New insight into the Germline Genes and CDR3 Repertoire of the TCRβ chain in *Chiroptera*

Hao Zhou[1], Jun Li[1], Longyu Liu[1], Dewei Zhou[1], Yingjie Wu[1], Xingliang Wang[2], Jiang Zhou[2], Qinqing Ma[1], Xinsheng Yao[1*], Long Ma[1*]

1：Department of Immunology, Center of Immunomolecular Engineering, Innovation & Practice Base for Graduate Students Education, Zunyi Medical UniversityZunyi, China

2：School of Life Sciences, Guizhou normal university, China

*：Correspondence：Xinsheng Yao: immunology@126.com, Long Ma: 164849671@qq.com

Keywords: Bat, TR loci annotation, Germline genes, TCRβ chain repertoire

Abstract：

Bats are natural reservoirs of many viruses, but their unique immune system allows them to co-exist with viruses and rarely manifest disease. Our understanding of the bat adaptive immune system is limited due to no database or tool that can process TCR High-throughput sequencing (HTS) data for bats and considerable diversity of the Chiroptera category. To explore our understanding of bats' t-cell immune system, we established a reference library of TCR-β germline genes based on the annotation of *Rhinolophus ferrumequinum* (Greater Horseshoe Bat, Rhinolophidae) (RF), *Phyllostomus discolor* (Pale spear-nosed bat, Phyllostomidae) (PD) and *Pipistrellus pipistrellus* (Common pipistrelle, Vespertilionidae) (PP). Further, we analyzed the *Rhinolophus affinis* (Intermediate horseshoe bat, Rhinolophidae) (RA) TCR-β repertoire by HTS. The bats' TCR-β repertoire was formed by rearranging of the V-D-J-C genes with a D-J/V-D deletion and insertion, showing a high diversity; the conserved amino acids suggest that bats' T-cell recognition of antigens is MHC-restricted. The sequencing and

analysis processes of immune repertoire established in this study can apply to the research of adaptive immune systems in bats and novel species.

Introduction

Bats (Chiroptera) are the second largest group of mammals in the world (*1*), The fact that they carry a large number of virulent zoonotic viruses but do not cause serious clinical symptoms of their own(*2-5*), has led to increasingly close attention to bats and their differences from the immune systems of other mammals.

Several differences between the bat and human or mouse innate immune systems were identified, such as activation of the PRR (Patterns Recognition Receptor) by RNA viruses, danger signals, or intracellular double-stranded DNA in humans or mice, resulting in activation of NLR-family pyrin domain containing 3 (NLRP3) or Interleukin-1β (IL-1β) inflammatory vesicles, in contrast to bats that inhibit the transcriptional initiation of NLRP3 proteins rendering them less functional, including Interferon-inducible protein AIM2 (AIM2) and Gamma-interferon-inducible protein Ifi-16 (IFI16), as well as reduced caspase-1 activity and IL-1β cleavage led to an overall reduction in inflammation(*6-9*); A recent study found that deletion of serine at the 358th key site of bat STING protein leads to inhibition of IFN secretion (*10*); Pavlovich et al. reported "high amplification and non-classical distribution of genes" at MHC-I loci in bats, respectively, and the authors suggested that the combination of these abundant non-classically distributed MHC-I type genes with highly expressed NKG2 (all NKG2 receptors in Egyptian fruit bats contain inhibitory interaction motifs) increased the activation threshold of NK cells and eventually led to a reduced response(*11, 12*). However, specific viral response studies in bats remain unclear. The main reasons are the lack of relevant reagents in cell biology studies and disease models of bats, in addition to this, isolating the virus and the diversity of bat species are also challenges(2 orders, 21 families, more than 1350 species) (*13*). The difficulty of

conducting bat-specific immune responses, coupled with the very limited information on the bat genome, has resulted in a near gap in bat characteristics from TCR and BCR.

With the development of Next Generation Sequencing (NGS), HTS technologies have been widely used for the analysis of the immune system, especially the immune repertoire (*14*). The TCR/BCR repertoire is a collection of all functional T/B cells in an individual's circulatory system, each characterized by its antigen-specific receptor. Deep sequencing of the repertoire has been extensively applied in the resolution of the adaptive immune system, assessment of individual health status, antibody development, and targeted detection and therapy (*15*). IMGT/HighV-QUEST, TRUST, MiXCR, etc., various tools have been used to analyze TCR/BCR repertoire(*16-18*).

In recent years, the bat1k project has published some high-quality bat genomic data, which has greatly helped our understanding of the bat immune system (*19, 20*), in particular, adaptive immunity studies of how bats respond to viral infections offer the possibility. However, gene annotation of the TR/IG germline in bats is currently extremely limited (*21, 22*). Together with the diversity of bat species, there is an urgent need for laboratories to work together to build a database that can be used to analyze the repertoire of bats and to resolve their differences from common mammals.

In our previous work, we completed the annotation of RF TRB loci, suggesting that the bat adaptive immune response is similar to that of humans or mice and that its TCR is rearranged from germline genes with high diversity (*22*). On this basis, we systematically annotated the TRB loci of the PD and PP, then compared the TRB loci of the three kinds of bats. Using the conserved TRBC gene, we aligned the TRBC Exon1 sequences of various bats, and designed primers to amplify the TCR β-chain CDR3 repertoire of RA by the 5’ rapid amplification of cDNA ends(5’race) method. Our study provides an important theoretical basis, a new research method, and a comparable database for the study of genetic evolution and adaptive immune response in *Chiroptera*.

Results

1.TR loci annotation and Reference directory establishment

The TRB loci analyzed by homologous genes and RSSs methods was recorded in Table 1 while exporting Genebank files (Supplementary Data 1). The three kinds of bat did not differ greatly in the structural distribution of whole TRB loci. All maintained the classical structure of mammals: multiple TRBV genes involved in rearrangements upstream, a one-three D-J-C clusters downstream, and a reverse TRBV gene after the last D-J-C cluster (Figure 1B, Supplementary Figure 1). Beyond that, the length distribution and orientation of TRB loci differed among the three bat species, but most of the regions were homologous (Figure 1C), with the main differences arising from internal duplication and deletion of the V gene family.

The numbers of PD, RF, and PP TRBV genes were 100 (including 2 pseudogenes whose families could not be identified), 30, and 45, respectively. A phylogenetic tree was constructed using nucleotide sequences of TRBV genes. Although each bat species has different V gene family lost, all genes could be found homologous to each other or in pigs and humans, and we did not find functional V gene families that could cluster separately in bats (Figure 2A). Internal duplication of family is the underlying cause of main differences in the number of germline genes, and this duplication process appears to have some degree of commonality. Regarding consistency of family duplication among species and overall numbers, Chiroptera is more similar to Artiodactyla in terms of the evolutionary direction of TRBV genes: random and large-scale, whereas carnivores are more similar to primates (Figure 2B). Except for the TRBV17 family, most of the TRBV and TRBJ gene families of the three bat species were more than 70% similar (Figure 2C). Here, we found a particular TRBV12 family of PD containing 37 members, close to almost all TRBV genes of PP and even more than the total TRBV genes of RF. Even in the known mammalian range, no TRBV multigene family is found with more than 30 members (the only one with 30 members is TRBV5 in goats)(*23*).

In the amino acid Multiple Sequence Alignment (MSA) of TRBV genes, we found that not only the conserved sites 23Cys, 41Trp, 89Leu and 104Cys existed, but also Gln at position 6, Tyr at position 42 was extremely conserved. Gln at position 44 in the PP was also highly conserved. In addition, most TRBV genes contained conserved CASS motifs at the end (Figure 3A, Supplementary Figure 2A). The number of pseudogenes is a key factor influencing the composition of the receptor pool, and a higher number of germline genes and more duplications within the family in PD implies an increased probability of gene mutation; therefore, the proportion of TRBV pseudogenes in PD, 17%, is the highest among the three bat species (7% in RF and 11% in PP), and even so, the number of 97 functional genes (V/D/J genes) in PD far exceeds that of RF and PP (Supplementary Table 2). The D, J, and C genes are named according to the location of the cluster. Like primates and carnivorous, the Chiroptera contains two D-J-C clusters. In terms of the composition of the number of genes in each cluster, the only difference between the three bat species is the number of J gene in the second cluster. All TRBJ genes in bats contain conserved FGNG motifs except for that TRBJ2-3 in RF (Figure 3B, Supplementary Figure 2B). As previously reported, all bats' TRBD genes are guanine-rich sequences, with the 23RSS and 12RSS downstream and upstream (Figure 3C). The RSS is no significant difference in these mammals and bats, the most conserved site are always the first four sites (CACA) of heptamer and three continuous adenosines in the middle position of nonamer (Figure 3D and Supplementary Figure 3).

