and human genomes. *Proceedings of the National Academy of Sciences* **100**, 11484–11489 (2003).
139. Osipova, E., Hecker, N. & Hiller, M. RepeatFiller newly identifies megabases of aligning repetitive sequences and improves annotations of conserved non-exonic elements. *Gigascience* **8**, 1–10 (2019).
140. Suarez, H. G., Langer, B. E., Ladde, P. & Hiller, M. chainCleaner improves genome alignment specificity and sensitivity. *Bioinformatics* **33**, 1596–1603 (2017).
141. 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).
142. 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 Evolutionary Biology 2019 19:1* **19**, 1–17 (2019).
143. Mirarab, S. & Warnow, T. ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. in *Bioinformatics* vol. 31 i44–i52 (2015).
144. Mirarab, S. *et al.* ASTRAL: Genome-scale coalescent-based species tree estimation. *Bioinformatics* **30**, 541–548 (2014).
145. 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).
146. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
147. Lemoine, F. *et al.* Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* **556**, 452–456 (2018).
148. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods 2017 14:6* **14**, 587–589 (2017).
149. Wang, K. *et al.* Molecular adaptation and convergent evolution of frugivory in Old World and neotropical fruit bats. *Mol. Ecol.* **29**, 4366–4381 (2020).

150. Dong, D. *et al.* The Genomes of Two Bat Species with Long Constant Frequency Echolocation Calls. *Mol. Biol. Evol.* **34**, 20–34 (2016).
151. Kosakovsky Pond, S. L., Frost, S. D. W. & Muse, S. V. HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2005).
152. 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).
153. 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**, W193–200 (2007).
154. Siepel, A. & Haussler, D. Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Mol. Biol. Evol.* **21**, 468–488 (2004).
155. Information and Likelihood Theory: A Basis for Model Selection and Inference. in *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (eds. Burnham, K. P. & Anderson, D. R.) 49–97 (Springer New York, 2002). doi:10.1007/978-0-387-22456-5_2.
156. O'Hara, R. & Kotze, J. Do not log-transform count data. *Nature Precedings* 1–1 (2010) doi:10.1038/npre.2010.4136.1.
157. Bürkner, P.-C. brms: An R Package for Bayesian Multilevel Models Using Stan. *J. Stat. Softw.* **80**, 1–28 (2017).
158. Carpenter, B. *et al.* Stan: A Probabilistic Programming Language. *J. Stat. Softw.* **76**, (2017).
159. Gelman, A., Goodrich, B., Gabry, J. & Vehtari, A. R-squared for Bayesian Regression Models. *Am. Stat.* **73**, 307–309 (2019).
160. Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *arXiv [stat.CO]* (2015).
161. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
162. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010).
163. Zheng, W. *et al.* LOMETS2: improved meta-threading server for fold-recognition and structure-based function annotation for distant-homology proteins. *Nucleic Acids Res.* **47**, W429–W436 (2019).
164. Jiang, Y. & Wang, X. Structural insights into the species preference of the influenza B virus NS1 protein in ISG15 binding. *Protein Cell* **10**, 681–687 (2019).
165. Durfee, L. A., Kelley, M. L. & Huibregtse, J. M. The basis for selective E1-E2 interactions in the ISG15 conjugation system. *J. Biol. Chem.* **283**, 23895–23902 (2008).
166. Reich, N. *et al.* Interferon-induced transcription of a gene encoding a 15-kDa protein depends on an upstream enhancer element. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 6394–6398 (1987).
167. Swaim, C. D. *et al.* Modulation of Extracellular ISG15 Signaling by Pathogens and Viral Effector Proteins. *Cell Rep.* **31**, 107772 (2020).
168. Rodrigues, C. H., Pires, D. E. & Ascher, D. B. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res.* **46**, W350–W355 (2018).
169. Rodrigues, C. H. M., Pires, D. E. V. & Ascher, D. B. DynaMut2: Assessing changes in stability and flexibility upon single and multiple point missense mutations. *Protein Sci.* **30**, 60–69 (2021).
170. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
171. Mirdita, M., Steinegger, M. & Söding, J. MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics* **35**, 2856–2858 (2019).
172. Mirdita, M. *et al.* Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res.* **45**, D170–D176 (2016).
173. 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).
174. Steinegger, M. *et al.* HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 473 (2019).

175. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021.10.04.463034 (2022) doi:10.1101/2021.10.04.463034.
176. Eastman, P. *et al.* OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **13**, e1005659 (2017).
177. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19–25 (2015).
178. Bauer, P., Hess, B. & Lindahl, E. *GROMACS 2022.1 Source code*. (2022). doi:10.5281/zenodo.6451564.
179. 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).
180. 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).
181. 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).
182. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
183. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
184. Nosé, S. & Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **50**, 1055–1076 (1983).
185. Glykos, N. M. Software news and updates. Carma: a molecular dynamics analysis program. *J. Comput. Chem.* **27**, 1765–1768 (2006).
186. 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. Algorithms* **5**, 475–504 (2006).
187. Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).
188. 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).
189. Banerjee, A. *et al.* Isolation, Sequence, Infectivity, and Replication Kinetics of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerg. Infect. Dis.* **26**, 2054–2063 (2020).
190. Banerjee, A. *et al.* Experimental and natural evidence of SARS-CoV-2-infection-induced activation of type I interferon responses. *iScience* **24**, 102477 (2021).
191. 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).
192. Spearman, C. THE METHOD OF 'RIGHT AND WRONG CASES' ('CONSTANT STIMULI') WITHOUT GAUSS'S FORMULAE. *Br. J. Psychol.* **2**, 227–242 (1908).
193. Kärber, G. Beitrag zur kollektiven Behandlung pharmakologischer Reihenversuche. *Naunyn-Schmiedebergs Archiv für experimentelle Pathologie und Pharmakologie* **162**, 480–483 (1931).

