Comparative genomics provides insights into chromosomal evolution and immunological adaptation in horseshoe bats

Check for updates

Shilin Tian [1,2,8], Junyu Si[1,8], Libiao Zhang[3,8], Jiaming Zeng[1], Xiangyi Zhang [1], Chen Huang[1], Gang Li [4], Caoqi Lei[1], Xuming Zhou [5], Rong Geng[6], Peng Zhou[6], Huan Yan [1], Stephen J. Rossiter [7] ✉ & Huabin Zhao [1] ✉

Horseshoe bats are natural hosts of zoonotic viruses, yet the genetic basis of their antiviral immunity is poorly understood. Here we generated two new chromosomal-level genome assemblies for horseshoe bat species (*Rhinolophus*) and three close relatives, and show that, during their diversification, horseshoe bats underwent extensive chromosomal rearrangements and gene expansions linked to segmental duplications. These expansions have generated new adaptive variations in type I interferons and the interferon-stimulated gene *ANXA2R*, which potentially enhance antiviral states, as suggested by our functional assays. Genome-wide selection screens, including of candidate introgressed regions, uncover numerous putative molecular adaptations linked to immunity, including in viral receptors. By expanding taxon coverage to ten horseshoe bat species, we identify new variants of the SARS-CoV-2 receptor ACE2, and report convergent functionally important residues that could explain wider patterns of susceptibility across mammals. We conclude that horseshoe bats have numerous signatures of adaptation, including some potentially related to immune response to viruses, in genomic regions with diverse and multiscale mutational changes.

Bats are the implicated natural reservoir hosts of several zoonotic viruses. Of all bats, horseshoe bats (family Rhinolophidae) show a high detection frequency of coronaviruses[1] and have the closest associations with betacoronaviruses within the subgenus *Sarbecoviridae*[2], which include SARS-like viruses. These bats number more than 100 congeneric species, divided between two clades (Afro-Palaearctic and Asian)[3] that diverged ~17 million years ago (Ma)[4]. The Asian horseshoe bat *Rhinolophus sinicus* was identified as the source of SARS-CoV[5,6]—the cause of the 2002–2003 pandemic—and, more recently, horseshoe bats have also been reported as probable hosts of SARS-CoV-2 (refs. 7,8). To date, viruses with the highest sequence identity (>95%) to SARS-CoV-2 have been sampled from *Rhinolophus* species in Laos (*Rhinolophus malayanus*, *Rhinolophus pusillus* and *Rhinolophus marshalli*)[7], with other close matches isolated from *Rhinolophus affinis* in China[8]. Large-scale viral surveillance of Chinese horseshoe bats further supports their role as hosts of SARS-like viruses, with ~9% of bats sampled showing asymptomatic infection[9].

[1]State Key Laboratory of Virology and Biosafety, Key Laboratory of Biodiversity and Environment on the Qinghai-Tibetan Plateau, Ministry of Education, Frontier Science Center for Immunology and Metabolism, Hubei Key Laboratory of Cell Homeostasis, College of Life Sciences, Wuhan University, Wuhan, China. [2]Novogene Bioinformatics Institute, Beijing, China. [3]Guangdong Key Laboratory of Animal Conservation and Resource Utilization, Guangdong Public Laboratory of Wild Animal Conservation and Utilization, Institute of Zoology, Guangdong Academy of Sciences, Guangzhou, China. [4]College of Life Sciences, Shaanxi Normal University, Xi'an, China. [5]CAS Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China. [6]Guangzhou National Laboratory, Guangzhou International Bio Island, Guangzhou, China. [7]School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK. [8]These authors contributed equally: Shilin Tian, Junyu Si, Libiao Zhang. ✉e-mail: s.j.rossiter@qmul.ac.uk; huabinzhao@whu.edu.cn

Despite close interactions between horseshoe bats and sarbecoviruses, the molecular mechanisms underlying horseshoe bat innate viral immunity are poorly understood. Comparisons of intestinal organoids from *R. sinicus* and humans show higher interferon (IFN) expression in *R. sinicus*, and showed stronger induction of type III IFNs and interferon-stimulated genes (ISGs) when challenged with a synthetic double-stranded RNA virus mimic, but not with the ancestral SARS-CoV-2 (ref. [10]). Both organoids also showed susceptibility to SARS-CoV-2 and other coronaviruses, with *R. sinicus* organoids showing lower viral replication. Evidence for permissiveness to certain viruses also comes from work on horseshoe bat angiotensin convert enzyme 2 (ACE2) receptors, which contain specific residues that allow virus binding and cell entry, while enhanced viral immunity is supported by a substitution (S358) in the STING (also known as MITA) protein of all bats[11]. Horseshoe bats also possess molecular changes that dampen inflammatory responses, including shared gene losses (for example, PYHIN[12,13] and NF-κB regulators[13,14]) and a newly reported lineage-specific deletion in the pro-inflammatory protein ISG15 (ref. [1]).

A more complete understanding of antiviral immunity in horseshoe bats has been hampered by a lack of comparative genome data. In particular, highly contiguous genome assemblies offer opportunities to uncover large-scale mutational processes, including segmental duplications (SDs) and chromosomal rearrangements, both of which have been linked to IFN clustering in mammals[15,16]. Such mutational processes might also be important in the diversification of horseshoe bats, which show marked karyotypic diversity among Asian species (14–26 chromosome pairs), but not among Afro-Palaearctic species (29 pairs)[17]. To pinpoint molecular adaptations that help explain antiviral responses and immunity in horseshoe bats, including large-scale mutational changes, here we generate chromosomal-level assemblies for horseshoe bats and their relatives, as well as low-coverage genome datasets spanning ten species. By combining comparative and population-based analysis with in vitro assays of immune function, we reveal new mechanisms and show that horseshoe bats are characterized by distinct putative immunological adaptations that have evolved via diverse and multiscale mutational changes across the genome.

Results and discussion
Genome assembly and SD identification
We generated chromosome-level genomes for two horseshoe bat species (*Rhinolophus pearsonii* and *R. sinicus*), and three sister taxa, comprising two roundleaf bats (*Hipposideros armiger* and *Hipposideros pratti*) and one false vampire (*Megaderma lyra*). We applied a modified assembly procedure that uses Hi-C interaction pairs to cluster nanopore long sequences with potential linkages[18,19] (Supplementary Table 1 and Supplementary Note 1). The final genomes ranged from 2.06 to 2.14 Gb, with 18 (*R. sinicus*) to 27 (*M. lyra*) chromosome pairs (Extended Data Table 1 and Supplementary Tables 2 and 3), and showed high completeness, accuracy and contiguity based on standard measures (Fig. 1a and Supplementary Tables 4–6).

Repetitive sequences constituted 31.19–31.78% of four chromosome-level genomes of horseshoe bats, which are higher than for Old World fruit bats (25.50–27.72%), but lower than for their closest relatives and other bat genomes[18] (Supplementary Table 7). A similar trend was previously reported for transposable element content based on an extended dataset of 248 mammalian genomes[20,21]. For each genome, we also identified SD sequences (>90% identity and >1 kb length), and found that genome SD content was proportionally lowest in the four *Rhinolophus* species (1.23–1.89%), followed by the two *Hipposideros* species (1.98% and 2.42%) and then *M. lyra* (3.72%) (Supplementary Table 8), although all values were substantially lower than equivalent values reported for humans (5.59%) and mice (4.94%)[22]. We classified non-redundant SDs as SD blocks, and found that blocks

were composed of an average of 31.39% long interspersed elements (LINEs) in *Rhinolophus* and an average of 34.90% LINEs in *Hipposideros* (Supplementary Table 9). Across these assemblies, pairs of SDs more commonly occurred on the same chromosome than on different chromosomes. The number of large-scale SD blocks (>100 kb) was lowest in *R. affinis* and highest in *R. ferrumequinum*, although all bat genomes contained markedly fewer than the human genome, consistent with genome downsizing.

Phylogenetic analyses
We identified an average of 21,833 protein-coding genes for five new assemblies (Supplementary Tables 10–12). Together with the genes from 15 other mammals, we obtained 6,658 single-copy orthologues to build a maximum-likelihood (ML) tree[23] (Extended Data Fig. 1). The resulting phylogenetic tree showed that Rhinolophidae and Hipposideridae diverged ~37.5 Ma, consistent with both molecular and fossil evidence[24–26]. Next, we analysed 1.02 Tb of new resequencing data from 91 individuals of 10 *Rhinolophus* species, alongside 18 published *H. armiger* resequencing datasets as outgroups[27] (Supplementary Table 13 and Supplementary Note 2). Using 5,494,189 single nucleotide polymorphisms (Supplementary Table 14), we built a neighbour-joining (NJ) tree[28], and found that both ML and NJ trees recovered a monophyletic clade of Asian horseshoe bats that diverged ~15.6 Ma from the greater horseshoe bat of the Afro-Palaearctic clade (Fig. 1b). Within the Asian clade, we also recovered known relationships with an early split of *R. pearsonii*.

Chromosome rearrangements during *Rhinolophus* diversification
Marked karyotype variation among the four *Rhinolophus* species (*R. affinis* 2*n* = 62, *R. ferrumequinum* 2*n* = 58, *R. pearsonii* 2*n* = 44 and *R. sinicus* 2*n* = 36) implies an association between chromosome rearrangement and speciation[29]. We reconstructed the ancestral *Rhinolophus* genome[30] and identified evolutionary breakpoint regions (EBRs), observing a high level of synteny among genomes (Extended Data Fig. 2). Using the phylogeny of four *Rhinolophus* species, five bat outgroups and the horse, we inferred 30 ancestral chromosomes for horseshoe bats at a resolution of 300 kb, spanning ~99% of each horseshoe bat genome (Fig. 2a and Supplementary Table 15). The ancestral genome shared 50 large-scale homologous synteny blocks (HSBs) with *R. affinis*, 41 with *R. pearsonii*, 42 with *R. sinicus* and 32 with *R. ferrumequinum*, with 19, 19, 24 and 3 EBRs, respectively (Supplementary Table 16). From these EBRs and HSBs we identified, respectively, 1, 9, 17 and 1 chromosome fusion events, and 2, 2, 2 and 0 fission events. The two fission events in the three Asian species appear to be shared, having occurred before their split with *R. ferrumequinum* (Fig. 2a). The fission sequences from the two ancestral chromosomes formed four independent chromosomes in *R. pearsonii* and *R. affinis*, and were involved in two chromosome fusions in *R. sinicus*. We verified the two chromosomal fissions by performing comparative analyses of the interaction signal map using Hi-C data between *R. ferrumequinum* and the other three *Rhinolophus* bats (Fig. 2b and Extended Data Fig. 3).

Taken together, *R. ferrumequinum* showed the fewest large-scale HSBs and EBRs, and only one chromosome fusion, implying its genome structure is most similar to that of the ancestral horseshoe bat. We also find evidence that horseshoe bat diversification was accompanied by chromosome fusions (Fig. 2a) that led to longer but fewer chromosomes, as seen in *R. pearsonii* and *R. sinicus*. By contrast, the X chromosome shows no large-scale chromosome fusions during *Rhinolophus* evolution, consistent with evolutionary conservation across mammals[31–33]. Finally, Gene Ontology (GO) analyses of EBR genes point to enrichment for roles in the transmission and processing of sensory information in the olfactory system in four horseshoe bats (Supplementary Table 17 and Supplementary Note 3).
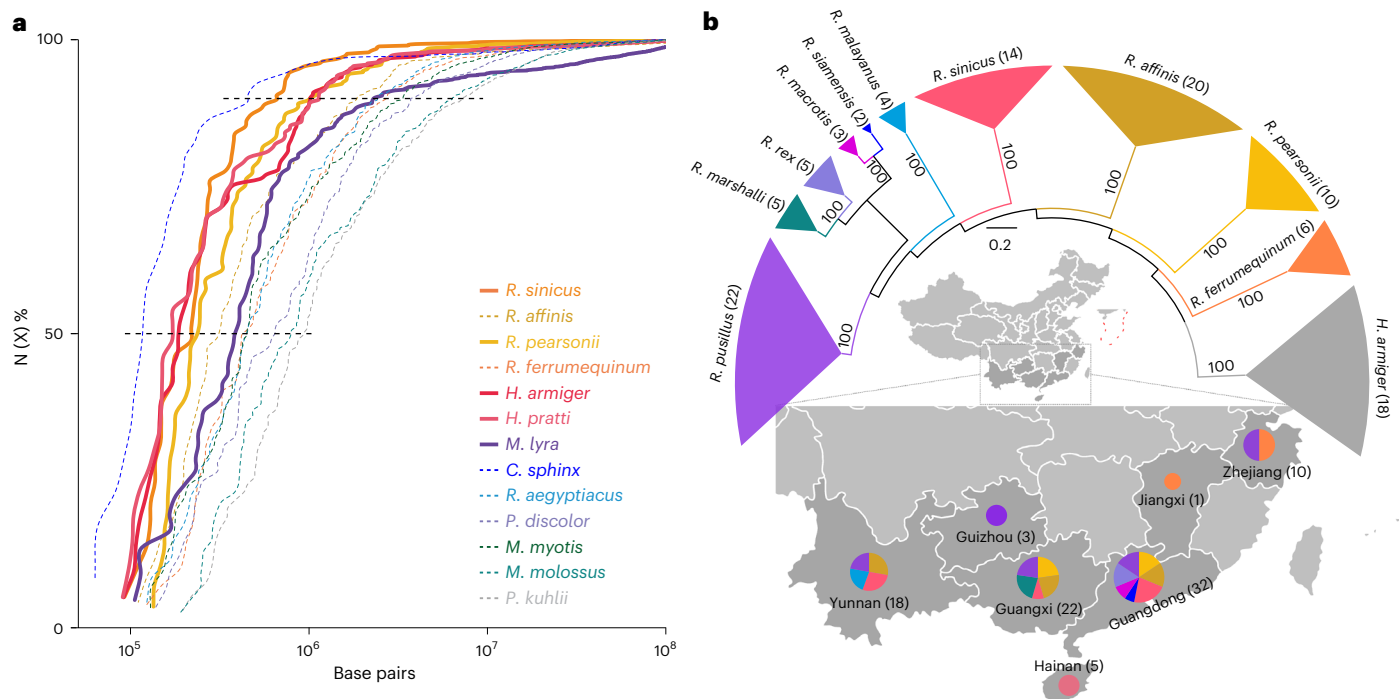
Fig. 1 | Genome assembly quality and phylogenetic relationships in the Rhinolophidae. a, Genome contiguity of new assemblies (solid lines) compared with seven other chromosome-level bat assemblies (broken lines). The x axis shows cumulative contig size, with combined lengths shown as a X% of total genome length. Horizontal black dashed lines indicate contig N50 and N90 values. **b**, NJ tree based on 91 genomes of ten *Rhinolophus* species collected from six provinces in China. Pie charts show the geographical distribution of samples obtained. All nodes had 100% bootstrap support. *C. sphinx*, *Cynopterus sphinx*; *M. molossus*, *Molossus molossus*; *M. myotis*, *Myotis myotis*; *P. discolor*, *Phyllostomus discolor*; *P. kuhlii*, *Pipistrellus kuhlii*. Credit: **b**, The Map of China was plotted by using the R package 'hchinamap', © Z. Cheng, 2019.

Repeats and SDs as a potential driver of genome evolution

Previous studies suggest EBRs and associated chromosomal rearrangements might stem from genomic instability caused by repetitive and duplicated elements[29,34–37]. We find that EBRs in *Rhinolophus* contain numerous repetitive sequences, particularly LINE L1 elements[35–37] (Supplementary Table 18). SDs might serve as a template for chromosomal rearrangements, as suggested for primates[29]. We related chromosome fusion and fission events, inferred from comparing HSBs positions in ancestral and extant taxa, to the location of SDs. We found that chromosome fusion regions in the *Rhinolophus* genomes were enriched for SDs compared with the overall genome (mean 55.48% versus 1.59%; Supplementary Table 18). Notably, two fixed fissions were located in the SD dense region of *R. ferrumequinum* (Fig. 2b).

To explore the potential role of SDs in karyotypic evolution in horseshoe bats we focused on *R. sinicus*, which shares several ancestral chromosomal rearrangements with other Asian species and might thus serve as a model for the group. In this taxon, chromosomes 12 and 14 originated from the fission of ancestral chromosome 14, whereas chromosomes 14 and 17 formed by the fission of ancestral chromosome 21 (Fig. 2b). We recorded abundant SDs in chromosomal regions adjacent to inferred fission sites in the ancestor. For example, 11 interchromosomal SDs occurred between the collinear regions on chromosomes 12 and 14, and 8 interchromosomal SDs between chromosomes 14 and 17 in *R. sinicus* (Supplementary Table 19). Interestingly, we also detected an extra-long intrachromosomal SD (~51 kb) in *R. sinicus* in the fusion region between ancestral chromosomes 14 and 21, and a long SD (~18 kb) in the fusion region between ancestral chromosomes 14 and 27. Although previous cytogenetic analyses have also reported chromosomal fusions in horseshoe bats, the mechanism for such chromosomal rearrangements is unclear[17,38,39]. Based on our findings, we suggest that horseshoe bat diversification involved chromosomal rearrangements that were probably facilitated by SDs, which provided a template for non-allelic homologous recombination events (Fig. 2c).

Expansion of the *ANXA2R* gene family in horseshoe bats

We examined gene family evolution using CAFE[40] and found 14 gene families (85 loci) showing evidence of expansion on the horseshoe bat ancestral branch (Viterbi *P* value <0.05; Extended Data Fig. 1). Despite this, horseshoe bats showed no overall difference in rate of gene family evolution compared with other bats. GO analyses of these 85 expanded genes identified 235 significant GO terms, of which 60 (25%) are related to immunity, including 'immune response' (GO:0006955, 23 genes of 1,614 genes in set) and 'cellular response to type I interferon' (GO:0071357, 6 of 65 genes) (Supplementary Table 20).