Regarding the MiXCR reference library of performing HTS data, most of the V and J genes of the three kinds of bat had more than 70% similarity, except the TRBV17. For V gene, we grouped each TRBV family of three kinds of bat into one family and TRBJ gene nomenclature is similar to that of alleles in IMGT (Supplementary Table 3). The reference library was produced based on the nucleotide sequences and results of region delineation by amino acid sequences MSA.

2. TCRβ chain Repertoire Building、Sequencing、Analyzing

The Cytochrome b (Cytb) sequencing results of three experimental bats had an average length of 488 bp after removing the bases with more heterochromatic peaks at the beginning and end, with a similarity of 98.5% (Cytochrome b). They were determined to be *Rhinolophus affinis* after BLAST according to the identity and score (Supplementary Table 4). Additionally, we compared the TRBC Exon1 nucleotide sequences of 15 kinds of bat and designed primers in the conserved regions to construct TCRβ chain repertoire by 5'RACE. The nucleotide sequence similarity of 16 TRBC Exon1 ranged from 77.66% to 98.9%, with many fully conserved regions suitable for repertoire constructing and sequencing using conserved regions for 5'RACE (Figure 4A).

The sequencing process are completed by ImmuQuad Company. Here is a peak between 600bp and 700bp in the RA samples, indicating a successful library build (Supplementary Figure 4). Next, the raw data of nine samples were processed by MiXCR.

On average, 4.27 million TCR reads and 24904 clonetype were obtained for each bat sample and 15 million TCR reads and 1.23 million clonetype were obtained for each mouse sample according to the 5'race method; 3.8 million TCR reads and 75000 clonetype were obtained for each human sample according to the m-PCR method. Although there were differences in the bat samples' sequencing depth, all met the CDR3 sequence depth analysis requirements for HTS sequencing (Table 2).

For the sequencing results of RA, to explore the degree of bat species similarity, we firstly compared the Cytb sequence of the three experimental bats with RF and PP (Figure 4B). The RF (87.9%) had higher similarity to RA than PP (74.6%). Besides, the TRBC fragment demonstrated similarity between RA with PD, RF and PP: 77.1%, 93.8%, 75.0% respectively (Figure 4C). After categorizing the sequencing results of the experimental samples RA, we found that most of the sequences of RA output sequences belonged to RF (Figure 4D). The comparison results further indicated that the sequences of RA were more similar to those of RF. Then, we demonstrated the completeness of the repertoire by

exporting 25 V and 14 J gene sequences from Bat1 samples (Figure 4E).

The similarity of the bat, human, and mouse CDR3 repertoire was assessed by the four indices of overlap coefficient, Jaccard, Morisita, and Tversky (Figure 5A), and it was found that the bat CDR3 overlap indices fluctuated widely, with three of them slightly higher than human but significantly lower than mouse. Based on the phylogenetic tree analysis, we analyzed the homologous TRBV gene usage and V-J pairing in bats, humans, and mice (Figure 5B, Figure 5C and Supplementary Figure 5). The homologous genes expressed were: TRBV3, TRBV13, TRB15, TRBV20, TRBV28, and TRBV30, among which TRBV15 and TRBV30 had similar expression ratios. Other homologous genes had some differences in access. In addition, we compared the usage of all TRBV genes in bats and humans (Figure 5F). The top 3 usage in bats were: TRBV12, TRBV18, and TRBV20, and humans were: TRBV13, TRBV28, and TRBV29. The TRBV2 and TRBV10 were not found in any of the three annotated bat genomes. TRBV9 and TRBV19 were not expressed in either bat1 and there was very low expression in bat2 and bat3 sample. The TRBV9 group was only found in PD, and all four TRBV9 members were pseudogenes. TRBV17 was very lowly expressed in all three bats. There are several possible explanations for these lowly expressed families: Firstly, the expression in RA itself this may be the case; in addition, because of the very low homology of the TRBV17 family in the three annotated bats, it is possible that TRBV17 also undergoes large mutations in the RA and therefore cannot be detected. This kind of mutation might lead to functional genes becoming pseudogenes, a situation that is very common across species. For example, the TRBV1 gene in bats is functional gene, while TRBV1 in humans is pseudogene. There are also cases where functional genes are not detected as expressed, for example, TRBV18 is barely expressed in humans but relatively highly expressed in bats; similar to TRBV23, TRBV24, TRBV26, etc. All functional TRBJ gene were expressed among the three species except TRBJ2-8 (pseudogene) in bats and TRBJ2-6 (pseudogene) in mice (Figure 5D). In the comparative TRBJ analysis, it was found that bats, human, and mouse all had high usage of

TRBJ2-1, TRBJ2-3, and TRBJ2-7 at the same time. In addition, we found that most of the J genes in the 3 species were used from the second D-J-C cluster (Figure 5E), a result that may suggest that the usage of genes during rearrangement is correlated with distance.

In the analysis of the motifs of the CDR3 region (5 AAs), we found that among the top ten high-frequency take motifs, four were shared in bats, humans, and mouses (Figure 6A, Figure 6B). We observed that most of the shared motifs among the three species were CASSN motifs. Since the CASS motifs in the V gene region belong to the antigen docking function directly related to the CDR3 region, the cell populations that play a major role in the repertoire of the three species may be similar. In another characterization of the CDR3 region, it was found that the length of the CDR3 region, which is similar to that of the mouse, has a bell-shaped distribution with 14 AAs as the main peak, whereas the human CDR3 region has a bell-shaped distribution with 15 AAs as the main peak (Figure 6C). To preliminarily assess the length effects caused by insertion and deletion of nucleotides in the CDR3 region, we calculated the insertion and deletion of V and J genes and found that human and mouse had similar deletion at the V'3 end. Bat and mouse had almost the same insertion at the V 3' end and the J gene 5' end (Figure 6E, 6D). The amino acids in the CDR3 region of the three species were very consistent, with all high frequencies taken from S, G, and A (Figure 6F).

Discussion

The lack of basic germline genetic information has hampered the development of immunological studies in bats. Although the importance of bats in virus transmission and specific physiological traits has been noticed, a comprehensive understanding of the bat immune system remains elusive. As early as 1991, researchers identified T, B, and Macrophage populations in Indian fox bats by electron microscopy, and the ratio of T/B cells in the spleen and lymph nodes was significantly higher than in mice (*24*) ; Julia et al. found that CD4+ T cells in fox bats dominate the blood lymph, while CD8+ T

cells dominate the spleen (*25*). These indicate that the T/B cell immune in bats plays a role, but the knowledge is limited.

In 2021, the TR/IG loci of the Greater horseshoe bat and the Egyptian fruit bat were annotated separately (*21, 22*), where TR is annotated from the chromosome level of the genome, IG is annotated using the Bacterial artificial chromosome (BAC). However, to date, no public database or tool has been used to analyze the TCR/BCR sequences for bats, and the TCR/BCR repertoire characterization of bats has not been reported. In order to establish a bat germline gene database, the priority work was a complete annotation of TR/IG loci. Unlike the usual genome annotation, which requires high genome quality due to the enormous span of TR/IG loci on the genome, the annotation at the scaffold or contig level results in the loss of most germline gene. In this study, we finally identified three bat species with high genomic quality from the 45 bat species in the existing database: RF (22), PD (Phyllostomidae), and PP. Second, the methodological approach to gene annotation, the central issue in creating germline gene databases is finding all potential genes as possible based on specific structural features. Due to the short sequence D (10-15 bp) and J (30-50 bp) genes, in order to avoid losing sequences from the homology search, we used RSS sequence search method at the same time. There may be non-classical 23 RSS (23+/-1 nt) to influence the final result. However, this kind of gene has minimal examples in human and mouse. Moreover, pseudogenes that lack RSS cannot involve in the rearrangement process. This means that these pseudogenes have no effect on the formation of the productive repertoire and pathogenic response(*26*).