Figures & Legends

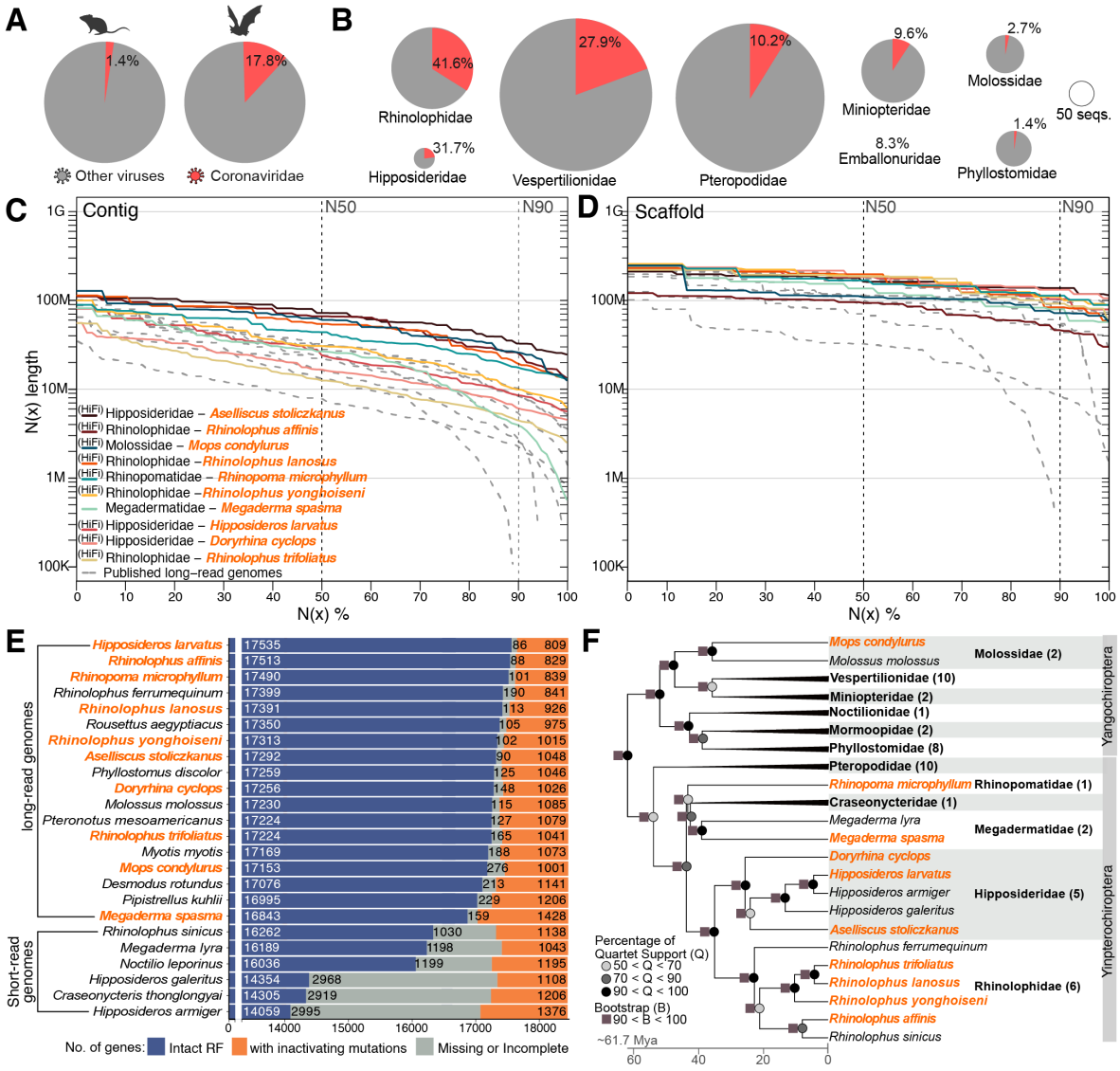

Figure 1: High-quality chromosome-level genome assemblies of ten bat species.
(A,B) Percent of viruses from the family Coronaviridae that were detected in rodents vs. bats (A) and in different bat families (B). Data was extracted from ZOVER [3] (last access, January 25, 2023), using the filter "Metagenomics/HTS" to include only metagenomic studies. Pie size is proportional to the number of viral sequences.
(C,D) Assembly contiguity visualized as N(x) graphs that show contig (C) or scaffold (D) sizes on the Y-axis, for which x percent of the assembly consists of contigs and scaffolds of at least that size. New bat assemblies are shown as solid colored lines, published long read-based assemblies are shown as dashed dotted grey lines. Legend is sorted by contig N50. N50 and N90 are indicated by dashed vertical lines.

(E) Status of 18,430 ancestral mammalian genes in our new and previous bat assemblies. Genes are classified by TOGA into those with an intact reading frame (RF, blue), with gene-inactivating mutations (orange), or missing or incomplete coding sequences due to assembly gaps or fragmentation (grey).
(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 in orange font in (E) and (F).