Eight expanded gene families (57.14%), containing 21 loci, had at least one copy located in an SD region in the ancestral *Rhinolophus*, a pattern also seen in species-specific branches (Supplementary Fig. 1). These included three loci related to viral immunity: *HS71B* (four to six copies), *BTNL* (three to eight gene members) and *ANXA2R* (three to six copies) (Supplementary Table 21). *HS71B* encodes Heat-shock 70 kDa, which might serve as a viral receptor and inhibit viral replication[41,42], whereas *BTNL* encodes a family of butyrophilin-like proteins, members of which can attenuate tissue-related inflammatory responses[43]. *ANXA2R* encodes the Annexin A2 receptor, an ISG that is known to suppress virus replication[44,45]. Intriguingly, most *ANXA2R* copies were seen in *R. sinicus*, which has been closely linked to SARS-like viruses. We also detected an independent SD-mediated expansion on the ancestral branch of Pteropodidae (Fig. 3a and Extended Data Fig. 4), a related group that also shows widespread molecular adaptations for viral tolerance[18]. To further explore a potential adaptive basis for *ANXA2R* copy number variation in horseshoe bats, we performed gene-species tree reconciliation[46] and detected multiple copies on the ancestral branches of the Asian (*n* = 3) and Afro-Palaearctic (*n* = 2) clades (Fig. 3b). We also
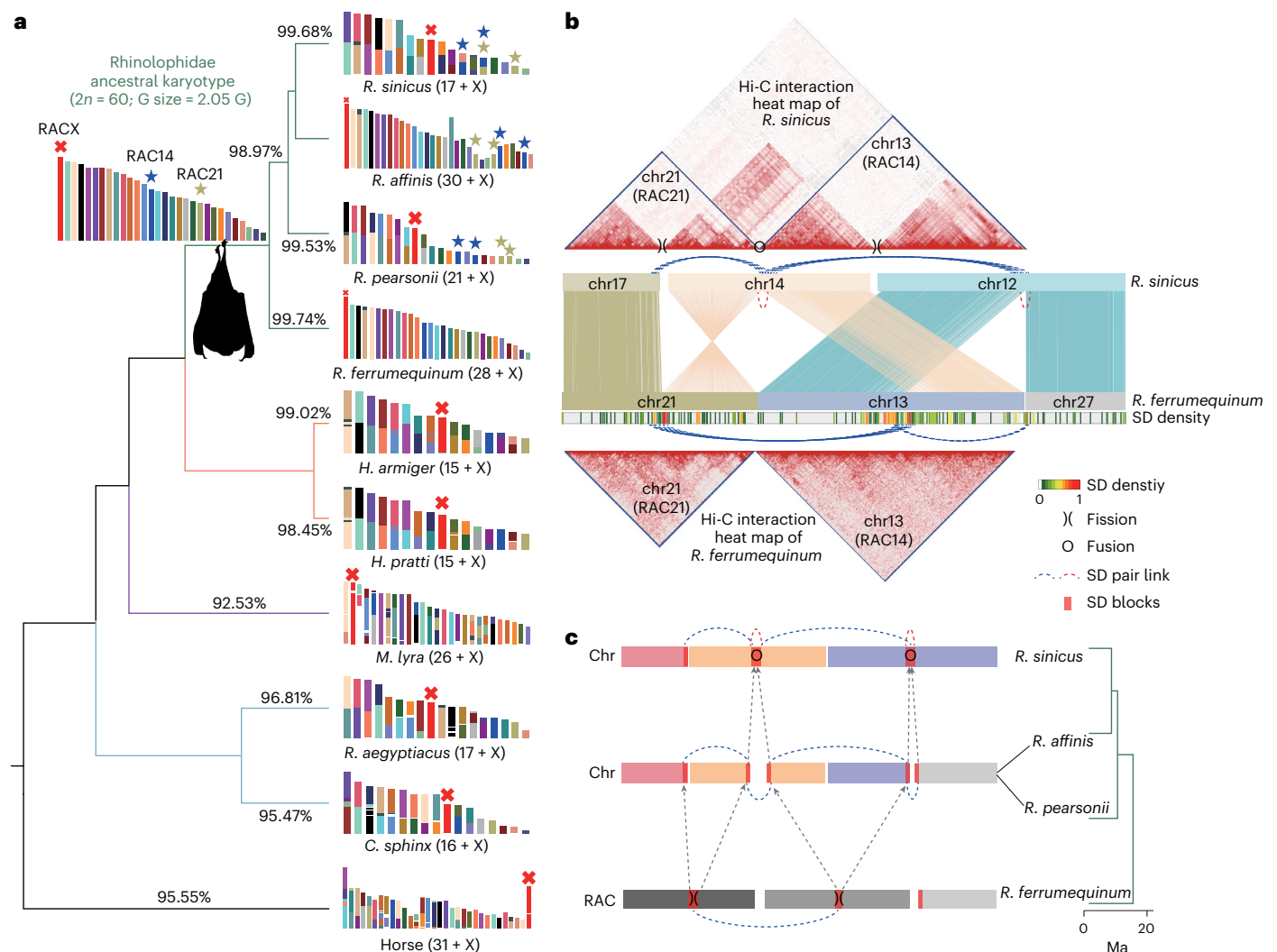
Fig. 2 | Chromosomal evolution in the Rhinolophidae. a, Cladogram of focal *Rhinolophus* species and outgroups showing the ancestral karyotype. The ancestral X chromosome (RACX, red cross) is shown, along with ancestral chromosomes 14 (RAC14, blue asterisk) and 21 (RAC21, green asterisk), both of which have undergone fission during species diversification. The different branch colours represent different bat families. The numerical values on the branch labels denote the percentage of each species covered by the ancestral genome. **b**, Chromosomal rearrangements in *R. sinicus*. The top panel displays a heat map of the interaction signal after aligning *R. sinicus* Hi-C data to the *R. ferrumequinum* genome. The middle panel shows the synteny and sequence characteristics related to the two ancestral chromosome fissions (asterisks in **a**)

between *R. sinicus* and *R. ferrumequinum*. Blue arcs indicate short SDs (<10 kb) and red arcs represent long SDs (>10 kb). Detailed information about these SDs is provided in Supplementary Table 19. The heat map is the density distribution of SDs in the *R. ferrumequinum* chromosomes. The bottom panel displays a heat map of the interaction signal after aligning the *R. ferrumequinum* Hi-C data to its genome. **c**, Potential model for chromosomal evolution in the Rhinolophidae in which SDs provide a template for non-allelic homologous recombination. Different chromosomes are depicted as different coloured blocks. chr, chromosome; G size, genome size. Credit: **a**, *Rhinolophus* silhouette from phylopic.org, M. Michaud, under a creative commons license CC0 1.0.

inferred more intact copies in horseshoe bats than either other bats (0 to 4 copies) ($P = 0.01$, Student's t-test) or other mammals (0 or 1 copy) ($P < 0.001$, Student's t-test). In addition, three putative *ANXA2R* genes with non-ATG start codons (that is, ACG) were identified in the family Hipposideridae (Fig. 3c). These non-ATG start codons may result in relatively weaker translation initiation efficiency compared to the standard ATG codons[47].

Previous work has shown that *ANXA2R* overexpression induces cell death in human cells[48]. Moreover, high levels of antibodies against its main ligand, AnxA2, have been linked to mortality in COVID-19 patients[49], suggesting a potential role of ANXA2R in the response to SARS-CoV-2 infection. Given that ANXA2R's function is well documented and its assay is established[44,45,48], we assessed the impact of varying ANXA2R expression on the replication of a SARS-CoV-2 virus-like particle, aiming to determine whether additional ANXA2R

copies might confer an enhanced antiviral state. We first established a Caco-2-N cell line, which expresses the nucleocapsid (N) protein of SARS-CoV-2, to support the replication of SARS-CoV-2 GFP/ΔN, a replication-competent SARS-CoV-2 virus-like particle in which the SARS-CoV-2 N gene is replaced with a green fluorescent protein (GFP) reporter gene[50]. Next, ANXA2R sequences for humans and *R. affinis*, along with a negative control using human-THO complex subunit 4 (Thoc4, as described previously[44]) were introduced into Caco2-N cells using a lentivirus vector, in which ANXA2R expression is regulated through a doxycycline-inducible gene expression system. In this expression system, higher doxycycline concentrations induce higher expression levels of ANXA2R[51]. After treatment with different doxycycline concentrations (0, 10, 100 µg ml⁻¹) for 24 h, the expression of ANXA2R and human-Thoc4 in Caco2-N cells was confirmed by immunofluorescence of the N-terminal haemagglutinin (HA)-tag
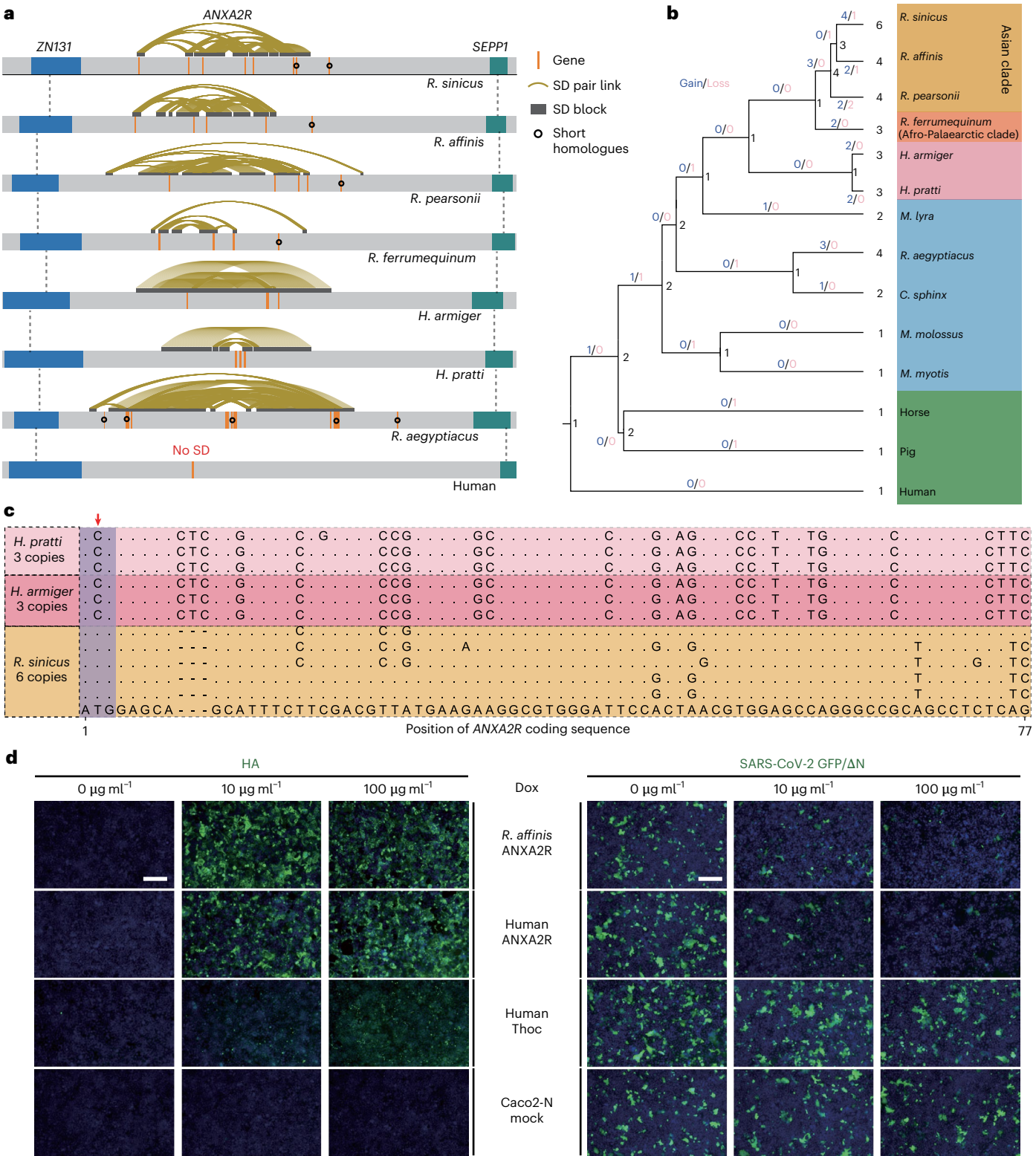
Fig. 3 | Expansion of the *ANXA2R* gene. a, Schematic showing location of the *ANXA2R* gene between *ZN131* (blue) and *SEPP1* (green) across all species examined. Positions of SD blocks (dark grey blocks connected with coloured lines) and short homologues (black circles) are shown. **b**, Cladogram showing *ANXA2R* gene evolution based on nine bats and three outgroups (human, pig and horse). Numbers of gene copies are given on the tips. Inferred gene gains (blue) and losses (pink) are shown on the nodes. **c**, Alignment of *ANXA2R* coding sequence for *R. sinicus, H. armiger* and *H. pratti*. The start codon for *R. sinicus* (purple box) with missense mutations in Hipposideridae (red arrow) are shown. **d**, Functional assays in *Rhinolophus* ANXA2R. The left panel displays

the expression of ANXA2R proteins on Caco2-N cells, detected by the N-terminal HA-tag, after treatment with a gradient increase in doxycycline concentrations (0, 10, 100 µg ml^{-1}) for 24 h. GFP indicates the expression of the HA-tag protein. The right panel demonstrates the reduced replication of SARS-CoV-2 GFP/ΔN (multiplicity of infection = 0.1) virus on Caco2-N cells 24 h post-infection. GFP fluorescence in cells represents the SARS-CoV-2 GFP/ΔN replication progeny. The cell nuclei are stained with Hoechst 33342 in blue. Human-Thoc4 serves as the control. Mock, Caco2-N cells. Scale bar, 200 µm. Results were consistent across three biological replicates.
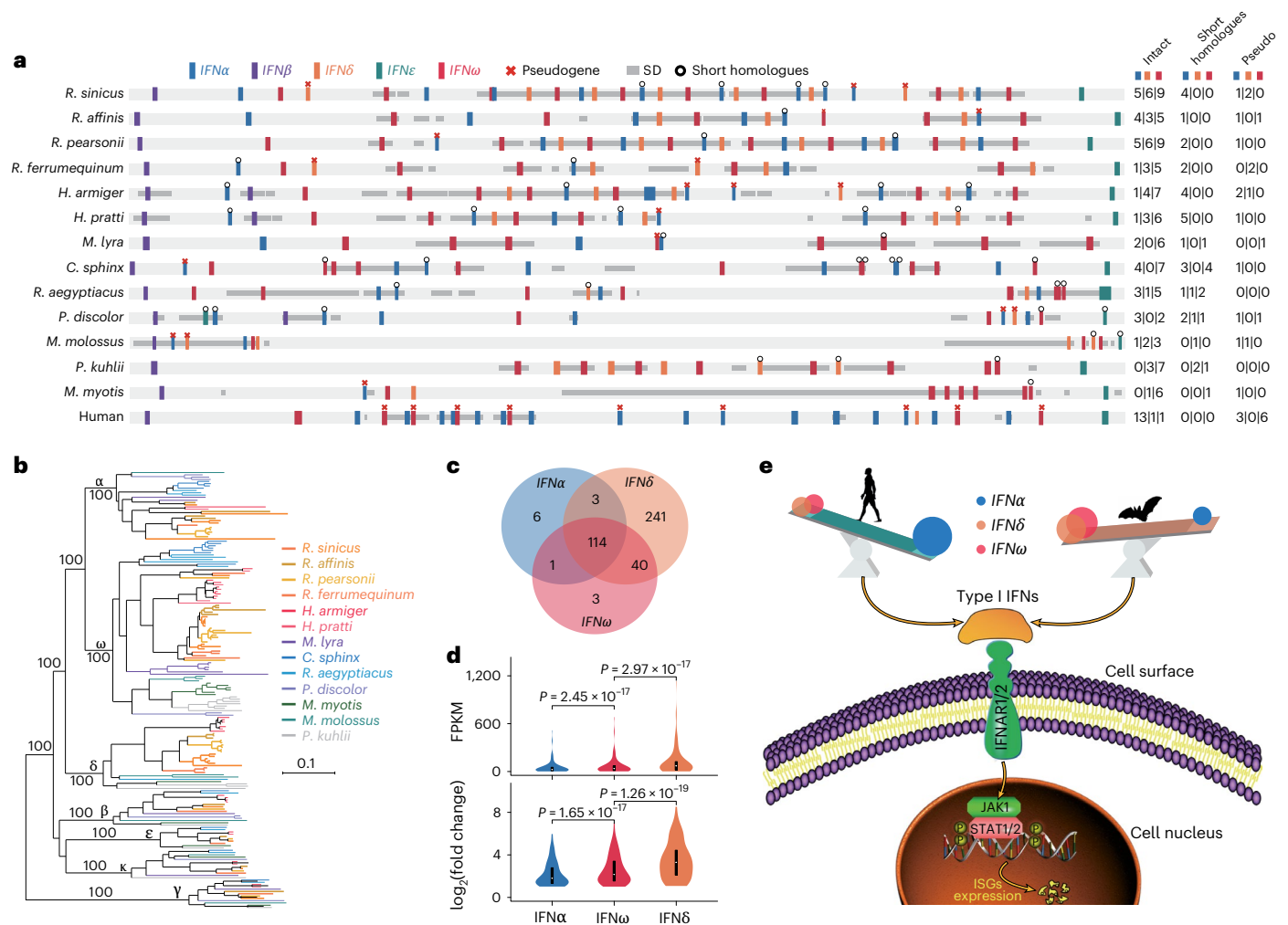
Fig. 4 | Diversity of type I IFN genes. a, Schematic showing distribution of type I IFNs in the 13 bat genomes and human outgroup. Left, positions of type I IFNs (coloured blocks), pseudogenes (red crosses), short homologues (black circles) and SDs (dark grey blocks) are shown. Right, summary counts for intact genes, short homologues and pseudogenes are given for *IFNα* (blue), *IFNδ* (orange) and *IFNω* (red). **b**, Phylogenetic ML tree of all type I IFNs in bat lineages. **c**, Venn diagram illustrating shared putative ISGs induced by *IFNα*, *IFNω* and *IFNδ*. **d**, Violin plot depicting the comparative expression levels and expression fold change for putative ISGs induced by *IFNα* ($n = 124$), *IFNω* ($n = 158$) and *IFNδ* ($n = 398$). Data are presented as the mean ± s.d. The centre dot represents the median; the box limits represent the upper and lower quartiles; the whiskers

represent 1.5× the interquartile range (IQR). Statistical significance was determined using a two-tailed Mann–Whitney *U*-test. **e**, Schematic diagram showing possible compensatory adaptation among type I IFN subfamilies in horseshoe bats versus humans. Large, medium and small balls represent high, medium and low copy numbers, respectively. Within each taxon, the tilt of the seesaw reflects differences between numbers of *IFNα* gene copies versus *IFNδ* and *IFNω* gene copies. The silhouettes of human and bat is from phylopic.org. FPKM, fragments per kilobase of transcript per million mapped reads. Credit: **e**, Human and bat silhouettes from phylopic.org, T. M. Keesey and Y. Wong, respectively, under a Creative Commons license CC0 1.0.

(Supplementary Fig. 2). Finally, we performed an experiment by exposing the doxycycline-treated cells to SARS-CoV-2 GFP/ΔN at a multiplicity of infection of 0.1. GFP fluorescence images were captured 24 h post-infection to analyse the potential antiviral effects of ANXA2R.

We observed lower GFP intensity (SARS-CoV-2 GPF/ΔN virus) with increased doxycycline concentration in Caco2-N cells expressing ANXA2R (either human or *R. affinis*), suggesting fewer cells contain detectable virus when ANXA2R expression is elevated (Fig. 3d). By contrast, no such GFP reduction was seen in the control. Therefore, increased dosage of ANXA2R may act to suppress viral replication, although expanding these experiments to encompass a broader range of viruses will be important to test this hypothesis.