Although several studies have reported the presence of varying degrees of replication of immune-related genes in Chiroptera, for the TCR, a large number of familial replication events were only discovered in Artiodactyla such as cattle (134 TRBV, 371 TRAV/DV)(*27, 28*)and goats（94 TRBV）(*23*), which are well known for their diversity in the T cell receptor pool. The (37 family members) super-amplification events of the TRBV12 family in PD and the considerable variation in TRBV gene

numbers (30 and 100) among Chiroptera have never been reported in mammals. On account of previous studies of primates: humans, rhesus monkeys, and crab-eating monkeys(*29-31*), carnivores: cats, dogs, ferrets (*32-34*) and three species of camels(*35*), species of the same order differed very little in TRB loci both in terms of loci structure and the number of germline genes. Phylogenetic analysis classified Chiroptera as a sister order to Carnivora and Artiodactyla (*36*). We found that the distribution of TRBV gene families in the three bat species was more similar to that of Artiodactyla, with random and massive replications. Based on the results of the annotation so far, the overall structure of the TRB loci, the conserved sites of germline gene, and the RSS do not differ much from those of usual mammals, except for the number of germline V genes. Interestingly, it has been reported that *Rhinolophus Hildebrandti* and *Pipistrellus pipistrellus* do not contain gut-associated lymphoid tissue (GALT) (*37, 38*). The chicken immunoglobulin system is a typical GALT species that relies primarily on Gene Conversion to generate receptor diversity. However, bats have not been reported to possess similar Gene Conversion-related in birds and rabbits (*39, 40*). T cells do not have SHM- or CSR-like mechanisms for receptor diversity, so the possibility of gene rearrangement is the predominant mechanism for generating TCR diversity. The difference in rearrangement likelihood makes it likely that TCRβ chain repertoire (naive) diversity is present in at least the three kinds of annotated bats.

The anchor residues in the TRBV gene are associated with MHC peptide binding. Although the MHC genes in bats are more diverse than in other mammalians, some research has been reported and showed bats' MHC genes have high similarity in sequences with human, mouse, dog, cow et al(*41-43*), the anchor residues in annotated TRBV and TRBJ genes are highly conserved, this finding supports the TCR recognition in bats are MHC restriction.

Here, we established a HTS sequencing-based method to analyze the bat TCR β chain repertoire. From the total number of TCR β chain CDR3 sequences, productive sequences, proportion of cloned sequences and compositional features of the 3 RA bats, and comparative analysis with the human and

mouse TCR β chain CDR3 repertoire. The three bat samples were able to obtain an average of 4.2 million reads, but the number of TCR clonetypes obtained for bat1 samples (3498) differed from bat2 and bat3 (25443, 45771). There are several possible explanations: Firstly, this bat sample may have little expression of TCRs. Second, since this is the first time that we performed repertoire building as well as sequencing of bats, the sequencing depth may have caused an impact, and increasing the number of sequenced samples would help to solve the first two problems due to the limited collection of samples; Third, regarding the impact of primers, we found that 25 TRBV genes as well as 14 TRBJ gene families were present in each experimental bat after the sequence output of each sample, suggesting that we have the first time to TCR β chain CDR3 in bats with very high completeness and accuracy, providing specific technical means and data analysis tools for designing and optimizing primers in TRBC of bats for HTS sequencing analysis of CDR3 group libraries of bats from different families using 5'race. The composition and characteristics of the TCR β chain CDR3 sequences of RA were firstly found to be very similar to the annotated genes of RF, with 99% of the C genes belonging to RF, and TRBV and TRBJ were also basically identical to the annotated V and J genes of RF. There is a high genetic conservation in the composition of specific immune response genes among different species of bats of the same family (RA and RF). Because of the large number of bat species, our first results provide a theoretical and feasible experimental protocol for conducting bat-specific immune response studies between different bat species of the same family.

We analyzed the usage and pairing patterns of TRBV, TRBJ genes, and V-J in bats, humans, and mice. Although there are more than 50 and 30 germline V genes in humans and mice, respectively, the specific immune response may rely on only 1 or a few of them (*44*). We first compared the homologous V genes of the 3 species, and in general, the V gene usage patterns were very different except for TRBV30, which is located in a specific position downstream of the second C gene in a position opposite to transcription in bats, humans, and mice (TRBV31), and several studies have reported that V

and J gene usage during rearrangement correlates with position in the locus (*45-47*) ,Interestingly, more than 60% of the J gene take in all 3 species is from the second D-J-C cluster. Human TRBV12-5 and mouse TRBV13-2 are frequently associated with a number of autoimmune diseases, and this class of TCRs occupies more than 50% of NK-T cells. In fact, mouse MAIT cells express a TCR-α chain usually consists of TRAV1 and TRAJ33, and the paired β chains are TRBV13 and TRBV19(*48-52*), In contrast, TRBV12 in bats and TRBV13 in mice were both far ahead in the usage of their respective V genes. Bats showed significant heterogeneity with humans in multiple TRBV usage, suggesting that there is a specific biased amplification in the evolution of specific immune response genes in the TCR of bats, which may be related to their evolutionary pressure or coevolutionary with viruses.

The bat, human, and mouse TRB CDR3 region AA usage is essentially the same and have high consistency in high-frequency motifs. In contrast, they show more significant differences in motif usage throughout the CDR3 region, suggesting that the V-terminal and J-front ends of the constituent mammalian CDR3 regions are more conserved and that the typical antigenic selection encountered in evolution may be relevant. The length of the TRB CDR3 region in bats and mice was consistent, both being shorter than that in humans, which we analyzed as being related to differences in insertion and deletion at the V'3 end and J5' end in the three species, in addition to differences in the evolutionary length of the D gene and V/J involvement in the CDR3 region genes; the overlap between individual bat CDR3 regions was higher than that of human, but significantly lower than that of mouse, suggesting higher concordance of CDR3 regions between different individual BALB/c mice with the same genetic background.

Our first annotation of the complete TRB loci in the RF, PD, and PP, as well as the first construction of the repertoire, sequencing, and MiXCR analysis database for the bat CDR3 repertoire, and the successful application in the bat (*Rhinolophus affinis*), have provided the opportunity to carry out the TCR gene annotation and specific immune response mechanism in bats. It provides a new theory,

feasible technical tools, and primary data for comparative analysis in studying the mechanism of virus tolerance in bats.

Materials and Methods

1.TRB locus annotation

Figure 1A shows the main experimental flow of this study. Based on the mammalian shared boundary genes for TRB, MOXD2 (Monooxygenase, DBH-like 2), and EPHB6 (EBH receptor), we localized the TRB chromosomal location in three bat species: *Phyllostomus discolor*: Chr10 (NC_040912: 85517683-85943353). *Rhinolophus ferrumequinum*: Chr26 (NC_046309: 7911840-8171343); *Pipistrellus pipistrellus*: Chr10 (LR862366.1: 26582722-26886962). Two methods were used in the germline gene annotation: 1. IMGT-LIGMotif (*53*): To find germline genes in the sequence by homology with known species, we selected representative animals of three different subjects for the search: human (primates), mouse (rodents), and pig (Artiodactyla); to prevent missing information due to species differences, we downloaded TRBV genes of all known species in IMGT and map them to the TRB loci of three bat species by Geneious Prime(Version 2022.2.1). 2. 12/23 RSS (Recombination Signal Sequence) scanning: similar method to previous studies(*22*). Briefly: we screened all RSS motif within the three bat TRB loci and searched for V/D/J genes upstream and downstream of the RSS. The number of germline genes obtained by several methods and the specific location information are summarized in Table 1. Following the IMGT guidelines (*54*), We labeled the essential components of each germline gene within the locus after annotation by Geneious Prime and exported the annotation file in Genebank format (Supplementary Data1). To compare the similarity of TRB loci of the three annotated bats, we performed a homology comparison based on 70% similarity by Easyfig(Version 2.2.5).