1569
1570

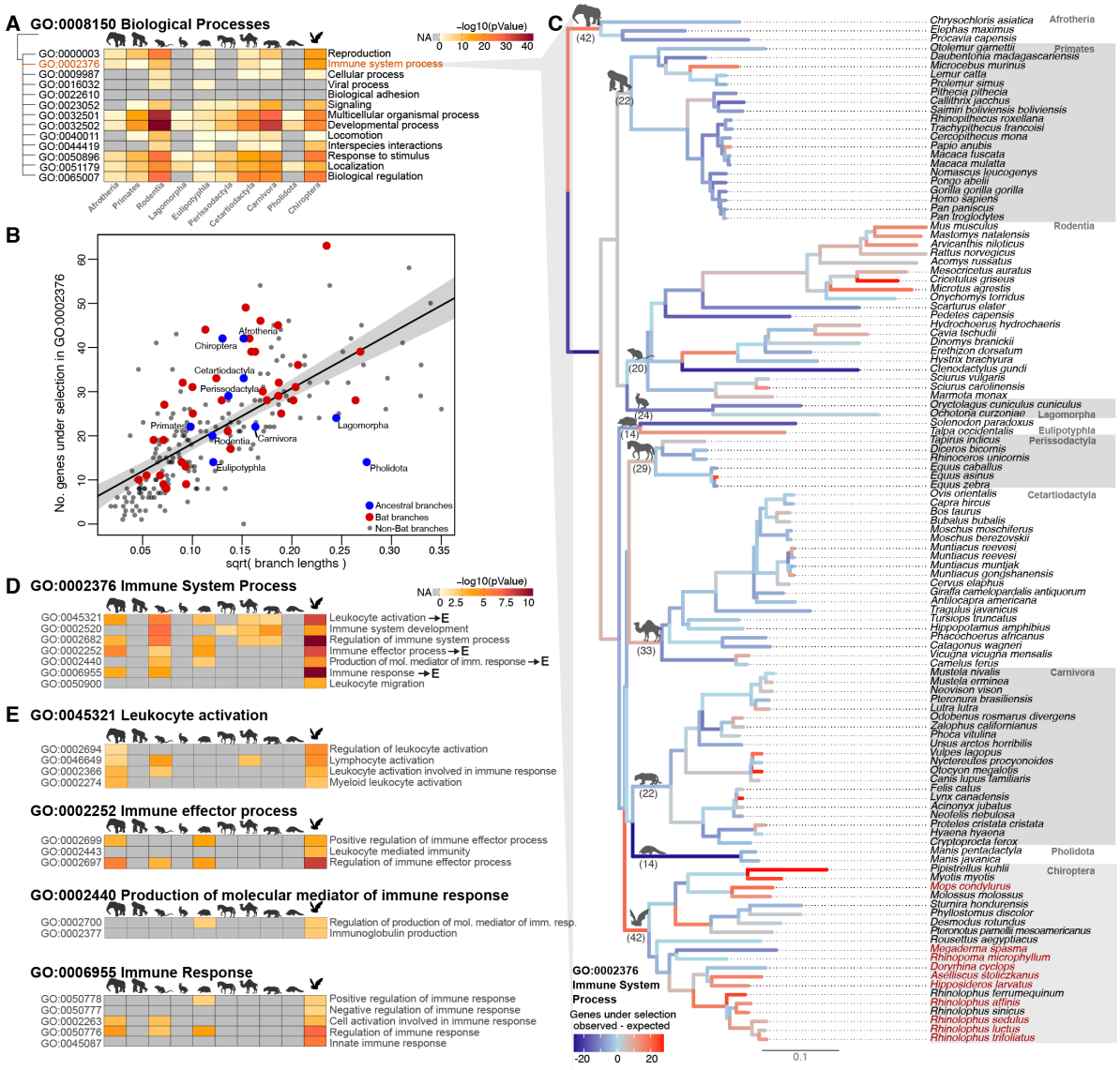

Figure 2: ***Selection of immune-related genes is most prevalent in bats.***

(A) Functional enrichments of genes under selection in mammalian groups (columns). All high-level gene ontology (GO) terms representing different biological processes were tested (rows). Only significant terms are shown.

(B) Linear regression model shows a significant correlation between the number of genes selected on a branch and the square-root of the branch lengths (substitutions per site).

(C) Per-branch signal of immune selection for the phylogeny of 115 mammals. Branches are color-coded based on the difference between the observed and expected number of selected genes annotated with the “immune system process” GO term. Expected numbers were calculated from the regression model shown in panel B. Observed gene number for each ancestral branch of the ten mammalian orders is listed below the branch. New Bat1K genomes are in red font.

(D, E) Enrichments of direct descendants (child terms) of GO “Immune system process” (D) and child terms thereof (E) for genes under selection in mammalian groups.