Potential genetic compensation in type I IFNs

We examined the evolutionary history of type I IFN gene families, which are strongly associated with antiviral immune responses across vertebrates[52]. Six type I IFN subfamilies (*IFNα*, *IFNβ*, *IFNδ*, *IFNε*, *IFNκ*

and *IFNω*) were present in all 13 chromosomal-scale bat genomes, located on one chromosome, as also reported for humans[53]. With the exception of *IFNκ*, all type I IFN genes were located in genomic regions rich in SD blocks, with *IFNβ* and *IFNε* in the outermost positions. Moreover, most bat *IFNω* and *IFNδ* genes were seen to occur in SDs, unlike in humans, in which only *IFNα* are reported to occur in SDs (Fig. 4a). Focusing on the four horseshoe bat genomes, we counted multiple intact gene copies for IFNα (one to five copies), IFNω (five to nine copies) and *IFNδ* (three to six copies), and single copies for the other subfamilies. Similar patterns of polymorphism were seen in Hipposideridae (two copies of IFNβ) and Pteropodidae. Our count of 5 *IFNω* genes and 1 *IFNδ* gene in the recently published genome[14] of *Rousettus aegyptiacus* disagrees with previous reports of 22 *IFNω* loci based on an earlier scaffold-level assembly (Raegyp2.0)[54]. Synteny analysis revealed that Raegyp2.0 contains numerous duplicated scaffold sequences carrying *IFNω* (Extended Data Fig. 5), suggesting that earlier counts might be overestimates.

To visualize IFN evolution in bats, we constructed a NJ tree for all type I IFN genes, using the *IFNγ* gene from type II IFN as an outgroup. Each subtype I IFN formed a well-supported monophyletic clade, consistent with duplications pre-dating bat diversification, as seen in other vertebrates (Fig. 4b). Using the accepted species tree, we applied CAFE to test for gene expansions and contractions and found significant contraction of *IFNα* loci on the ancestral branch of all bats, alongside expansions of *IFNδ* and *IFNω* loci on the ancestral branch of Rhinolophidae, and in *R. sinicus* and *R. pearsonii* (Extended Data Fig. 6). We also found evidence of an independent expansion of *IFNβ* in Hipposideridae (Fig. 4a). By generating a reconciliation tree[46], we inferred a marked gain in intact *IFNδ* and *IFNω* loci at the base of Rhinolophidae followed by further lineage-specific gains in all *Rhinolophus* species, with *R. sinicus* and *R. pearsonii* each seen to possess the highest numbers of copies (nine *IFNω* and six *IFNδ*), followed by *R. affinis* and *R. ferrumequinum* (five *IFNω* and three *IFNδ*) (Extended Data Fig. 7).

To test whether the diversification of type I IFNs in horseshoe bats has been potentially driven by molecular adaptation, we also built gene trees for each type I IFN subfamily. We applied the branch-site test[55] to the ancestral branch of horseshoe bats—which consistently formed a monophyletic group—and found that 14.39% of sites for *IFNδ* were under positive selection (ω2 (i.e. the ω for the foreground branch) = 4.44; $P = 2.46 \times 10^{-5}$), 16.96% for *IFNω* (ω2 = 6.55, $P = 3.63 \times 10^{-5}$) and 13.40% for *IFNβ* (ω2 = 7.14, $P = 3.78 \times 10^{-4}$) (Supplementary Table 22). We also applied a clade model to these three gene trees, and found divergent selection pressures between Rhinolophidae and other bats, with levels of positive selection consistently higher in the former (Supplementary Table 22).

Finally, to examine the effect of horseshoe bat IFNα, IFNδ and IFNω on gene expression, we analysed published transcriptome data from stimulated RfKT cells (immortalized *R. ferrumequinum* kidney cells)[56], and compared expression profiles of the induced putative ISGs (Supplementary Note 5 and Supplementary Fig. 3). Comparisons of each treatment group (*n* = 3) with a control revealed consistent patterns in which 86.3–91.1% differentially expressed genes (DEGs) were upregulated (Supplementary Tables 23–25). IFNδ induced most putative ISGs (*n* = 398) followed by IFNω (*n* = 158) and IFNα (*n* = 124) (Fig. 4c,d). The majority of the 114 putative ISGs induced by all three IFNs showed the same upward trend in expression from IFNα to IFNω to IFNδ based on expression levels (*n* = 87, 87 of 114 or 76.32%) and fold changes (*n* = 92, 92 or 114 or 80.70%) (Supplementary Fig. 4). GO analysis of putative ISGs induced by IFNα, IFNδ and IFNω revealed 626, 812 and 1,249 significant terms, respectively, with 510 shared terms (Supplementary Tables 26 and 27). For shared GO terms, we ranked each list based on corrected *P* values. We then ranked by the sum of the three ranks (Supplementary Table 28) and found a strong signal of antiviral immunity throughout, with the top ten terms including 'defence response to virus' (GO:0051607) and 'innate immune response' (GO:0045087). However, despite the role of IFNs in inflammasome activation[57–59], the highest term associated with inflammation ('inflammatory response'; GO:0006954) was ranked 303rd.

Our results imply that horseshoe bats have undergone an adaptive expansion of their *IFNδ* and *IFNω* loci, following an ancient contraction of *IFNα* loci. Loss of some *IFNα* loci in bats might seem surprising

given it is the largest type I IFN subfamily in many vertebrates[60–62]; however, our functional data suggest that, in horseshoe bats, IFNα has relatively weaker antiviral activity compared with that of IFNδ and IFNω. Thus, *IFNδ* and *IFNω* subfamily expansions in horseshoe bats might represent a form of genetic compensation (Fig. 4e). Indeed, *IFNδ* activates the JAK-STAT pathway to produce antiviral proteins by binding to type I IFN receptors[63], whereas *IFNω* shows similar antiviral and immunomodulatory functions to *IFNα*[62,64,65], based on its expression and ability to induce ISGs.

Widespread molecular adaptations in single-copy immune genes

To assess whether adaptive gene duplications for antivirus immunity are complemented by molecular adaptations in single-copy genes, we used PAML[66] to analyse 6,658 orthologues in our comparative genome dataset (Extended Data Fig. 1). We identified 523 genes that showed evidence of both positive selection and rapid evolution on the ancestral horseshoe bat branch (Supplementary Table 29). GO analysis of these genes revealed 197 significant terms, including three related to immunity, ranked 13th, 58th and 152nd respectively: 'viral process' (GO:0016032; 40 of 528 genes; *P* < 0.001), 'immune system process' (GO:0002376; 92 of 2,462 genes; *P* < 0.001) and 'virus receptor activity' (GO:0001618; 8 of 80 genes; *P* = 0.03) (Supplementary Table 30). We repeated these analyses for three additional bat families, Hipposideridae, Pteropodidae and Vespertilionidae, and recorded similar percentages (84.48–89.83%) of lineage-specific genes, but lower percentages (79.55–85.45%) of lineage-specific immune-related genes for all three GO terms (Supplementary Fig. 5). Moreover, gene sets for other families showed no enrichment for 'virus receptor activity' (GO:0001618) (Supplementary Table 31).

Among the 523 genes showing positive selection and rapid evolution in horseshoe bats were several candidates previously linked to pathogenic viruses, including bat zoonoses. *C5aR1* encodes the C5a complement factor, a key component of the complement cascade (C5a–C5aR1 axis) that is critical for sensing and clearance of pathogens and inflammation[67,68], and which appears to induce inflammation following infection by Middle East respiratory syndrome[69,70] and SARS-CoV-2 (ref. 71). All ten horseshoe bats showed three unique replacements in C5aR1 (199Q/K, 278T and 322M) (Extended Data Fig. 8a). The mutations at E199 residue, known to reduce human C5aR1's binding affinity to C5a[72,73], are predicted to alter physicochemical properties and induce conformational changes in the horseshoe bat C5aR1 (Extended Data Fig. 8b). We therefore speculate that the two specific E199 substitutions in *Rhinolophus* bats alter ligand binding affinity of the complement cascade, potentially attenuating inflammatory responses to viruses.

Eight of the 523 genes under selection in horseshoe bats encode receptors for pathogenic viruses (Supplementary Table 32). Ephrin-B2 (EFNB2) is a functional receptor for Hendra[74], ACE2 mediates cellular entry of both SARS-CoV[75] and SARS-CoV-2 (ref. 76), and neural cell adhesion molecule 1 (NCAM1) is a potential receptor for the rabies virus[77]. Similarly, scavenger receptor class B member 2 (SCARB2) has been confirmed as a receptor for enterovirus 71 (ref. 78), and supports efficient viral entry in horseshoe bats[79]. Among the receptors, ACE2 was previously reported to show positive selection in horseshoe bats

Fig. 5 | Evaluation of ACE2 critical sites determining SARS-CoV-2 binding and entry among species and individuals. a, Amino acids in the ACE2 RBD that interact with SARS-CoV-2, based on human RBD/SARS-CoV-2 (PDB 6M0J). Variable residues shown in red were examined in the assays. **b**, Immunofluorescence assays of bat ACE2 orthologues and related mutants in HEK293T cells showing differences in protein expression levels (top), binding efficiency with a purified recombinant SARS-CoV-2 RBD human IgG Fc fusion protein (middle) and cell entry of a SARS-CoV-2 S protein-bearing VSV (bottom). All scale bars are marked in the diagram. Results were consistent across two biological replicates. **c**, Ability of the rACE2 and related mutants to support

the entry of coronavirus pseudotypes. HEK293T cells expressing the rACE2 and their mutants were infected with SARS-CoV-2 using luciferase. Infection was analysed at 20 h post-infection. Error bars are presented as the mean ± s.d. (*n* = 3 technical replicates). Statistical significance was determined using a two-tailed Student's *t*-test and shown as a *P* value. Results were consistent across two biological replicates. **d**, Binding affinities between ACE2 and SARS-CoV-2 RBD shown for different protein concentrations. The *y* axis shows the real-time binding response. $K_{D,app}$ is the apparent binding affinity (Methods). Rhipea2-MUT contains two replacements (H41Y and E42Q). Species name abbreviations follow Supplementary Table 13. MUT, mutant-type; WT, wild-type.

a

b

c

d
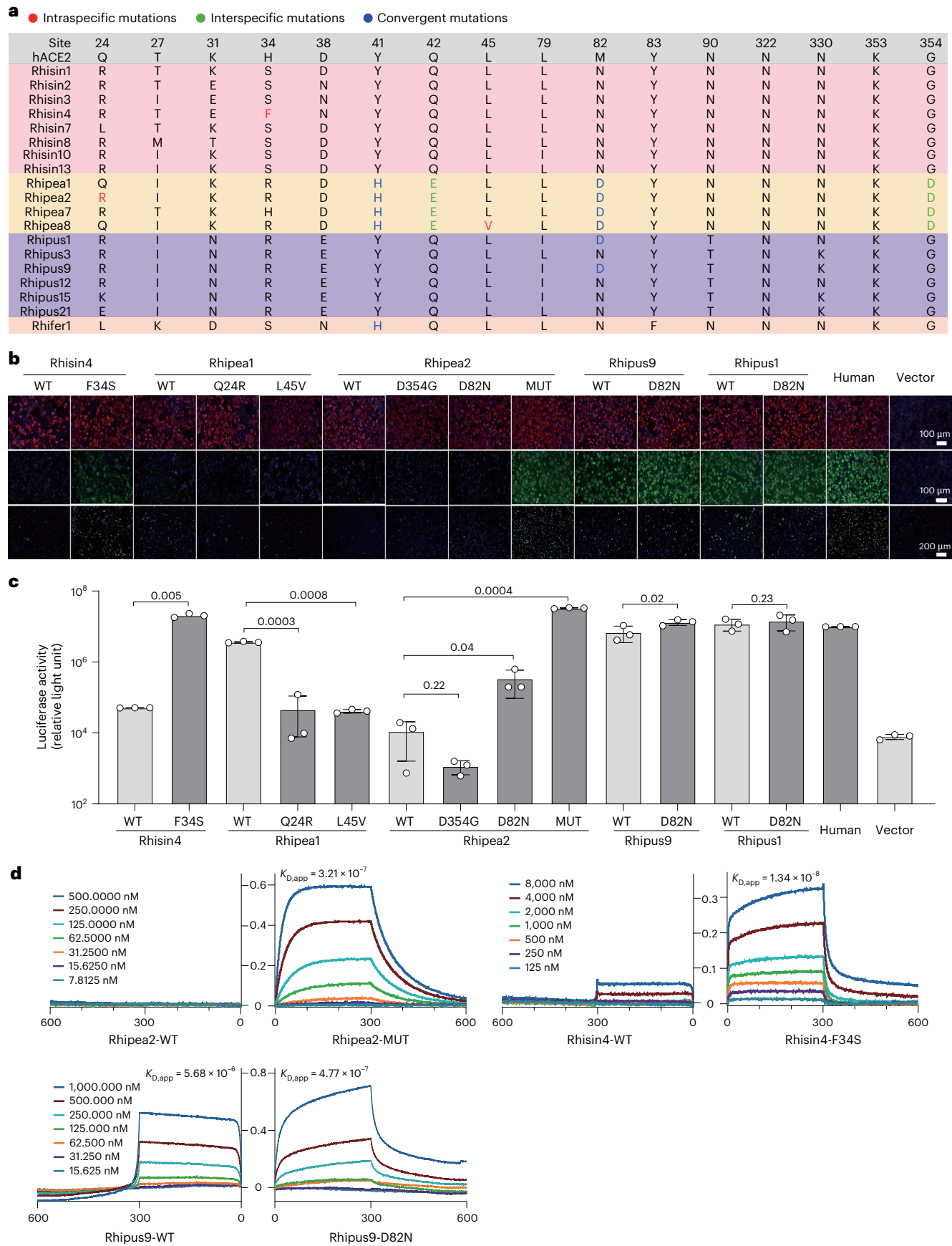
as well as functional variation with respect to sarbecoviruses[80]. To build on earlier findings for ACE2, we compared gene sequences in our extended dataset of 91 *Rhinolophus* genomes and recorded 16 sites that are known to interact directly with the receptor binding domain (RBD) of SARS-CoV-2 (ref. 81) (Supplementary Table 33). These sites combined to form 30 haplotypes across 10 species, of which 21 (70%) have not been described elsewhere (Supplementary Table 34).

Functional assays performed on all 30 ACE2 variants revealed consistently strong SARS-CoV-2 RBD binding in six species (*n* = 11 variants), no binding in *R. ferrumequinum* (*n* = 1) and *R. pearsonii* (*n* = 4), and variable binding in two species (*R. sinicus* and *R. pusillus*) (Supplementary Note 6, Supplementary Fig. 6 and Extended Data Fig. 9a,b). We also tested the influence of ACE2 phenotype on cellular entry of a SARS-CoV-2 S pseudotyped vesicular stomatitis virus (VSV), and found efficient infection in all but three variants from *R pearsonii* and one from *R. sinicus* (Supplementary Fig. 7 and Extended Data Fig. 9c). We pinpointed 16 ACE2 residues that correlated with intra- and interspecific phenotypic variation in our assays. Replacements associated with potentially reduced RBD binding were seen in *R. sinicus* (S34F) and *R. pearsonii* (Q24R and L45V) as well as between species (Y41H, Q42E, N82D and G354D; Fig. 5a). Among these, Y41 and Q42 were previously reported to affect hydrogen bond formation in the ACE2/SARS-CoV-2 spike complex in bats[82]. Finally, we examined all 16 sites for evidence of positive selection (bayes empirical bayes probability value > 0.95) based on our PAML results and found one site (S34) in common to both sets (Supplementary Table 35).

To confirm the importance of these residues in explaining the observed phenotypic variation in ACE2–SARS-CoV-2 RBD interactions, we used site-directed mutagenesis to generate ACE2 single and double mutants, and compared these with their respective wild-type proteins in cell-based functional assays (Supplementary Note 7). RBD binding and SARS-CoV-2 entry were both increased by introducing single mutations into individuals of *R. sinicus* and *R. pusillus* and double mutations in *R. pearsonii*. Similarly, other single mutations in *R. pearsonii* resulted in altered SARS-CoV-2 virus entry (Fig. 5b,c). Enhanced RBD binding in mutants was also confirmed using bio-layer interferometry assays, which showed higher affinity than the respective wild-types (Fig. 5d). Finally, to determine the potential molecular basis of altered RBD binding, we modelled the structure of the ACE2–SARS-CoV-2-spike complex (Supplementary Note 7). We found that in both the single mutant of *R. sinicus* (F34S) and double mutant of *R. pearsonii* (H41Y and E42Q), the increased binding was associated with a reduction in distance between ACE2 and the SARS-CoV-2 spike protein, alongside the formation of an additional hydrogen bond with the residues in the RBD region (Supplementary Figs. 8 and 9).

Our results provide compelling evidence of differences in susceptibility to SARS-CoV-2 among horseshoe bat species, as well as among individuals within species, supporting insights from fewer species[82–86]. In particular, the ACE2 receptor of *R. pearsonii* appears to show less susceptibility to SARS-CoV-2 infection when compared with other Asian species. Interestingly, the associated replacements H41 and E42 also occur in New World primates, whereas V45 occurs in some mongooses and civets (Supplementary Table 36), and it is noteworthy that these lineages have also been predicted to show lower susceptibility to SARS-CoV-2 compared with other lineages in their respective orders[81]. Further cases of molecular convergence are also seen in our results, with Y41H and N82D—associated with attenuated viral binding—showing independent evolution in *R. pearsonii* and *R. ferrumequinum*[82], and in *R. pearsonii* and *R. pusillus*, respectively. Because high susceptibility may predispose certain taxa to infection, comparing the effects of different ACE2 genotypes may help reveal potential reservoir hosts[81,83]. Although previous surveys suggest that the coronaviruses with the closest RBD sequence similarity to SARS-CoV-2 are found in *R. malayanus* and *R. pusillus* (BANAL-52 and BANAL-103, respectively)[7], we report that both of these species possess ACE2 variants that facilitate binding and invasion of SARS-CoV-2 (Extended Data Fig. 9).