2. Germline gene naming and TRBV families' cluster

Family naming was performed for each annotated gene according to the IMGT guidelines: for V genes, nucleotide similarity more significant than 75% was considered the same family; a biological evolutionary tree was constructed with other species of TRBV genes to identify specific V gene families, and we downloaded human, mouse, and pig TRBV genes from IMGT-GeneDB (https://www.imgt.org/genedb/) (Accession number is shown in Supplementary Table 1), and performed nucleotide multiple sequence alignment of TRBV genes by MEGA7(Version 7.0.26) to construct a phylogenetic tree by Neighbor-joining method, and the principles of gene selection were:1. Only functional gene and ORF, 2. One gene per family. We also compared twelve species of four mammalian subjects' TRBV gene family distributions to understand the evolutionary direction of the V gene, including Artiodactyla: cattle, sheep, and pigs; Carnivores: dogs, cats, and ferrets; and primates: humans, crab-eating monkeys, and rhesus monkey (https://www.imgt.org/IMGTrepertoire/). D, J, and C genes were named according to the position of the cluster to which they belonged on the locus.

3. Reference library establishment

The amino acid sequences of the germline V and J genes were aligned by Geneious Prime(Clustal Omega), with priority alignment of three conserved sites: 23CYS, 41TRP, and 104CYS, based on the known structure and specific positions of mammalian TRBV(*54*). The structural domains were classified according to the results of the alignment, V: FR1 (1- 26), CDR1 (27-38), FR2 (39-55), CDR2 (56-65), FR3 (66-104), CDR3 (105-117), and J: FR4 (phenylalanine until the end of the sequence of the FGNG motif). The functional description of germline genes was referenced to IMGT, based on stop codons, RSS, conserved amino acids, and splice sites. The pseudogenes as well as the ORFs interpretation are described in Supplementary Table 2. Further we exported the germline genes nucleic sequence and their structural domain information as MiXCR identifiable reference gene library files. This library can be used to analyze bat HTS sequencing TRB data.

4.Sample collection

Bat samples were collected in XIshui County, Zunyi City, Guizhou Province with the help of Zhou's group. Muscle tissues were used for genotype identification and then the Cytb gene was amplificated with designed primers (Supplementary Table 5), and the PCR products were sent to Sangon Biotech (Shanghai, China) for sanger sequencing. The genotype determination of bats was made based on the E-value as well as sequence similarity. Finally, three RA were selected in this study. The spleen tissues were used for RNA extraction and TCRβ chain library construction. 3 5-month-old BALB/c mice were obtained from the animal center of Zunyi Medical University, the spleen tissues were used for RNA extraction and TCRβ chain library construction. The molecular cells from 5 ml peripheral blood of 3 individuals were collected from the Affiliated Hospital of Zunyi Medical University. The above study was approved by the Animal Protection and Ethics Committee of Zunyi Medical University.

5.TCR-β chain repertoire sequencing and analysis

Fifteen known bats' TRBC genes were aligned by their nucleotide sequences, then primers for TCR-β library construction were designed in the conserved region, and repertoires of RA were prepared by the 5'RACE method (Figure 4). The 5'race (rapid amplification of cDNA ends) library construction and sequencing were completed by ImmuQuad Company (Hangzhou, China). The 5'RACE TCR-β repertoires construction and sequencing work of mice was done by SEQHEALTH Company (Wuhan, China). The TCR-β repertoires of humans were prepared ourselves with multiplex PCR library construction (Supplementary Table 5), and the sequencing work was done by BGI Company, Shenzhen, China.

All raw data can be found in NCBI_PRJNA877449 (SRX17468175, SRX17468176, and SRX17468177). The sequencing results of bats were assembled and presented by MiXCR based on the library constructed from the annotated genes, and the sequencing results of humans and mice were also routinely pipeline analyzed by MiXCR (Version 3.0.13). The MiXCR output data of the nine samples are shown in Table 2 and Supplementary Table 6. To check the quality of reads, we exemplify the

sequences of bat1 sample, which containing 25 TRBV genes as well as 14 TRBJ genes. The evolutionary relationship between the sequenced samples and the three annotated bats by sequence attribution, Cytb, and TRBC gene alignment. Since this was the first time that HTS sequencing and repertoire construction were performed, subsequent analyses were performed using Productive-Clonetype sequences in order to exclude the effects of repertoire construction and sequencing, physiological health status, and to assess TCR-β repertoires homogeneity and heterogeneity among the three species. We performed the following analyses: 1. Repertoire overlap index statistics among the three species; 2. TRBV gene and TRBJ gene usage, V-J gene pairing, and conserved motif tracking; 3. CDR3 length and AA usage, nucleotide deletion and insertion.

6. Statistical analysis

The figures were draw with GraphPad Prism (Version 8.0.2), R package "ggplot2", and R package "Venn Diagram". Data analysis was performed by R studio (v3.3.3) and GraphPad Prism. P-values were calculated with the aid of the t test. $P<0.05$ was considered statistically significant.

Acknowledment

We thank Jiang Zhou and Xingliang Wang of Guizhou Normal University for their help in field bat collection; we thank Hangzhou Aimu Anthracite for library construction and high-throughput sequencing of bats; we thank XXX for sequencing three bat genomes and providing data sharing. This study was supported by the National Natural Science Foundation of China (31860257&81860300) and the Guizhou Provincial Hundred Talent Fund [No. (2018) 5637].

Author contribution, Conflict of interest, Ethical statement

Xinsheng Yao and Long Ma completed the experimental design, data analysis, and paper writing, etc. Long Ma and Hao Zhou completed the experiments, data analysis, and graph production, etc. Hao

Zhou, Jun Li, Longyu Liu, Dewei Zhou, Yingjie Wu, Xingliang Wang, Jiang Zhou, Qinqin, Mada, and others participated in sample collection, data analysis, and TRB CDR3 library construction in human and mouse.

There is no conflict of interest in this paper, and the experimental sample collection of bats and humans were approved by the ethics committee of Zunyi Medical University, and the informed consent of volunteers was obtained for the collection of human peripheral blood samples, and all sample collection and experiments were performed in strict accordance with biosafety requirements.

Reference

1. C. J. Burgin, J. P. Colella, P. L. Kahn, N. S. Upham, How many species of mammals are there? *Journal of Mammalogy* **99**, 1-14 (2018).
2. M. Letko, S. N. Seifert, K. J. Olival, R. K. Plowright, V. J. Munster, Bat-borne virus diversity, spillover and emergence. *Nat Rev Microbiol* **18**, 461-471 (2020).
3. M. Ruiz-Aravena *et al.*, Ecology, evolution and spillover of coronaviruses from bats. *Nat Rev Microbiol* **20**, 299-314 (2022).
4. P. Zhou *et al.*, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273 (2020).
5. K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, R. F. Garry, The proximal origin of SARS-CoV-2. *Nat Med* **26**, 450-452 (2020).
6. G. Goh *et al.*, Complementary regulation of caspase-1 and IL-1beta reveals additional mechanisms of dampened inflammation in bats. *Proc Natl Acad Sci U S A* **117**, 28939-28949 (2020).
7. M. Ahn *et al.*, Dampened NLRP3-mediated inflammation in bats and implications for a special viral reservoir host. *Nat Microbiol* **4**, 789-799 (2019).

407 8. A. T. Irving, M. Ahn, G. Goh, D. E. Anderson, L. F. Wang, Lessons from the host defences of
408 bats, a unique viral reservoir. *Nature* **589**, 363-370 (2021).
409 9. M. Ahn, J. Cui, A. T. Irving, L. F. Wang, Unique Loss of the PYHIN Gene Family in Bats
410 Amongst Mammals: Implications for Inflammasome Sensing. *Sci Rep* **6**, 21722 (2016).
411 10. J. Xie *et al.*, Dampened STING-Dependent Interferon Activation in Bats. *Cell Host Microbe*
412 **23**, 297-301 e294 (2018).
413 11. S. S. Pavlovich *et al.*, The Egyptian Rousette Genome Reveals Unexpected Features of Bat
414 Antiviral Immunity. *Cell* **173**, 1098-1110 e1018 (2018).
415 12. H. E. Randolph, L. B. Barreiro, Holy Immune Tolerance, Batman! *Immunity* **48**, 1074-1076
416 (2018).
417 13. C. G. Sotero-Caio, R. J. Baker, M. Volleth, Chromosomal Evolution in Chiroptera. *Genes (Basel)*
418 **8**, (2017).
419 14. A. Mori *et al.*, Next generation sequencing: new tools in immunology and hematology.
420 *Blood Res* **48**, 242-249 (2013).
421 15. A. Six *et al.*, The past, present, and future of immune repertoire biology - the rise of
422 next-generation repertoire analysis. *Front Immunol* **4**, 413 (2013).
423 16. S. Li *et al.*, IMGT/HighV QUEST paradigm for T cell receptor IMGT clonotype diversity and
424 next generation repertoire immunoprofiling. *Nat Commun* **4**, 2333 (2013).
425 17. D. A. Bolotin *et al.*, MiXCR: software for comprehensive adaptive immunity profiling. *Nat*
426 *Methods* **12**, 380-381 (2015).
427 18. L. Song *et al.*, TRUST4: immune repertoire reconstruction from bulk and single-cell
428 RNA-seq data. *Nat Methods* **18**, 627-630 (2021).
429 19. D. Jebb *et al.*, Six reference-quality genomes reveal evolution of bat adaptations. *Nature*
430 **583**, 578-584 (2020).