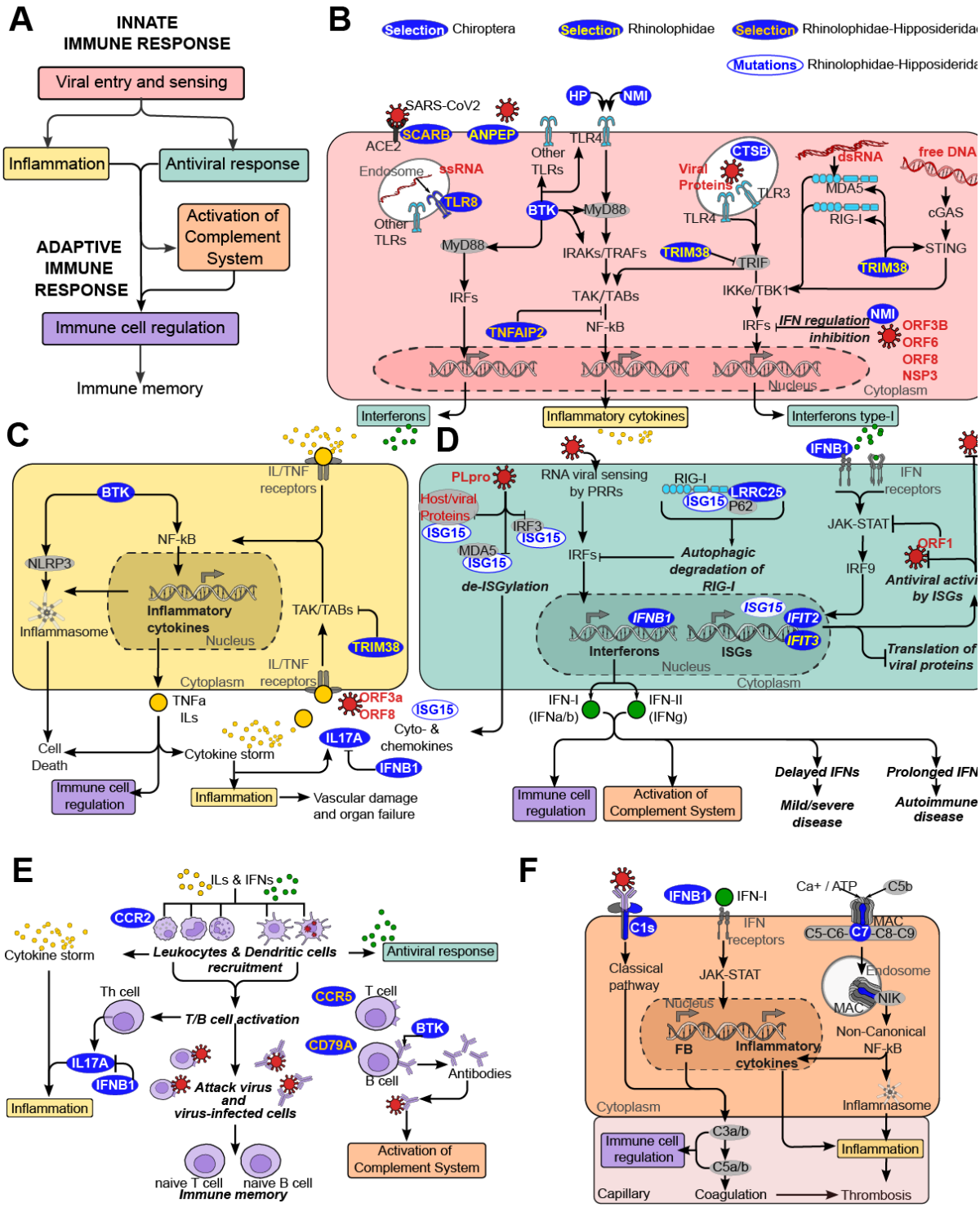

Figure 3: Genes under selection are involved in immune responses during viral infections.

(A) Overview of biological processes involved in a synchronized immune response triggered by viral infections. (B-F) Schematic showing how ISG15 and genes under selection in bats (highlighted in blue) 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). Colored backgrounds correspond to the processes in (A).