Genetic introgression in Rhinolophidae

Recent genome studies show that molecular adaptations underpinning virus immunity can also be acquired via adaptive introgression among related taxa[87–90]. To test for genetic introgression among the ten *Rhinolophus* species, we ran ABBA–BABA tests[91] and identified 38 four-taxon combinations with genome-wide excess of shared derived alleles (Fig. 6a and Supplementary Note 8, Supplementary Figs. 10 and 11 and Supplementary Tables 37 and 38). Focusing on the four-taxon combination (*R. sinicus*, *R. pearsonii*, *R. ferrumequinum*, *H. armiger*) with the highest excess of shared derived alleles, we identified 436 'introgressed windows', containing 735 genes (Fig. 6b). Introgressed windows showed lower genetic differentiation compared with the wider genome (fixation index (F_{ST}) = 0.64 versus 0.67, P = 2.01 × 10^{-7}, Mann–Whitney U-test) (Supplementary Fig. 12) and were significantly enriched for selective sweep signals from horseshoe bats of the Asian clade (Supplementary Table 39). To test for possible adaptive introgression, we examined whether genes in putative introgressed windows show functional enrichment[92]. After accounting for spatial clustering of functionally related genes, we detected 43 significant GO terms for biological processes, including 21 (~50%) related to immune function (Fig. 6c and Supplementary Table 40). Notably, we found that 186 (25%) of putative introgressed genes were also positively selected genes (PSGs) based on site model tests[66], including 40 PSGs significantly associated with two immune-related GO terms (P = 0.04, χ^2 test; Supplementary Tables 41 and 42). GO analysis of the 186 PSGs, using all 735 introgression loci as the background, revealed 53 significant GO terms, including 'viral genome replication' (GO:0019079) and 'regulation of inflammatory response' (GO:0050727) (Supplementary Table 43). Thus horseshoe bats might have undergone adaptive introgression of immune-related genes, including loci with roles in altering responses to viruses.

Conclusions

Our analyses indicate that horseshoe bats have evolved numerous putative immunological adaptations via diverse and multiscale mutational changes. SDs underlying chromosomal rearrangements during diversification have generated new expansions of the ANXA2R gene family alongside type I IFNs. Cell-based assays confirm that higher ANXA2R dosage inhibits viral replication in human cells, and that IFNδ and IFNω invoke more potent immune responses than the contracted IFNα, pointing to possible genetic compensation. At a finer scale, we uncover putative molecular adaptations in diverse immune genes, including some acquired through introgression. Finally, we report numerous previously undescribed ACE2 variants, and show that critical sites associated with reduced susceptibility to SARS-CoV-2 have evolved convergently across horseshoe bats.

Fig. 6 | Putative genetic introgression in Rhinolophidae. a, D statistic for ten *Rhinolophus* species when using *H. armiger* species as outgroup. We showed the 38 introgression patterns that have genome-wide excess of shared derived alleles with exhibiting the significant D values (values of D statistic, |*Z*-score| > 3). The error bars represent the mean ± s.e.m. The red asterisk marks the highest excess of ABBA sites. **b**, f_d values tested for the four-taxon introgression pattern (*R. sinicus*, *R. pearsonii*, *R. ferrumequinum*, *H. armiger*). The cut-off for the top 5% f_d values is shown by a horizontal broken line. **c**, Top ten GO terms among all significantly enriched results of introgressed genes are shown. The *P* value was test by using the hypergeometric test, and we applied the family-wise error rate based on 10,000 random permutations to correct for multiple testing. *R. rex*, *Rhinolophus rex*.

a

b

c
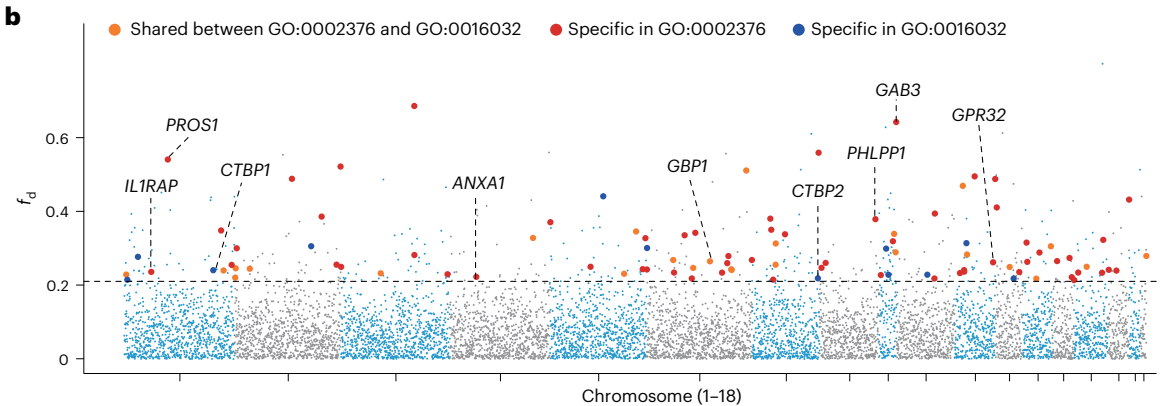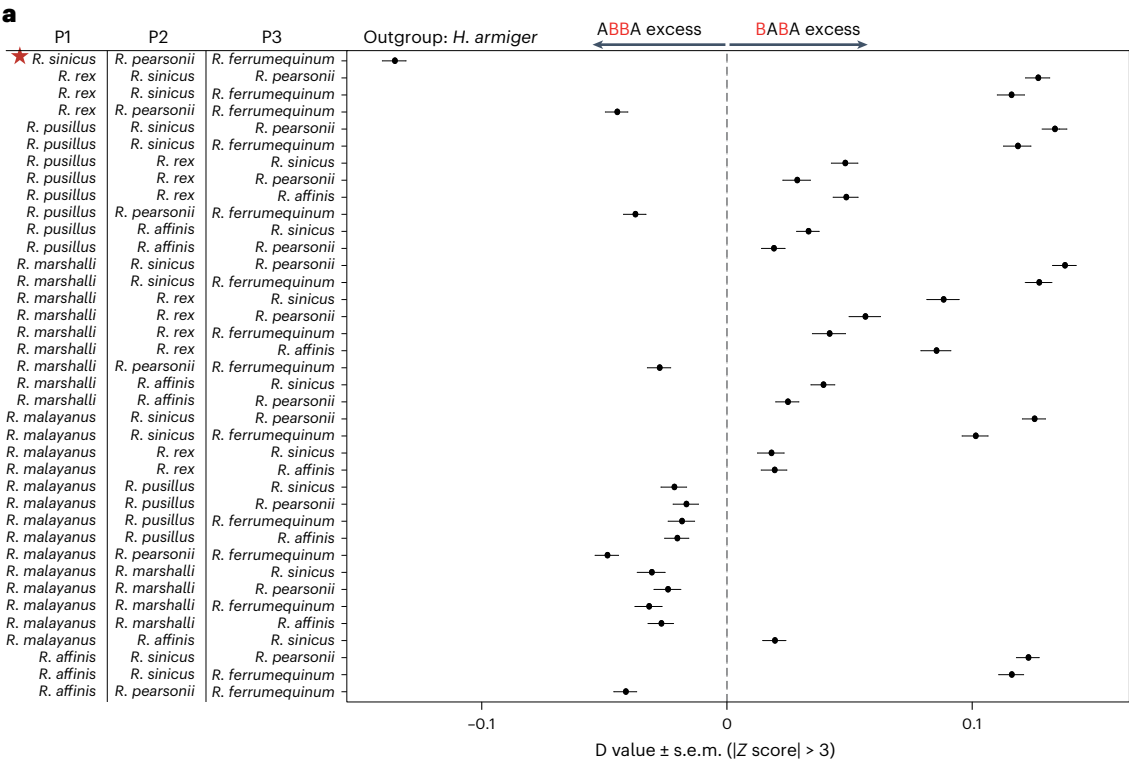
Methods

Sampling and sequencing

We collected individual females of *R. sinicus*, *H. armiger* and *H. pratti* from Xianning City, Hubei Province, China, one female of *R. pearsonii* from the Nanling Mountains, Guangdong Province, China, and one female of *M. lyra* from Huizhou City, Guangdong Province, China. All the bats were adults. All bats were captured with mist nets in caves, then each was placed into a separate clean cloth bag and transported to the temporary laboratory. Field sampling of these bats was approved by the Guangdong Institute of Applied Biological Resources (GIABR) of the Institute of Zoology Guangdong Academy of Sciences (GIABR20200810). DNA was isolated from liver tissue using Qiagen Genomic DNA extraction kits and was tested for quality before use in genomic libraries.

For short reads we constructed short insert (350 bp) libraries with ~1.5 μg of genomic DNA using the TruSeq Nano DNA HT Sample Preparation Kit (Illumina), and these were sequenced to generate 150 bp paired-end data on an Illumina NovaSeq platform. Raw reads were filtered for low-quality reads using the fastp software[93]. For long reads, we treated DNA with the NEBNext FFPE DNA Repair Mix (M6630) and the NEBNext End repair/dA-tailing Module (E7546) in accordance with the manufacturer's instructions. The resulting library was run on a PromethION sequencer (Oxford Nanopore Technologies), and reads with Phred quality scores of <7 were removed before assembly.

For each species, we also constructed Hi-C libraries using liver cells. Cells were fixed with formaldehyde and lysed to obtain cross-linked DNA. Following digestion with the restriction enzyme DpnII, the sticky ends were biotinylated and proximity-ligated to form chimeric junctions. Finally, fragments of 300–500 bp were enriched and processed into paired-end Hi-C libraries, which were sequenced on an Illumina NovaSeq platform to generate 150 bp paired-end reads. Data were filtered using fastp (v.0.23.2)[93] and used for subsequent chromosome anchoring.

Genome assembly

First, we conducted *k*-mer frequency analysis to estimate the genome size of each bat using Jellyfish (v.2.1.3)[94] by setting the *k*-mer length to 17 bp. Next, we used a recently described improved assembly pipeline[18] to assemble chromosomal-level genomes. Briefly, the initial contigs were assembled using NextDenovo v.2.5.0 (https://github.com/Nextomics/NextDenovo) and corrected with NextPolish v.1.4.0 (ref. 95). Linked contigs were then clustered by calculating Hi-C interaction frequencies using an agglomerative hierarchical clustering algorithm implemented in ALLHiC software (v.0.9.13)[96] based on the initial contigs as a reference. Next, we realigned all high-quality nanopore reads to the initial contigs and performed a local assembly after extracting optimal mapped reads of each contig group. This resulted in a final set of contigs. Finally, the Hi-C interaction pairs were again used to anchor these contigs onto chromosomes using the ALLHiC algorithm[96]. Finally, we used Juicebox v.1.22 (https://github.com/aidenlab/Juicebox) to visualize and adjust the placement and orientation of contigs that showed no chromatin interactions.

Repetitive sequences identification

Repetitive sequences were annotated using a combination of homology searching and ab initio prediction. Homology-based prediction was performed using RepeatMasker (v.4.0.5)[97] and RepeatProteinMask (v.4.0.5) with default parameters to search against the Repbase library[98]. For ab initio prediction, we built a de novo repeat library based on combining the results from four tools (LTR FINDER v.1.0.7 (ref. 99), RepeatScout v.1.0.5 (ref. 100), PILER v.3.3.0 (ref. 101) and RepeatModeler v.1.0.8 (ref. 102)) that were applied to all five genomes with default parameters. The de novo repeat library was used to mask repetitive sequences in each genome with RepeatMasker (v.4.0.5)[103]. In addition, tandem repeats were also predicted using Tandem Repeats Finder[104].

SD identification

To identify SDs, the soft-masked genome sequence was used to conduct self-genome alignments through the package LASTZ (v.1.04.15)[105], allowing a maximum simultaneous gap of 100 bp. Overlapping fragments were removed, and the non-redundant alignments were obtained. Genomic regions with alignment lengths >1 kb and identities of >90% were considered as SDs. To benchmark our pipeline, we ran it on a human chromosome assembly (GRCh38) and found an almost identical SD content as previously reported (5.59% versus 5.54%, respectively)[22].

Gene family cluster analysis

For gene family cluster analyses, we used proteins from 20 focal mammals (Supplementary Table 10). We first selected the longest protein-coding sequences of each gene and filtered out transcripts of <150 bp. Next, protein sequences were clustered into paralogous and orthologous sequences using OrthoFinder (v.2.3.1) with default parameters[106]. This process yielded 22,536 gene families, including 6,681 single-copy orthologues. Subsequently, we minimized the impact of multiple sequence alignment errors and divergent regions by applying the Gblocks (v.0.91b) package[107], and the high-quality alignments were obtained from single-copy orthologues with PRANK (v.170427)[108]. Finally, we discarded alignments shorter than 150 nucleotides and used the remaining 6,658 single-copy orthologues for subsequent analysis.

Resequencing of 91 horseshoe bats

For resequencing, we collected wing membrane biopsies from 91 *Rhinolophus* individuals from 10 species, sampled at sites across Guangxi, Guangdong, Yunnan, Zhejiang, Hainan and Guizhou provinces in China. This was approved by the Institute of Zoology Guangdong Academy of Sciences (GIABR20200810). All the bats were adults. All bats were captured with mist nets in caves, then each was placed into a separate clean cloth bag and transported to the temporary laboratory. All individuals were released at their point of capture after sampling. Genomic DNA was extracted using TIANamp genomic DNA kits (Tiangen Ltd). Adapter ligation and clean up was performed using MGIEasy DNA Adapters kits and a polymerase chain reaction was carried out on purified adapter-ligated DNA and cleaned-up again using magnetic beads. After quality control using Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific), purified polymerase chain reaction products were denatured and ligated to generate single-strand circular DNA libraries. Libraries were used to generate an average of 11.20 Gb of 150 bp paired-end reads per sample on a MGI DNBSEQ-T7 platform (MGI Tech Co., Ltd). In addition, we downloaded paired-end read datasets (mean 34.83 Gb) for 18 *H. armiger* individuals[27] from the NCBI database (accession number PRJNA309742).

Phylogeny construction and divergence time estimation

We performed ML phylogenetic reconstruction based on our set of 6,658 loci for 20 focal mammals in RaxML (v.8.2.12)[23]. We applied a GTR+GAMMAX substitution model, as indicated by MODELTEST (v.0.1.7)[109] with 1,000 bootstrap replicates. To generate a time-calibrated tree, we estimated divergence times in MCMCTree (v.4.5)[66]. We included two fossil constraints: the divergence times between mouse and human (61.5–100.5 Ma) and horse and human (95.3–113 Ma)[110] as well as calibration points obtained from the Time-Tree database[111] for the divergence between pig and cow (54 Ma), dog and cat (62 Ma) and horse and cow (75 Ma).

GO enrichment analysis

We generated a tailored GO database for the *R. sinicus* using the g:Profiler database[112], which synchronizes quarterly with Ensembl (current Ensembl 110) and obtains functional data from geneontology.org. Specifically, we conducted a BLASTX analysis[113] with an E value of 1×10^{-5}, comparing the protein-coding gene sequences of

R. sinicus with those of humans and two bat species (*Pteropus vampyrus* and *R. ferrumequinum*) already included in the g:Profiler database. The resulting alignments of each pair of species were utilized as input for the MCScanX package, with a requirement for at least three genes in a syntenic block. This facilitated the identification of orthologous gene pairs. Following that, we used the curated GO term mapping file for these three species in the g:Profiler database to illustrate the mapping relationships between *R. sinicus* genes and GO terms. A total of 18,483 *R. sinicus* protein-coding genes were successfully associated with 22,406 GO terms.

Based on the acquired GO annotation file of *R. sinicus*, we generated a standard file in gene matrix transpose format. This file was then uploaded to the g:Profiler web server (http://biit.cs.ut.ee/gprofiler/) as the query database, with the *R. sinicus* gene set designated as the background for performing gene enrichment analyses. Fisher's one-tailed test was used for statistical significance, and the *P* values were corrected for multiple testing by the Benjamini–Hochberg method. All the GO enrichment analyses were used in subsequent analyses unless stated otherwise.

Expansion and contraction analysis

Using our dataset of 22,536 gene families, we tested for gene family expansion and contraction across 20 mammals (Extended Data Fig. 1) using CAFE (v.5.0)[40]. We used a time tree (Extended Data Fig. 1) as input, and estimated a global evolutionary rate (λ) of 0.0056. We first identified gene families showing evidence of rapid evolution (family-wide *P* value <0.05). From these, we subsequently identified gene families showing significant shifts on the ancestral branch of Rhinolophidae (Viterbi *P* value <0.05). We also tested whether horseshoe bats showed a different overall rate of gene family evolution by comparing a two-rate model ($\lambda 1$ bats, $\lambda 2$ other mammals) to a three-rate model ($\lambda 3$ horseshoe bats). These models did not differ in likelihood (*P* > 0.05, likelihood ratio test), probably reflecting the fact that sister lineages of horseshoe bats also show gene expansions (Extended Data Fig. 1).

Reconstruction of gene repertoire evolution

We predicted gene gains and losses based on the reconciliation method in NOTUNG (v.2.9)[46], by comparing the species tree and the gene tree. We used a species tree topology and branch lengths based on the time tree (Extended Data Fig. 1). To build the gene tree, we first obtained gene sequence alignments from multiple species using the MUSCLE program (v.3.8.31)[114], and then inferred an ML tree in RaxML[23] fitting a GTR+GAMMA model with 1,000 bootstrap replicates.

Selection associated with the evolution of type I IFNs

To test whether natural selection drives the diversification of type I IFNs in horseshoe bats, we used CodeML in PAML (v.4.10.6)[66]. We ran tests of: (i) positive selection; and (ii) diversifying selection on the gene trees of three subfamilies (IFNδ, IFNω and IFNβ) of type I IFN. For positive selection, we labelled the ancestral branch of the Rhinolophidae as the foreground and compared the modified branch-site model A with site class $\omega 2 > 1$ with a null model in which the same site is under purifying selection or neural evolution. For diversifying selection, we labelled the Rhinolophidae as a foreground clade and compared clade model C (CmC) with its null model (M2a_rel)[115,116]. The CmC model accounts for divergence by estimating distinct ω ratios for two or more clades (for example, $\omega 2$ and $\omega 3$ for two clades). By contrast, the M2a_rel model is derived from CmC by imposing a single constraint that equates the ω ratios ($\omega 2 = \omega 3$). For both tests, we assessed significance with a likelihood ratio test with *P*-values calculated following a χ^2 distribution.

Analysis of RNA sequencing for RfKT cells

We filtered raw reads to remove adapters and low-quality reads using the fastp software[93]. High-quality reads were then aligned to the *R. ferrumequinum* reference genome using HISAT2 software (v.2.2.1)[117].

Subsequently, we used HTSeq (v.2.0.3) to quantify the read numbers mapped to each gene[118]. The fragments per kilobase of transcript per million mapped reads were computed for each gene. DEGs were identified using DESeq2 (v.1.42.0) in the R package (v.4.2.0)[119]. The resulting *P* values underwent adjustment using Benjamini and Hochberg's method to control the false discovery rate (FDR). Genes with an absolute fold change ≥2 and an FDR-adjusted *P* value ≤0.05 were considered significant DEGs. All significant DEGs resulting from IFN stimulation were considered to be putative ISGs.