431 20. A. Rhie *et al.*, Towards complete and error-free genome assemblies of all vertebrate
432 species. *Nature* **592**, 737-746 (2021).
433 21. P. A. Larson *et al.*, Genomic features of humoral immunity support tolerance model in
434 Egyptian rousette bats. *Cell Rep* **35**, 109140 (2021).
435 22. H. Zhou, L. Ma, L. Liu, X. Yao, TR Locus Annotation and Characteristics of Rhinolophus
436 ferrumequinum. *Front Immunol* **12**, 741408 (2021).
437 23. F. Giannico *et al.*, The expansion of the TRB and TRG genes in domestic goats (Capra hircus)
438 is characteristic of the ruminant species. *BMC Genomics* **21**, 623 (2020).
439 24. P. D. a. M.-P. L. Jérôme Lane, From IMGT-ONTOLOGY to IMGT/LIGMotif: the IMGT®
440 standardized approach for immunoglobulin and T cell receptor gene identification and
441 description in large genomic sequences. *BMC Bioinformatics* **11**, 1-16 (2010).
442 25. V. Giudicelli, M. P. Lefranc, Imgt-Ontology 2012. *Front Genet* **3**, 79 (2012).
443 26. C. A. Sarkar SK, Analysis of immunocompetent cells in the bat, Pteropus giganteus:
444 isolation and scanning electron microscopic characterization. *Dev Comp Immunol* **15**,
445 423-430 (1991).
446 27. J. M. Martinez Gomez *et al.*, Phenotypic and functional characterization of the major
447 lymphocyte populations in the fruit-eating bat Pteropus alecto. *Sci Rep* **6**, 37796 (2016).
448 28. K. B. a. G. E. W. Dale A.Ramsden, Conservation of sequence in recombination signal
449 sequence spacers. *Nucleic Acids Research* **22**, 1785-1796 (1994).
450 29. T. Connelley, J. Aerts, A. Law, W. I. Morrison, Genomic analysis reveals extensive gene
451 duplication within the bovine TRB locus. *BMC Genomics* **10**, 192 (2009).
452 30. K. D. Timothy K Connelley, Cassandra W Longhi and W Ivan Morrison, Genomic analysis
453 offers insights into the evolution of the bovine TRA/TRD locus. *BMC Genomics* **15**, 1-17
454 (2014).

455 31. S. Jaiswal *et al.*, Identification and characterization of the T cell receptor (TCR) repertoire
456 of the Cynomolgus macaque (Macaca Fascicularis). *bioRxiv*, 1-29 (2022).
457 32. P. Pegorier *et al.*, IMGT(R) Biocuration and Comparative Study of the T Cell Receptor Beta
458 Locus of Veterinary Species Based on Homo sapiens TRB. *Front Immunol* **11**, 821 (2020).
459 33. H. Y. Greenaway *et al.*, Extraction and characterization of the rhesus macaque T-cell
460 receptor beta-chain genes. *Immunol Cell Biol* **87**, 546-553 (2009).
461 34. M. Mineccia *et al.*, New insight into the genomic structure of dog T cell receptor beta (TRB)
462 locus inferred from expression analysis. *Dev Comp Immunol* **37**, 279-293 (2012).
463 35. A. Radtanakatikanon *et al.*, Topology and expressed repertoire of the Felis catus T cell
464 receptor loci. *BMC Genomics* **21**, 20 (2020).
465 36. B. Gerritsen *et al.*, Characterization of the ferret TRB locus guided by V, D, J, and C gene
466 expression analysis. *Immunogenetics* **72**, 101-108 (2020).
467 37. R. Antonacci, M. Bellini, G. Linguiti, S. Ciccarese, S. Massari, Comparative Analysis of the
468 TRB Locus in the Camelus Genus. *Front Genet* **10**, 482 (2019).
469 38. G. Tsagkogeorga, J. Parker, E. Stupka, J. A. Cotton, S. J. Rossiter, Phylogenomic analyses
470 elucidate the evolutionary relationships of bats. *Curr Biol* **23**, 2262-2267 (2013).
471 39. S. Strobel, J. A. Encarnacao, N. I. Becker, T. E. Trenczek, Histological and histochemical
472 analysis of the gastrointestinal tract of the common pipistrelle bat (Pipistrellus
473 pipistrellus). *Eur J Histochem* **59**, 2477 (2015).
474 40. A. N. M. a. J. N. MAINA, The morphology of the intestine of the insectivorous horseshoe bat
475 (Rhinolophus hildebvandti, Peters): a scanning electron and light microscopic study. *Afr. J.*
476 *Ecol* **32**, 158-168 (1994).
477 41. V. A. Claude-Agnès Reynaud, Hélène Grimal, Jean-Claude Weill,, A hyperconversion
478 mechanism generates the chicken light chain preimmune repertoire. *Cell* **48**, 379-388

479 (1987).
480 42. M. RG., Diversification of rabbit VH genes by gene-conversion-like and hypermutation
481 mechanisms. *Immunol Rev* **162**, 49-54 (1998).
482 43. Z. Qu *et al.*, Structure and Peptidome of the Bat MHC Class I Molecule Reveal a Novel
483 Mechanism Leading to High-Affinity Peptide Binding. *J Immunol* **202**, 3493-3506 (2019).
484 44. T. Qurkhuli, N. Schwensow, S. D. Brandel, M. Tschapka, S. Sommer, Can extreme MHC class I
485 diversity be a feature of a wide geographic range? The example of Seba's short-tailed bat
486 (Carollia perspicillata). *Immunogenetics* **71**, 575-587 (2019).
487 45. J. H. Ng *et al.*, Evolution and comparative analysis of the bat MHC-I region. *Sci Rep* **6**, 21256
488 (2016).
489 46. S. J. Turner, P. C. Doherty, J. McCluskey, J. Rossjohn, Structural determinants of T-cell
490 receptor bias in immunity. *Nat Rev Immunol* **6**, 883-894 (2006).
491 47. T. O. Hansen, A. B. Lange, T. Barington, Sterile DJH rearrangements reveal that distance
492 between gene segments on the human Ig H chain locus influences their ability to rearrange.
493 *J Immunol* **194**, 973-982 (2015).
494 48. L. Fu *et al.*, A comprehensive profiling of T- and B-lymphocyte receptor repertoires from a
495 Chinese-origin rhesus macaque by high-throughput sequencing. *PLoS One* **12**, e0182733
496 (2017).
497 49. J. E. Park *et al.*, A cell atlas of human thymic development defines T cell repertoire
498 formation. *Science* **367**, (2020).
499 50. R. Alli, Z. M. Zhang, P. Nguyen, J. J. Zheng, T. L. Geiger, Rational design of T cell receptors
500 with enhanced sensitivity for antigen. *PLoS One* **6**, e18027 (2011).
501 51. C. J. Cibotti R, Pannetier C, Delarbre C, Vergnon I, Kanellopoulos JM, Kourilsky P, Public and
502 private V beta T cell receptor repertoires against hen egg white lysozyme (HEL) in

nontransgenic versus HEL transgenic mice. *J Exp Med* **180**, 861-872 (1994).

52. T. M. Haqqi, Gary D. Anderson, Subhashis Banerjee, and Chella S. David, Restricted Heterogeneity in T-Cell Antigen Receptor Vβ Gene Usage in the Lymph Nodes and Arthritic Joints of Mice. *Proc Natl Acad Sci U S A* **89**, 1253-1255 (1992).