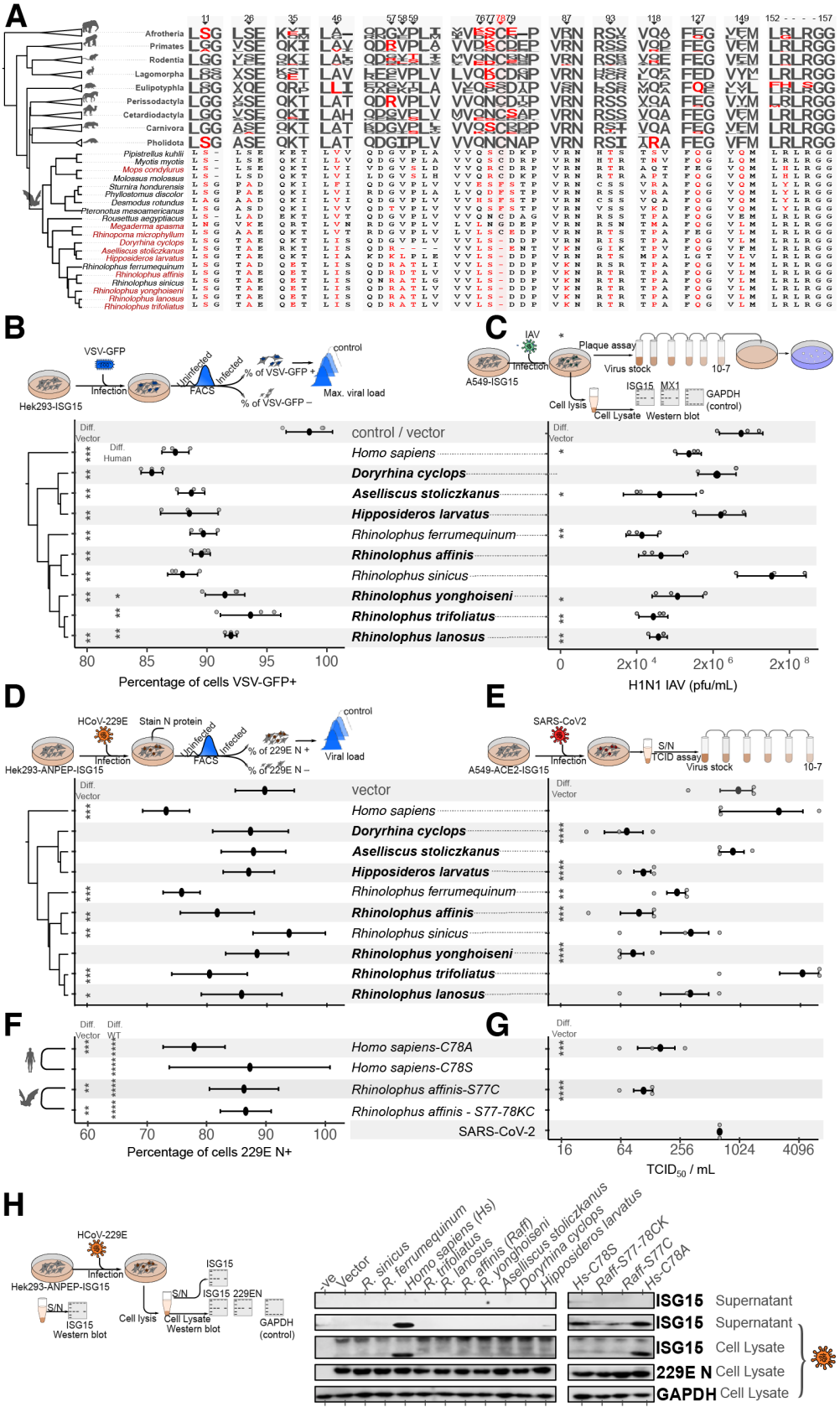

Figure 4: ***Altered antiviral capacity of bat-ISG15 against viruses.***

(A) Alignment of ISG15 residues that are important for protein function, numbered according to human ISG15. Residues in red font differ from the consensus. The conserved Cys78 is deleted in all rhinolophid and hipposiderid bats. Sequence logos visualize protein conservation of other mammalian groups.
(B) Percentage of transfected HEK293 cells that are positive for GFP-tagged Vesicular Stomatitis Virus (VSV-GFP) at 16 hr post-infection, as measured by FACS. Cells were transfected with an empty vector or an IRES-mCherry vector containing ISG15 of different species.
(C) Plaque assay of stable ISG15-transduced A549 cells infected with Influenza A virus (IAV, H1N1/PR8 strain). Viral plaques were determined as pfu/ml for the 10^{-7} dilution. Direct overlay images are shown in Supplementary Figure 26.
(D) Percentage of stable ANPEP-expressing, ISG15 (or vector transfected) +ve HEK293 cells that are positive for N-protein after human coronavirus 229E (HCoV-229E) infection, as measured by FACS.
(E) $TCID_{50}$ assays measuring viral production after 3 days in A549-ACE2 cells that were transfected with vector or ISG15 constructs and infected with SARS-CoV-2. Higher $TCID_{50}$ values indicate a higher viral production.
(F/G) Viral infection assay with HCoV-229E (F) and SARS-CoV-2 (G) transfecting cells with mutant *Homo sapiens* or *Rhinolophus affinis* ISG15 constructs that remove or restore the Cys78 residue.
(H) Western blot of ISG15 transfected HEK293-ANPEP cells infected with HCoV-229E (MOI 0.1, 3 days). In supernatants, ISG15 is not detected prior to infection. Only human or Cys78-mutant bat ISG15 is detected in supernatants post-infection. Cell lysates post-infection show free ISG15 (lower band) and ISGylated proteins (upper bands). HCoV-229E N protein and GAPDH are shown as controls. Photos are representative images of three individual experiments. Uncropped images are provided in Supplementary Figures 29-33, Western blot quantifications in Supplementary Figure 34.
Data are presented as mean (solid oval) and standard error (bars), showing individual data points of three biological replicates as grey circles (B-G).
Significant differences to the vector control, or to wild-type ISG15 were determined with a two-tailed t-test and is indicated with * P< 0.05, **P< 0.01, ***P< 0.001, ****P< 0.0001. All data are provided in Supplementary Tables 10,11,14,15,17, and 18.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplement.pdf
- SupplementaryTables.xlsx