Positive selection and rapid evolution analyses

To identify genes that have undergone (i) positive selection and (ii) rapid evolution in horseshoe bats, we estimated selection pressure in 6,658 single-copy orthologues using CodeML[66]. We labelled the ancestral branch of the Rhinolophidae as the foreground. For positive selection, we compared the modified branch-site model A with its null model as previously described. For rapid evolution, we compared a two-ratio branch model in which the foreground branch can evolve at a faster rate than the background to model M0 in which ω is fixed across the tree. We ran likelihood ratio tests and adjusted *P* values for the FDR to correct for multiple tests. Using the same strategy, we also identified genes showing signatures of molecular adaptation on the ancestral branches of Hipposideridae, Pteropodidae and Vespertilionidae (two species per clade). Gene sets were analysed for GO enrichment as previously described.

Cell-based assays for ACE2

Cell cultures and plasmids. We maintained Vero E6 (CRL-1586; ATCC) and HEK293T cells (CRL-1586; ATCC) in Dulbecco's modified Eagle's medium (DMEM; Gibco) supplemented with 10% fetal bovine serum (FBS), 2.0 mM L-glutamine, 110 mg l^{-1} sodium pyruvate, and 4.5 g l^{-1} D-glucose. I1-Hybridoma (CRL-2700) producing a monoclonal antibody targeting VSV glycoprotein was maintained in minimum essential medium (Gibco) with Earle's salts and 2.0 mM L-glutamine. All cells were cultured at 37 °C with 5% CO_2 and regularly passaged every 1–2 days.

We synthesized the DNA sequences of human codon-optimized ACE2 orthologues and their mutations, fused with a C-terminal 3× FLAG-tag (DYKDHDGDYKDHDIDYKDDDDK), which were subcloned into a lentiviral transfer vector (pLVX-IRES-puro) through the EcoRI and BamHI restriction sites (F-primer: CTCGAGCTTTTGGAGTACGT; R-primer: GCGGCCGCTCACTTGTCGTC). The DNA sequence of human codon-optimized SARS-CoV-2 S protein (Wuhan-Hu-1, GenBank MN908947) was cloned into the pCAGGS vector with a C-terminal 18 amino acid deletion to improve VSV pseudotyping efficiency, and the D614G mutation was introduced into the SARS-CoV-2-S coding sequence to enhance in vitro infection efficiency. In addition, plasmids were generated to express coronavirus RBD–immunoglobulin G (IgG) Fc fusion proteins, with the coding sequences of SARS-CoV-2 RBD (amino acids 331–530) inserted into the pCAGGS vector to express fusion proteins with C-terminal human Fc (IgG1) and an N-terminal CD5 secretion leading sequence (MPMGSLQPLAT-LYLLGMLVASVL). We also created plasmids to express ACE2 ectodomain (amino acids 18–740) proteins, which were also inserted into the pCAGGS vector and expressed as fusion proteins with C-terminal Twin-Strep-tag and an N-terminal CD5 secretion leading sequence (MPMGSLQPLATLYLLGMLVASVL).

SARS-CoV-2 pseudovirus production. To produce pseudoviruses containing spike proteins from SARS-CoV-2, we followed a published protocol with minor modifications[120]. HEK293T cells were transfected with plasmids expressing S proteins using Lipofectamine 2000 (Biosharp). After 24 h, the transfected cells were infected with VSV-deactivated glycoprotein (dG)-firefly luciferase (FLuc)-enhanced green fluorescent protein (EGFP) (1 × 10^6 the 50% tissue culture infectious dose per ml) diluted in DMEM. The mixture was incubated on a shaker for 2 h at 37 °C,

and then replenished with DMEM containing anti-VSV-G monoclonal antibody (I1, 0.1 µg ml⁻¹). After another 24 h, the supernatant containing the SARS-CoV-2 pseudovirus was harvested and clarified by centrifugation at 13,523×g for 5 min at 4 °C, and stored at −80 °C. The 50% tissue culture infectious dose of pseudovirus was determined using a serial dilution-based infection assay on BHK-21-hACE2 cells, and calculated according to the Reed–Muench method.

ACE2 expression assay. To evaluate expression levels of ACE2 variants, we used an immunofluorescence assay based on targeting a C-terminal 3× FLAG-tag. HEK293T cells were transiently transfected with Lipofectamine 2000 and seeded into poly-lysine pretreated 96-well plates at a cell density of 5×10^5 per ml (100 µl per well), and cultured for 48 h. The cells were fixed with 4% paraformaldehyde for 10 min at room temperature, permeabilized with 0.2% Triton X-100–phosphate-buffered saline (PBS) for 10 min at room temperature, and blocked with 1% BSA at 37 °C for 30 min. The cells were then incubated with a mouse monoclonal antibody targeting the FLAG-tag (M2; Sigma, catalogue no. F1804A-5MG) at 1:1,000 dilution in 1% BSA–PBS at 37 °C for 1 h. After three rounds of washing with PBS, cells were incubated with Alexa Fluor 594-conjugated goat anti-mouse IgG (Thermo Fisher Scientific, catalogue no. A32742) at 1:1,000 dilution in 1% BSA–PBS at room temperature for 30 min. Nuclei were stained blue with Hoechst 33342 (1:5,000 dilution in PBS) and images captured using a fluorescence microscope (MI52-N; Mshot).

Protein expression, purification and binding assay. Recombinant SARS-CoV-2-RBD-hFc proteins and ACE2 ectodomains (with a strep tag) were produced through the transient transfection of HEK293T cells with Lipofectamine 2000. The transfected cells were cultured in SMM 293-TIS Expression Medium (serum-free, without L-glutamine; Sino Biological). The supernatant containing the recombinant proteins was collected at days 2, 4 and 6 post-transfection, the RBD was purified using Protein A/G Plus Agarose (Thermo Fisher Scientific) and the ACE2 ectodomains were purified using Strep-Tactin XT 4Flow high-capacity resin (IBA). The protein concentration was determined using the BCA protein determination kit (EpiZyme). The proteins were analysed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis with Coomassie blue staining. HEK293T cells overexpressing ACE2 orthologues were generated by transfecting ACE2 coding sequences (pLVX-EF1a-Puro; GENEWIZ) into HEK293T cells using Lipofectamine 2000 (Biosharp). After 48 h, the cells were incubated with RBD–hFc protein (4 µg ml⁻¹) diluted in growth medium for 0.5 h at 4 °C. Cells were washed twice with DMEM and then incubated with Alexa Fluor 488-conjugated goat anti-human IgG (Thermo Fisher Scientific, catalogue no. A11013) at 1:1,000 dilution in DMEM with 2% FBS for 30 min at 4 °C. The nucleus was stained blue with Hoechst 33342 (1:5,000 dilution in PBS). Images were captured with a fluorescence microscope (MI52-N; Mshot).

Pseudotype entry assay. The HEK293T cell lines were transfected with various ACE2 orthologues and inoculated with SARS-CoV-2 pseudotyped viruses in DMEM with 10% FBS after 24 h. At 16 to 24 h after infection, images of infected cells with GFP expression were acquired using a fluorescence microscope (MI52-N). Intracellular luciferase activity was determined using a Bright-Glo Luciferase Assay Kit (Promega) and measured using either a SpectraMax iD3 Multi-well Luminometer (Molecular Devices) or a GloMax 20/20 Luminometer (Promega).

Bio-layer interferometry binding assays. The apparent binding affinity ($K_{D,app}$) between the RBD and ACE2 was measured using the Octet RED96 system (ForteBio). The buffer for analysis was PBS with 0.05% Tween 20. The RBD (10 µg ml⁻¹) was captured on ProA biosensors, followed by binding of ACE2 ectodomains at twofold serial dilutions from 500 or 8,000 nM for 500 s and then dissociated in PBS with 0.05%

Tween 20 for another 500 s. The kinetics were performed in a 1:1 model using Octet Analysis Studio v.12.2.0.20 (ForteBio). Mean $K_{D,app}$ values were determined by averaging all binding curves that matched the theoretical fit with an R^2 value ≥0.95. Methods followed ref. 83.

Statistical analysis
For data summary, we calculated means with standard errors or standard deviation. Statistical analyses were performed using Prism 7 software (GraphPad, v.9.4.1). Differences were assessed using two-tailed Mann–Whitney U-tests (two independent samples) or paired two-tailed t-tests (related samples).

Genetic introgression analysis
To test for potential genetic introgression among populations of different *Rhinolophus* species, we first calculated genome-wide D-statistics in the qpDstat program in AdmixTools (v.7.0.2)[91]. Standard errors were estimated by jackknifing[121] and a Z-score of ≥3 standard errors was used to indicate statistical significance. To test for potential introgression in specific parts of the genome, we also calculated the related statistic f_d (an improved version of the statistic originally designed to measure genome-wide admixture proportions) for 100 kb non-overlapping sliding windows across the genome[122] using the General tools for genomic analyses (v.0.4)[123].

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

Data availability
New genome sequence data for the five bats are deposited at the Genome Sequence Archive in National Genomics Data Center (https://ngdc.cncb.ac.cn) under accession code CRA018832. Genome assemblies are deposited at the NGDC GenBank under accessions GWHFDMV00000000.1 (https://ngdc.cncb.ac.cn/gwh/Assembly/86071/show, *R. sinicus*), GWHFDMW00000000.1 (https://ngdc.cncb.ac.cn/gwh/Assembly/86072/show, *R. pearsonii*), GWHFDMX00000000.1 (https://ngdc.cncb.ac.cn/gwh/Assembly/86073/show, *H. armiger*), GWHFDMY00000000.1 (https://ngdc.cncb.ac.cn/gwh/Assembly/86073/show, *H. pratti*) and GWHFDMZ00000000.1 (https://ngdc.cncb.ac.cn/gwh/Assembly/86075/show, *M. lyra*). RNA-seq data of RfKT cells can be accessed at the NGDC (accession number: PRJCA023723). SARS-CoV-2 S protein sequence was obtained from GenBank (accession number: MN908947). Human ACE2 protein sequence was obtained from GenBank (accession number: NP_001358344.1). The data files used for the population genomics analyses, gene family analysis and PAML analysis are available via Figshare at https://doi.org/10.6084/m9.figshare.27612597 (ref. 124).

Code availability
The code and pipelines used for the analyses are available via Zenodo at https://doi.org/10.5281/zenodo.13690583 (ref. 125) and GitHub (https://github.com/SLbio/Comparative-genomics-of-horseshoe-bats).

References

1. Morales, A. E. et al. Bat genomes illuminate adaptations to viral tolerance and disease resistance. *Nature* https://doi.org/10.1038/s41586-024-08471-0 (2025).
2. Anthony, S. J. et al. Global patterns in coronavirus diversity. *Virus Evol.* **3**, vex012 (2017).
3. Hutson, A., Rossiter, S., Csorba, G. & Burgin, C. in *Handbook of the Mammals of the World: Bats* (eds Wilson, D. E. & Mittermeier, R. A.) 260–332 (Lynx, 2019).
4. Foley, N. M. et al. How and why overcome the impediments to resolution: lessons from rhinolophid and hipposiderid bats. *Mol. Biol. Evol.* **32**, 313–333 (2015).

5. Hu, B. et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog.* **13**, e1006698 (2017).

6. Ge, X. Y. et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **503**, 535–538 (2013).

7. Temmam, S. et al. Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature* **604**, 330–336 (2022).

8. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).

9. Lau, S. K. et al. Ecoepidemiology and complete genome comparison of different strains of severe acute respiratory syndrome-related *Rhinolophus* bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events. *J. Virol.* **84**, 2808–2819 (2010).

10. Liu, X. et al. Analogous comparison unravels heightened antiviral defense and boosted viral infection upon immunosuppression in bat organoids. *Signal Transduct. Target. Ther.* **7**, 392 (2022).

11. Xie, J. et al. Dampened STING-dependent interferon activation in bats. *Cell Host Microbe* **23**, 297–301 e4 (2018).

12. 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).

13. Moreno Santillan, D. D. et al. Large-scale genome sampling reveals unique immunity and metabolic adaptations in bats. *Mol. Ecol.* **30**, 6449–6467 (2021).

14. Jebb, D. et al. Six reference-quality genomes reveal evolution of bat adaptations. *Nature* **583**, 578–584 (2020).

15. Woelk, C. H., Frost, S. D., Richman, D. D., Higley, P. E. & Kosakovsky Pond, S. L. Evolution of the interferon alpha gene family in eutherian mammals. *Gene* **397**, 38–50 (2007).

16. Walker, A. M. & Roberts, R. M. Characterization of the bovine type I IFN locus: rearrangements, expansions, and novel subfamilies. *BMC Genomics* **10**, 187 (2009).

17. Mao, X. et al. Karyotype evolution in *Rhinolophus* bats (Rhinolophidae, Chiroptera) illuminated by cross-species chromosome painting and G-banding comparison. *Chromosome Res.* **15**, 835–848 (2007).

18. Tian, S. et al. Comparative analyses of bat genomes identify distinct evolution of immunity in Old World fruit bats. *Sci. Adv.* **9**, eadd0141 (2023).

19. Zhang, C. et al. High-quality genome of a modern soybean cultivar and resequencing of 547 accessions provide insights into the role of structural variation. *Nat. Genet.* **56**, 2247–2258 (2024).

20. Osmanski, A. B. et al. Insights into mammalian TE diversity through the curation of 248 genome assemblies. *Science* **380**, eabn1430 (2023).

21. Paulat, N. S. et al. Chiropterans are a hotspot for horizontal transfer of DNA transposons in Mammalia. *Mol. Biol. Evol.* **40**, msad092 (2023).

22. She, X., Cheng, Z., Zollner, S., Church, D. M. & Eichler, E. E. Mouse segmental duplication and copy number variation. *Nat. Genet.* **40**, 909–914 (2008).

23. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

24. Teeling, E. C. et al. A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* **307**, 580–584 (2005).

25. Lavery, T. H., Leung, L. K. P. & Seddon, J. M. Molecular phylogeny of hipposiderid bats (Chiroptera: Hipposideridae) from Solomon Islands and Cape York Peninsula, Australia. *Zool. Scr.* **43**, 429–442 (2014).

26. Hao, X., Lu, Q. & Zhao, H. A molecular phylogeny for all 21 families within Chiroptera (bats). *Integr. Zool.* **19**, 989–998 (2023).

27. Dong, D. et al. The genomes of two bat species with long constant frequency echolocation calls. *Mol. Biol. Evol.* **34**, 20–34 (2017).

28. Vilella, A. J. et al. EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.* **19**, 327–335 (2009).

29. Samonte, R. V. & Eichler, E. E. Segmental duplications and the evolution of the primate genome. *Nat. Rev. Genet.* **3**, 65–72 (2002).

30. Kim, J. et al. Reconstruction and evolutionary history of eutherian chromosomes. *Proc. Natl Acad. Sci. USA* **114**, E5379–E5388 (2017).

31. Ross, M. T. et al. The DNA sequence of the human X chromosome. *Nature* **434**, 325–337 (2005).

32. Bellott, D. W. & Page, D. C. Reconstructing the evolution of vertebrate sex chromosomes. *Cold Spring Harb. Symp. Quant. Biol.* **74**, 345–353 (2009).

33. Skinner, B. M. et al. The pig X and Y chromosomes: structure, sequence, and evolution. *Genome Res.* **26**, 130–139 (2016).

34. Wu, S. et al. Extensive genomic rearrangements mediated by repetitive sequences in plastomes of Medicago and its relatives. *BMC Plant Biol.* **21**, 421 (2021).

35. Ma, J. et al. Reconstructing contiguous regions of an ancestral genome. *Genome Res.* **16**, 1557–1565 (2006).

36. Zhou, X. et al. Whole-genome sequencing of the snub-nosed monkey provides insights into folivory and evolutionary history. *Nat. Genet.* **46**, 1303–1310 (2014).

37. Fan, H. et al. Chromosome-level genome assembly for giant panda provides novel insights into Carnivora chromosome evolution. *Genome Biol.* **20**, 267 (2019).

38. Ao, L. et al. Karyotype relationships of six bat species (Chiroptera, Vespertilionidae) from China revealed by chromosome painting and G-banding comparison. *Cytogenet. Genome Res.* **115**, 145–153 (2006).

39. Sotero-Caio, C. G., Baker, R. J. & Volleth, M. Chromosomal evolution in Chiroptera. *Genes (Basel)* **8**, 272 (2017).

40. De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).

41. Phillips, A. M. et al. Host proteostasis modulates influenza evolution. *eLife* **6**, e28652 (2017).

42. Reyes-Del Valle, J., Chavez-Salinas, S., Medina, F. & Del Angel, R. M. Heat shock protein 90 and heat shock protein 70 are components of dengue virus receptor complex in human cells. *J. Virol.* **79**, 4557–4567 (2005).

43. Abeler-Dorner, L., Swamy, M., Williams, G., Hayday, A. C. & Bas, A. Butyrophilins: an emerging family of immune regulators. *Trends Immunol.* **33**, 34–41 (2012).

44. Hubel, P. et al. A protein-interaction network of interferon-stimulated genes extends the innate immune system landscape. *Nat. Immunol.* **20**, 493–502 (2019).

45. Ma, Y. et al. Annexin A2 (ANXA2) interacts with nonstructural protein 1 and promotes the replication of highly pathogenic H5N1 avian influenza virus. *BMC Microbiol.* **17**, 191 (2017).

46. Chen, K., Durand, D. & Farach-Colton, M. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J. Comput. Biol.* **7**, 429–447 (2000).

47. Andreev, D. E. et al. Non-AUG translation initiation in mammals. *Genome Biol.* **23**, 111 (2022).

48. Xiong, Y. et al. Annexin II receptor induces apoptosis independent of Annexin II. *Apoptosis* **18**, 925–939 (2013).

49. Zuniga, M. et al. Autoimmunity to annexin A2 predicts mortality among hospitalised COVID-19 patients. *Eur. Respir. J.* **58**, 2100918 (2021).

50. Ju, X. et al. A novel cell culture system modeling the SARS-CoV-2 life cycle. *PLoS Pathog.* **17**, e1009439 (2021).

51. Das, A. T., Tenenbaum, L. & Berkhout, B. Tet-On systems for doxycycline-inducible gene expression. *Curr. Gene Ther.* **16**, 156–167 (2016).

52. Secombes, C. J. & Zou, J. Evolution of interferons and interferon receptors. *Front. Immunol.* **8**, 209 (2017).

53. Whaley, A. E., Meka, C. S., Harbison, L. A., Hunt, J. S. & Imakawa, K. Identification and cellular localization of unique interferon mRNA from human placenta. *J. Biol. Chem.* **269**, 10864–10868 (1994).

54. Pavlovich, S. S. et al. The Egyptian Rousette genome reveals unexpected features of bat antiviral immunity. *Cell* **173**, 1098–1110 e18 (2018).

55. Yang, Z. & Rannala, B. Molecular phylogenetics: principles and practice. *Nat. Rev. Genet.* **13**, 303–314 (2012).

56. Geng, R. et al. Unconventional IFNω-like genes dominate the Type I IFN locus and the constitutive antiviral responses in bats. *J. Immunol.* **213**, 204–213 (2024).