53. Y. M. K Masuda, J Cui, T Ito, T Tokuhisa, Y Takahama, H Koseki, K Tsuchida, T Koike, H Moriya, M Amano, M Taniguchi, Phenotypes and invariant alpha beta TCR expression of peripheral V alpha 14+ NK T cells. *The Journal of Immunology* **158**, 2076-2082 (1997).

54. H. F. Koay *et al.*, Diverse MR1-restricted T cells in mice and humans. *Nat Commun* **10**, 2243 (2019).

Figure legends

Supplementary Fig 1. Sequence maps of PD and PP TRB loci exported by Geneious Prime software. (A) TRB loci of Phyllostomus discolor (B) TRB loci of Pipistrellus pipistrellus

Supplementary Fig 2. Geneious Prime software output of PD germline genes. (A) Amino acid sequence comparison of 100 TRBV genes of PD. (B) Amino acid sequence comparison of 14 TRBJ genes of PD.

Supplementary Fig 3. Comparison of TRBV gene 23RSS and TRBJ gene 12RSS in bat, human, mouse and cattle.

Supplementary Fig 4. TCRβ chain repertoire building peaks of the three RA samples.

Supplementary Fig 5. Analysis of V-J pairing (the order is Bat2, Human2 and Mouse2, Bat3, Human3 and Mouse3, respectively).

Supplementary Table 1 Accession number of TRBV gene phylogenetic tree construction and TRBC gene sequence alignment in this study.

Supplementary Table 2 Functional description of the annotated germline genes and analysis.

Supplementary Table 3 MiXCR normalized reference library gene nomenclature constructed for three kinds of bats.

Supplementary Table 4 Cytb sequencing of three bat samples and the BLAST result.

Supplementary Table 5 Cytb primers for bat muscle amplification and primers for human PBMC TCR β chain repertoire construction.

Supplementary Table 6 Output of MiXCR for nine samples from bat, human and mouse.

Supplementary Data 1 Genebank files for Phyllostomus discolor, Rhinolophus ferrumequinum, and Pipistrellus pipistrellus

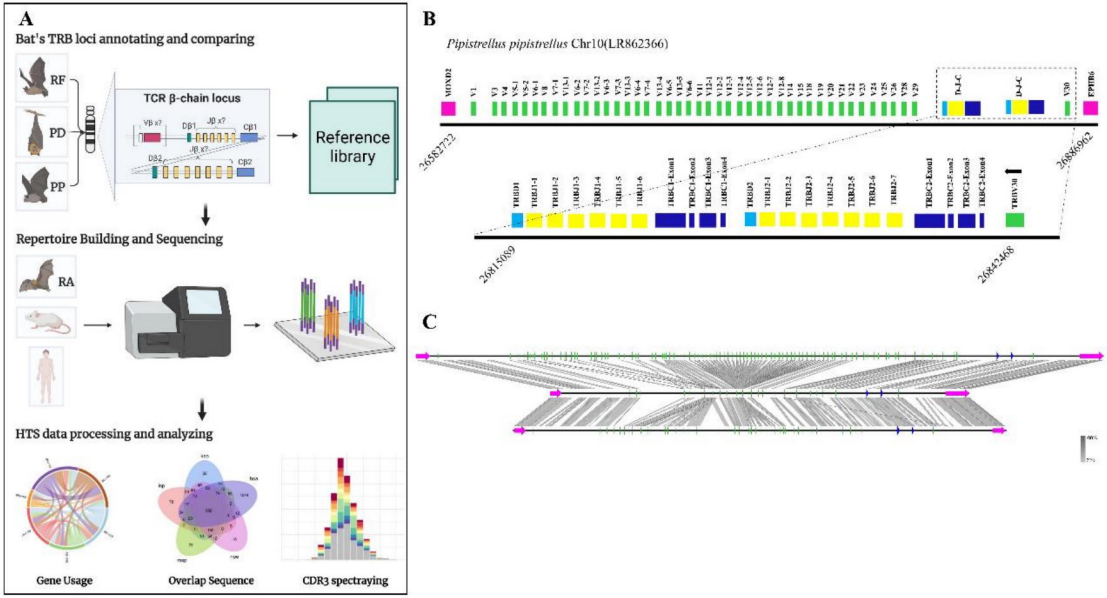


Figure1-TRB loci annotation and correlation analysis of three bat species.
(A) Overview of this study design.
(B) PP TRB loci on chromosone 10. The colored boxes represent each gene annotated, namely 44 TRBV genes, 2 TRBD genes, 13 TRBJ genes, and 2 TRBC genes.
(C) Genome homology comparison map of TRB loci for the three kinds of annotated bat.

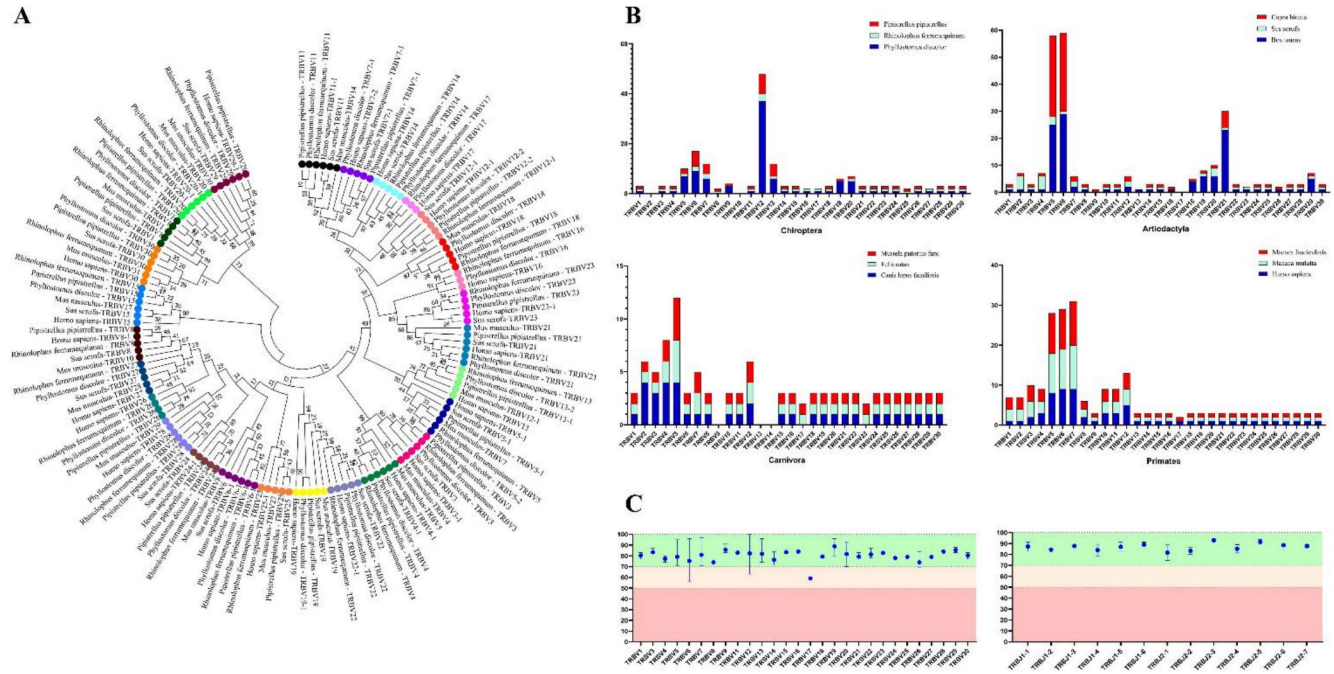


Figure2-Analysis of TRBV gene homology in bats with different species.
(A) NJ-phylogenetic tree constructed from functional TRBV gene nucleotide sequences of three kinds of bat, human, mouse and pig, with each color representing one family and bootstrap is 1000.
(B) Comparison of TRBV gene family distribution in three mammalian order: Carnivora, Artiodactyla, and Carnivora (https://www.imgt.org/ IMGTrepertoire/) with 3 kinds of Chiroptera.
(C) Nucleotide sequence similarity comparison of TRBV and TRBJ genes in 3 annotated bat species.

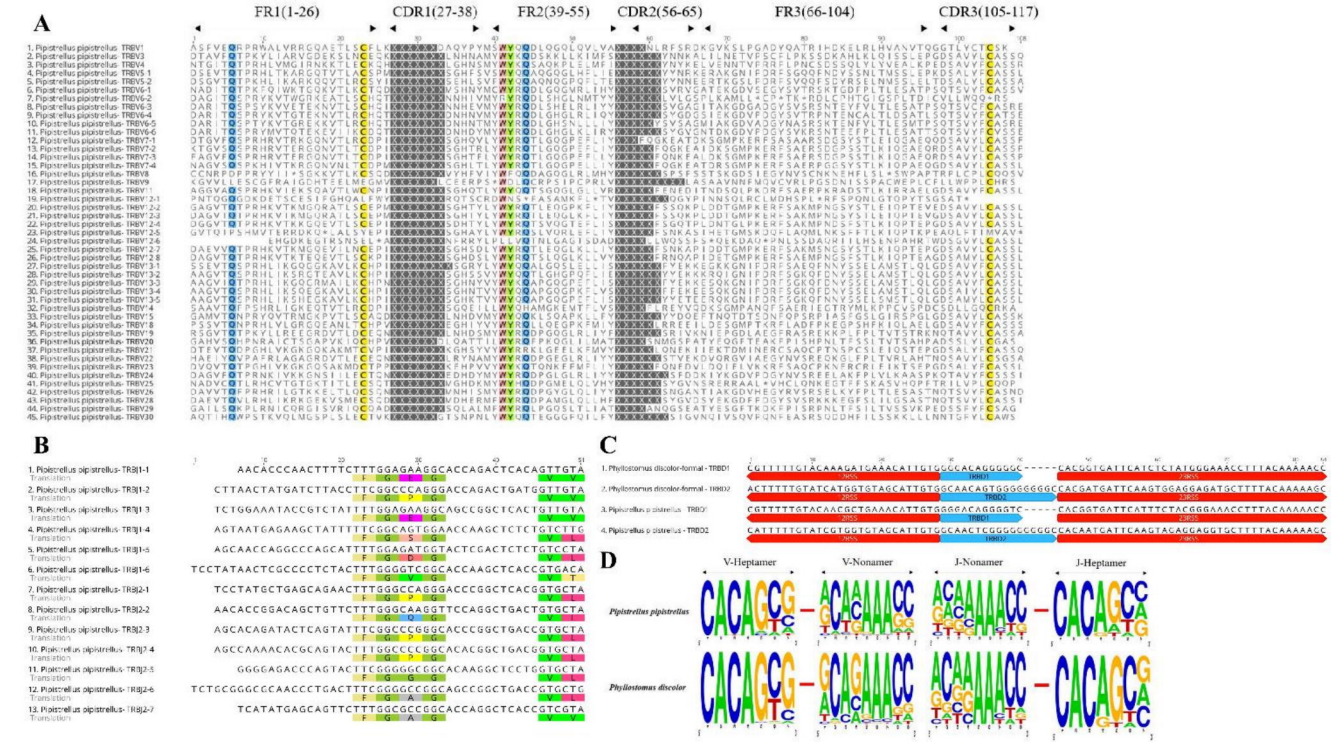


Figure3-PP annotated germline genes display.

(A) 45 TRBV amino acid sequences are shown with a consistency threshold of 90%.

(B) The nucleotide and amino acid sequences of 13 TRBJ genes are shown, and the conserved FGNG motifs are translated.

(C) Four TRBD genes with their 23RSS and 12RSS nucleotide sequences are showed.

(D) Comparison of conserved sites of TRBV'23RSS and TRBJ'12RSS in PD and PP.

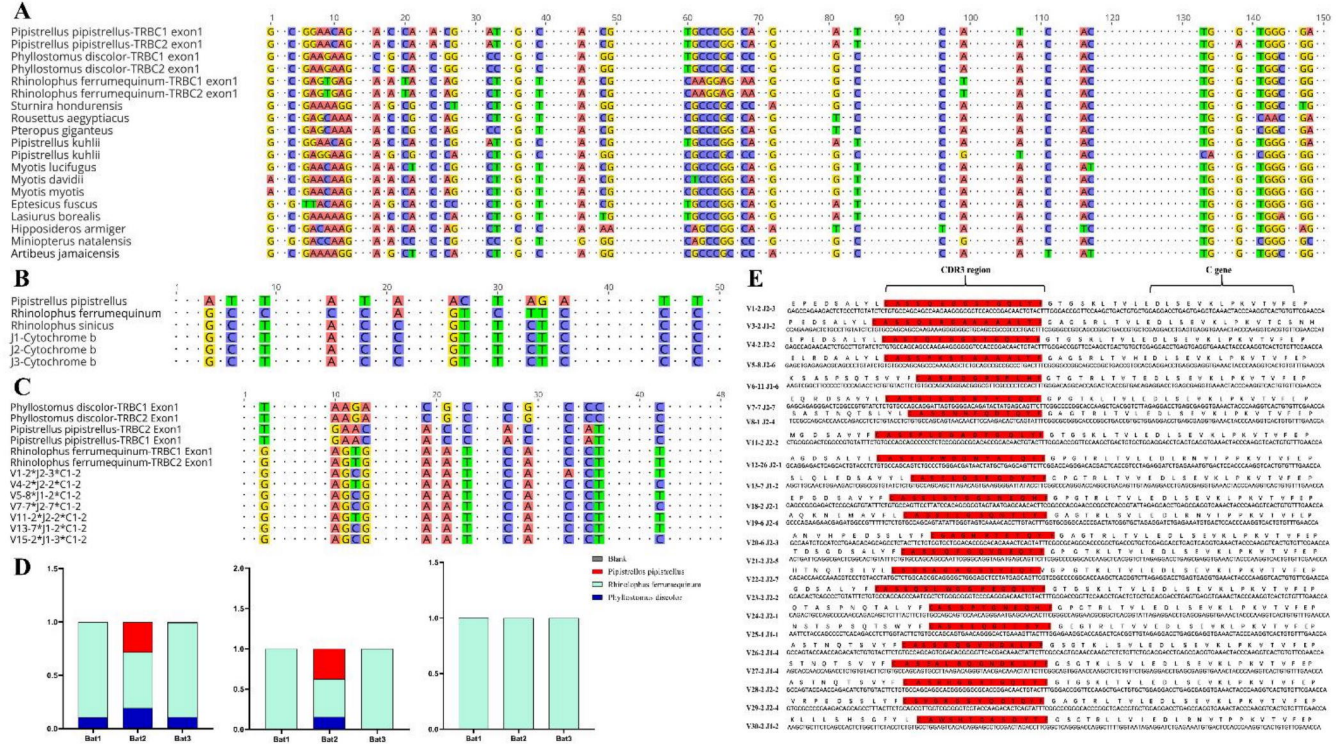


Figure4-Analysis of homology between RA samples and annotated bat species.