57. Kopitar-Jerala, N. The role of interferons in inflammation and inflammasome activation. *Front. Immunol.* **8**, 873 (2017).

58. Gonzalez-Navajas, J. M., Lee, J., David, M. & Raz, E. Immunomodulatory functions of type I interferons. *Nat. Rev. Immunol.* **12**, 125–135 (2012).

59. Banerjee, A. et al. Novel insights into immune systems of bats. *Front. Immunol.* **11**, 26 (2020).

60. Scheben, A. et al. Long-read sequencing reveals rapid evolution of immunity- and cancer-related genes in bats. *Genome Biol. Evol.* **15**, evad148 (2023).

61. Sang, Y., Rowland, R. R., Hesse, R. A. & Blecha, F. Differential expression and activity of the porcine type I interferon family. *Physiol. Genomics* **42**, 248–258 (2010).

62. Xu, L., Yang, L. & Liu, W. Distinct evolution process among type I interferon in mammals. *Protein Cell* **4**, 383–392 (2013).

63. Zhao, X. et al. Cloning and characterization of porcine interferon-delta-related genes identified by genomic database screening. *J. Interferon Cytokine Res.* **32**, 378–385 (2012).

64. Roberts, R. M., Liu, L., Guo, Q., Leaman, D. & Bixby, J. The evolution of the type I interferons1. *J. Interferon Cytokine Res.* **18**, 805–816 (1998).

65. Thomas, C. et al. Structural linkage between ligand discrimination and receptor activation by type I interferons. *Cell* **146**, 621–632 (2011).

66. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

67. Ricklin, D., Hajishengallis, G., Yang, K. & Lambris, J. D. Complement: a key system for immune surveillance and homeostasis. *Nat. Immunol.* **11**, 785–797 (2010).

68. Guo, R. F. & Ward, P. A. Role of C5a in inflammatory responses. *Annu. Rev. Immunol.* **23**, 821–852 (2005).

69. Jiang, Y. et al. MERS-CoV infection causes brain damage in human DPP4-transgenic mice through complement-mediated inflammation. *J. Gen. Virol.* **102**, 001667 (2021).

70. Jiang, Y. et al. Complement receptor C5aR1 inhibition reduces pyroptosis in hDPP4-transgenic mice infected with MERS-CoV. *Viruses* **11**, 39 (2019).

71. Carvelli, J. et al. Association of COVID-19 inflammation with activation of the C5a–C5aR1 axis. *Nature* **588**, 146–150 (2020).

72. Monk, P. N., Barker, M. D., Partridge, L. J. & Pease, J. E. Mutation of glutamate 199 of the human C5a receptor defines a binding site for ligand distinct from the receptor N terminus. *J. Biol. Chem.* **270**, 16625–16629 (1995).

73. Feng, Y. et al. Mechanism of activation and biased signaling in complement receptor C5aR1. *Cell Res.* **33**, 312–324 (2023).

74. Bonaparte, M. I. et al. Ephrin-B2 ligand is a functional receptor for Hendra virus and Nipah virus. *Proc. Natl Acad. Sci. USA* **102**, 10652–10657 (2005).

75. Li, W. et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450–454 (2003).

76. Hoffmann, M. et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280 e8 (2020).

77. Thoulouze, M. I. et al. The neural cell adhesion molecule is a receptor for rabies virus. *J. Virol.* **72**, 7181–7190 (1998).

78. Zhou, D. et al. Unexpected mode of engagement between enterovirus 71 and its receptor SCARB2. *Nat. Microbiol.* **4**, 414–419 (2019).

79. Chen, P. et al. Molecular determinants of enterovirus 71 viral entry: cleft around GLN-172 on VP1 protein interacts with variable region on scavenge receptor B 2. *J. Biol. Chem.* **287**, 6406–6420 (2012).

80. 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).

81. 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. USA* **117**, 22311–22322 (2020).

82. Liu, K. et al. Cross-species recognition of SARS-CoV-2 to bat ACE2. *Proc. Natl Acad. Sci. USA* **118**, e2020216118 (2021).

83. 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**, e00902-20 (2020).

84. Li, P. et al. The *Rhinolophus affinis* bat ACE2 and multiple animal orthologs are functional receptors for bat coronavirus RaTG13 and SARS-CoV-2. *Sci. Bull. (Beijing)* **66**, 1215–1227 (2021).

85. Zhang, H. L. et al. Evaluating angiotensin-converting enzyme 2-mediated SARS-CoV-2 entry across species. *J. Biol. Chem.* **296**, 100435 (2021).

86. Li, P. et al. Effect of polymorphism in *Rhinolophus affinis* ACE2 on entry of SARS-CoV-2 related bat coronaviruses. *PLoS Pathog.* **19**, e1011116 (2023).

87. Sams, A. J. et al. Adaptively introgressed Neandertal haplotype at the OAS locus functionally impacts innate immune responses in humans. *Genome Biol.* **17**, 246 (2016).

88. Garcia-Erill, G. et al. Warthog genomes resolve an evolutionary conundrum and reveal introgression of disease resistance genes. *Mol. Biol. Evol.* **39**, msac134 (2022).

89. Howard-McCombe, J. et al. Genetic swamping of the critically endangered Scottish wildcat was recent and accelerated by disease. *Curr. Biol.* **33**, 4761–4769 e5 (2023).

90. Foley, N. M. et al. Karyotypic stasis and swarming influenced the evolution of viral tolerance in a species-rich bat radiation. *Cell Genom.* **4**, 100482 (2024).

91. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).

92. Prufer, K. et al. FUNC: a package for detecting significant associations between gene sets and ontological annotations. *BMC Bioinformatics* **8**, 41 (2007).

93. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

94. Marcais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**, 764–770 (2011).

95. Hu, J., Fan, J., Sun, Z. & Liu, S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**, 2253–2255 (2020).

96. Zhang, X., Zhang, S., Zhao, Q., Ming, R. & Tang, H. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nat. Plants* **5**, 833–845 (2019).

97. Smit, A., Hubley, R. & Green, P. RepeatMasker Open-4.0. 2013–2015 (2015).

98. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).

99. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).

100. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).

101. Edgar, R. C. & Myers, E. W. PILER: identification and classification of genomic repeats. *Bioinformatics* **21**, i152–i158 (2005).

102. Smit, A. & Hubley, R. R. RepeatMasker Open-1.0 (2008).

103. Bergman, C. M. & Quesneville, H. Discovering and detecting transposable elements in genome sequences. *Brief Bioinform.* **8**, 382–392 (2007).

104. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).

105. Harris, R. S. Improved pairwise alignmnet of genomic DNA. *PSU* https://www.bx.psu.edu/~rsharris/lastz/README.lastz-1.04.15.html (2007).

106. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).

107. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).

108. Veidenberg, A., Medlar, A. & Loytynoja, A. Wasabi: an integrated platform for evolutionary sequence analysis and data visualization. *Mol. Biol. Evol.* **33**, 1126–1130 (2016).

109. Posada, D. & Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818 (1998).

110. Benton, M. J. & Donoghue, P. C. Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* **24**, 26–53 (2007).

111. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).

112. Kolberg, L. et al. g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res.* **51**, W207–W212 (2023).

113. Mount, D. W. Using the basic local alignment search tool (BLAST). *CSH Protoc.* **2007**, pdb.top.17 (2007).

114. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

115. Bielawski, J. P. & Yang, Z. A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J. Mol. Evol.* **59**, 121–132 (2004).

116. Weadick, C. J. & Chang, B. S. An improved likelihood ratio test for detecting site-specific functional divergence among clades of protein-coding genes. *Mol. Biol. Evol.* **29**, 1297–1300 (2012).

117. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

118. Putri, G. H., Anders, S., Pyl, P. T., Pimanda, J. E. & Zanini, F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* **38**, 2943–2945 (2022).

119. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

120. Nie, J. et al. Quantification of SARS-CoV-2 neutralizing antibody by a pseudotyped virus-based assay. *Nat. Protoc.* **15**, 3699–3715 (2020).

121. Busing, F. M., Meijer, E. & Leeden, R. V. D. Delete-m jackknife for unequal m. *Stat. Comput.* **9**, 3–8 (1999).

122. Martin, S. H., Davey, J. W. & Jiggins, C. D. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32**, 244–257 (2015).

123. Martin, S. General tools for genomic analyses. *GitHub* https://github.com/simonhmartin/genomics_general (2019).

124. Tian, S. Comparative-genomics-of-horseshoe-bats. *figshare* https://doi.org/10.6084/m9.figshare.27612597 (2024).

125. Tian, S. Comparative genomics of horseshoe bats. *Zenodo* https://doi.org/10.5281/zenodo.13690582 (2024).

Acknowledgements

Author contributions

H.Z. conceived and designed the research. S.T. designed and performed analyses. J.S., J.Z., X. Zhang, X. Zhou, C.H., C.L. and H.Y. performed the functional experiments. L.Z. collected the samples. G.L. provided the genome assembly of *R. affinis*. R.G. and P.Z. provided RNA-seq data of RfKT cells. S.T., S.J.R. and H.Z. analysed data, discussed results and wrote the manuscript. All authors have read and approved the paper.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41559-025-02638-2.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-025-02638-2.

Correspondence and requests for materials should be addressed to Stephen J. Rossiter or Huabin Zhao.

Peer review information *Nature Ecology & Evolution* thanks Zijun Xiong and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
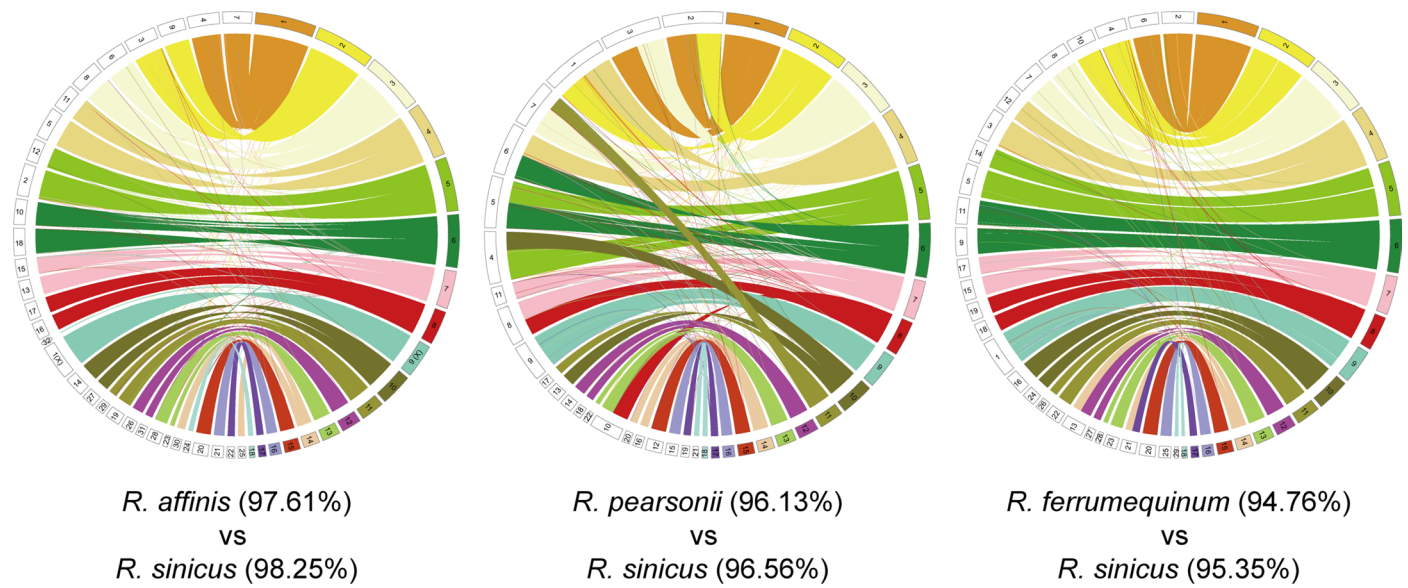
Extended Data Table 1 | Global summary of genome assemblies for six bats

| Species	R. sinicus	R. affinis*	R. pearsonii	H. armiger	H. pratti	M. lyra
Karyotype (2n)	36	62	44	32	32	54
Genome size (Gb)	2.06	2.09	2.08	2.12	2.13	2.02
Contig N50 (Mb)	45.68	31.54	41	52.97	58.05	25.26
Contig Number	104	587	176	265	271	883
Scaffold N50 (Mb)	182.9	93.07	172.78	174.32	175.14	94.71
Anchoring onto chromosomes (%)	99.95	99.67	99.63	99.74	99.97	99.88
Gene Number	22,650	21,220	21,211	22,249	21,669	21,986
Average intron length per gene (bp)	4,307	4,167	4,011	4,378	4,235	3,572
Repeat ratio (%)	31.19	31.41	31.98	34.73	34.51	32
Quality Value (QV)	42.68	41.98	42.37	42.39	42.48	41.43
Complete BUSCOs (%)	95.7	95.4	95.6	94.5	94.3	95.4

Note: * indicates *R. affinis* assembly have been published in GenBank (Accession Number: JAUKPG000000000).
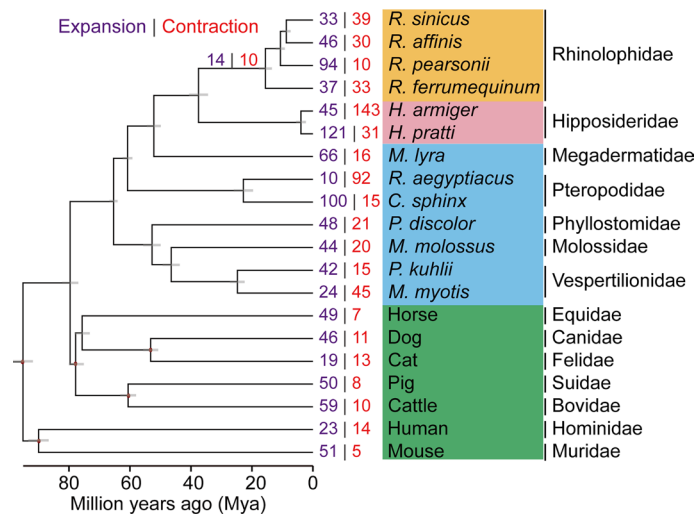
Extended Data Fig. 1 | Maximum likelihood (ML) tree of 13 bats and 7 other mammals, showing numbers of expanded (purple) and contracted (red) gene families.
Grey blocks indicate confidence intervals for divergence times for all nodes. Orange shading corresponds to Rhinolophidae (horseshoe bats), pink to Hipposideridae, blue to other bats, and green to non-bat mammals.