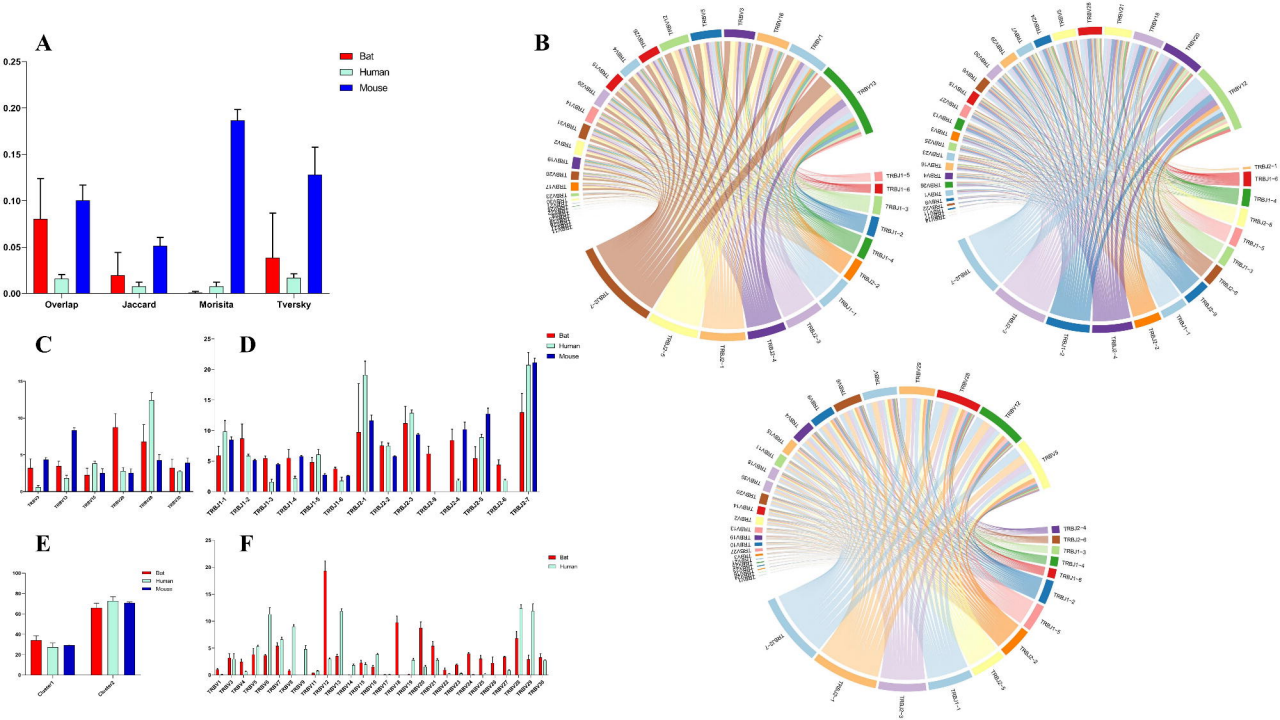
(A) The TRBC Exon1 nucleotide sequence alignment results of a total of 15 species of Chiroptera, were used to design amplification primers for the 5'RACE method of the RA.

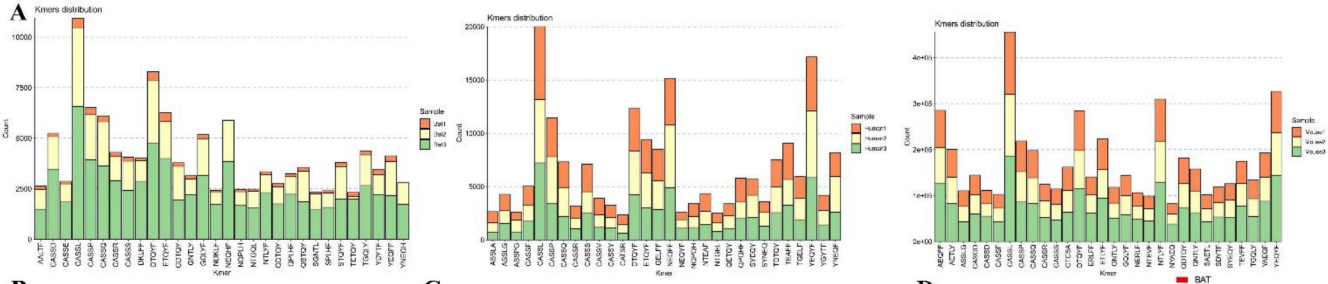
(B) Nucleotide sequence comparison of Cytb sequencing results of the 3 bat samples with PD, RF, and PP.

(C) The partial fragment of C gene from bat1 samples were compared with those of the 3 annotated bats.

(D) Attribution statistics of the sequencing results of the three RA samples in the annotated three kinds of bat.

(E) The 25 TRBV genes and 14 TRBJ genes output in bat1 sample.


A
Bat
Human
Mouse
0.25
0.20
0.15
0.10
0.05
0.00
Overlap
Jaccard
Morisita
Tversky
B
C
D
E
Cluster1
Cluster2
F



B

Bat	Human	Mouse
CASSD	CASSL	AEQFF
CASSL	CASSP	AETLY
CASSP	CASSQ	CASSL
CASSQ	DTQYF	CASSP
CASSR	ETQYF	CASSQ
DKLFF	GELFF	DTQYF
DTQYF	NEQFF	ETLYF
ETQYF	TEAFF	NTLYF
GQLYF	YEQYF	YAEQF
NEQHF	YNEQF	YEQYF

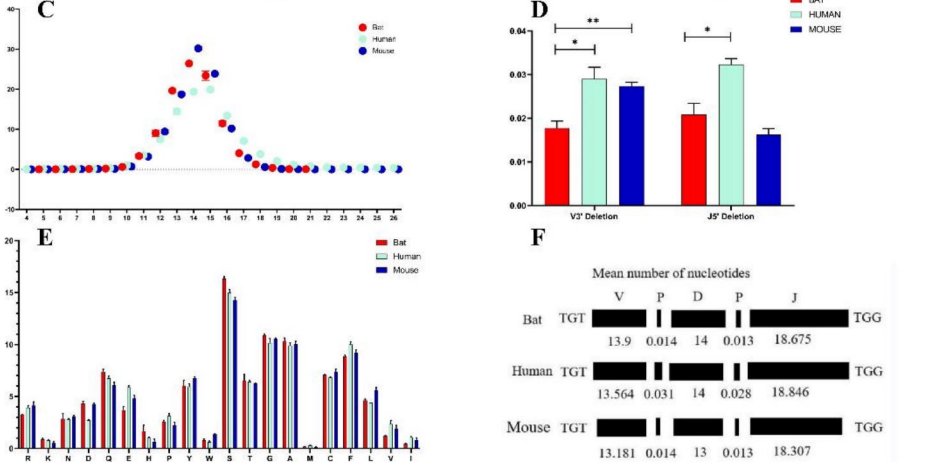


Figure6-Comparative analysis of the CDR3 region features of the TCRβ chain repertoire in bat, human and mouse.
(A) Top 30 5 AA length motif usage analysis.
(B) Example of top 10 motif usage with shared motifs labeled in the same color.
(C) CDR3 region length analysis.
(D) Comparison of deletion at the 3' end of V gene and 5' end of J gene.
(E) AA usage of the CDR3 region.
(F) Composition of the CDR3 region.

Table1 Number of germline genes retrieved by homologous genes and RSS method

Species and Chromosome accession	Phyllostomus discolor Chr12(NC_040912.2)	Rhinolophus ferrumequinum Chr26(NC_046309)	Pipistrellus Pipistrellus Chr10(LR862366.1)
12/23RSS	96V、2D、13J	30V、2D、14J	43V、2D、13J
IMGT/LIGMotif(Homo Sapiens)	99V、8J	30V、2D、12J	44V、9J
IMGT/LIGMotif(Mus musculus)	92V、10J	30V、2D、13J	44V、9J
IMGT/LIGMotif(Sus scrofa)	97V、10J	30V、2D、14J	44V、9J
Geneious Prime Mapping	64V、2D、	30V、14J	34V、13J
Total Number of Germline Genes	100TRBV 、 2TRBD 、 13TRBJ 、 2TRBC	30TRBV、2TRBD、15TRBJ、2TRBC	45TRBV、2TRBD、13TRBJ、2TRBC
Organization of TRB loci	Vβ(99)-Dβ(1)-Jβ(6)-Cβ(1)-Dβ(1)-Jβ(7)-Cβ(1)- Vβ(1)	Vβ(30)-Dβ(1)-Jβ(6)-Cβ(1)-Dβ(1)-Jβ(9)-Cβ(1)- Vβ(1)	Vβ(44)-Dβ(1)-Jβ(6)-Cβ(1)-Dβ(1)-Jβ(7)-Cβ(1)-Vβ(1)

Table2 Sequence statistics of a total of 9 samples from bats, humans and mice output by MiXCR

Sample ID	Total Sequence	Total Sequence Clonetype	Analysis Sequence Productive	Analysis Sequence Clonotype
Bat1	2557414	3498	2381182	3234
Bat2	4949687	25443	4199657	22971
Bat3	5303900	45771	4912794	41950
Human1	3844196	76589	3432143	76588
Human2	3872836	75533	3469357	75530
Human3	3931149	75037	3538969	75034
Mus1	17381732	1334514	7776887	1077810
Mus2	14093768	1233928	7604800	1073215
Mus3	15174528	1150536	9456533	1147763