R. affinis (97.61%)
vs
R. sinicus (98.25%)

R. pearsonii (96.13%)
vs
R. sinicus (96.56%)

R. ferrumequinum (94.76%)
vs
R. sinicus (95.35%)

Extended Data Fig. 2 | Chord diagram depicting genome synteny between R. sinicus and each of R. affinis, R. pearsonii, and R. ferrumequinum. In each diagram, *R. sinicus* is shown on the right-hand side. Syntenic blocks are connected, with percentages indicating the overall synteny rates across the genomes.
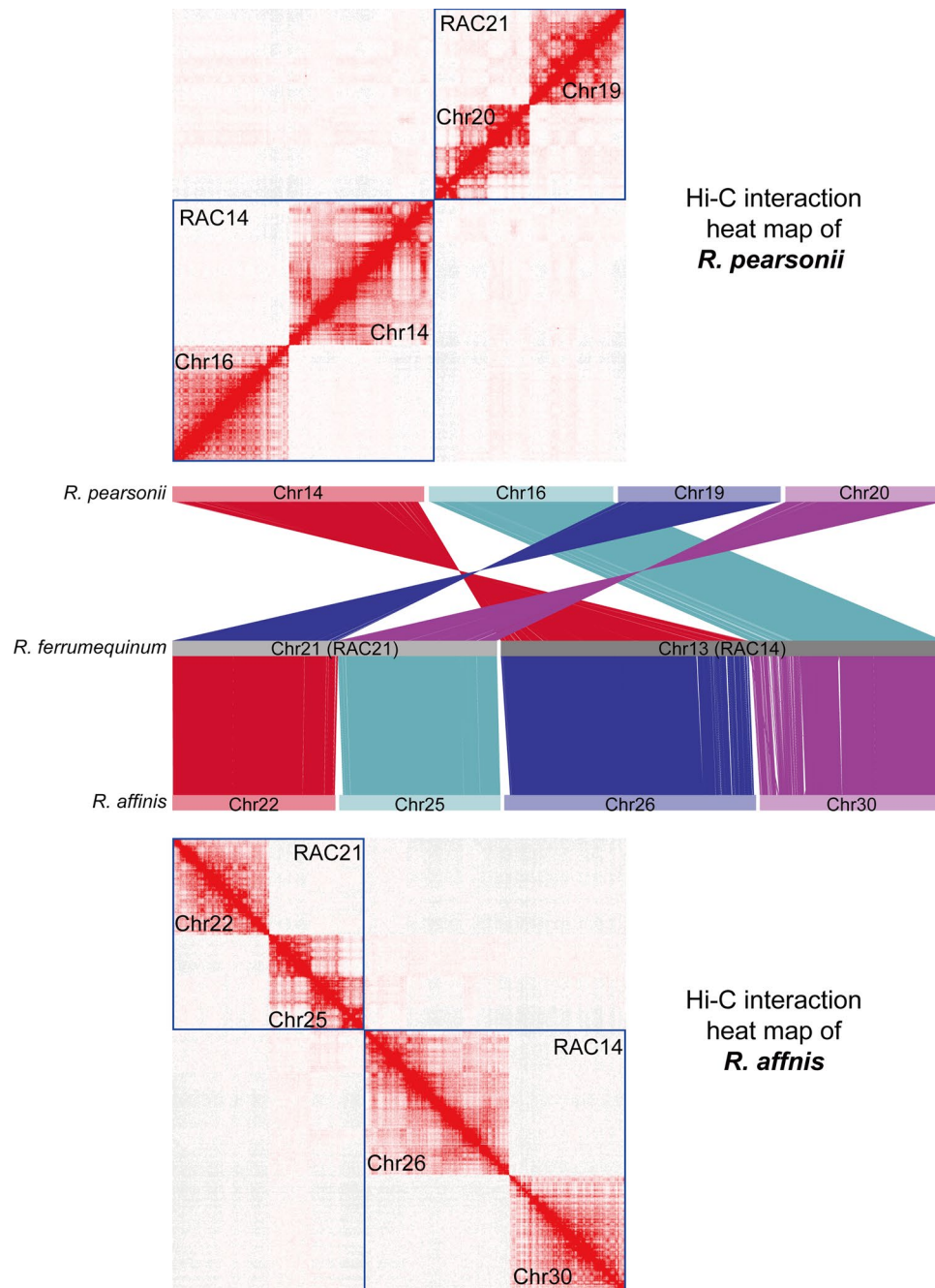
Extended Data Fig. 3 | Tracing chromosomal evolution of Rhinolophidae.
The top panel displays the heat map of the interaction signal after aligning the *R. pearsonii* Hi-C data to the *R. ferrumequinum* genome. The middle panel shows the collinearity of the sequences related to two ancestral chromosome fission events between *R. pearsonii* and 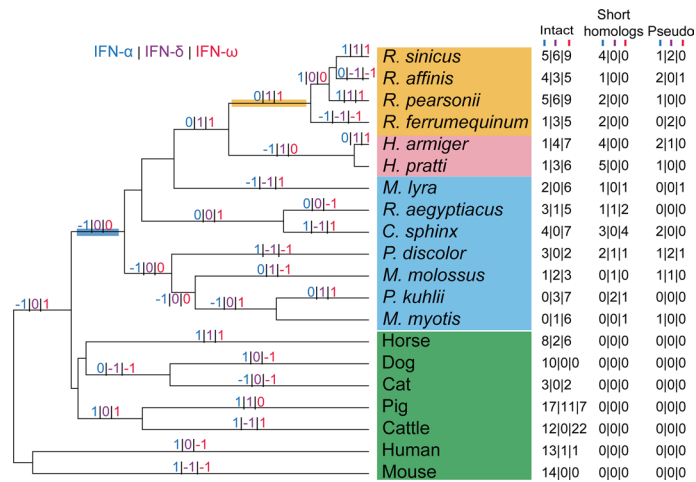*R. ferrumequinum*, and between *R. affinis* and *R. ferrumequinum*, respectively. The bottom panel displays the heat map of the interaction signals after aligning the *R. affinis* Hi-C data to the *R. ferrumequinum* genome. Abbreviations: RAC, *Rhinolophus* ancestral chromosome; Chr, Chromosome.
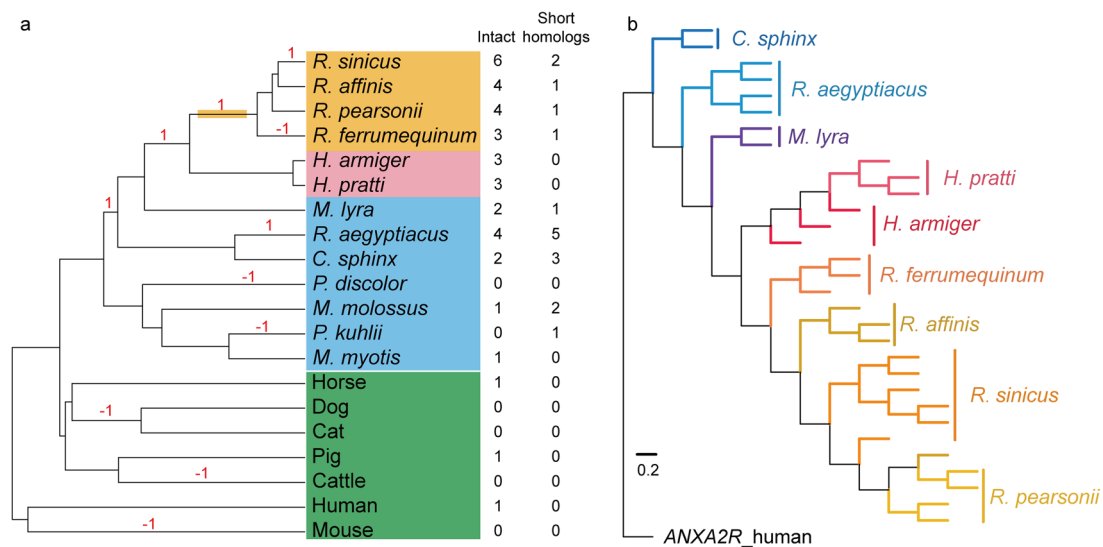
Extended Data Fig. 4 | Evolution of *ANXA2R* genes. (a) The significant change of ANXA2R loci in each branch. "1" represents expansion, "−1" denotes contraction, no marks mean no change. Orange shading corresponds to horseshoe bats (Rhinolophidae), pink to roundleaf bats (Hipposideridae), blue to other bats, and green to non-bat mammals. **(b)** Phylogenetic tree analysis of intact ANXA2R genes of Yinpterochiroptera bats. The human ANXA2R sequence is used as an outgroup. All nodes received 100% bootstrap support.
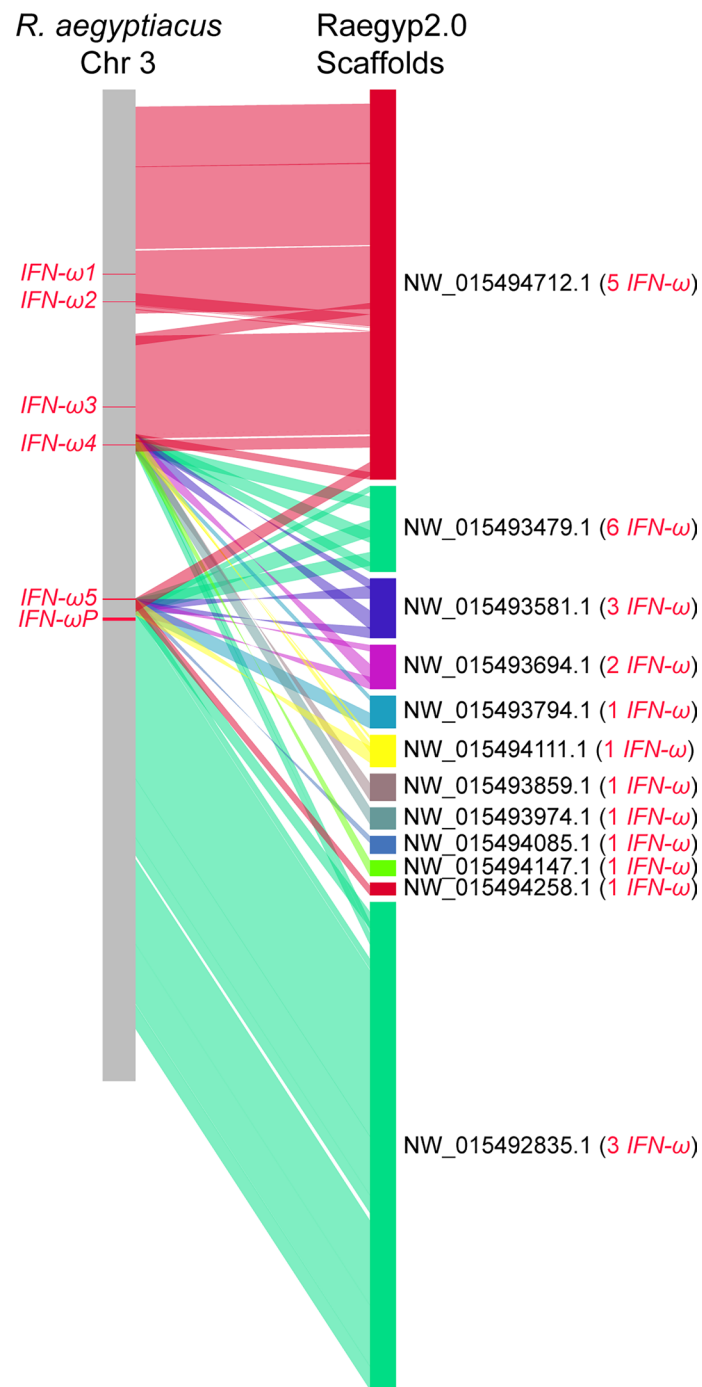
Extended Data Fig. 5 | Collinear analysis of *R. aegyptiacus* between the chromosome-level genome and the scaffold-level genome (Raegyp2.0). There were many duplicated scaffold sequences where IFN-ω was located when examining.

	Intact	Short homologs	Pseudo
R. sinicus	5|6|9	4|0|0	1|2|0
R. affinis	4|3|5	1|0|0	2|0|1
R. pearsonii	5|6|9	2|0|0	1|0|0
R. ferrumequinum	1|3|5	2|0|0	0|2|0
H. armiger	1|4|7	4|0|0	2|1|0
H. pratti	1|3|6	5|0|0	1|0|0
M. lyra	2|0|6	1|0|1	0|0|1
R. aegyptiacus	3|1|5	1|1|2	0|0|0
C. sphinx	4|0|7	3|0|4	2|0|0
P. discolor	3|0|2	2|1|1	1|2|1
M. molossus	1|2|3	0|1|0	1|1|0
P. kuhlii	0|3|7	0|2|1	0|0|0
M. myotis	0|1|6	0|0|1	1|0|0
Horse	8|2|6	0|0|0	0|0|0
Dog	10|0|0	0|0|0	0|0|0
Cat	3|0|2	0|0|0	0|0|0
Pig	17|11|7	0|0|0	0|0|0
Cattle	12|0|22	0|0|0	0|0|0
Human	13|1|1	0|0|0	0|0|0
Mouse	14|0|0	0|0|0	0|0|0

Extended Data Fig. 6 | Evolutionary expansions and contractions of type I IFN genes across 20 focal mammals. "1" represents an expansion event relative to its ancestral clades, "−1" denotes a contraction event relative to its ancestral clades. "0" and no marks mean no change. We used the conditional likelihoods to test the statistical significance for each lineage (P < 0.05). Orange shading corresponds to horseshoe bats (Rhinolophidae), pink to roundleaf bats (Hipposideridae), blue to other bats, and green to non-bat mammals.
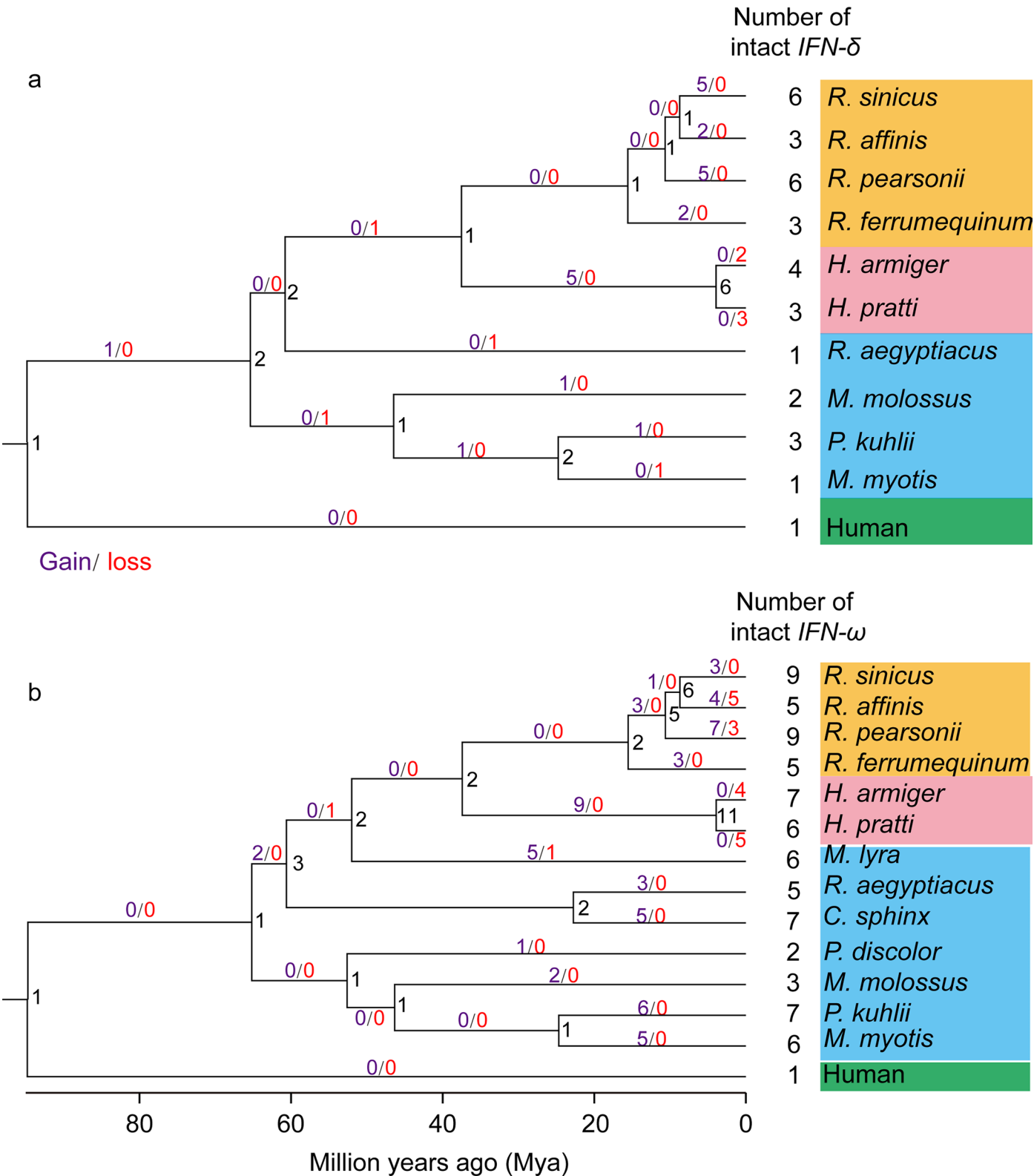
Extended Data Fig. 7 | Gene gain and loss events in the examined bats. The intact *IFN-δ* (a) and *IFN-ω* (b) gene repertoires of bats, with humans serving as the outgroup. Gene gain and loss events are mapped to the species tree, marked by purple and red numbers, respectively. Orange shading corresponds to horseshoe bats (Rhinolophidae), pink to roundleaf bats (Hipposideridae), blue to other bats, and green to humans.
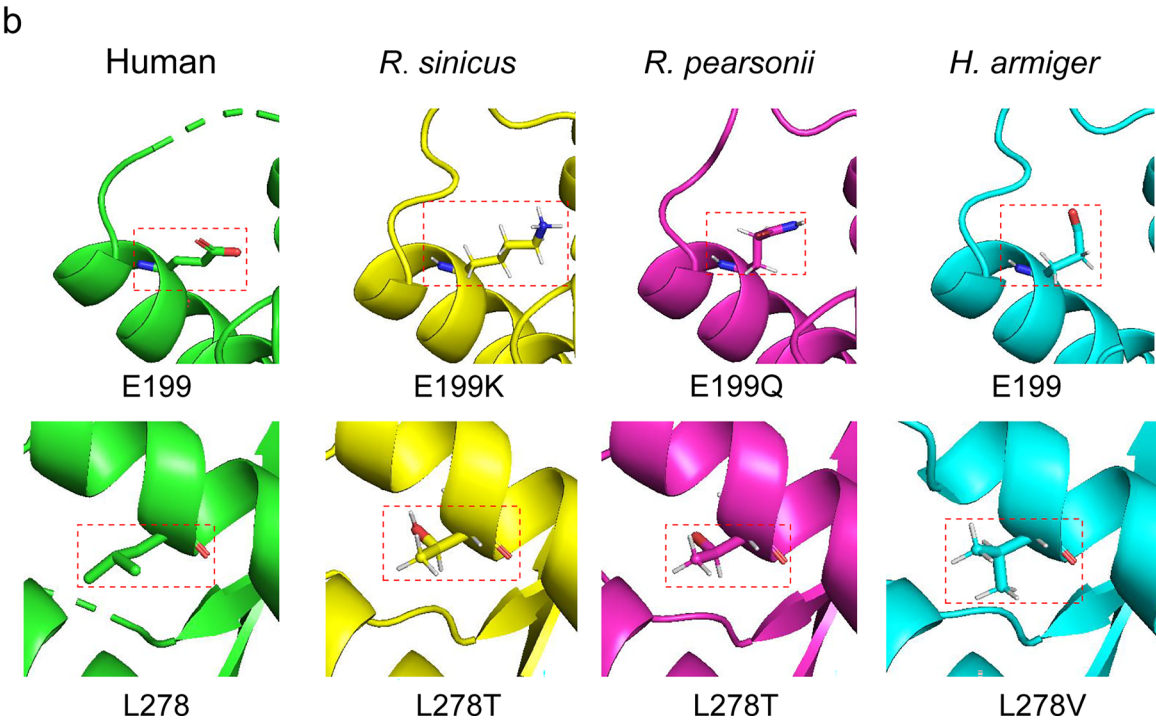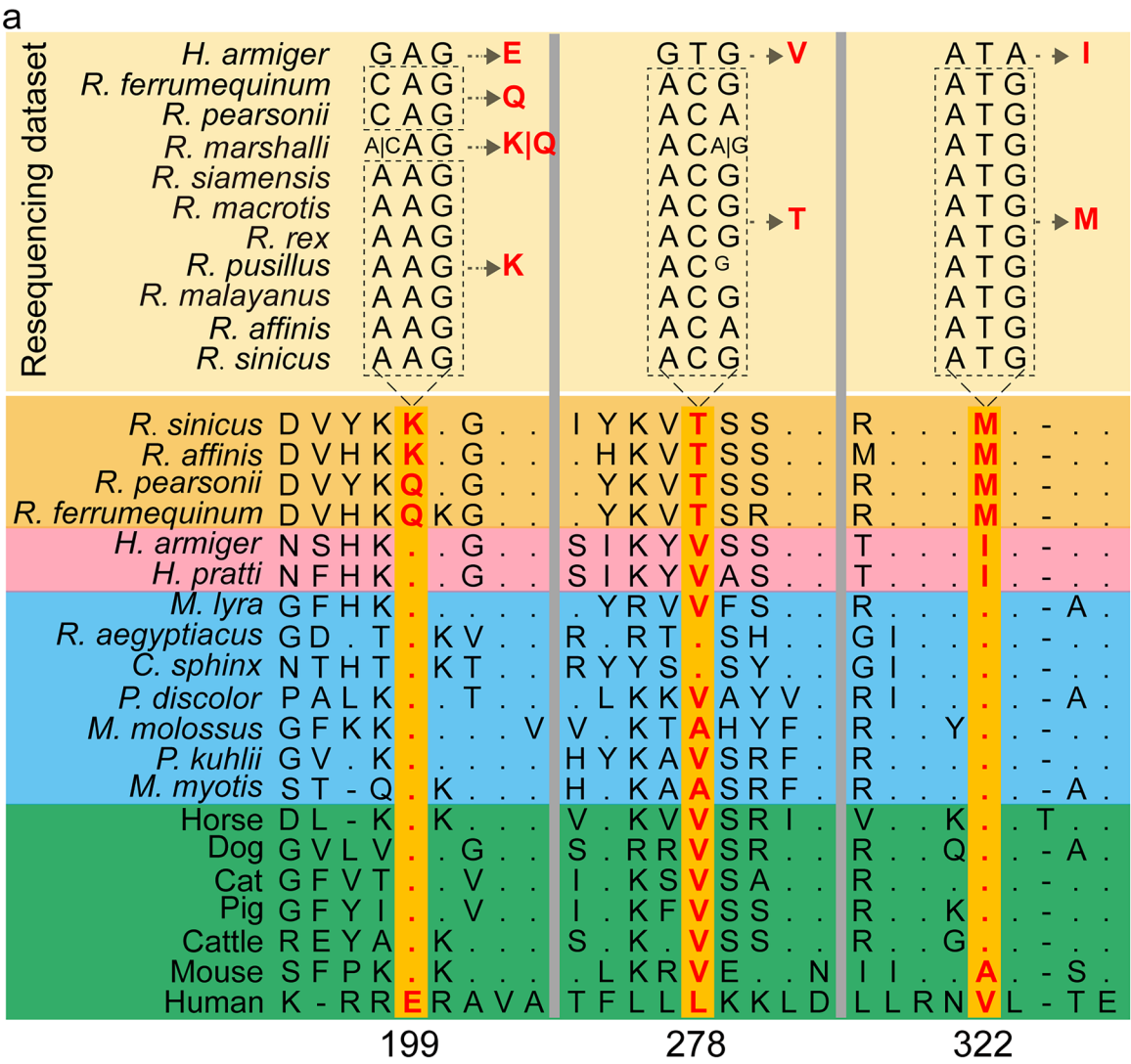
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Molecular evolutionary changes in C5aR1 protein.
(**a**) Alignment of C5aR1 protein sequences. The bottom panel shows the alignment for 20 mammals, with dots representing amino acids identical to the human sequence, and dashes denoting alignment gaps. The top panel shows the genotypes for three Rhinolophus-specific residues in 10 *Rhinolophus* populations, with H. armiger for comparison. We assessed the predicted physicochemical impact and found that E199Q (*R. ferrumequinum* and *R. pearsonii*) and L278T (all Rhinolophus bats) alters hydrophilic affinity, and E199K (*R. sinicus* and *R. affinis*) alters negative charge. (**b**) 3D-structure predictions of C5aR1 proteins for humans, *R. sinicus*, *R. pearsonii*, and *H. armiger*.
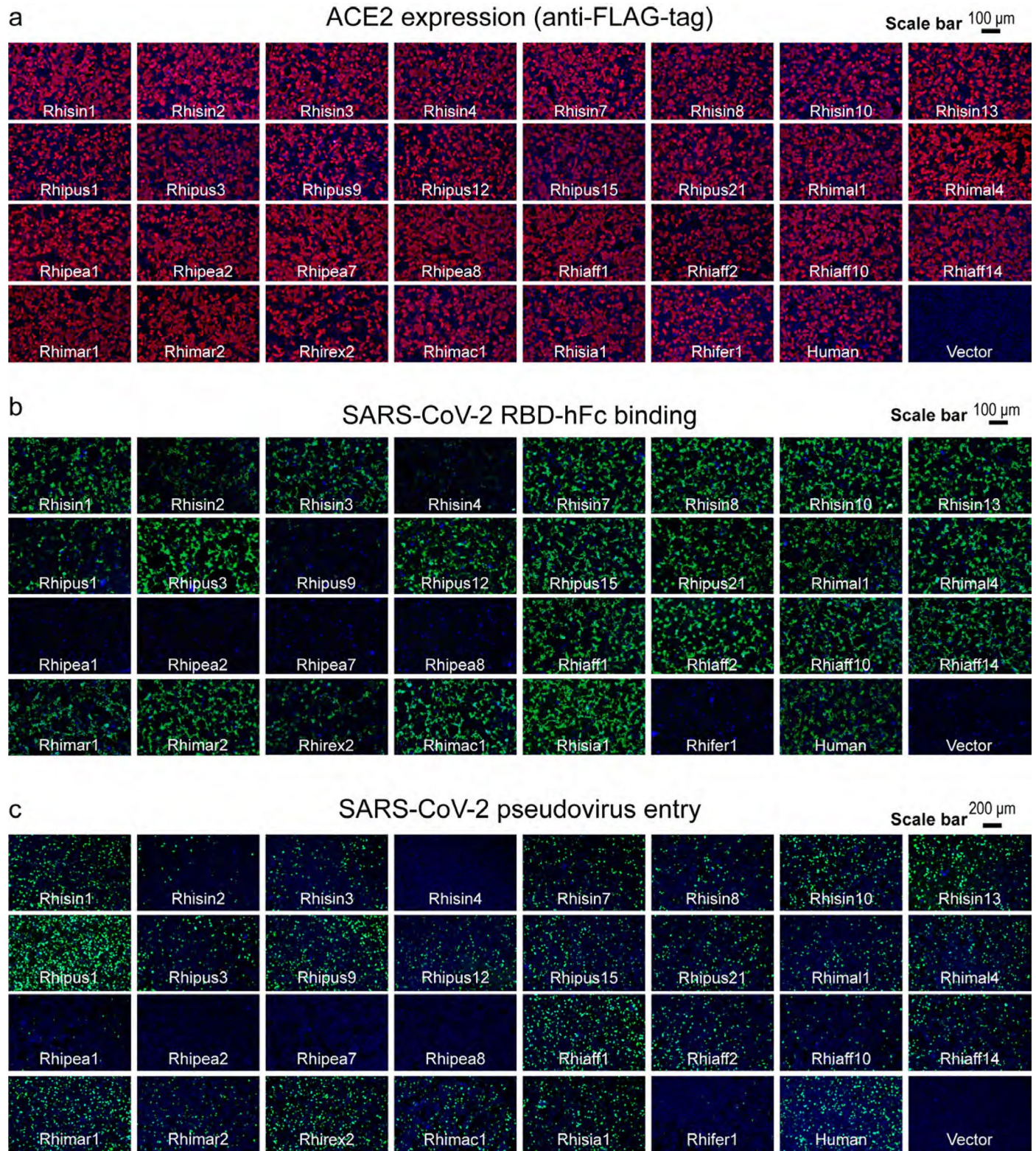
Extended Data Fig. 9 | Functional assays in *Rhinolophus* ACE2. (**a**) Expression of *Rhinolophus* ACE2 orthologues. We conducted immunofluorescence of intracellular *Rhinolophus* ACE2 expression level by detecting the C-terminal 3×FLAG-tag. The label 'Human' represents human ACE2-expressing HEK293T cells, and the 'Vector' indicates the HEK293T control cells. The scale bar is shown. One time for this experiment. (**b**) Assessment of the interaction between various *Rhinolophus* ACE2 orthologues and the SARS-CoV-2 RBD (receptor binding domain). Results were consistent across two biological replicates. (**c**) Characterization of *Rhinolophus* ACE2 orthologues mediating entry of SARS-CoV-2 viruses as shown using intracellular EGFP (Enhanced Green Fluorescent Protein). One time for this experiment.

nature portfolio

Corresponding author(s): Huabin Zhao, Stephen J. Rossiter

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

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection

1. Sequencing data
Nanopore sequences were collected from PromethION sequencer (Oxford Nanopore Technologies, UK); Paired-end short reads and Hi-C reads for assembly were from Illumina NovaSeq platform; 91 Rhinolophus resequencing data were sequenced on a MGI DNBSEQ-T7 platform (MGI Tech Co., Ltd, Shenzhen)

2. Experimental data
The luciferase activity are measured by Spectra Max iD3 multi-well Luminometer (Molecular Devices, United States), GloMax 20/20 Luminometer (Promega Corporation, United States).The OD450 absorbance signals are measured by Varioskan LUX Multi-well Luminometer (Thermo Fisher).Fluorescent images were captured with a fluorescence microscope (Mshot MI52-N).Bio-Layer interferometry assays were performed on Octet RED96 system (ForteBio, Menlo Park, CA).

Data analysis

The code and pipelines used for the analyses are available at Zenodo (https://doi.org/10.5281/zenodo.13690583) and GitHub (https://github.com/SLin191/Comparative-genomics-of-horseshoe-bats).
All software used are as follows:
fastp (version 0.23.2) was utilized for filtering Illumina sequencing data; Jellyfish (v2.1.3) was utilized for k-mer frequency analysis; NextDenovo (v2.5.0) was utilized for genome assembly; NextPolish (v1.4.0) was utilized for correcting initial contigs; ALLHiC (v0.9.13) was utilized for clustering linked contigs; Juicebox (v1.22) was utilized to visualise and adjust the placement and orientation of contigs; RepeatMasker (v4.0.5), RepeatProteinMask (v4.0.5), LTR FINDER (v1.0.7), RepeatScout (v1.0.5), PILER (v3.3.0) and RepeatModeler (v1.0.8) were utilized for repetitive sequences identification; BUSCO analysis (v5.4.2), Merqury (v1.3) were utilized for assessing assembly quality; blast (v2.14.0) was utilized for alignment; Solar (v0.9.6) was utilized for Sorting Out Local Alignment Results; Tophat (v2.0.13) was utilized for transcription data alignments; Cufflinks (v2.1.1) was utilized for predicting the isoform structure; Trinity (v2.1.1) was utilized for transcription

assembly; wise2 (v2.4.1) was used to generate gene models; PASA (v2.5.2) was used to assemble spliced alignments; Augustus (v2.5.5), GlimmerHMM (v3.0.1) and SNAP (v1.0), GeneID (v1.4) and GeneScan (v1.0) were utilized for predicting gene models; EVidenceModeler (v1.1.1) was used to integrate all gene model; LASTZ (v 1.04.15) was utilized for genome alignment; OrthoFinder (v2.3.1) was utilized for clustering paralogous and orthologous sequences; Gblocks (v0.91b) was utilized for minimizing the impact of multiple sequence alignment errors; PRANK (v.170427) was utilized for yielding high-quality alignments; BWA-MEM algorithm (v0.7.17-r1188) was utilized for alignment; Picard (v2.18.15) marked duplicate alignments; (GATK, v4.1.2.0) was used to identify SNPs; ANNOVAR (v2019-10-24) annotated SNPs; TreeBest (v1.92) constructed a neighbor joining tree; DESCHRAMBLER (v1.0) reconstruct the Rhinolophus ancestral karyotype; MODELTEST (v0.1.7) was used for selecting the best-fit model of evolution for DNA and protein alignments; RaxML (v8.2.12) was used for constructing maximum likelihood (ML) tree; MCMCTree (v4.5) was used for estimating divergence times; CAFE (v5.0) tested for gene family expansion and contraction; NOTUNG (v2.9) predicted gene gains and losses; MUSCLE (v 3.8.31) was used to generate multiple gene sequence alignments; PAML (v4.10.6) was used to test positive selection and rapid evolution genes; HISAT2 (v2.2.1), HTSeq (v2.0.3), DESeq2 (v1.42.0), R package (v4.2.0) were utilized for analysis of RNA-seq; AdmixTools (v7.0.2) was used to calculate genome-wide D-statistics; General tools for genomic analyses (v0.4) was used to calculate the related statistic fd; SweeD (v4.0.0) was used for genome-wide selective sweep analyses; GOfuncR package (v1.26.0) was utilized the Family-Wise Error Rate (FWER); GraphPad Prism (v9.4.1) was used for statistical analyses.The Bio-Layer interferometry data were analyzed by Octet® 958 Analysis Studio v12.2.0.20 (ForteBio, Menlo Park, CA).

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

New genome sequence data for the five bats are deposited at the Genome Sequence Archive (GSA) in National Genomics Data Center (NGDC) (https://ngdc.cncb.ac.cn) under accession code CRA018832. Genome assemblies are deposited at the NGDC GenBank under accessions GWHFDMV00000000.1 (R. sinicus), GWHFDMW00000000.1 (R. pearsonii), GWHFDMX00000000.1 (H. armiger), GWHFDMY00000000.1 (H. pratti) and GWHFDMZ00000000.1 (M. lyra). RNA-seq data of RfKT cells can be accessed at the NGDC (Accession Number: PRJCA023723). SARS-CoV-2 S protein sequence was obtained from GenBank (Accession Number: MN908947). Human ACE2 protein sequence was obtained from GenBank (Accession Number: NP_001358344.1). The data files used for the population genomics analyses, gene family analysis, and PAML analysis are available in the figshare repository (https://doi.org/10.6084/m9.figshare.27612597).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	n/a
Reporting on race, ethnicity, or other socially relevant groupings	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

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

Field-specific reporting

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

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

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

Ecological, evolutionary & environmental sciences study design

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

Study description	We generate chromosomal-level assemblies for horseshoe bats and their relatives, as well as low-coverage genome datasets spanning 10 horseshoe bat species. We assessed the impact of varying ANXA2R expression on the replication of a SARS-CoV-2 virus-like particle. We conducted three experimental assays (expression, RBD affinity and viral entry efficiency) to compare mutant versus

| Research sample | wildtype ACE2 proteins in 293T cells |
| | For whole-genome assemblies, we collected one female individual of each species of R. sinicus, H. armiger, and H. pratti from Xianning City, Hubei Province, China, one female of R. pearsonii from the Nanling Mountains, Guangdong Province, China, and one female of M. lyra from Huizhou City, Guangdong Province, China. All bats were captured with mist nets in caves. For resequencing, we collected wing membrane biopsies from 91 individuals from 10 horseshoe bat species, sampled at sites across Guangxi, Guangdong, Yunnan, Zhejiang, Hainan, and Guizhou provinces in China. Additionally, we downloaded paired-end read datasets (mean 34.83 Gb) for 18 H. armiger individuals from the NCBI database (accession PRJNA309742) |

| Sampling strategy | We collected 94 individuals from 10 Asian horseshoe bat species. In comparison, we generated chromosome-level genomes for three sister taxa of these horseshoe bat species, comprising two roundleaf bats (Hipposideros armiger and H. pratti) and one false vampire (Megaderma lyra). |

| Data collection | Nanopore sequences were collected from PromethION sequencer (Oxford Nanopore Technologies, UK); Paired-end short reads and Hi-C reads for assembly were from Illumina NovaSeq platform; 91 horseshoe bat resequencing data were sequenced on a MGI DNBSEQ-T7 platform (MGI Tech Co., Ltd, Shenzhen) |

| Timing and spatial scale | All the samples in our study were collected during the period from 2020 to 2023. |

| Data exclusions | No data were excluded from analysis. |

| Reproducibility | Experiments were repeated 2~4 biological replicates,each yielding similar results. Most of the assay were conducted three times with biological triplicates. our results are highly reproducible as most of them are well-established cell-based in vitro assays with strict controls. All result described in this study could be replicated or reproduced. |

| Randomization | We used 30 haplotypes ACE2s across the 10 horseshoe bat species for RBD binding and viral entry according to the 25 sites previously reported to interact directly with the RBD of SARS-CoV-2.Therefore ACE2 species randomization is not relevant to our study. |

| Blinding | No relevant blinding. |

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

| Field conditions | Sampling locations are spread across various provinces in China, but the field laboratory stations in each location are fully equipped with the equipment required for the experiments to ensure that the experiment is carried out normally. In this study, temperature and rainfall had no impact on the research results. |

| Location | We sampled at sites across Guangxi, Guangdong, Yunnan, Zhejiang, Hainan, Hebei and Guizhou provinces in China. |

| Access & import/export | The research design follows the experiences from peers worldwide. Meanwhile, it seriously considers subjects in terms of protecting wild animals (e.g. quantity, habitats),minimizing the hurt of experiments to animals, and respecting lives, striving to achieve the harmony between human beings and nature. |

| Disturbance | No relevant disturbance. |

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used	Primary antibodies: Anti-Flag Mouse Sigma, F1804/clone M2 (immunofluorescence assay,1:1,000); Anti-HA-tag mAb (Cat: M180-3, MBL) (immunofluorescence assay,1:1,000). Secondary antibodies: Alexa Fluor 594-conjugated goat anti-mouse IgG, Goat, Thermo Fisher Scientific, A32742 (immunofluorescence assay, 1:1,000); Alexa Fluor 488-conjugatedgoat anti-human IgG, Goat Thermo Fisher Scientific, A11013 (binding assay,1:1,000); Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed, Alexa Fluor Plus 488, Thermo Fisher Scientific, A32723 (immunofluorescence assay,1:1,000). Neutralizing antibodies: Anti-VSVG, Mouse ATCC: I1-Hybridoma (CRL-2700) cell line.
Validation	Mouse anti-FLAG® mAb, M2 Sigma-Aldrich (IFA,1:1,000). Vendor validated highly sensitive and specific detection of FLAG fusion proteins by WB/IP/IHC/IF/ICC. Optimized for single banded detection of FLAG fusion proteins in mammalian, plant, and bacterial expression systems. https://www.sigmaaldrich.cn/CN/en/product/sigma/f1804 Anti-HA-tag mAb (Cat: M180-3, MBL) (immunofluorescence assay,1:1,000) ; Vendor validated application for WB/IP/FCM/FCM. https://www.mblbio.com/bio/g/dtl/A/?pcd=M180-3

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	HEK293T (CRL-3216), Vero E6 cells (CRL-1586); I1-Hybridoma(CRL-2700) were purchased from ATCC. Caco2-N cells was generated from our own laboratory; The BHK-21-hACE2 cell line is a stable hACE2-expressing BHK-21 cell line, constructed by transducing the BHK-21 (CCL-10) cell line, purchased from ATCC, with a lentiviral vector encoding the human ACE2 (hACE2) gene.
Authentication	None of the cell lines used were authenticated.
Mycoplasma contamination	Cell lines used in this study were tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified line was used.

Animals and other research organisms

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

Laboratory animals	The study did not involve laboratory animals.
Wild animals	For whole-genome assemblies, we collected one female individual of each species of R. sinicus, H. armiger, and H. pratti from Xianning City, Hubei Province, China, and one female of R. pearsonii from the Nanling Mountains, Guangdong Province, China, and one female of M. lyra from Huizhou City, Guangdong Province, China. For genome resequencing, we collected wing membrane biopsies from 91 individuals from 10 horeshoe bat species, sampled at sites across Guangxi, Guangdong, Yunnan, Zhejiang, Hainan, and Guizhou provinces in China (Supplementary Table 12). All the bats were adults. All bats were captured with mist nets in caves, then each was placed into a separate clean cloth bag and transported to the temporary laboratory. All individuals were released at their sites of capture after sampling.
Reporting on sex	This study does not involve gender-related analyses.
Field-collected samples	The bats were periodically misted with water while in cloth bags to maintain humidity and were kept in a cool indoor environment with the temperature controlled at around 25°C, consistent with the natural light cycle. For the five bats used in the genome assembly study, they were euthanized, and liver tissue was extracted during dissection. The remaining carcasses were sterilized in an autoclave for centralized disposal. For the bats used in the resequencing study, wing membrane samples were taken, placed in cryovials, quickly immersed in liquid nitrogen, and later transferred to a -80°C freezer upon returning to the laboratory. The live bats were released at the original capture site during the night.
Ethics oversight	Field sampling of these bats was approved by the Guangdong Institute of Applied Biological Resources (GIABR) of the Institute of Zoology Guangdong Academy of Sciences (GIABR20200810).

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

Plants

Seed stocks

n/a

Novel plant genotypes

n/a

Authentication

n/a